15th International Conference on Spatial Information Theory

COSIT 2022, September 5–9, 2022, Kobe, Japan

Edited by

Toru Ishikawa
Sara Irina Fabrikant
Stephan Winter

Editors

Toru Ishikawa
Toyo University, Tokyo, Japan
toru.ishikawa@iniad.org

Sara Irina Fabrikant
University of Zurich, Switzerland
sara.fabrikant@geo.uzh.ch

Stephan Winter
University of Melbourne, Australia
winter@unimelb.edu.au

ACM Classification 2012
Information systems → Geographic information systems; Applied computing → Transportation

ISBN 978-3-95977-257-0

Published online and open access by
Schloss Dagstuhl – Leibniz-Zentrum für Informatik GmbH, Dagstuhl Publishing, Saarbrücken/Wadern, Germany. Online available at https://www.dagstuhl.de/dagpub/978-3-95977-257-0.

Publication date
September, 2022

Bibliographic information published by the Deutsche Nationalbibliothek
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the Internet at https://portal.dnb.de.

Digital Object Identifier: 10.4230/LIPIcs.COSIT.2022.0

ISBN 978-3-95977-257-0 **ISSN 1868-8969** **https://www.dagstuhl.de/lipics**

LIPIcs – Leibniz International Proceedings in Informatics

LIPIcs is a series of high-quality conference proceedings across all fields in informatics. LIPIcs volumes are published according to the principle of Open Access, i.e., they are available online and free of charge.

ISSN 1868-8969

https://www.dagstuhl.de/lipics

Contents

15th International Conference on Spatial Information Theory (COSIT 2022).
Editors: Toru Ishikawa, Sara Irina Fabrikant, and Stephan Winter
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Short Papers

Preface

Established in 1993[1], COSIT is a biennial international conference series concerned with theoretical aspects of space and spatial information. COSIT grew out of a series of workshops, NATO Advanced Study Institutes and NSF specialist meetings, all concerned with cognitive and applied aspects of representing large scale space, particularly geographic space. In these meetings, the need for a well-founded theory on spatial information processing was identified, and COSIT was formed in order to provide the platform for the intensive interdisciplinary scientific exchange on this theory. Since its inception, COSIT continuously attracted like-minded scientists from computer science, geography, cognitive science, philosophy and geomatics. Out of this interdisciplinary community grew journals such as Spatial Cognition and Computation (Taylor and Francis) and the open access Journal of Spatial Information Science. Originally, COSIT proceedings were published in Springer's Lecture Notes in Computer Science series. Since 2017 COSIT proceedings appear open-access in the Leibniz International Proceedings in Informatics series.

The first two COSIT conferences, 1993 and 1995, were held in Europe, to then turn into a truly international enterprise, held every second year in principle between locations in Europe and America. With Kobe, Japan, as the host city 2022, the 15th International Conference on Spatial Information Theory has been located for the first time in Asia. COSIT 2022 aimed to rekindle the well established COSIT spirit after a long, pandemic-caused hibernation. This spirit is based on the serendipity of a single-track meeting, of in-depth discussions of innovative and significant recent contributions, and of gathering new collaborations of interdisciplinary research amongst on-site participants.

International borders not yet open, COSIT 2022 dared to bet on a chiefly on-site event, which was supported by the distance in time and space from both Europe and America: A hybrid participation was thus only offered for the early morning or late evening paper presentation sessions. However, at the time of writing this Preface, we do not know yet whether border closures will be lifted in time for the conference. At least this volume of proceedings will document in writing the current state of research in the broad scientific communities dealing with spatial information theory.

Despite various uncertainties COSIT 2022 aims again to bring together scientists with various disciplinary backgrounds to extensively discuss the interdisciplinary state-of-the-art in theoretical aspects of space and spatial information, in order to advance geographic information science and its emerging research frontiers. This aim is facilitated by the on-site presentation and discussion of a restricted number of papers and posters -– the most innovative and significant recent contributions -– rather than papers covering incremental advances in one field. The conference offered three (refereed) submission tracks: vision papers, full papers, and short papers, which are collected in this volume of proceedings.

We received 13 full paper submissions, 33 short paper submissions and 4 vision paper submissions. The latter category, introduced at COSIT 2019, aimed at rigorously researched and argued agenda-setting ideas that identify emerging research frontiers, and/or societally relevant research problems that are complex or hard to solve.

[1] In 1992, Andrew Frank organized a successful international conference on "GIS–From Space to Territory: Theories and Methods for Spatio-Temporal Reasoning", in hindsight referred to as COSIT Zero.

15th International Conference on Spatial Information Theory (COSIT 2022).
Editors: Toru Ishikawa, Sara Irina Fabrikant, and Stephan Winter
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

All submissions were thoroughly reviewed, in the large majority by at least three Program Committee members. Nine full papers, 18 short papers and two vision papers were selected for this volume (approx. 58% acceptance rate), and further five short papers were selected for presentation only.

The breadth of the topics in this volume also reflects the breadth of the disciplines involved in fundamental research related to geographic information theory. Excitingly, traditional research topics such as space-time representations, spatial relations, navigation, or (strong) spatial cognition are still alive and well. Empirical research on how to extract and analyze spatial information from rapidly growing user-generated online multimedia databases, for example, produced in a citizen science context, has clearly emerged as a new and popular research frontier in the field. Meanwhile, "big picture" theories and human behavioral studies have recently yielded fewer contributions (although still represented herein), despite being of great value to this interdisciplinary field.

Organizing any conference is not possible without the commitment, effort, and help of many people involved. The conference organizing team, fully aware of these uncertain times for on-site conferencing and its impact on budgeting, submission numbers, and participation, particularly wishes to thank the numerous researchers submitting their best ideas in hope of on-site discussions, the program committee thoroughly reviewing submissions in short time frames, the local organizing team for their flexibility and courage of planning an exciting conference, the keynote speakers willing to travel despite their full calendars and current travel risks, and ... all the participants at the conference for their lively discussions, should the conference indeed take place as planned.

The conference organizers gratefully acknowledge the trust and support of the conference sponsors to make COSIT 2022 happen in Kobe, Japan. COSIT 2022 is supported by the Association of Geographic Information Laboratories in Europe (AGILE), Platinum Sponsorship, the American Association of Geographers (AAG), Bronze Sponsorship, Taylor & Francis with the International Journal of Geographical Information Science, Bronze Sponsorship, the Toyo University's Department of Information Networking for Innovation and Design (INIAD) with its collaboration hub for University and Business (cHUB), the Kwansei Gakuin University, the Geographic Information Systems Association of Japan, and the Association of Japanese Geographers.

Tokyo, 24 May 2022

Toru Ishikawa
Toyo University, Tokyo, Japan
General Chair, COSIT 2022

Sara Irina Fabrikant
University of Zurich, Switzerland
Program Chair, COSIT 2022

Stephan Winter
University of Melbourne, Australia
Program Chair, COSIT 2022

What Do You Mean You're *in* Trafalgar Square? Comparing Distance Thresholds for Geospatial Prepositions

Niloofar Aflaki[1]
Massey Geoinformatics Collaboratory, Massey University, Auckland, New Zealand

Kristin Stock
Massey Geoinformatics Collaboratory, Massey University, Auckland, New Zealand

Christopher B. Jones
School of Computer Science and Informatics, Cardiff University, UK

Hans Guesgen
Massey Geoinformatics Collaboratory, Massey University, Auckland, New Zealand

Jeremy Morley
Ordnance Survey, Southampton, UK

Yukio Fukuzawa
School of Natural and Computational Sciences, Massey University, Auckland, New Zealand

Abstract

Natural language location descriptions frequently describe object locations relative to other objects (*the house near the river*). Geospatial prepositions (e.g.*near*) are a key element of these descriptions, and the distances associated with proximity, adjacency and topological prepositions are thought to depend on the context of a specific scene. When referring to the context, we include consideration of properties of the relatum such as its feature type, size and associated image schema. In this paper, we extract spatial descriptions from the Google search engine for nine prepositions across three locations, compare their acceptance thresholds (the distances at which different prepositions are acceptable), and study variations in different contexts using cumulative graphs and scatter plots. Our results show that adjacency prepositions *next to* and *adjacent to* are used for a large range of distances, in contrast to *beside*; and that topological prepositions *in*, *at* and *on* can all be used to indicate proximity as well as containment and collocation. We also found that reference object image schema influences the selection of geospatial prepositions such as *near* and *in*.

2012 ACM Subject Classification Computing methodologies → Natural language processing

Keywords and phrases contextual factors, spatial descriptions, acceptance model, spatial template, applicability model, geospatial prepositions

Digital Object Identifier 10.4230/LIPIcs.COSIT.2022.1

Funding This work is partly funded through by an Ordnance Survey (UK) PhD scholarship.

1 Introduction

In natural language location descriptions, people tend to describe their location or that of a point of interest (POI) using relative relation terms (e.g. *the house beside the park* describes the location of the house relative to the park). These types of location descriptions contain three essential elements: the locatum (the object for which the location is being described);

[1] Corresponding author

the relatum (used as a reference location for describing a locatum) and the spatial relation term (which specifies the relation in space between the locatum and relatum) [35] (*the house <locatum> beside <spatial relation term> the park <relatum>*).

Relative spatial descriptions are a common method for describing location in human communication, and accurate automated interpretation of these kinds of expressions can be of critical importance for many applications. For example during emergency events, they may describe the location of stranded people or dangerous conditions as in *there is a fire in the house on Victoria street, next to the Coffee Club cafe* [39, 13]. The availability of large amounts of text on web sites, blogs and social media motivates the development of methods to automatically interpret and generate such natural language relative location descriptions.

Most of the previous work on georeferencing relative spatial descriptions has focused on toponym recognition and disambiguation [20, 21, 17, 16, 18] without taking into account the role of spatial relation terms. For example, in a description such as *behind the Shell building*, consideration of the preposition *behind* improves the accuracy of the georeference that would be obtained if only the place name were used. To consider geospatial prepositions[2], it is necessary to understand the locations (relative to the relatum) in which a given spatial relation term may validly be used (e.g., how near does a locatum have to be to a relatum for *near* to be an appropriate spatial relation term?). To address this kind of question, a number of models have been developed for specific spatial prepositions, known as acceptance models, applicability models or spatial templates [26, 9, 31, 2, 11, 37, 32, 42, 6, 28, 3, 4]. These models are often probabilistic or predictive, describing areas in which a given preposition is highly suitable, compared to others where it may be borderline.

We address two gaps in the previous research. Firstly, previous work has mostly focused on the task of developing models for small numbers of individual prepositions in isolation. Here, we compare the acceptance thresholds (the distances at which a preposition is acceptable) for different prepositions in the English language in order to study their semantic similarities and differences. Secondly, the importance of contextual factors on location interpretation has been emphasised in a number of previous works [12, 15, 27, 38, 24, 41, 8, 28], but the comparative impact of context on prepositions has not been widely studied empirically. We address these gaps by comparing the use of the nine prepositions across three well-known landmarks in London, UK (Trafalgar Square, Buckingham Palace, and Hyde Park) taking account of context with particular reference to the nature of the relatum and its associated image schema. We address two research questions of: **RQ1**: How do distances between relata and locata that are acceptable differ between geospatial prepositions? **RQ2**: How important is context in the use of geospatial prepositions?

In Section 2, we discuss relevant previous work, while Section 3 defines the data extraction method; Section 4 presents the results and Section 5 provides a discussion and findings.

2 Previous work

2.1 Acceptance models

Spatial acceptance models define the areas in which a given preposition may be applied, relative to a relatum, and have been investigated for several purposes, including location prediction [2, 4], selection of an appropriate preposition for a description (language generation) [6, 28] and georeferencing [9, 3]. Chang et al. [2] and Yu and Siskind [42] used

[2] We consider a spatial preposition to be *geospatial* if the relatum is a geographical object.

acceptance models to draw a spatial scene in 3D using textual descriptions and to find objects in videos in an indoor environment respectively. Malinowski and Fritz [22] and Lan et al. [19] used deep learning and other machine learning models (CNN and latent ranking SVM) to retrieve specific objects in image configurations, relying on spatial acceptance models. However, they did not compare individual spatial prepositions or consider contextual factors, and the studies were image-based without consideration of geographic space. In a series of studies Hall et al. [10, 9, 11] used location descriptions from the Geograph website and human subjects experiments to quantify the distances and angles for projective and proximity spatial relations, and to create spatial templates for the purpose of generating and interpreting natural language photo captions. The scale varied between urban and rural locations and corresponds to the environmental and geographical spaces of Montello [25]. A study of the distances associated with the relation *near* was conducted in Derungs and Purves [5] based on evidence of n-grams mined from the web. They considered relations between places that were either cities or points of interest, and hence environmental and geographical spaces, and found the distances were typically between nearest neighbours and random. Our work considers a relatively wide range of spatial prepositions and focuses on the environmental scale using data scraped from web page sources.

2.2 Effects of contextual factors on spatial preposition use

The importance of context in selection and interpretation of spatial prepositions has been well documented. Herskovits [12] identified geometric configuration, use types, salience, relevance, tolerance and typicality as important in determining whether a preposition would apply in a given situation or not. Tyler and Evans [36] counted context and other elements of spatial descriptions such as the locatum and relatum as important factors for understanding spatial prepositions such as *over*. Stock and Yousaf [34] used a Linguistically Augmented Geospatial Ontology (LAGO) to extract five contextual factors for locata and relata and calculated the similarity between them using WordNet [40], showing that the consideration of contextual factors improves the ability to identify semantically similar descriptions. Wallgrun et al. [38] used a similar method to us to extract content from internet search engines, and demonstrated the variation in distances associated with three proximal spatial relations according to the mode of transport. Collell et al. [4] used spatial templates to predict the location of objects in photos using contextual factors such as embeddings (which describe linguistic context) and size of locatum. However, their focus was on spatial relations in the form of verbs and they did not address prepositions. Stock and Hall [33] identified a broad range of factors that affect the interpretation of spatial descriptions including: proximity and locatum and relatum characteristics such as liquid/solid and image schema. However, they did not explicitly describe the ways that context influences the interpretation of spatial prepositions or the distance between locatum and relatum.

3 Data extraction method

We used Google searches to extract descriptions from the World Wide Web that contained three elements: locatum, geospatial preposition and relatum. We used place names for the locatum and relatum from OpenStreeMap[3], identified coordinates of the relatum and locatum used with a specific preposition and calculated the distance between them for

[3] `https://www.openstreetmap.org/`

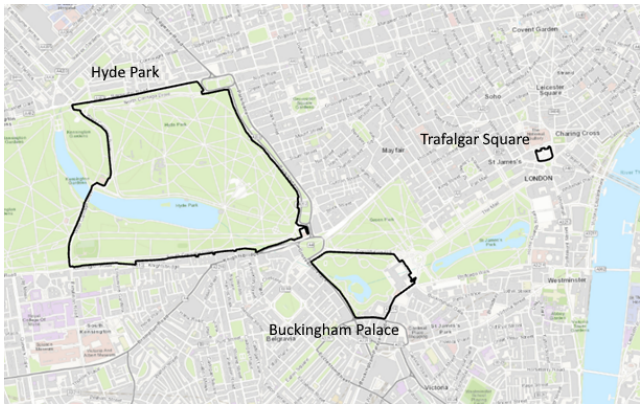


Figure 1 Locations of Hyde Park, Buckingham Palace and Trafalgar Square in London.

further analysis. We use the frequency of mentions of a locatum with a particular geospatial preposition-relatum combination as a proxy for the applicability of that geospatial preposition. For example, a search for *Green Park next to Buckingham Palace* returned a count of 83 mentions (which we refer to as frequency). We consider that this frequency of use indicates that the *next to* preposition is more acceptable for the Green Park-Buckingham Palace locatum-relatum pair than for some other pair of place names with a lower frequency.

In order to compare the influence of context, we selected three popular tourist attractions in the London area (Figure 1) as relata with a variety of scales and feature types: Trafalgar Square (area: $18040m^2$, perimeter: $954m$), Hyde Park (area: $1388013m^2$, perimeter: $5629m$), and Buckingham Palace (the building and its grounds)(area: $201240m^2$, perimeter: $1997m$). We used the following steps to extract the data for each relatum:

1. We used the OpenStreetMap export service to extract all places in the general area of the three relata using a bounding box that covered a large section of central London.
2. From the set returned in Step 1, we identified those features that had centroids within a specified distance of the centroids of each relatum. A maximum distance of 2km was used for Trafalgar Square and Buckingham Palace, and 3km for the much larger Hyde Park. These distances were selected to retrieve a manageable number of locata but with the aim of encompassing typical extents of acceptable use of the prepositions. Our results indicate that the selected distance ranges were sufficient in most cases (see Section 4.2). We only extracted point and polygon geometries and defer consideration of linear objects to a later study.
3. We manually checked and excluded place names returned from Step 2 that had multiple instances (for example, *McDonald's* has multiple branches across the London area) in Google Maps, to avoid ambiguity regarding the coordinate location of the locata.
4. After Step 3, we had around 800 locata for each relatum. Google search counts were used to identify the 100 most frequently mentioned places as candidate locata for each relatum (though it is acknowledged that these counts might be approximate due to Google's search algorithm).
5. We generated query triples combining each of the 100 locata for a given relatum, each of the prepositions and the relatum itself, enclosed with quotation marks (e.g. *"National Gallery near Trafalgar Square"*), using Python's Beautiful Soup library [29] to run a query for each triple and scrape the URL and excerpt from the retrieved page.
6. We ran a version of the previous step in which a wildcard character was included before the preposition, to accommodate the common presence of verbs.

7. We manually reviewed the output returned from the previous steps in order to remove mentions that were repetitive, non-spatial or outside the geographical area. We also examined outliers and removed invalid expressions manually (e.g. expressions that referred to buildings that had moved since the text was written, which occurred particularly for historical documents online).
8. We combined the frequencies of mentions for searches with and without the wildcard character.
9. We retrieved the geometry for each locatum and relatum from OpenStreetMap, along with the category (e.g., tourism), type (e.g., attraction) and centroid coordinates (if a polygon).
10. Finally, we calculated the shortest distance between the location of each locatum and relatum using point geometries (for points) or boundary geometries (for polygons).

This analysis process resulted in a total of 1970, 1523 and 1746 mentions (across all locata and all nine prepositions) for Buckingham Palace, Hyde Park and Trafalgar Square respectively.

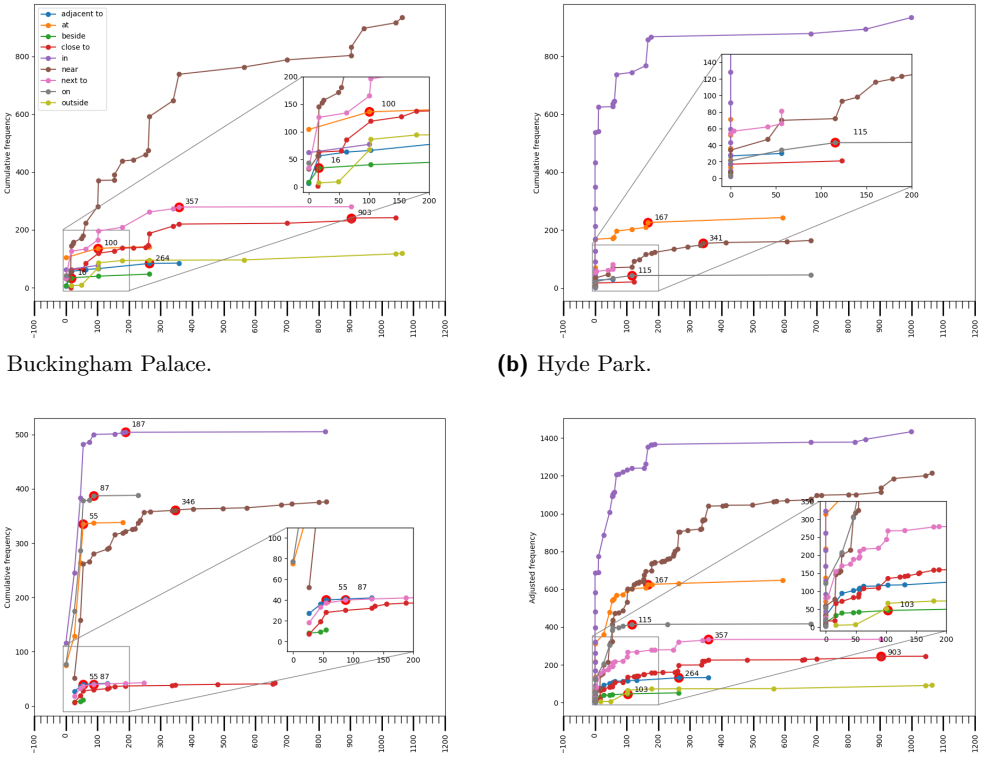
4 Results

4.1 RQ1: How do acceptable distances between relata and locata differ between prepositions?

Figure 2 shows the cumulative frequency graphs [14] for each preposition for each of the three sites. Each coloured line is a preposition, as labelled in the legend. The points on the graphs represent locata and are positioned on the x axis using the distance between the relatum and locatum geometries. The vertical axis shows cumulative frequencies, being the total of all mentions of a given locatum plus all locata at shorter (closer) distances. We use cumulative frequency graphs because they provide a clearer picture of the behaviour of each preposition in comparison to raw frequency graphs in which individual locata can obscure the visualisation. Note that the point at which each curve flattens is the distance beyond which there are very few new mentions, so we consider this to be the acceptance threshold for each preposition.

On the graphs in Figure 2, acceptance thresholds for each preposition for each relatum are marked with large red dots (Figure 2 (a-c)). They are identified as follows. Starting from the right-most point on each line for a given preposition in the graph, we move progressively leftward, point by point, to identify the first point for which the slope of the edge connecting that point to the point to its right exceeds 5°, and consider this point the acceptance threshold. If the right-most edge (between the last two points) exceeds 5°, we consider that we have insufficient data to identify the acceptance threshold. If none of the edges for a given preposition has a slope exceeding 5° (they are all relatively flat), we calculate the average slope for all edges for the preposition line, and repeat the above process to find the right-most point for which the slope of the edge connecting that point to the point to its right exceeds the average slope. The figure of 5° was selected through a process of trial and error with a range of other angles, with 5° best identifying the point at which the lines flatten consistently across all of the prepositions.

In addition to the graphs for each relatum, we present an aggregated cumulative frequency graph Figure 2(d), in which we sum the data for all relata, and adjust frequency values to account for varying site popularity. Some relata are more popular than others (i.e., attract more mentions in social media in total). So, if more popular sites are not adjusted for popularity, their values will have a disproportionately large influence on the shape of the graph that combines the results of all relata. We therefore scale down the mention count for



(a) Buckingham Palace.

(b) Hyde Park.

(c) Trafalgar Square.

(d) d) All relata aggregated.

Figure 2 Cumulative frequency graphs for each relatum and aggregated (Euclidean distances on x axis and frequencies on y axis).

these more popular sites so that all relata have an equal total adjusted number of mentions, this being equal to the minimum total across all relata (in this case it is for Hyde Park, which has 1523 mentions). The reason for doing this rather than normalising (adjusting to values between 0 and 1) is that it still gives some indication of the number of mentions in absolute terms (i.e. it is clear if there are very few mentions) (Equation 1).

$$adjfreq(r_i, l_i) = \frac{freq(r_i, l_i)}{\sum_{i=1}^{n} freq(r_i, l_i..n)} * min(\sum_{i=1}^{n} freq(r_i..n, l_i..n)) \quad (1)$$

$freq(r_i, l_i)$ indicates the frequency of a given preposition for $locatum_i(l_i)$ and $relatum_i(r_i)$, $\sum_{i=1}^{n} freq(r_i, l_i..n)$ indicates the sum of the frequencies of the prepositions across all locata for $relatum_i$ and $min(\sum_{i=1}^{n} freq(r_i..n, l_i..n))$ indicates the sum of the frequencies of the prepositions across all the locata for the relatum that has the minimum sum of the preposition frequencies (in this case Hyde Park). The numbers of mentions for Trafalgar Square and Buckingham Palace were higher than those for Hyde Park, and were adjusted down accordingly.

The data show some clear patterns across the prepositions, with the range in acceptance thresholds for each site summarised in Figure 3. We were unable to identify an aggregate threshold for *in* and *near* because the aggregated data did not reach a point of flattening. The *close to* preposition has the highest acceptance threshold, with (903m) for Buckingham Palace and for the aggregated data, but there is insufficient data to establish a threshold

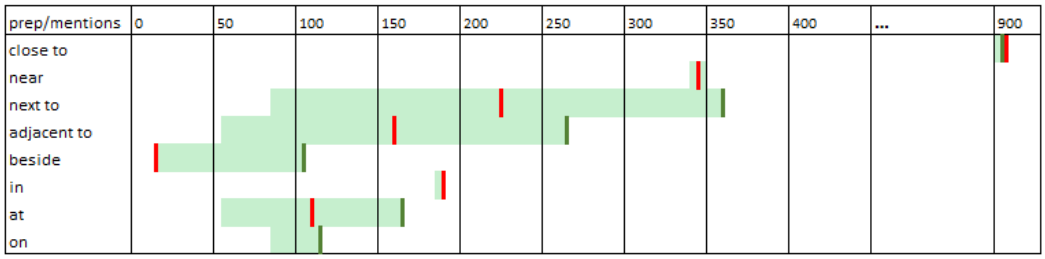


Figure 3 Acceptance threshold ranges for each preposition (red vertical line is mean, green vertical line is aggregate).

for the other two relata (*close to* is a relatively infrequently used preposition, so we have few mentions in the data). The *near* preposition also has a high threshold: 341m for Hyde Park and 347m for Trafalgar Square. Furthermore, we see *near* being used infrequently for much greater distances than the threshold (700-800m for Hyde Park and Trafalgar Square). The acceptance threshold for Buckingham Palace (and for the aggregated data) is much higher (>1100m), and the graph does not level off for this site, indicating that the acceptance threshold is beyond our last data point.

Of the prepositions that convey notions of adjacency, *next to* has the highest threshold for the aggregate data (357m). However, the range in thresholds between 358m for Buckingham Palace and 87m for Trafalgar Square has some substantial overlap with the range for *adjacent to* (265m to 55m, with an aggregate data threshold of 264m). *Beside* has a much smaller threshold, being 103m for the aggregate data, and 17m for Buckingham Palace (*beside* did not appear in our data for Hyde Park, and only infrequently, with distances up to 50m, for Trafalgar Square). From this evidence, we postulate that *beside* is typically limited to much closer locations than *next to* and *adjacent to*, both of which are used for locations within a closer range than the proximity prepositions *near* and *close to*, but more data is needed to confirm this. *Outside* only appeared for the Buckingham Palace site, but no acceptance threshold could be determined as the slope between its rightmost two points exceeds 5°.

Moving to the containment and collocation prepositions *in*, *at* and *on*; surprisingly, the acceptance thresholds for *in* appear to be large, being beyond our last data point for the aggregate data and for Hyde Park, and to a lesser extent, for Trafalgar Square (187m), even though, given that our distances are measured boundary to boundary, we might expect distances of zero (the locatum inside the relatum). For Buckingham Palace, only two locata were used with the *in* preposition. The first one is within its boundary (The Royal Mews) and the second is 100m away (Victoria Memorial). While the latter is located on the site of Buckingham Palace, it was not within the boundary geometry we extracted from OpenStreetMap. The Hyde Park data is affected by the location of the neighbouring Kensington Gardens, which people sometimes refer to as Hyde Park. For example, the description *Princess Diana Memorial Playground in Hyde Park* appears, but the playground is on the east side of Kensington Gardens, 900m outside the boundary of Hyde Park. In the case of Trafalgar Square, the most distant locata within the acceptance threshold are Her Majesty's Theatre (187m) and the Nigerian High Commission (180m). We expect that the reason for using *in Trafalgar Square* for these two locations is that Trafalgar Square is a well-known landmark in the area. It appears that in natural language location descriptions, the geographic boundaries of well-known landmarks may be "stretched", but more work is needed to validate this. We also see an unusual outlier for Trafalgar Square: the Methodist

Central Hall. It is marked with a dashed line on the graph as we suspect that it is an error. The source description reads "*16/09/2015 – The largest air raid shelter in England was at the Methodist Central Hall in Trafalgar Square which could hold 2,000 people each night.*" (Westminster Reporter, September 2015, Issue 120, page 21). Methodist Central Hall is 821m from Trafalgar Square and other documentation indicates it was associated with a large air raid shelter, but we have no evidence that either was in Trafalgar Square.

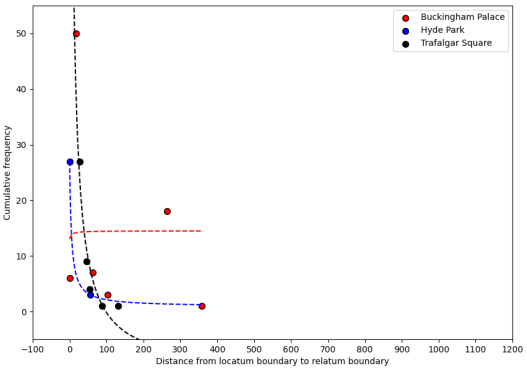
The *at* preposition has acceptance thresholds in a similar range to the *in* preposition, but with the lowest acceptance threshold of the three topological prepositions, being for Trafalgar Square. The *on* preposition has thresholds in a similar range (87m-116m) with a threshold for the aggregate of 115m, being the lowest for the three containment and collocation prepositions.

4.2 RQ2: How important is context in the use of geospatial prepositions?

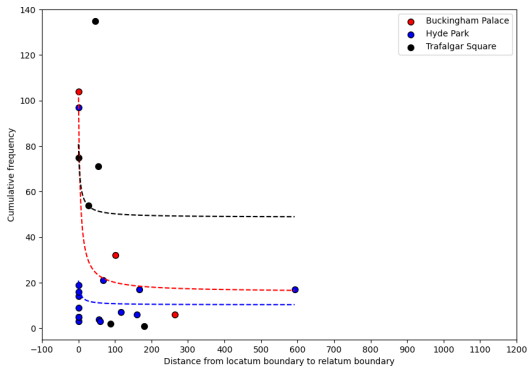
Our three relata were specifically selected for their difference in size and feature type (Figure 1) to enable investigation of contextual differences that relate to the nature of the relatum. Figure 4(a-h) compares the three relata for each of the prepositions using scatter plots and regression lines for each relatum. We use a reciprocal, linear regression equation to plot the regression line (y=1/x), and we refer to these regression lines as acceptance profiles, as they show the distances at which the preposition is highly acceptable, and those at which it becomes less so. This provides more information than the acceptance threshold alone. We do not plot the *outside* preposition, as we only identified its use with the Buckingham Palace relatum, and in RQ2 our focus is on comparison of the context.

Several of the prepositions show clear similarity across all three sites, including *next to* and *close to*. The curves across the three relata for *next to* are very similar, the main difference being in frequency of mentions, which is discussed further below. The ranges of the data points vary for the three sites, with Buckingham Palace having low frequency mentions for more distant locata, while in contrast Hyde Park only uses *next to* with locata that are relatively close (up to 56m). For smaller distances, the *adjacent to* and *beside* preposition graphs are similar to those of *next to*. Both are used with Buckingham Palace for distances up to 250m, but the most frequent uses across the other two sites are less than 100m. We see a very similar pattern for *close to*, with Buckingham Palace attracting mentions out to approximately 1km, Trafalgar Square to 660m (albeit very low frequency) and Hyde Park to 123m. We consider it likely that the larger distances associated with Buckingham Palace are influenced by the ambiguity in the specific size/area of the relata: the entire grounds, or only the palace building itself. In this analysis, we used the entire grounds, since public access to the grounds and palace is limited, so it is less likely that mentions would refer only to the palace. Our comparative analysis confirmed this, with calculations based on the use of the palace geometry alone returning less consistent results than for the combined palace and grounds.

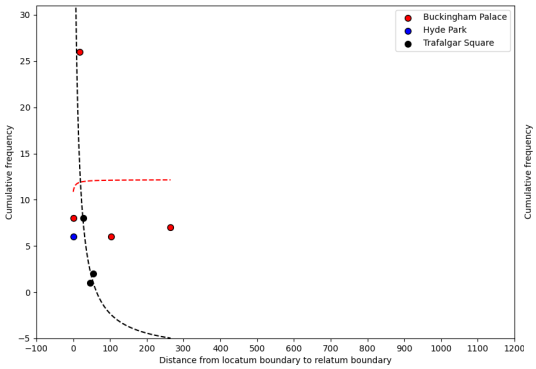
Despite the differences among the three sites discussed above, the highest frequency uses of *close to* and *next to* (the point at which the steeper sections of the graph level out being less than 100m in both cases) are consistent across all three sites, suggesting that context has limited impact on the most common uses of these prepositions. While the *near* graph is similar to that of *close to*, it does show some variation between the three relata: namely the absence of high frequency use at very low distances for Buckingham Palace and Hyde Park. It appears that while *near* may be used for Trafalgar Square for locations very close to (or at) the Square, this does not apply to the other two relata, in contrast to *close to*, which



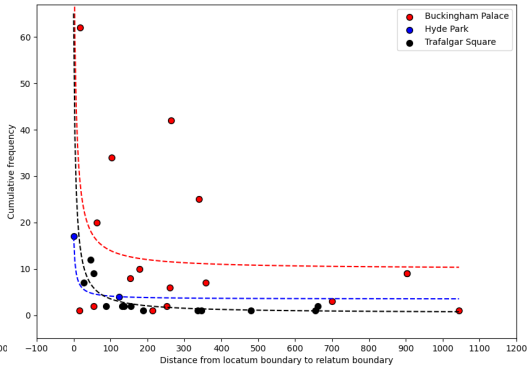
(a) adjacent to.



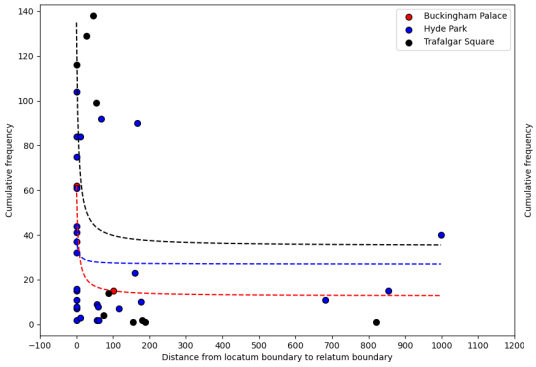
(b) at.



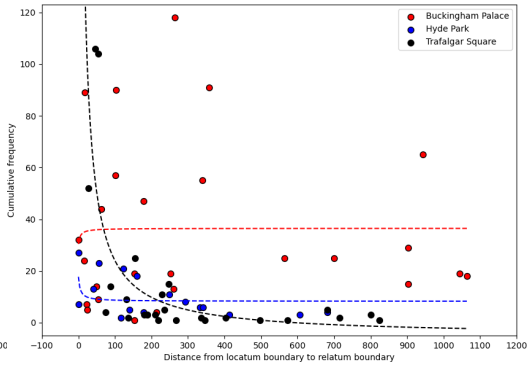
(c) beside.



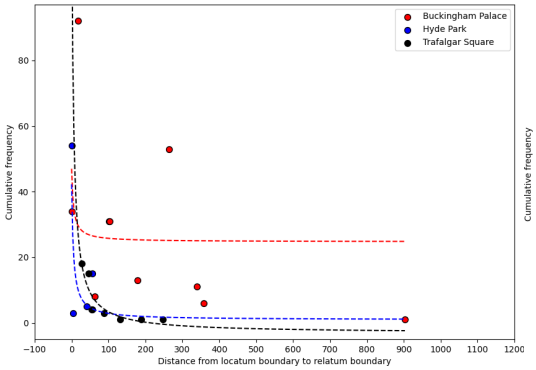
(d) close to.



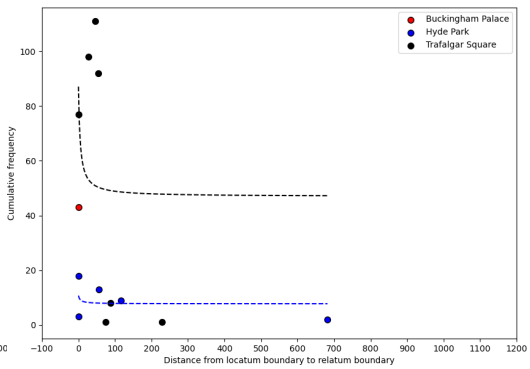
(e) in.



(f) near.



(g) next to.



(h) on.

Figure 4 Scatter plot and regression lines for the frequency of each preposition across all three sites.

is applied at very small distances for all three relata. Trafalgar Square is a smaller feature, with arguably vaguer boundaries, than the other two relata, which may explain the more liberal use of *near* in that case.

The *in* preposition shows similar patterns across all three sites. Hyde Park and Trafalgar Square have a much greater range of mentions (going up to 1km), but all of the curves flatten at a distance within approximately 50m of the relata. It is interesting to note that, counterintuitively, the *in* preposition is used relatively frequently with locata much further from the relatum than the *next to* and *close to* prepositions, and clearly for objects that are well outside the boundary of the relatum. For example, "*Baglioni Hotel in Hyde Park*" is 700 metres away from the boundary of Hyde Park.

Like *in*, the *at* preposition shows flattening at distances very close to the relatum, but this distance is greater for Buckingham Palace than for the other two relata. This may be due to the closed nature of Buckingham Palace (public access is strictly controlled, being limited in timing, volume of visitors and area of access, and requiring payment) compared to the other sites. Thus, the description *I'm at Buckingham Palace* could mean that the speaker is outside the Palace gates, while this is less likely (but still possible) for the other two relata. *At* is used at a much greater distance for Hyde Park than for the other two relata, but this may be related to the Kensington Gardens effect described above.

Use of the *on* preposition is much more frequent for Trafalgar Square than for Hyde Park and is only used once for Buckingham Palace in our data set, with distance zero. This is likely the result of image schema, with squares and plazas being more frequently associated with a platform schema than parks or palaces ([23, 24]). However, we do not see a similar pattern for the *in* and *at* prepositions, which are commonly used for parks and similar types of objects. Trafalgar Square is frequently used with *in*, *at* and *on*, suggesting that a range of different image schemata are suitable for this feature type, while parks and palaces are more limited.

Across all of the prepositions, we see much greater variation between relata at the outer extremes of acceptability of the prepositions. That is, many of the prepositions studied are used less frequently for quite large distances for some sites more than others; while the most frequent uses are much more uniform across the sites, despite the differences in size, feature type and level of urban construction among the relata. Generally, the proximity (*near*, *close to*) and adjacency (*next to*, *adjacent to*, *beside*) prepositions are more frequently used for greater distances for Buckingham Palace than the other two relata; while *in*, *on* and *at* are most frequently used for Trafalgar Square across all distances. Like *in*, the *at* preposition shows flattening at distances very close to the relatum, but this distance is greater for Buckingham Palace than for the other two relata.

5 Discussion

5.1 Acceptance thresholds (RQ1)

Our findings show that among the proximity/adjacency prepositions, *near* and *close to* have the highest acceptance threshold distances from the relatum with *near* > 1100m and *close to* = 903m. *Next to* and *adjacent to* were used for distances between 55-358m, and the smallest threshold for proximity/adjacency prepositions was for *beside* (17m for the Buckingham Palace) but there was not enough data on the other two sites to confirm this finding. Carlson and Covey [1] ran a human subjects experiment to estimate the distance associated with some spatial prepositions such as *next to*, *beside* and *near*. Similar to our findings, their research showed that *beside* and *next to* are not associated with relatum size and these prepositions

specify smaller distances than *near*. Fisher and Orf [7] also reviewed the interpretation of *near* and *close* in a university campus area and found that people did not use these terms for buildings that were very close to the student centre (relatum), but instead those that were further away. Hall et al.'s [11] density models of spatial prepositions such as *near*, *between*, *at the corner*, *at* and *next to* also show a larger distance for *near* and smaller distances for *at* and *next to*. Skoumas et al. [32] also confirmed that the use of *near* is not restricted only to locations close to the relatum, while *at* and *next to* were used for areas close to the relatum.

5.2 Contextual factors (RQ2)

The analysis in Section 4.2 highlights a number of observations regarding contextual variations in the use of spatial prepositions. Firstly, we note that *near* is used less frequently for locata very close to the relatum for two of the three relata, the exception being Trafalgar Square. This 'doughnut effect' was less evident in the analysis of Hall et al. [9], but was identified by Fisher and Orf [7] who claimed that this might be due to the similarity in place names or functions. More research is needed to confirm this.

Secondly, we noted in Section 4.2 the likely impact of image schema on the use of the containment and collocation prepositions. This has been identified by other researchers, who identified the use of the *on* preposition when a platform schema is used, or *in* for a container schema ([23, 24]). However, our results identify a variation in image schemata applied to different feature types. Some of our relata were strictly subject to a single image schema (e.g., Hyde Park with the container schema, indicated by the use of the *in* preposition in preference to *at* or *on*), while others were more promiscuous (notably Trafalgar Square, which uses all three of these prepositions liberally, suggesting platform, container and possibly other image schemata such as link are appropriate).

Thirdly, we note that the *outside* preposition is only used with Buckingham Palace in our data (see Figure 2(a)). While the *outside* preposition would normally be associated with the container image schema, we also see low frequency use of *in* (also associated with the container image schema) with Buckingham Palace compared to the other two relata. However, this may be due to the access limitations reducing the frequency of mentions for Buckingham Palace. In contrast, *in* is the most frequently used containment or collocation preposition used for Hyde Park, but *outside* is not used for this relatum in our data. This suggests that *outside* requires a stronger form of containment than *in*, with Buckingham Palace being a stronger container than Hyde Park.

In addition to the influence of generalisable characteristics of different relata on the acceptance profiles and thresholds of prepositions, our data shows that individual contexts can influence the use of prepositions, as in the case of Hyde Park and the likely influence of the neighbouring Kensington Gardens. This suggests that general models of preposition applicability, even if they are able to incorporate a rich range of contextual factors such as feature type, accessibility or image schema, are still likely to be limited in accuracy, as they are unable to capture these individual nuances. Also in the Hyde Park/Kensington Garden example, the influence of familiarity on the use of prepositions and the associated selection of locata to describe a location, reported in [41], is confirmed by our research. Thus people appear to use the reference objects whose names they are more familiar with, and this may influence the acceptance profiles and thresholds of prepositions in specific contexts.

These results also show that while there are distances that are acceptable for a given preposition across all three of our relata, which we might refer to as "ranges of agreement", there are also outer extremes of those ranges that are only used in certain circumstances, for only one or two relata, and depending on context. Both Hyde Park and Buckingham Palace

have data points at much greater distances than Trafalgar Square, usually with low frequency (for example, the *next to* preposition is used at around 900m for Buckingham Palace), but with greater frequency at much lower distances for all three relata. We thus consider it likely that the acceptance areas for geospatial prepositions follow prototype theory [30] in having a range of exemplars, and outliers, which are at least partly determined by context.

5.3 Locatum popularity

In addition to the impact of relata on the selection of appropriate geospatial prepositions, we acknowledge the impact of locatum popularity on our dataset. It is likely that people chose the most popular or salient locata when describing a scene, in favour of less noticeable objects. While this may result in bias, since we are looking at the same three relata for all prepositions, the comparison between the relata is still valid, and we see that for a given locatum, a mixture of different prepositions is selected, rather than a single preposition. It would however be of interest in future work to examine more closely the possible influence of variation in popularity of the locata as well as variation in their properties such as size and type.

6 Conclusion

In this paper, we investigated the acceptable distances for a set of prepositions used in geospatial situations, and the impact of context on those distances. We used Google searches to extract locatum, preposition and relatum triples for three sites in London (Trafalgar Square, Buckingham Palace, and Hyde Park) and analysed the frequency of mentions of specific locata with each preposition and relatum. Our experiment led to a number of findings. Firstly, proximity/adjacency geospatial prepositions such as *near* and *close to* are used for larger distances than *adjacent to*, *next to* and *beside*. *Adjacent to* and *next to* are used for relatively large distance ranges, while *beside* is confined to small distances. Also, *in*, *on* and *at* are not only used for locata that are inside the relata, but sometimes when there is a short distance between locata or relata (depending on the context). Secondly, degree of adherence to a particular image schema varies depending on feature type: some use a single schema, while others use a variety depending on context. Thirdly, the use of the *outside* preposition relies on a notion of strength of containment which may depend on other contextual factors such as accessibility.

It is acknowledged that this study is limited in being confined to three relata and to scales relating only to an urban space. Thus future work is needed to consider a wider range of relata within different geo-spatial contexts. However there are challenges in such work with regard to obtaining adequate volumes of good quality data for individual spatial prepositions when used with multiple occurrences of specific combinations of named locatum and relatum as was done here. As part of future studies it would also be of interest to investigate relationships between locatum feature type and acceptance distance, as well as the relationship between image-schema and feature types and flexibility in the use of image-schema by a given feature type.

References

1 L. A. Carlson and E. S. Covey. How far is near? inferring distance from spatial descriptions. *Language and Cognitive Processes*, 20(5):617–631, 2005.

2 A. Chang, M. Savva, and C. D. Manning. Learning spatial knowledge for text to 3d scene generation. In *EMNLP*, pages 2028–2038, 2014.

3 H. Chen, S. Winter, and M. Vasardani. Georeferencing places from collective human descriptions using place graphs. *Journal of Spatial Information Science*, 17:31–62, 2018.

4 G. Collell, L. Van Gool, and M.-F. Moens. Acquiring common sense spatial knowledge through implicit spatial templates. In *Proc. AAAI 2018*, 2018.

5 Curdin Derungs and Ross S. Purves. Mining nearness relations from an n-grams web corpus in geographical space. *Spatial Cognition & Computation*, 16(4):301–322, 2016. `doi:10.1080/13875868.2016.1246553`.

6 S. Du, X. Wang, C. Feng, and X. Zhang. Classifying natural-language spatial relation terms with random forest algorithm. *Int. J. Geographical Information Science*, 31(3):542–568, 2017.

7 P. F. Fisher and T. M. Orf. An investigation of the meaning of near and close on a university campus. *Computers, Environment and Urban Systems*, 15(1-2):23–35, 1991.

8 N. Gronau, M. Neta, and M. Bar. Integrated contextual representation for objects' identities and their locations. *Journal of Cognitive Neuroscience*, 20(3):371–388, 2008.

9 M. Hall, P. Smart, and C.B. Jones. Interpreting spatial language in image captions. *Cognitive Processing*, 12(1):67–94, 2011.

10 M. M. Hall and C. B. Jones. Quantifying spatial prepositions: an experimental study. In *Proc. 16th ACM SIGSPATIAL*, pages 1–4, 2008.

11 M.M. Hall, C.B. Jones, and P. Smart. Spatial natural language generation for location description in photo captions. In *COSIT*, pages 196–223. Springer, 2015.

12 Annette Herskovits. Semantics and pragmatics of locative expressions. *Cognitive science*, 9(3):341–378, 1985.

13 Y. Hu and J. Wang. How do people describe locations during a natural disaster: an analysis of tweets from hurricane harvey. *arXiv preprint*, 2020. `arXiv:2009.12914`.

14 Arthur J Jelinek. Use of the cumulative graph in temporal ordering. *American Antiquity*, 28(2):241–243, 1962.

15 M. Johnson. *The body in the mind: The bodily basis of meaning, imagination, and reason*. University of Chicago Press, 2013.

16 E. Kamalloo and D. Rafiei. A coherent unsupervised model for toponym resolution. In *Proc, WWW Conference*, pages 1287–1296, 2018.

17 Morteza Karimzadeh. Performance evaluation measures for toponym resolution. In *Proc. GIR Workshop*, pages 1–2, 2016.

18 T. Kew, A. Shaitarova, I. Meraner, J. Goldzycher, S. Clematide, and M. Volk. Geotagging a diachronic corpus of alpine texts: Comparing distinct approaches to toponym recognition. In *Proc. Workshop on Language Technology for Digital Historical Archives*, pages 11–18, 2019.

19 Ti. Lan, W. Yang, Y. Wang, and G. Mori. Image retrieval with structured object queries using latent ranking svm. In *European Conf. Computer Vision*, pages 129–142. Springer, 2012.

20 Jochen L Leidner. *Toponym resolution in text: Annotation, evaluation and applications of spatial grounding of place names*. Universal-Publishers, 2008.

21 M. D Lieberman and H. Samet. Adaptive context features for toponym resolution in streaming news. In *Proc. ACM SIGIR*, pages 731–740, 2012.

22 M. Malinowski and M. Fritz. A pooling approach to modelling spatial relations for image retrieval and annotation. *arXiv preprint*, 2014. `arXiv:1411.5190`.

23 D.M. Mark. Cognitive image-schemata for geographic information: Relations to user views and gis interfaces. In *GIS/LIS*, volume 89 (2), pages 551–560, 1989.

24 D.M. Mark and A.U. Frank. Experiential and formal models of geographic space. *Environment and Planning B: Planning and Design*, 23(1):3–24, 1996.

25 Daniel R. Montello. Scale and multiple psychologies of space. In Andrew U. Frank and Irene Campari, editors, *Spatial Information Theory A Theoretical Basis for GIS*, pages 312–321, Berlin, Heidelberg, 1993. Springer Berlin Heidelberg.

26 R. Moratz and T. Tenbrink. Spatial reference in linguistic human-robot interaction: Iterative, empirically supported development of a model of projective relations. *Spatial cognition and computation*, 6(1):63–107, 2006.

27 D.G. Morrow and H.H. Clark. Interpreting words in spatial descriptions. *Language and Cognitive Processes*, 3(4):275–291, 1988.

28 G. Platonov and L. Schubert. Computational models for spatial prepositions. In *Proc. Workshop on Spatial Language Understanding*, pages 21–30, 2018.

29 Leonard Richardson. Beautiful soup documentation. *April*, 2007.

30 E. Rosch. Cognitive representations of semantic categories. *J. Experimental Psychology*, 104(3):192, 1975.

31 S. Schockaert, M. De Cock, and E. E. Kerre. Location approximation for local search services using natural language hints. *International Journal of Geographical Information Science*, 22(3):315–336, 2008. doi:10.1080/13658810701626277.

32 G. Skoumas, D. Pfoser, A. Kyrillidis, and T. Sellis. Location estimation using crowdsourced spatial relations. *ACM Trans. Spatial Algorithms and Systems*, 2(2):1–23, 2016.

33 K. Stock and M. Hall. The role of context in the interpretation of natural language location descriptions. In *COSIT*, pages 245–254. Springer, 2017.

34 K. Stock and J. Yousaf. Context-aware automated interpretation of elaborate natural language descriptions of location through learning from empirical data. *Int. J. Geographical Information Science*, 32(6):1087–1116, 2018.

35 L. Talmy. How language structures space. In *Spatial Orientation*, pages 225–282. Springer, 1983.

36 A. Tyler and V. Evans. *The semantics of English prepositions: Spatial scenes, embodied meaning, and cognition*. Cambridge University Press, 2003.

37 M. Vasardani, L.F. Stirling, and S. Winter. The preposition at from a spatial language, cognition, and information systems perspective. *Semantics and Pragmatics*, 10:3, 2017.

38 Jan Oliver Wallgrün, Alexander Klippel, and Timothy Baldwin. Building a corpus of spatial relational expressions extracted from web documents. In *Proceedings of the 8th workshop on geographic information retrieval*, pages 1–8, 2014.

39 D. Wu and Y. Cui. Disaster early warning and damage assessment analysis using social media data and geo-location information. *Decision support systems*, 111:48–59, 2018.

40 Z. Wu and M. Palmer. Verb semantics and lexical selection. *arXiv preprint*, 1994. arXiv:cmp-lg/9406033.

41 X. Yao and J.-C. Thill. How far is too far?–a statistical approach to context-contingent proximity modeling. *Trans. in GIS*, 9(2):157–178, 2005.

42 H. Yu and J.M. Siskind. Sentence directed video object codiscovery. *Int. J. Computer Vision*, 124(3):312–334, 2017.

I Can Tell by Your Eyes! Continuous Gaze-Based Turn-Activity Prediction Reveals Spatial Familiarity

Negar Alinaghi
Geoinformation, TU Wien, Austria

Markus Kattenbeck
Geoinformation, TU Wien, Austria

Ioannis Giannopoulos
Geoinformation, TU Wien, Austria
Institute of Advanced Research in Artificial Intelligence (IARAI), Wien, Austria

Abstract

Spatial familiarity plays an essential role in the wayfinding decision-making process. Recent findings in wayfinding activity recognition domain suggest that wayfinders' turning behavior at junctions is strongly influenced by their spatial familiarity. By continuously monitoring wayfinders' turning behavior as reflected in their eye movements during the decision-making period (i.e., immediately after an instruction is received until reaching the corresponding junction for which the instruction was given), we provide evidence that familiar and unfamiliar wayfinders can be distinguished. By applying a pre-trained XGBoost turning activity classifier on gaze data collected in a real-world wayfinding task with 33 participants, our results suggest that familiar and unfamiliar wayfinders show different onset and intensity of turning behavior. These variations are not only present between the two classes –familiar vs. unfamiliar– but also within each class. The differences in turning-behavior within each class may stem from multiple sources, including different levels of familiarity with the environment.

2012 ACM Subject Classification Computing methodologies → Activity recognition and understanding; Computing methodologies → Supervised learning by classification

Keywords and phrases Spatial Familiarity, Gaze-based Activity Recognition, Wayfinding, Machine Learning

Digital Object Identifier 10.4230/LIPIcs.COSIT.2022.2

Supplementary Material *Dataset (train and test)*: `https://doi.org/10.48436/f0chy-11p06`

Acknowledgements We would like to thank Antonia Golab (Vienna University of Technology) for collecting the valuable data used for this work.

1 Introduction

Wayfinding as a general concept is outlined as "the most prominent real-world application of spatial cognition" [40]. It is an ongoing decision-making process consisting of several tasks, each of which requires cognitive resources [3]. The cognitive demand for each task heavily relies on environmental and user-related features. Theoretical reasoning (e.g. [40, 14, 15]) and empirical evidence (e.g. [31, 23, 25, 12] suggest that *familiarity with the environment* as a particular state of spatial cognition plays an important role in the wayfinding decision-making process. Our previous research on the prediction of wayfinders' turning activity, as an actual realization of a series of spatial decisions, introduced *familiarity* as the most important/ influential feature for the prediction model [2].

OSM Open Street Map
ML Machine Learning
POI Point of Interest

Motivated by our previous finding [2], presenting *familiarity with the environment* as the most important feature in turning activity prediction, we decided to more thoroughly look into this feature during the decision-making process for the turning activity. After receiving any instruction, wayfinders try to match the conveyed spatial information to the physical environment and decide for their turning action (i.e., whether to turn left/right or continue straight ahead). Trying to understand the undergoing cognitive processes in this *matching-to-action* phase may reveal aspects of spatial cognition, including spatial familiarity. Understanding the behavioral correlates of familiarity in this phase of decision-making is a valuable theoretical contribution in the spatial familiarity domain and provides a fertile ground for cognitively engineered spatial systems, e.g., more context-aware navigation aid systems.

In this paper we take a step into this direction: We demonstrate that familiar and unfamiliar wayfinders, exhibit differing behaviors while reacting on navigational instructions to decide whether to turn or continue straight (this we call "turning-activity"). On this account, we monitor the wayfinders' gaze behavior in relation to their turning-activity, i.e., during the matching-to-action phase of the decision-making process, using a mobile eye tracker and a high-precision GNSS receiver. We report on an in-situ pedestrian wayfinding study ($N = 33$), during which participants walked two routes: one located in an area with which they were familiar, whereas another route located in an area they were unfamiliar with. The familiarity with the environment was reported by participants in a multi-step procedure online study prior to the in-situ experiment (see Section 3.1 for more details). We adapted the pre-trained XGBoost model taken from our previous research [2] and applied it to the gaze behavior of these two groups of wayfinders, within the matching-to-action phase.

Our analysis provide evidence that detecting gaze-behavioral differences concerning the turn activity reveals the familiarity of the wayfinders as a binary measure. In addition to that, as the distribution of probabilities predicting turn vs. no turn activities varies within both groups, familiar and unfamiliar participant, alike, our findings may lead to a potential proxy for the estimated level of familiarity.

2 Related Work

Given the aforementioned research goals, we first review previous findings on familiarity aspects of wayfinding and continue with wayfinding studies that examine gaze behavior during decision-making. Finally, we report on the prior work on using Machine Learning (ML) methods for activity classification based on eye-tracking data.

2.1 Familiarity Aspects in Wayfinding and Beyond

"We reason that a navigator's search behavior and search strategy will be heavily influenced by their degree of familiarity with the environment." [40]. Investigating the effect of familiarity on wayfinding and environmental learning, as well as understanding the variables that contribute to the development of a sense of familiarity, constitute the majority of theoretical research in the domain of spatial familiarity [1]. Up until now, however, we lack a comprehensive conceptual definition for "*spatial familiarity*" [15]. As a consequence, assessing and modeling familiarity continue to be open research problems. The majority of attempts taking familiarity into consideration rely on self-reported measures, e.g., customized questionnaires and sketched maps (see e.g. [6, 19]) and these studies most often use a binary definition of environmental familiarity (see e.g., [43, 25, 2]). Approaches exploiting behavioral data benefit from various sources ranging from mobile phone GPS tracks (see.

e.g., [37]) to social media (e.g., [38]) and recently also gaze-behavior ([25]). Having said this, empirical studies addressing environmental familiarity have been conducted both indoors [18], outdoors [25]) and in virtual environments [12], alike.

Only recently, has the classification of binary familiarity based on behavioral correlates gained momentum. For instance, Gokl and colleagues [12], reported 51.87% to 65.70% accuracy for a binary classification on familiarity performed on behavioral data collected from an avatar-based VR study. Savage and colleagues in [34], used place-visits in Foursqure application and combined it with user-contextual information taken from Facebook profiles and GPS trajectories to score the level of familiarity with places using Bayesian techniques. Liao and colleagues in a very recent paper [25], performed a Random Forest binary classification (data was initially collected on a Likert-like scale) for familiarity using the gaze-behavior collected in a real-world navigation task with 38 participants and reported 70% to 81% accuracy.

2.2 Gaze-based Activity Recognition

As one of the most promising ways of gaining access to people's cognitive state, gaze, has received a great deal of attention in the GIScience research domain (see [21, p. 2–9] for an overview): Its utility has been explored in many applications, for instance, tailoring wayfinding assistance systems to individual's demands (see e.g. [11]), and numerous wayfinding studies conducted in real-world or virtual reality environments ([35], [8] and [39]).

Having said this, the importance of gaze in task prediction has seen great interest for decades, starting with the seminal works of Yarbus [41] and Buswell [5]. For instance, Kiefer and colleagues [20] performed gaze analysis for automatic recognition of map users' activities by introducing novel gaze features based on string sequence analysis, and reported 78% accuracy in classifying six common map activities (free exploration, global and focused search, route planning, line following, and polygon comparison). Bulling and colleagues [4], also applied a Support Vector Machine (SVM) on gaze features for recognizing five office-related activities (e.g. copying a text, reading a printed paper etc.) in an eight-participant study and reported an average precision of 76.1%. In addition to task prediction, researchers have tried to predict the underlying cognitive or even demographic aspects which are reflected in various tasks. For instance, Henderson and colleagues [17], tried to predict the cognitive workload of 12 participants performing various tasks e.g., scene memorization, text reading etc., using a multivariate pattern classification on the recorded eye movements with 68% to 89% accuracy. Galdi and colleagues [9], studied the applicability of gaze analysis for gender and age categorization by applying Adaboost and SVMs on data collected from 112 different observers.

Similar recent research in our domain includes the work of Alinaghi and colleagues [2]. The authors trained an XGBoost classifier on gaze data acquired from a real outdoor wayfinding study. They reported the best performing model (with 91% of accuracy for predicting the turning activity of wayfinders with three classes of No-Turn, Turn-Left, and Turn-Right) trained with the gaze data from the last three seconds before the turning action was performed. Liao and colleagues [24] reported an accuracy of 67% using a Random Forest classifier on 38 recorded eye movements during navigation, to predict five tasks, namely self-localization, environment target search, map target search, route memorization, and walking to a destination.

2.3 Machine Learning Models for Activity Recognition

The related work reviewed so far, already indicates the prevalence of SVMs and tree-based ML techniques in performing classification tasks based on eye movement data. For instance, Bulling and colleagues [4], and Shiga and colleagues [36] both used SVMs for classification of general office tasks and everyday tasks based on gaze data and reported reasonable results. Liao and colleagues reported promising results by applying Random Forest both for task recognition ([24]), and familiarity prediction ([25]). Liu and colleagues' results from Random Forest in [26], where they examined differences in gaze behavior related to the regularity of road patterns and signage, are also encouraging.

Although SVMs and tree-based techniques provide promising results in gaze-analysis, recent studies report higher accuracies achieved by ensemble learning, e.g., Gradient Boosted Trees, across different domains. For instance, in their gaze-based turning activity prediction, Alinaghi and colleagues [2], compared SVM-RBF, CART, and Random Forest in a pilot testing phase and reported that XGBoost achieved higher accuracy. Similarly, Zhang and colleagues [42] provide evidence that XGBoost-based indoor activity recognition algorithm outperforms both other ensemble learning classifiers and single classifiers in indoor activity classification (achieving 84.41% accuracy). In a different domain, Mathur and colleagues [27] found XGBoost superior to eight ML models as well as two deep learning models (long short-term memory networks and temporal convolutional networks), with an accuracy of 69% and an AUC of 72% when detecting users' empathy while listening to a narrative robot.

3 Detecting Familiarity Based on Turn Activity

With a similar approach as [17, 9], where trained models for specific task recognition problems were deployed to predict user-related characteristics, we have also used a pre-trained turning-activity classifier for familiarity detection. This section provides information on our analysis for familiarity detection from turning activity behavior. We start with a short description of the human-subject in-situ study as we use the same dataset as [2] which was collected to address several research questions. We, then, move on to explain our feature extraction approach. Finally, we describe the prediction model that we deployed for our familiarity detection and provide details on how we came upon the familiarity patterns by monitoring the turning activity behavior during the matching-to-action phase.

3.1 Data Collection

The data used in this paper is part of a larger data collection effort in 2020 addressing human spatial behavior in real-world wayfinding scenarios and is first presented in [13][1]. In a within-subject design study, participants were required to walk two routes each, one of which was located in an area they were familiar with, while they were unfamiliar with the other. Environmental familiarity was collected during an online study: Participants were asked to indicate areas in which they would be able to find their way easily without any kind of wayfinding assistance This was used as proxy for being familiar. Subsequently, participants were asked to pinpoint familiar places therein. Two of these places were randomly chosen based on the condition of being between 0.9 and 1.3 km apart in order to ensure a reasonable duration of the experiment, and participants provided their preferred route between these

[1] Parts of the data used in the current paper, will be made available at: `https://geoinfo.geo.tuwien.ac.at/resources/` (DOI: `10.48436/f0chy-11p06`).

places. All participants walked the familiar route they provided and were randomly assigned an unfamiliar route which was located in a polygon with no overlaps with the polygons a participant indicated as familiar.

For either route, participants had to find their way by means of auditory, German-language, landmark-based[2], turn-by-turn route instructions, which were provided to them on demand and as many times as they requested. In order to ensure the demand of route instructions when walking a familiar route, participants were instructed that they might be guided on a route different to the one they actually provided. Participants' behavior during the experiment was monitored with a mobile eye-tracking device (PupilLabs Invisible; 200Hz recording frequency), high precision GNSS receiver (PPM 10-xx38) that tracked their position in time and a head-mounted Inertial Measurement Unit (IMU), whose data we do not use in this study. Additionally, participants were given a small clicker-device, to request an instruction by simply pressing a button. In total, $N = 52$ people (27 female and 25 males, $M(age) = 26$ years, $SD(age) = 8.3$) participated in the outdoor experiments resulting in $N = 104$ trials out of which $N = 32$ were retained for further processing in this paper (18 trials: equipment malfunction or participants not cooperating; 54 trials: not suitable given our research question, see Section 3.3 for details).

3.2 Data Preparation

To obtain the gaze-related data at each decision-point, we first matched each of the Open Street Map (OSM)-driven junctions along a route to their corresponding GPS point by drawing a ray oriented along the segments of the given intersection and intersecting it with the GPS trace. As the result of this step, we obtained the corresponding timestamps at which each junction was reached.

Subsequently, we approximated the *Start* points (i.e. *when an instruction was given for the first time*) and *End* points (i.e. *when a junction, for which the instruction was given, was perceived as junction*) of the matching-to-action phase in our analyses. From now on, we refer to each pair of these points as *test-sample*. Figure 1 A, visualizes these points. Start points were found by synchronizing the clicker-output with GPS timestamps. This point is depicted as a green speaker icon in Figure 1 A. The yellow icons in this figure indicate the points in time when the same instruction was requested for a second time (again, based on the synced clicker and GPS timestamps). Since the perception of decision *points* in reality do not exactly equate to the junction point (e.g., we do not decide to turn left to a street right in the middle of a crossing), we model the decision *situations* by using OSM building footprints to find all intersection boundaries[3]. These boundaries are presented in Figure 1 A, as hatched polygons located at each junction (orange: *No-Turn*; blue: *Turn*). From now on, we will refer to the intersection point of the GPS trace with this boundary as *decision-point*. Figure 1 A represents these decision-points with green and blue circles denoting the start and end of each segment in the test-sample.

[2] Points of Interest (POIs) were used as landmarks and chosen according to the algorithm described in [33]. For a detailed description on how these were double-checked, see [13]. Route instruction pattern: Turn left [imperative] at Cafe Ritter. [landmark and, hence, location of turning action].

[3] If no buildings were located at a junction, we found the boundary of the intersection by using a threshold of $3.75secs$ from the projected junction point (i.e. $\approx 5m$, based on $4.5km/h$ avg. walking speed [22].

3.3 Sliding-Window Feature Extraction

As the highest accuracy for the XGBoost turn-activity classifier in our previous research [2] was reported when trained three seconds before the actual turn occurred, we selected three seconds as our window size. The step distance was also set to three seconds, avoiding any overlaps between windows. This is based on findings suggesting that non-overlapping windows deliver comparable recognition accuracy while majorly reducing the required training computations and memory usage [7].

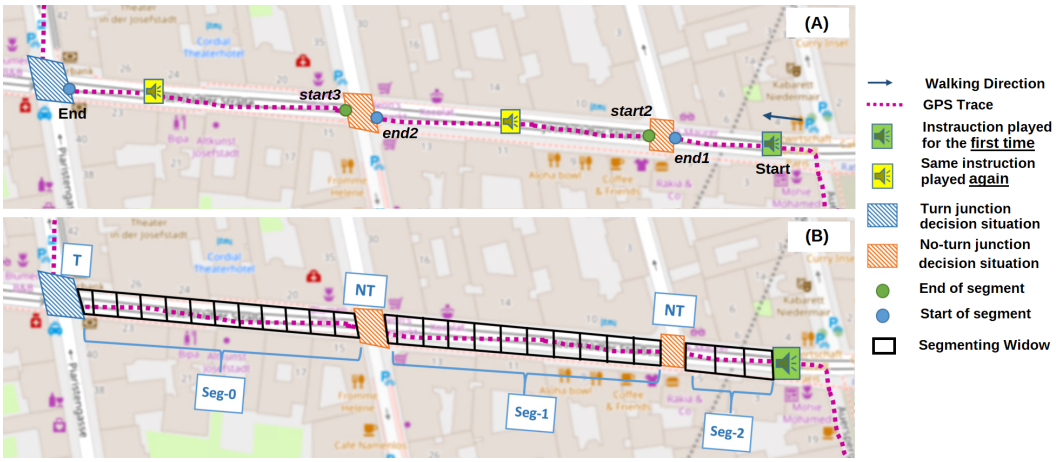


Figure 1 Figure **A**, representing one *test-sample*, visually defines the *Start* and *End* points of the test-sample as well as the middle *start(i)* and *end(i)* points of the segments. As schematically presented in **B**, each test-sample is then segmented with a sliding window approach into non-overlapping windows of three seconds duration. Since we want to analyze the results segment-wise, we group the windows belonging to one segment by segment id as *Seg-i*.

With the method explained in 3.2, we extracted all the *Start* and *End* points per turning junctions in all trials, with respect to two main conditions. First, as the experimental design allowed participants to ask for instructions as often as they needed, we decided to exclude all the repetitions and only keep the first point in time when the instruction was given. Second, test-samples should at least have one *No-Turn* junction (denoted in Figure 1 B as *NT*). This decision is made to have enough analysis time and more decision situations in the matching-to-action phase. This ensures that we always have at least two segments in each test-sample. The naming convention for identifying these segments is as follows: *Seg-**n*** with n starting from zero always indicating the final segment right before the *Turn* junction. This setting, i.e., starting each test-sample always from the initial instruction point and excluding the data before that, ensures a clean test-sample and is inline with the experimental design in which both un/familiar groups did not know the route in advance and had to relay on the given turn-by-turn instructions to reach the destination. Figure 1 B, represents one of these clean test-samples for which two *No-Turn* (*NT*) junctions were passed before the *Turn* (*T*) junction was reached. This results in three segments with ids starting from *Seg-0*, to *Seg-2* which is the first segment right after the instruction was given. After selecting the test-samples according to these conditions, we ran a non-overlapping sliding window (with $window - size = 3sec$ and $step - size = 3sec$) approach for segmenting the eye-tracking data. Figure 1 B, visually present this approach.

3.4 Detecting Familiarity by XGBoost Turning Activity Classifier

The trained XGBoost turning-activity prediction model from our previous research [2] is modified and used for familiarity detection purpose in this paper. XGBoost is one of the Gradient Boosted Tree algorithm implementations, which allows parallel tree boosting for unfolding very complex patterns in a highly efficient and scalable manner.

The model was trained on 31 features including 28 fixation- and saccade-based gaze feature (i.e., fixation frequency, min/max/mean/variance of fixation- duration/dispersion/dispersionX and dispersionY; and saccade frequency, min/max/mean/variance of saccade-amplitude/duration, skewness of saccade amplitude, and g-l ratio which is the ratio between long and short saccades), 2 environmental features (i.e., number and skweness of street segments at each junction) and 1 user-related feature (i.e., familiarity with the environment as a binary measure) for 1335 junctions acquired from the same dataset used in [2] (see Section 3.1). In that paper [2], the highest accuracy of turning-activity prediction (91.4%) was achieved when the model was trained with the data from the last three seconds before the turning action. By analyzing the SHAP values, we came upon *familiarity* as the most important feature for the model.

As explained in Subsection 3.3, we segmented the data within each test-sample into three seconds windows to use this model. To investigate the effect of familiarity on turning activity, we customized this pre-trained model in two ways: We modeled the turning activity as a binary measure: *Turn* vs. *No-Turn*, i.e. we did not distinguish between left and right turns. In addition to that, we turned the model into a *probabilistic classifier* in order to gain the probability distribution over the two classes. It has been suggested in statistics that such *posterior probabilities* are required "when a classifier is making a small part of an overall decision or the outputs need to be combined for the overall decision" [32]. Our following familiarity analysis also requires the investigation of these small parts of an overall classification decision.

Therefore, the design decisions presented here are based on our initial assumption that the expected variations in the posterior probabilities can be considered as a proxy of spatial familiarity regarding the matching-to-reaction phase of the turning activity. In order to visually inspect patterns for *familiar* and *unfamiliar* cases, we plotted these probabilities in the matching-to-action phase.

4 Results

This section provides the results of our familiarity detection based on turning activity behavior during the matching-to-action phase. The two subsections outline the results from three different angles: the primary outcome of our analysis representing the overall difference between *familiar* and *unfamiliar* wayfinders with regards to their turn-activity behavior is presented in subsection 4.1; subsection 4.2, however, sheds light on the results of a case-wise between class comparison (familiar vs unfamiliar) and the result of a case-wise within class comparison, alike.

4.1 Overall Familiarity Detection

Figure 2 represents the median percentage of windows in which a turn is predicted, aggregated along each segment across all trials. To analyze the computed probabilities across all cases, we aggregated the results by computing the median percentage of windows belonging to *Turn* class in each segment. This percentage represents the number of windows for which the

probability of class *Turn* was higher than the probability of class *No-Turn*. Since the segments (i.e., street segments belonging to different routes) within each test-sample have different lengths, we computed the percentages *segment-wise*. This ensures that the final result is normalized and comparable on one scale. When interpreting the figure, it is important to keep in mind that test-samples vary concerning the number of segments they have: For instance, depending on the route and the first time someone asked for instruction, one test-sample ended up having three segments while another included four. However, the second condition we had set for selecting test-samples assured the presence of at least two segments (i.e., *Seg-0* and *Seg-1*): As explained above, we had chosen the test-samples in a way such that at least one *No-Turn* junction was passed before reaching the *Turn* junction for which the turn-instruction was given. In Figure 2, the notation n represents the number of trials for which we have calculated the median percentage.

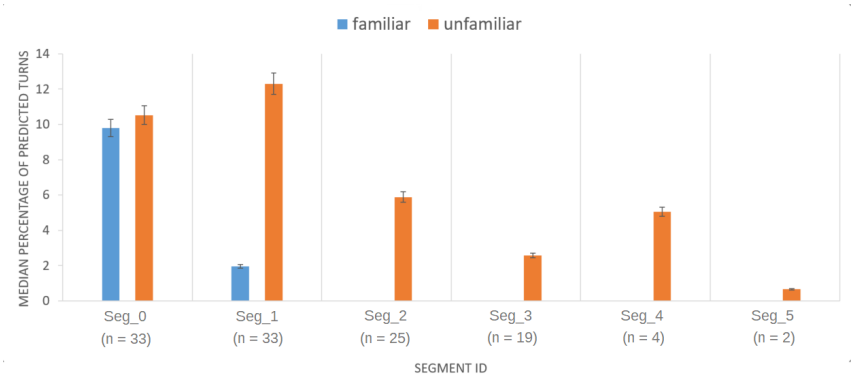


Figure 2 This plot represents the aggregated results, measured by the median percentage of windows per segment for which the class *turn* is predicted, for the two targeted groups: *Familiar* vs. *Unfamiliar* wayfinders. Note that not all the test-samples have all the segments and as a result the same number of time windows (due to the difference in the number of junctions and instruction-click points per route). The number of test-samples per segment is denoted by the letter n and error bars indicate the 95% confidence limit. The results suggest that familiar wayfinders are more confident in matching the turn instruction to the spatial environment.

4.2 Case-wise Comparisons: Between- and Within-Class

Figure 3 illustrates the resulting plots for a test-sample taken from one route walked by a familiar and an unfamiliar wayfinder. Since familiar and unfamiliar routes traveled by wayfinders differ in terms of the street network, length, urban complexity, etc., we compare turning behavior on a single route walked by two different wayfinders. The *x-axis* in Figure 3 represents multiple information for clearer understanding: the window-id, constellation of instruction-point (depicted in green) and junctions (distinguishing *Turn* and *No-Turn* junctions with blue and orange), and the segments outlined in purple for the last segment immediately before the turning junction and yellow for other segments. To visualize the turn behavior of each wayfinder, the probability of classes *Turn* (orange) and *No-Turn* (blue) within the matching-to-action phase are given for each window. The difference between familiar and unfamiliar wayfinders is apparent in this plot: The onset and the intensity of the variations in probabilities for the two plots are within different segments before the actual turn.

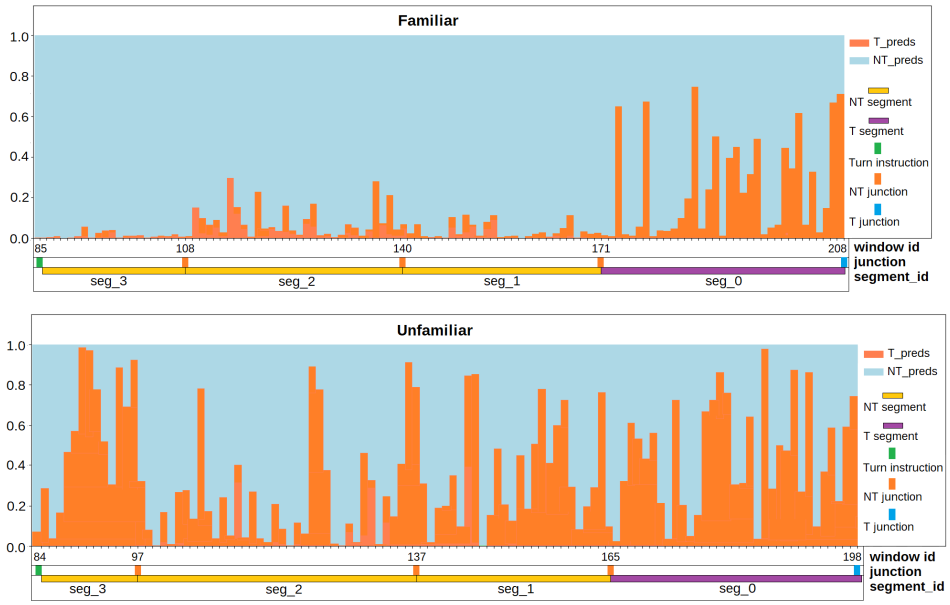


Figure 3 This figure contrasts the probability plots of one familiar and one unfamiliar wayfinder for a single route. The plots indicate that the variation in turning activity probability (i.e., turning-behavior) starts at different stages between the two classes: The familiar wayfinder demonstrates more variations only in the last segment before the turn-junction (*Seg-0*), while the unfamiliar wayfinder exhibit this behavior closely after receiving the instruction (*Seg-3*). Along the x-axis, a schematic view of the test-sample in terms of the constellation of instruction point, NT as well as T junctions is shown. Note that although the two diagrams represent the same route, they cannot be fully aligned due to different instruction-click times and wayfinders' walking speeds.

By investigating the probability plots per trial within each class (familiar and unfamiliar), we observed different distribution of turning-activity probabilities. The onset and intensity of turning behavior are equally different within familiar and unfamiliar wayfinders. As the in-situ experiment was designed so that different un/familiar wayfinders walked different routes, disentangling the impact of the route and individual differences might not be feasible. Coincidentally, however, one route was walked by two different unfamiliar wayfinders. This gave us a single case where the environment is fixed, and we could relate the patterns of turning behavior to individual differences. Figure 4, illustrates the probability plots for this case. The plots show differences concerning the distribution of the turning activity probabilities even within one familiarity class. A similar pattern is also present in both familiar and unfamiliar classes when different routes are considered. This is particularly remarkable for familiar cases: While the vast majority of cases showed a variation of probabilities only on the final segment, there were also cases showing these variations sooner, e.g., in *Seg-1* or *Seg-2*.

5 Discussion

We analyzed turning activity behavior in the *matching-to-reaction* phase, which starts right after an instruction is given until the corresponding junction is reached. We interpret the variations in turn probabilities as an indicator of *turning-behavior* stemming from the given instruction. We provide evidence that familiar wayfinders show a different behavior during

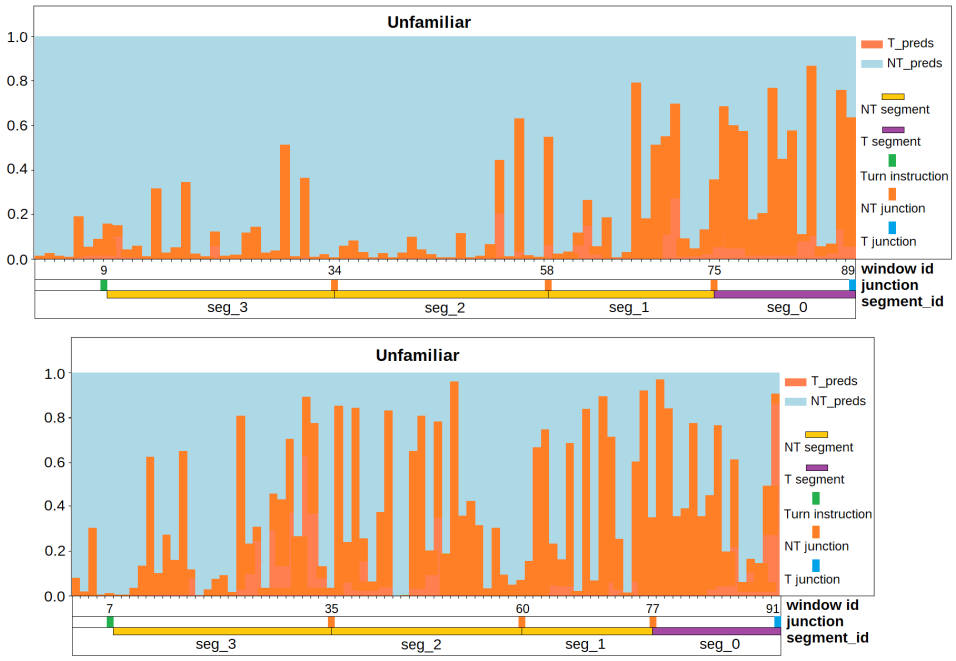


Figure 4 This figure shows the probability plots of an exceptional case in which two unfamiliar wayfinders coincidentally walked a single route. The plot indicates that the variation in turning activity probability (i.e., turning-behavior) starts at different stages. Along the x-axis, a schematic view of the test-sample in terms of the constellation of instruction point, NT as well as T junctions is shown.

the matching-to-reaction phase (high median percentage of predicted turns only in the final segment before the turning junction) of the decision-making compared to unfamiliar wayfinders (higher rate of turning behavior compared to the familiar group in almost all segments once an instruction was received).

While familiar and unfamiliar wayfinders can be distinguished reasonably well based on their gaze behavior before turns, our data suggest considerable within-class diversity. This was revealed when the results within each class were analyzed case-wise. Within both familiar and unfamiliar groups, no two cases represented the turn-behavior precisely at the same time or with the same intensity. We discuss three different potential explanations:

Spatial Environments: Each route was walked twice by one familiar and one unfamiliar person, i.e., there were 16 different routes considered in this study. Differences in within-class cases might, hence, stem from the environmental differences among the routes, e.g., urban configuration, segment (e.g., length) and junction characteristics e.g., number of segments, etc.), POIs, etc.

Levels of Familiarity: Theoretical reasoning [28, 10] suggests that there are different levels of spatial familiarity. Consequently, differences in turning-behavior during the matching-to-action phase may be considered as a potential indicator of different levels of familiarity. However, collecting ground truth data for these levels is still an open research question. Researchers tend to use either the number of years a person has lived in a city as an indicator (see e.g. [30]) or collect level of familiarity data using custom-designed questionnaires (see, e.g. [25] who collected self-report familiarity ratings on a 7-point scale but re-classed it to binary measure for their analysis.)

Users' Spatial Abilities: A considerable body of literature stresses the importance of individuals' spatial abilities in all spatial tasks including wayfinding [29, 16]. Thus, another possible explanation for the difference in turning activity could be individual differences in spatial abilities, which could act as a moderator for the effects of spatial familiarity.

All of these explanations can be valid and the differences observed within each class may stem from one or many of these factors. However, we have slight evidence that highlights the priority of *Levels of Familiarity* more than others. As presented in Figure 4, even for the single case in the unfamiliar class for which two different wayfinders walked the same route, we can observe these individual variations in turning-behaviors. Although this one case is not considered as a representative sample, observing a case like this, with a fixed spatial environment for both wayfinders, reinforces the finding that the variations in turn-behaviors within each class may stem from the levels of familiarity (see Future Work below).

6 Conclusion and Future Work

In this paper, we provide evidence that familiarity of wayfinders can be detected by analyzing their gaze behavior during the matching-to-action phase of decision-making for turning activity. We draw this conclusion based on the analysis of gaze data that has been collected during an in-situ wayfinding experiment by a customized pre-trained XGBoost turning activity classifier. The classification results indicate a distinguishable pattern between these two groups regarding their turning-behavior. Within each group, however, we also observe unique patterns of turning-behavior. We discussed this observation with respect to three possible explanations: the impact of the spatial environment, different levels of familiarity, and users' spatial abilities. Each of these factors may account for the within-group differences in observed turn-behavior. However, a single-case observation hints that spatial environment may not be the most important factor, as in this case, two different wayfinders, both unfamiliar, walked the same route. Hence, the results of the current study open the door to predicting, modeling, and hopefully defining spatial familiarity on a continuous scale. This leaves room for further investigations (in both controlled and uncontrolled settings) concerning all of these factors in general and different levels of familiarity in particular. This investigation will be fostered by the fact that spatial ability and spatial environment can both be fairly controlled in experimental designs. For instance, to disentangle the user and environmental effects, it would be interesting to conduct an experiment with relatively similar routes for each wayfinder, so that each wayfinder can be assigned a comparable familiar and unfamiliar route concerning the environmental factors. Such a setting allows for further within-class analysis, e.g., the gaze-pattern differences between familiar and unfamiliar routes walked by the same person. Another experiment for keeping the environment factor fixed would be to select a single route and recruit only familiar or unfamiliar participants. Adding more behavioral data sources may as well be worthwhile to consider in this research endeavor.

References

1 Linda P. Acredolo. The familiarity factor in spatial research. *New Directions for Child and Adolescent Development*, 1982(15):19–30, 1982.

2 N. Alinaghi, M. Kattenbeck, A. Golab, and I. Giannopoulos. Will you take this turn? gaze-based turning activity recognition during navigation. In *Proc. of the 11th Intl. Conf. on Geographic Information Science (GIScience 2021)-Part II*, 2021.

3 G. L Allen. Spatial abilities, cognitive maps, and wayfinding. *Wayfinding Behavior: Cognitive Mapping and Other Spatial Processes*, 4680, 1999.

4 A. Bulling, J. A Ward, H. Gellersen, and G. Tröster. Eye movement analysis for activity recognition using electrooculography. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(4):741–753, 2010.

5 G. T. Buswell. *How people look at pictures: a study of the psychology and perception in art.* Univ. Chicago Press, 1935.

6 H. Couclelis, R. G Golledge, N. Gale, and W. Tobler. Exploring the anchor-point hypothesis of spatial cognition. *J.of Environmental Psychology*, 7(2):99–122, 1987.

7 A. Dehghani, O. Sarbishei, T. Glatard, and E. Shihab. A quantitative comparison of overlapping and non-overlapping sliding windows for human activity recognition using inertial sensors. *Sensors*, 19(22):5026, 2019.

8 W. Dong, H. Liao, B. Liu, Z. Zhan, H. Liu, L. Meng, and Y. Liu. Comparing pedestrians' gaze behavior in desktop and in real Envs. *Cartography & GIScience*, 47(5):432–451, 2020.

9 C. Galdi, H. Wechsler, V. Cantoni, M. Porta, and M. Nappi. Towards demographic categorization using gaze analysis. *Pattern Recognition Letters*, 82:226–231, 2016.

10 N. Gale, R. G Golledge, W. C Halperin, and H. Couclelis. Exploring spatial familiarity. *The Professional Geographer*, 42(3):299–313, 1990.

11 I. Giannopoulos. *Supporting Wayfinding Through Mobile Gaze-Based Interaction.* PhD thesis, ETH Zurich, 2016.

12 L. Gokl, M. Mc Cutchan, B. Mazurkiewicz, P. Fogliaroni, and I Giannopoulos. Towards urban environment familiarity prediction. *Adv. in Cartography and GIScience of the ICA*, 2(5), 2019.

13 Antonia Golab, Markus Kattenbeck, Georgios Sarlas, and Ioannis Giannopoulos. It's also about timing! when do pedestrians want to receive navigation instructions. *Spatial Cognition & Computation*, pages 1–33, 2021.

14 R. G Golledge. Place recognition and wayfinding: Making sense of space. *Geoforum*, 23(2):199–214, 1992.

15 I. M Harms, B. RD Burdett, and S. G Charlton. The role of route familiarity in traffic participants' behaviour and transport psychology research: A systematic review. *Transportation Research Interdisciplinary Perspectives*, 9:100331, 2021.

16 M. Hegarty and D. Waller. *Individual differences in spatial abilities.* Cam. Univ. Press, 2005.

17 J. M Henderson, S. V Shinkareva, J. Wang, S. G Luke, and J. Olejarczyk. Predicting cognitive state from eye movements. *PloS One*, 8(5):e64937, 2013.

18 C. Hölscher, T. Meilinger, G. Vrachliotis, M. Brösamle, and M. Knauff. Up the down staircase: Wayfinding strategies in multi-level buildings. *J. of Env. Psychology*, 26(4):284–299, 2006.

19 T. Ishikawa and D. R Montello. Spatial knowledge acquisition from direct experience in the environment: Individual differences in the development of metric knowledge and the integration of separately learned places. *Cognitive Psychology*, 52(2):93–129, 2006.

20 P. Kiefer, I. Giannopoulos, and M. Raubal. Using eye movements to recognize activities on cartographic maps. *Proc of SIGSPATIAL 2013*, pages 478–481, 2013.

21 P. Kiefer, I. Giannopoulos, M. Raubal, and A. Duchowski. Eye tracking for spatial research: Cognition, computation, challenges. *Spatial Cognition & Computation*, 17(1-2):1–19, 2017.

22 R. V Levine and A. Norenzayan. The pace of life in 31 countries. *J.of Cross-Cultural Psychology*, 30(2):178–205, 1999.

23 R. Li and A. Klippel. Wayfinding behaviors in complex buildings: The impact of environmental legibility and familiarity. *Environment and Behavior*, 48(3):482–510, 2016.

24 H. Liao, W. Dong, H. Huang, G. Gartner, and H. Liu. Inferring user tasks in pedestrian navigation from eye movement data in real-world environments. *International J.of Geographical Information Science*, 33(4):739–763, 2019.

25 H. Liao, W. Zhao, C. Zhang, W. Dong, and H. Huang. Detecting individuals' spatial familiarity with urban environments using eye movement data. *CEUS*, 93:101758, 2022.

26 B. Liu, W. Dong, Z. Zhan, S. Wang, and L. Meng. Differences in the gaze behaviours of pedestrians navigating between regular and irregular road patterns. *ISPRS Int J.of Geo-Information*, 9(1), 2020.

27 L. Mathur, M. Spitale, H. Xi, J. Li, and M. J Matarić. Modeling user empathy elicited by a robot storyteller. In *9th Intl. Conf. on Affective Computing and Intelligent Interaction (ACII)*, pages 1–8, 2021.

28 F. Meneses and A. Moreira. Using gsm cellid positioning for place discovering. In *2006 Pervasive Health Conf. and Workshops*, pages 1–8. IEEE, 2006.

29 D. R Montello, K. L Lovelace, R. G Golledge, and C. M Self. Sex-related differences and similarities in geographic and environmental spatial abilities. *Annals of the Association of American geographers*, 89(3):515–534, 1999.

30 Raffaella Nori and Laura Piccardi. I believe I'm good at orienting myself... But is that true? *Cognitive Processing*, 16(3):301–307, 2015.

31 M. J O'Neill. Effects of familiarity and plan complexity on wayfinding in simulated buildings. *J.of Environmental Psychology*, 12(4):319–327, 1992.

32 J. Platt et al. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. *Advances in Large Margin Classifiers*, 10(3):61–74, 1999.

33 A. Rousell and A. Zipf. Towards a landmark-based pedestrian navigation service using OSM data. *ISPRS Int J.of Geo-Information*, 6(3), 2017.

34 N Savage et al. Seems familiar: An algorithm for inferring spatial familiarity automatically, accessed feb. 5, 2013. *Computer Science Department, University of California, Santa Barbara, CA*, 2013.

35 V. Schnitzler, I. Giannopoulos, C. Hölscher, and I. Barisic. The interplay of pedestrian navigation, wayfinding devices, and Enval features in indoor settings. In *Proc of ETRA 2016*, pages 85–93, 2016.

36 Y. Shiga, T. Toyama, Y. Utsumi, K. Kise, and A. Dengel. Daily activity recognition combining gaze motion and visual features. *Adjunct Proc of UbiComp 2014*, pages 1103–1111, 2014.

37 M. van Haeren and W. Mackaness. The influence of familiarity on route choice: Edinburgh as a case study. In *Proc. of GISRUK Conf., Leeds*, 2015.

38 W. Wang. Using location-based social media for ranking individual familiarity with places: a case study with foursquare check-in data. In *Progress in Location-Based Services 2014*, pages 171–183. Springer, 2015.

39 F. Wenczel, L. Hepperle, and R. von Stülpnagel. Gaze behavior during incidental and intentional navigation in an outdoor Env. *Spatial Cognition & Comp.*, 17(1-2):121–142, 2017.

40 J. M Wiener, S. J Büchner, and C. Hölscher. Taxonomy of human wayfinding tasks: A knowledge-based approach. *Spatial Cognition & Computation*, 9(2):152–165, 2009.

41 A. L. Yarbus. *Eye movements and vision*. Springer, 2013.

42 W. Zhang, X. Zhao, and Z. Li. A comprehensive study of smartphone-based indoor activity recognition via xgboost. *IEEE Access*, 7:80027–80042, 2019.

43 Z. Zhou, R. Weibel, and H. Huang. Familiarity-dependent computational modelling of indoor landmark selection for route communication: a ranking approach. *International J.of Geographical Information Science*, pages 1–33, 2021.

Automatically Discovering Conceptual Neighborhoods Using Machine Learning Methods

Ling Cai[1]
Center for Spatial Studies, University of California, Santa Barbara, CA, USA
STKO Lab, Department Geography, University of California, Santa Barbara, CA, USA

Krzysztof Janowicz
Universität Wien, Austria
Center for Spatial Studies, University of California, Santa Barbara, CA, USA

Rui Zhu
Center for Spatial Studies, University of California, Santa Barbara, CA, USA
School of Geographical Sciences, University of Bristol, Bristol, UK

Abstract

Qualitative spatio-temporal reasoning (QSTR) plays a key role in spatial cognition and artificial intelligence (AI) research. In the past, research and applications of QSTR have often taken place in the context of declarative forms of knowledge representation. For instance, conceptual neighborhoods (CN) and composition tables (CT) of relations are introduced explicitly and utilized for spatial/temporal reasoning. Orthogonal to this line of study, we focus on bottom-up machine learning (ML) approaches to investigate QSTR. More specifically, we are interested in questions of whether similarities between qualitative relations can be learned from data purely based on ML models, and, if so, how these models differ from the ones studied by traditional approaches. To achieve this, we propose a graph-based approach to examine the similarity of relations by analyzing trained ML models. Using various experiments on synthetic data, we demonstrate that the relationships discovered by ML models are well-aligned with CN structures introduced in the (theoretical) literature, for both spatial and temporal reasoning. Noticeably, even with significantly limited qualitative information for training, ML models are still able to automatically construct neighborhood structures. Moreover, patterns of asymmetric similarities between relations are disclosed using such a data-driven approach. To the best of our knowledge, our work is the first to automatically discover CNs without any domain knowledge. Our results can be applied to discovering CNs of any set of jointly exhaustive and pairwise disjoint (JEPD) relations.

2012 ACM Subject Classification Computing methodologies → Machine learning; Computing methodologies → Knowledge representation and reasoning; Computing methodologies → Temporal reasoning; Computing methodologies → Spatial and physical reasoning

Keywords and phrases Qualitative Spatial Reasoning, Qualitative Temporal Reasoning, Conceptual Neighborhood, Machine Learning, Knowledge Discovery

Digital Object Identifier 10.4230/LIPIcs.COSIT.2022.3

Funding *Krzysztof Janowicz*: Funded by the National Science Foundation – OIA (2033521).

Acknowledgements We want to thank Yutao Zhou for discussion on experiment design/ visualization.

1 Introduction

Since the 90s, Qualitative Spatio-Temporal Reasoning (QSTR) has attracted attentions from researchers and practitioners in several fields, such as geographical information science, artificial intelligence and cognitive science [6, 10, 17, 13, 26, 14]. Aside from the clear connection to human representations and linguistic communication of the spatial configuration of our environment, QSTR has numerous advantages over its quantitative counterpart [16].

[1] Corresponding author

Representing qualitative information by using symbols and developing calculi to infer unknown qualitative information is the key to QSTR. Different sets of qualitative spatial relations (such as directional and topological relations) along with a system of qualitative caculi are developed [9], among which reasoning over topological relations becomes the most well-established area in QSTR.

As far as regions are concerned, the most well-known formalizations for qualitative topological relations are - the Region Connection Calculus (RCC-8) [22] and the 9-Intersection Model (9-IM) [4, 10]. Both arrive at the same conclusion that there exist eight base topological relations between regions in 2D space, although they are developed independently during the earlier 90s [2]. Those relations form the foundation for a variety of qualitative spatial reasoning techniques [8, 10, 11, 12]. Two major (and interconnected) lines of works are: (1) Composition Tables (CTs) (i.e., transitivity tables), which store possible resulting relations arising from the composition of two relations [1, 22, 24, 23]. (2) Conceptual Neighborhood Graphs (CNGs), which formalize transitions between relations. Conceptual neighbors of a relation are defined as a set of relations that can be directly transformed into/from the relation by deforming (e.g., moving and scaling) the related entities continuously (in a topological sense) [15]. In a CNG, relations are modeled as nodes and an undirected edge is established between two neighboring relations (see Figure 3f). CNGs play an essential role for reasoning with uncertain or incomplete information [14], and have been used in research of cognitive similarity assessment [19, 18] and modeling of linguistic spatial terms [7]. In addition to topological relations, composition tables, and conceptual neighborhoods have also been developed for reasoning over temporal relations [1, 15].

Those reasoning methods follow a top-down manner, which usually requires (noise-free) explicit domain knowledge. On the contrary, success in data-driven Machine Learning (ML) approaches, which are insensitive to noise and good at dealing with incomplete information as well as uncertainty, provides new opportunities to study QSTR from a bottom-up perspective. ML models rely solely on training data to discover patterns/rules that can be implicitly used for reasoning rather than explicitly injecting domain knowledge into the model. However, the question of why they succeed and whether they are able to (re)discover theories, here in the sense of rule sets or CNGs, is unexplored.

In this paper, we propose a graph-based approach to investigate similarities of qualitative relations from a bottom-up perspective. Particularly, we are interested in how the similarities derived from ML methods are related to classic theoretical studies (e.g., on conceptual neighborhoods). By conducting extensive experiments on synthetic data regarding spatial reasoning (here, *RCC-8* relations) and temporal reasoning (here, Allen's thirteen interval relations), we are able to demonstrate that ML models can automatically discover conceptual neighborhood graphs. In addition, experiment results showcase that such graphs can be easily discovered by ML methods even when limited data are available for training. Moreover, the similarities of relations are mostly asymmetric, which echos the findings in [19] from a perspective of cognitive assessment. Furthermore, patterns observed in asymmetric similarities of relations are disclosed. To the best of our knowledge, we are the first to automatically discover conceptual neighborhood graphs of qualitative relations from a bottom-up perspective by analyzing ML methods. In theory, our approach can be used to discover CNGs for any calculus with jointly exhaustive and pairwise disjoint (JEPD) relations.

The remainder of this paper is structured as follows: Section 2 introduces background about how to perform QSTR by using machine learning methods. Section 3 elaborates on the proposed graph-based approach to discover similarities among relations. Section 4 describes the generation of synthetic data, evaluation metrics, and reports experimental results. Section 5 discusses our findings and points out the direction for future studies.

2 Background

In this section, we introduce preliminaries of ML methods to achieve QSTR. We summarize notations and abbreviations we use in this paper in Table 1 for quick reference.

Table 1 Terms and their abbreviations used in this paper.

Terms (abbrev.)		
Qualitative Spatio-temporal Reasoning (QSTR)	Conceptual Neighborhood Graphs (CNGs)	
Machine Learning (ML)	Artificial Intelligence (AI)	
Knowledge Graphs (KGs)	Knowledge Graph Embedding (KGE)	
Composition Tables (CTs)	Jointly Exhaustive and Pairwise Disjoint (JEPD) relations	
RCC-8 Relations	**IR-13 Relations**	
disconnected (dc)	before (<)	after (>)
externally connected (ec)	meets (m)	met-by (mi)
partially overlapping (po)	overlaps (o)	overlapped-by (oi)
tangentially proper part (tpp)	during (d)	contains (di)
tangentially proper part inverse (tppi)	starts (s)	started-by (si)
non-tangentially proper part (ntpp)	finishes (f)	finished-by (fi)
non-tangentially proper part inverse (ntppi)	equal (=)	
equal (eq)		

2.1 Qualitative Representation of Relations

In this paper, we store binary relations between entities in form of triples. A triple of the form $\langle s, r, o\rangle$ represents an entity *subject* that has a *relation* to another entity *object*. For instance, the statement that a house is externally connected (ec) to a park can be represented as $\langle house, ec, park\rangle$. A set of such tripled is called a knowledge graph (KG). In our paper, a KG is a simple directed graph, consisting of entities being modeled as nodes and relations between them being modeled as labels of edges. Formally, it can be represented as $G = (V, E)$, where V and E are the set of nodes/entities and edges with relations being labels, respectively.

2.2 Relation Prediction Task

We will focus on a task known as relation prediction, namely inferring the relation between two entities based on other information. It is equivalent to answering the query $\langle s, ?r, o\rangle$. Examples include: *what is the topological relation between Los Angeles and Santa Monica?* or *what is the temporal relation between the Battle of Trafalgar and the Napoleonic Wars?*

2.2.1 Symbolic Reasoning Methods

Traditionally, symbolic representations are adopted to represent entities and relations, on top of which qualitative calculi are developed to perform reasoning tasks. For instance, CTs along with path-consistency algorithms are often used to infer missing relation between entities [24]. Given that (*property A, tangential proper part (tpp), park B*) and (*park B, disconnect (dc), house C*), we are able to infer that (*property A, disconnect (dc), house C*) by checking the CT of *RCC-8*. Usually such top-down approaches (which are based on qualitative calculi) fall into the group of symbolic reasoning. Despite their great success in qualitative reasoning in the past, such approaches are faced with noticeable limitations. For instance, they are sensitive to erroneous information or noise. Moreover, they can only be applied to a limited range of reasoning tasks, do not scale well over large datasets, and cannot be easily applied in combination with numeric approaches [25].

2.2.2 Knowledge Graph Embedding Methods

Knowledge Graph Embedding (KGE) methods are an embedding technique in ML that has been empirically proven to be effective in reasoning in a subsymbolic way.

Generally speaking, the goal of KGE methods is to learn subsymbolic representations of entities and relations in a high-dimensional continuous vector space while preserving the connectivity between entities and relations from KGs. Typically, developing a KGE model requires the following three components.

(I) The first is to randomly initialize subsymbolic representations for each entity/relation in a high-dimensional continuous vector space. By doing so, each entity/relation is initialized as a high-dimensional vector (a.k.a embedding or subsymbolic representation) and can be viewed as a point in such high-dimensional vector space. The vector space could be Euclidean space, Hyperbolic space, Spherical space, etc., which vary between different KGE models. The embedding of an entity v, or a relation r, can be expressed as $\mathbf{v} \in \mathbb{U}^d$, or $\mathbf{r} \in \mathbb{U}^d$, where $\mathbb{U}$ denotes the vector space and d is its dimension.

(II) a scoring function is required to measure the likelihood of a triple being positive (i.e., a true statement). Various KGE models specify different scoring functions. For instance, TransE [3], the first KGE model, assumes that for a triple $\langle s, r, o\rangle$, the relation r is a transformation operator in a vector space, which translates the subject s to the object o. Thus the embedding of an object entity $\mathbf{o}$ should be equivalent to the resulting embedding of a subject entity $\mathbf{s}$ being translated by the relation $\mathbf{r}$ in the vector space. Then the distance between the embedding of the object entity and the resulting entity can be used as a scoring function: $score(s, r, o) = \|\mathbf{s} + \mathbf{r} - \mathbf{o}\|$. Thus, triples that are present in KGs (i.e., positive triples) will obtain a lower score while triples that are not present will gain a higher score.

(III) an objective function is needed for training through a process of optimization. A commonly used way of constructing such an objective function is by contrasting scores obtained by positive triples with those of negative triples. Often, the objective function is built upon the task of entity prediction (namely answering queries such as $\langle ?s, r, o\rangle$ or $\langle s, r, ?o\rangle$). For each positive triple $\langle s, r, o\rangle$, a number of negative triples (e.g., k) are generated by switching the subject s and/or the object o with other randomly selected entities (e.g., s_i or o_i). Then an objective function $\mathcal{L}$ can be defined to minimize scores for positive triples while maximizing scores for negative ones:

$$\mathcal{L} = -log\ \sigma(\gamma - score(s, r, o)) - \frac{1}{k}\sum_{i=1}^{k} log\ \sigma(score(s_i, r, o) - \gamma) \tag{1}$$

where σ is the sigmoid function and γ is a pre-specified hyper-parameter as a margin. $\langle s_i, r, o\rangle$ is a negative sample of $\langle s, r, o\rangle$.

After a number of iterative optimization over the training data, minimizing the objective function yields embeddings (representations) for all entities and relations in the KG. The optimized KGE model then can be used in various downstream tasks, such as entity prediction relation prediction, and triple classification. A plethora of KGE models have been developed in the the past years, e.g., [3, 20] and various scoring functions have been used (refer to [27] for more details).

Here we elaborate on how to perform relation prediction (i.e., answering a query $\langle s, ?r, o\rangle$) by using trained KGE methods, since they are closely related to our approach discussed in Section 3. Concretely, we enumerate all possible relations ($r' \in \mathcal{R}$) to replace $?r$ individually

and then sort these relations by $score(s, r', o)$ in an ascending/descending order. Finally the embedding method regards the relation ranked first as the correct answer to the query. The table on the left in Figure 1 shows examples of ordered sets of relations produced by a trained KGE model regarding different testing queries.

3 Knowledge Graph Embedding Methods as Knowledge Miner

Similarity is one of the most commonly used measures to examine relationships of objects. For instance, domain experts introduce conceptual neighbors to indicate similar qualitative relations [15]. Likewise, in this section we introduce an approach to examine similarities between qualitative relations by analyzing trained KGE models from a bottom-up perspective. There are two steps in this approach - initial construction of a relation graph and its refinement.

The first question is how to derive similarities between any two relations in the set $\mathcal{R}$ from a trained KGE model. Our assumption is that it would be difficult for a trained embedding model to distinguish relations that are similar in a topological sense. That is, in terms of the task of relation prediction, for a testing query (geometry A, ?r, geometry B) (whose target answer is *externally connected* (*ec*)), we hypothesize the embedding-based model may yield similar scores for (geometry A, ***ec***, geometry B) and (geometry A, *partially overlap* (***po***), geometry B), because *po* and *ec* are topologically similar. Put differently, the sorted set of predicted relations reveals structural similarities among relations in the sense that similar relations are more easily confused in relation prediction (see Figure 1).

Based on this assumption, we initiate a graph in which vertices are different types of relations. For each testing query $\langle s, ?r, o\rangle$, a directed edge is established from the correct relation to either the relation ranked at first (top 1) or second (top 2) in the ordered list of relations. Such a choice relies on whether the relation at top 1 is the correct relation or not. When the correct is ranked at top 1, we do not introduce a loop. Instead, a directed edge starting from the correct relation to the relation at Top 2 is added. If the relation at Top 1 is not the correct, then an edge is built from the correct relation to Top 1. The resulting graph is a *directed* graph, whose edges originate from the correct relation to a relation identified as most similar to the correct by the KGE model. In a directed edge, we use terms - head and tail - to refer to the source and the target of an edge, respectively. The direction of edges reflects which candidate relation (tail) is similar to the target relation (head). Note that by such a distinction, we are able to examine the asymmetric similarities between relations.

The graph constructed above only illustrates which relations are considered as similar by a KGE model, but does not quantify similarities between relations. Here, we design a weighting function to quantify these similarities. Specifically, the weight of an edge is estimated as the proportion of the number of edges from a head to a tail relation over the total number of edges originating from the head. This function can be formulated as follows:

$$weight(r_i \to r_j) = \frac{count(r_i \to r_j)}{\sum_{r' \in \mathcal{R}} count(r_i \to r')} \tag{2}$$

where $count(r_i \to r_j)$ is the cardinality of edges originating from r_i (head) to r_j (tail) (with shortest paths). An example of the construction process is shown in Figure 1.

So far, we obtain a directed and weighted graph, which reveals the similarities between different relations; see Figure 2a. We observe that this graph is almost complete (i.e., any two relations/vertices are connected via an edge), because eventually any two relations are likely to be thought of as similar by a KGE model. However, not all these similarities are significant; for instance many edges only have marginal weights (e.g., 0.01). In order to extract significant relationships from the initial relation graph, the next step is to prune insignificant edges to get a refined graph.

id	queries	correct answer	sorted list
1	<obj 1, ?r, obj 3>	**tpp**	tpp|**po**|ec|dc|eq|ntpp|tppi|ntppi
2	<obj 4, ?r, obj 5>	**dc**	dc|**po**|ec|tpp|tppi|ntpp|eq|ntppi
3	<obj 45, ?r, obj 56>	**po**	**ntpp**|po|tpp|ec|dc|tppi|ntppi|eq
4	<obj 25, ?r, obj 6>	**ec**	**tpp**|ec|ntpp|po|dc|tppi|ntppi|eq
5	<obj 51, ?r, obj 9>	**ntpp**	**tpp**|ntpp|po|ec|dc|eq|ntppi|tppi
6	<obj 7, ?r, obj 11>	**tppi**	tppi|**ec**|ntppi|po|dc|tpp|ntpp|eq
7	<obj 15, ?r, obj 1>	**ntppi**	ntppi|**tppi**|po|ec|dc|eq|tpp|ntpp
8	<obj 22, ?r, obj 5>	**eq**	eq|**tppi**|tpp|po|ec|ntppi|ntpp|dc
9	<obj 17, ?r, obj 51>	**tppi**	**ntppi**|ntpp|ec|po|dc|tpp|ntpp|eq
...	...	...	...

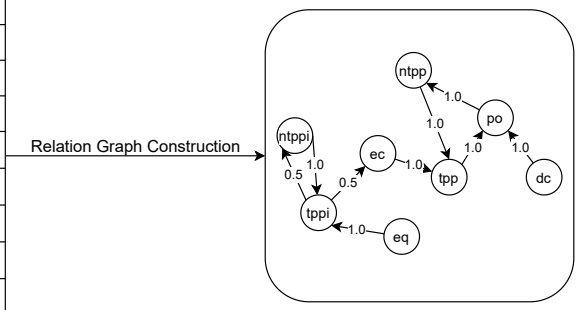


Figure 1 Relation Graph Construction. Here nine queries are used as examples and the *sorted list* column shows relations sorted by a scoring function from a KGE method. Each relation is represented as a vertex in the graph and edges are established from the correct answer (column 3) to the relation in bold in the sorted list. Weights are calculated by using Eq. 2.

Intuitively one could enumerate different thresholds (for instance, by gradually increasing a threshold (i.e., 0.0, 0.05, 0.1,..., 1.0)) to cut off edges whose weights are insignificant. Then one can terminate the enumeration process by manually checking whether the refined graph is aligned with our domain knowledge/cognition. However, without enough domain knowledge, it is hard to conclude which graph is meaningful and this means the proposed solution is not truly bottom-up. In order to reduce human intervention in the refinement process, we define a condition to automatically terminate the enumeration. The condition is based on the naive fact that *all relations/vertices must be preserved/connected in the graph after the refinement*, since our focus in this paper is on the relationships of all *relations*. Based on graph theory, such a fact boils down to ensuring that there is always one connected component in the graph after the refinement. Therefore, we can gradually increase thresholds by constant margins (e.g., 0.05) until the initial graph is no longer one connected component. In summary, in the process of refining relation graphs, we generate a number of candidate thresholds (within the range of $(0.0, 1.0)$ and a step of 0.05) in an ascending order and find the maximal threshold that leads to only one connected component in the graph, which is regarded as the refined relation graph.

4 Experiments

In this section, we introduce the synthetic data we use to test our method, the evaluation metrics used for graph similarity measure, and present experimental results. Although theoretically our proposed approach can be applied to any set of JEPD relations to automatically discover a graph of relations, we focus on *RCC-8* and *IR-13* here.

4.1 Data Preparation

Since real-life datasets are usually incomplete, we generate sets of synthetic data for the purpose of demonstration. Specifically, we choose rectangles as primitive geographical entities for *RCC-8* relations and closed-intervals as primitive temporal entities for *IR-13* relations.

To generate rectangles, we first set up a main area, in which rectangles should be located. By default, the main area is set to be a 15×15 unit square with the origin being its bottom-left corner. Then we randomly generate pairs of points within the square and each pair of points

compose the top-left corner and the bottom-right corner of a rectangle[2]. Finally, we compute *RCC-8* relations between any two rectangles to generate synthetic spatial relation triples. Likewise, we generate a number of closed-intervals on the x-axis within the range [0, 500]. Specifically, we randomly select two integers from the range and use the smaller one as the beginning of an interval and the bigger one as the ending of the interval. Then we compute the *IR-13* relations between any two intervals to generate synthetic temporal relation triples. We call the set of all synthetic triples as *complete synthetic data*.

However, without any prior knowledge, it is hard to decide how many rectangles/intervals should be generated within the given main area/line segment. Meanwhile, the number of rectangles/intervals generated in the same area/line may affect discovered relation graphs. Therefore, we independently generate several sets of synthetic triples for both the *RCC-8* and the *IR-13* relations with different number of rectangles/intervals (i.e., [64, 128, 256, 512, 1024]). These sets of triples have different densities of rectangles/intervals. The proportions of different relations generated with respect to different numbers of rectangles/intervals are shown in Table 2.

Table 2 Relation proportions of *RCC-8* (on the left) and *IR-13* (on the right) regarding different numbers of rectangles/intervals *N=64, 128, 256, 512 or 1024*. All values are multiplied by 100.

N	*64*	*128*	*256*	*512*	*1024*
dc	43.5	43.4	42.7	42.8	42.3
ec	12.2	11.9	11.8	11.5	11.8
eq	1.6	0.8	0.4	0.2	0.1
ntpp	1.2	1.4	1.1	1.5	1.5
ntppi	1.2	1.4	1.1	1.5	1.5
po	35.6	34.8	37.4	36.4	36.7
tpp	2.4	3.1	2.8	3.1	3.1
tppi	2.4	3.1	2.8	3.1	3.1

N	*64*	*128*	*256*	*512*	*1024*
$<$	18.7	14.9	16.1	16.4	16.5
$=$	1.6	0.8	0.4	0.2	0.1
$>$	18.7	14.9	16.1	16.4	16.5
d	15.7	17.2	16.4	16.2	16.9
di	15.7	17.2	16.4	16.2	16.9
f	0.1	0.1	0.1	0.1	0.1
fi	0.1	0.1	0.1	0.1	0.1
m	0	0.1	0.1	0.1	0.1
mi	0	0.1	0.1	0.1	0.1
o	14.6	17.1	16.9	16.9	16.1
oi	14.6	17.1	16.9	16.9	16.1
s	0.1	0.2	0.1	0.1	0.1
si	0.1	0.2	0.1	0.1	0.1

4.2 Experiment Settings

We choose *HyperRotatE* [5] as the embedding model to learn subsymbolic representations of entities and relations, thanks to its ability of modeling the composition of relations (which is relevant to composition tables) and tree-like graph structures (which is useful for modeling transitive relations (e.g., *ntpp*)). This model also contains the three components mentioned in Section 2.2.2 and has a different scoring function. We use the original implementation of *HyperRotatE* to learn embeddings for entities and relations[3]. Hyper-parameters used for the *RCC-8* and the *IR-13* relations include learning rates: 0.05 (for the *RCC-8* relations) and 0.1 (for the *IR-13* relations), batch sizes: 1024 for both, negative samples: 64 (for the *RCC-8* relations) and 32 (for the *IR-13* relations), and dimensions: 110 (for the *RCC-8* relations)

[2] We ensure that each rectangle is valid. For example, if the two points in a pair align along the same axis, we will remove this pair.

[3] `https://github.com/HazyResearch/KGEmb`

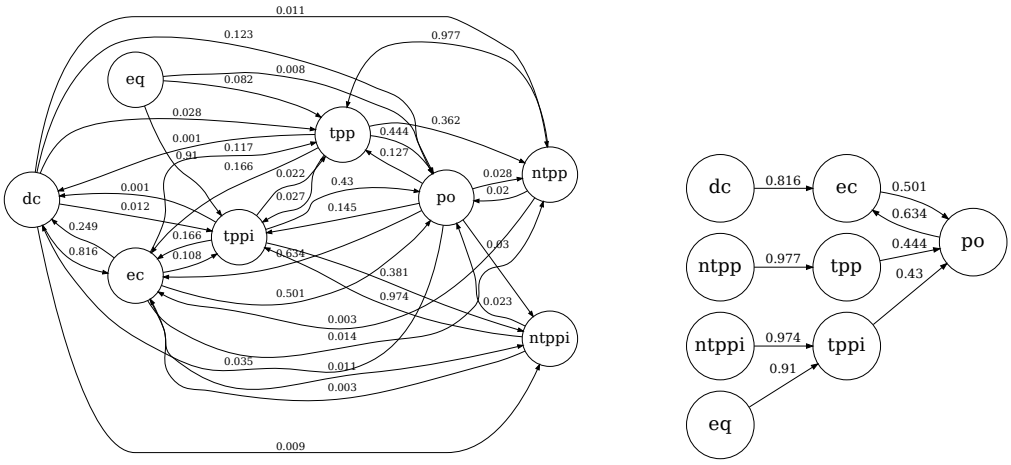
and 18 (for the *IR-13* relations). For *IR-13* relations, we use the same hyper-parameters for all synthetic data. For *RCC-8* relations, we increase the dimension of the embedding space to 200 when the number of entities is 512 or 1024[4]. In the experiment, we train *HyperRotatE* and then perform relation prediction over the *complete synthetic data* by default [5].

4.3 Evaluation Metrics

In order to quantify the differences between the learned relation graph and from CNGs, we introduce three metrics to measure commonality and difference. One solution is to convert graphs to sets of edges (each edge consists of a pair of relations) and to use set operations for quantification. Three metrics can be defined: (1) *False Recall* (i.e., number of false positives): the number of edges that are in our generated graph but not in CNGs (set difference). (2) *True Recall* (i.e., number of true positives): the number of edges that are in both our generated graph and CNGs (set intersection). (3) *Failed Recall* (i.e., number of false negatives): the number of edges that are not in our generated graph but in CNGs (set difference). Clearly, a graph that is similar to CNGs should have a low *False Recall*, a high *True Recall*, and a low *Failed Recall*.

4.4 Experimental Results

In this section, we first show direct results from our approach introduced in Section 3. Figure 2 illustrates (a) the initial relation graph resulting from the construction steps and (b) the refined relation graph after pruning. Next we report main findings based on the refined relation graph.



(a) Initial Relation Graph.

(b) Refined Relation Graph.

Figure 2 Examples of initial/refined relation graphs produced by our approach.

[4] When the number of entities increased to 512/1024, the model's performance greatly deteriorated. We assume the performance is compromised due to lack of learnable parameters. Thus, we increase the dimensions to provide more learnable parameters for our models to learn.

[5] Note that we do not tune these hyper-parameters but choose them by empirical experiences. It is worthwhile to investigate the impact of hyper-parameters on the experiment results in the future.

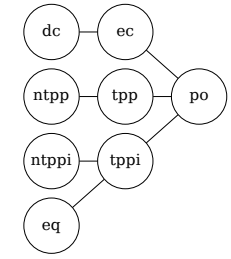
1. Relation graphs automatically discovered by our approach are well-aligned with CNGs for both *RCC-8* and *IR-13* relations. Figure 2b implies that our refined relation graph resembles conceptual neighborhood graphs (Figure 3f). This motivates us to examine how similar our refined relation graphs are CNGs from the literature and whether this is merely a coincidence. In order to make our refined graphs comparable with CNGs, we convert the refined graphs into undirected and unweighted relation graphs (*UU-RGs*).

Figure 3 and Figure 4 report the results for *RCC-8* and *IR-13* relations, respectively, with different number of entities being considered. Noticeably, our approach discovers stable relation graphs for both *RCC-8* and *IR-13* relations. In specific, relation graphs for *RCC-8* remain almost unchanged with an increasing number of rectangles and relation graphs for *IR-13* begin to be fixed (except for the *equal* relation) when the number of intervals is 256. This observation also aligns with the statistics shown in Table 2, in which the relation proportions become relatively stable when the number of entities reaches 256. This indicates that the KGE model is mainly affected by the proportion of relations in the synthetic data. Moreover, by comparing Figure 3a, 3b, 3c, 3d and 3e with Figure 3f (or comparing Figure 4c, 4d and 4e with Figure 4f), we can observe that the discovered relation graphs are well-aligned with the CNGs which are defined in the literature (see Figure 3f and Figure 4f), except for differences around the *equal* relation(i.e., "eq" and "="). This observation demonstrates the ability of ML models in learning domain knowledge purely from data and the effectiveness of our approach in automatically discovering relationships of JEPD relations (*RCC-8* and *IR-13* as examples here). *This demonstrates that conceptual neighborhood graphs can be reproduced from data without any domain knowledge/inductive bias.*

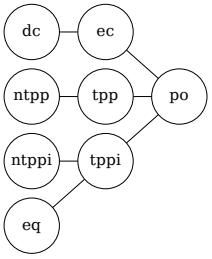
As for the differences around the *equal* relation, one explanation is the lack of enough *equal* relations in our synthetic data. Because we randomly generate rectangles/intervals within a given area/segment, it is relatively rare to yield two rectangles/intervals that have the same geometry. As a result, most *equal* relations are just self-equivalent (e.g., $\langle s, eq, s\rangle$), which in fact does not provide enough useful information for the model to learn. Hence, we do not consider this a shortcoming of the model.

2. Similarities of relations are asymmetric and certain relations are more similar. Several patterns in asymmetric similarities of relations are also disclosed. In this experiment, we examine similarities of relations, which are quantified by weights in Eq. 2. We extract a subgraph from our initial relation graphs (see Figure 2a) that contain edges presented in the theoretical CGNs except for edges that are connected to the *equal* relation (since the *equal* relation is not well-reproduced). We set the number of entities to 1024 and run the *HyperRotatE* model for 20 times to obtain average weights/similarity scores. The extracted subgraphs for *RCC-8* and *IR-13* relations are illustrated in Figure 5.

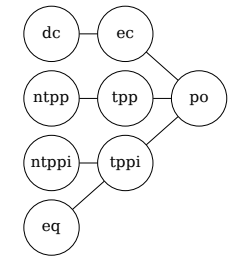
Apparently, we can observe that similarities of relations are *asymmetric*. In other words, the statement that a is similar to b differs from that b is similar to a (a and b are relations). For instance, the similarity between *dc* and *ec* is 0.903 while the inverse similarity is 0.215. Namely, *dc* is more similar to *ec* while *ec* is less similar to *dc*. In fact, Figure 5a shows *ec* is most similar to *po*, and both *dc* and *po* are most similar to *ec*. Meanwhile, we find that *ec* and *po* are more similar in general with higher similarities of 0.556 and 0.654. Additionally, there exist similar patterns between relations and their inverses in terms of their asymmetric similarities to other relations. For instance, Figure 5b shows $<$ is most similar to *m* and *m* is most similar to *o*. In terms of their inverse relations, $>$ is most similar to *mi* and *mi* is most similar to *oi*. Moreover, in Figure 5a, *ntpp* is most similar to *tpp* and *ntppi* is most similar to *tppi*. Similar patterns are shown between *d–>f–>oi* and *di–>fi–>o*, as well as between



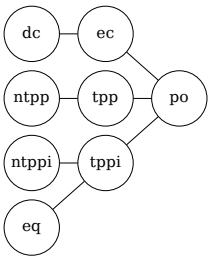
(a) 64.



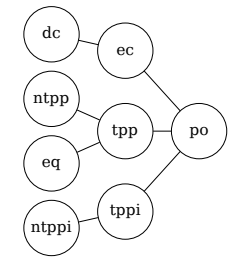
(b) 128.



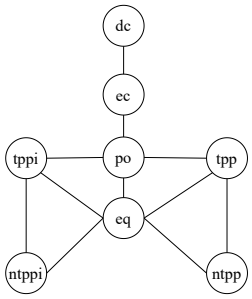
(c) 256.



(d) 512.

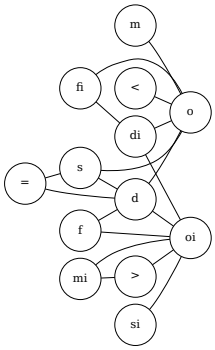


(e) 1024.

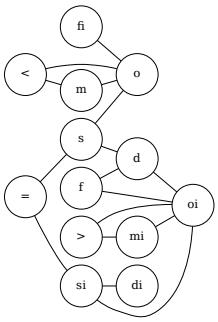


(f) CNG.

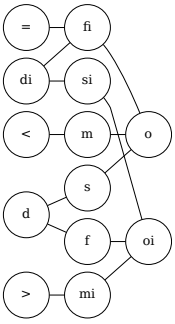
Figure 3 The relation graph of the RCC-8 relations w.r.t. different number of rectangles.



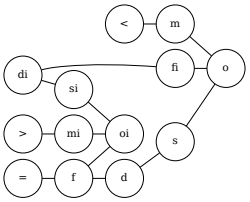
(a) 64.



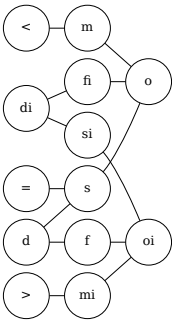
(b) 128.



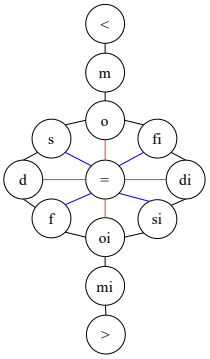
(c) 256.



(d) 512.



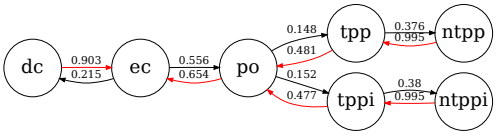
(e) 1024.



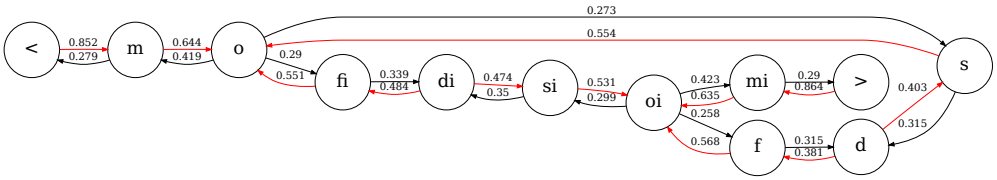
(f) CNG.

Figure 4 The relation graph of the IR-13 relations w.r.t. different number of intervals.

d–>s->o and *di–>si–>oi*. Another interesting observation is that all neighboring relations of the *overlapping* relation (i.e., *po* in *RCC-8* and *o* and *oi* in *IR-13*) are most similar to the *overlapping* relation (see the arrows that point to the *overlapping* relation). By contrast, in Figure 5b, both *d* and *di* are most similar to their neighboring relations (the red arrows around them leave out of them). Interestingly, similarity assessments in the cognitive science literature have been shown to be highly non-symmetric as well due to differences in (feature) alignment. For instance, Klippel et al. disclosed that the similarity between *RCC-8* relations vary from different scenarios (such as hurricane, cannon and geometry). In addition, Mark et al. also found that some topological relations indeed are conceptually more similar to others [21].



(a) Similarities for *RCC-8* relations.



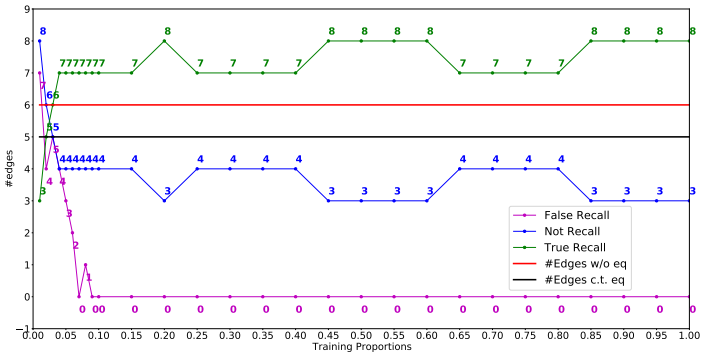
(b) Similarities for *IR-13* relations.

Figure 5 Asymmetric similarities of relations. For two edges between two vertices, the edge with a larger weight is highlighted in red.

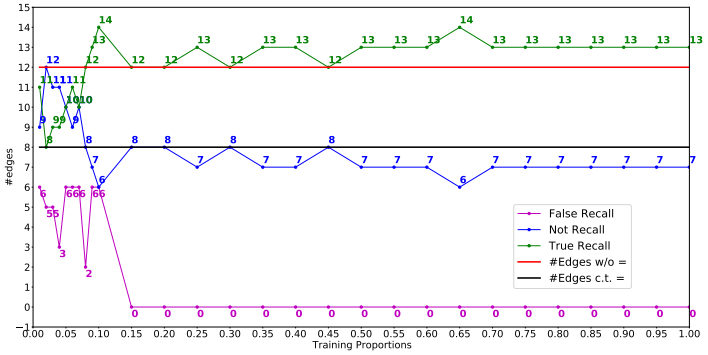
3. Even with limited training data (i.e., as low as 15% of the *complete synthetic data*), *HyperRotatE* is still capable of reproducing CNGs. Finally, we are interested in the question of how much training data is needed for *HyperRotatE* to reproduce CNGs. In order to answer this question, we extract subsets of the *complete synthetic data* with different proportions and use the three metrics introduced in Section 4.3 to evaluate the commonality and difference between *UU-RGs* and CNGs. Experiment results with the number of entities being 256 are shown in Figure 6. Red lines and black lines are theoretical references, indicating numbers of edges that are connected to the *equal* relation and that are not in theoretical CNGs, respectively. Clearly, regarding *RCC-8* relations, when more than 10% of the *complete synthetic data* are used for training, *HyperRotatE* is able to reproduce CNGs with stable recalls. Specifically, when the proportion is larger than 10%, *False Recall* continues to be 0, *True Recall* is either 7 or 8 and *Failed Recall* is either 3 or 4. Noticeably, *True Recall* is always above the red line (i.e., 6 – the theoretical number of edges that are not connected to "=" in the CNG) and *Not Recall* is close to the black line (i.e., 5 – the theoretical number of edges that are connected to "=" in the CNG). That is, the relation graphs (except for conceptual neighbors of the *equal* relation) is well-aligned with the theoretical CNGs even when only 10% of the *complete synthetic data* are available. Similar observations are shown for *IR-13* relations; see Figure 6b; however, the same pattern is observed when the training proportion is larger than 15%. In summary, *HyperRotatE* is a robust knowledge miner, which succeeds in discovering CNGs even with limited training data.

5 Conclusion

In this work, we presented a graph-based approach to examine similarities among *RCC-8* and *IR-13* relations in neighborhood graphs since they are important to spatio-temporal reasoning and spatial queries. In contrast to traditional approaches that heavily rely on top-down techniques and rule sets, we address this problem in a bottom-up manner without the need of any domain knowledge. Specifically, we focus on the task of relation prediction; namely to answer the query $\langle s, ?r, o \rangle$. Our rationale is that it would be difficult for machine learning methods to distinguish relations that are topologically similar when predicting



(a) Quantitative comparison for *RCC-8* relations.



(b) Quantitative comparison for *IR-13* relations.

Figure 6 Quantitative comparison between *UU-RGs* and CGNs. *UU-RGs* are reproduced w.r.t. different proportions of the *complete synthetic data* as training data. Lines in red denote the number of edges that are not connected to the *equal* relation in CNGs and lines in black denotes the opposite.

missing relations between two entities. Therefore, we can pull similar relations out of the relation prediction task, and then use the proposed method to construct a graph to examine the structure among relations. Our experiments on synthetic data about *RCC-8* and *IR-13* relations reveal that (1) the extracted relation graphs are well-aligned with conceptual neighborhood graphs introduced in [15] and [10] except for neighboring relations of the *equal* relation. We believe this may be caused by a lack of enough *equal* relations in generated training data, which is left for future work; that (2) similarities of relations are asymmetric, and patterns in asymmetric similarities of relations are the same as those in their inverse relations; and that (3) the presented embedding models are robust in mining qualitative spatial and temporal knowledge (i.e., CNGs), even with limited training data.

Theoretically, our approach could be applied to any calculus with JEPD relations [6] to automatically discover CNGs. We believe our research would benefit theoretical studies of CNGs in general and contribute to a broader field, such as geospatial artificial intelligence, by promoting a deeper understanding of what machines really learn from data in a bottom-up manner. In the future, we plan to study whether such CNGs will be preserved when realistic data (particularly when non-spatial information is also considered) are used at training.

References

1 James F Allen. Maintaining knowledge about temporal intervals. *Communications of the ACM*, 26(11):832–843, 1983.

2 Roland Billen and Nico Van de Weghe. Qualitative spatial reasoning. *International Encyclopaedia of Human Geography*, pages 12–18, 2009.

3 Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko. Translating embeddings for modeling multi-relational data. In *NIPS*, pages 1–9, 2013.

4 Lejdel Brahim, Kazar Okba, and Laurini Robert. Mathematical framework for topological relationships between ribbons and regions. *Journal of Visual Languages & Computing*, 2015.

5 Ines Chami, Adva Wolf, Da-Cheng Juan, Frederic Sala, Sujith Ravi, and Christopher Ré. Low-dimensional hyperbolic knowledge graph embeddings. In *Proceedings of the 58th Annual Meeting of ACL*, pages 6901–6914, 2020.

6 Anthony G Cohn and Jochen Renz. Qualitative spatial representation and reasoning. *Foundations of Artificial Intelligence*, 3:551–596, 2008.

7 Matthew P Dube and Max J Egenhofer. An ordering of convex topological relations. In *GIScience*, pages 72–86. Springer, 2012.

8 Max J Egenhofer. Deriving the composition of binary topological relations. *Journal of Visual Languages & Computing*, 5(2):133–149, 1994.

9 Max J Egenhofer. The family of conceptual neighborhood graphs for region-region relations. In *GIScience*, pages 42–55. Springer, 2010.

10 Max J Egenhofer and Khaled K Al-Taha. Reasoning about gradual changes of topological relationships. In *TMSTRGS*, pages 196–219. Springer, 1992.

11 Max J Egenhofer and Robert D Franzosa. Point-set topological spatial relations. *IJGIS*, 5(2):161–174, 1991.

12 Max J Egenhofer and John Herring. Categorizing binary topological relations between regions, lines, and points in geographic databases. *The*, 9(94-1):76, 1990.

13 Andrew U Frank. Qualitative spatial reasoning about distances and directions in geographic space. *Journal of Visual Languages & Computing*, 3(4):343–371, 1992.

14 Christian Freksa. Qualitative spatial reasoning. In *Cognitive and linguistic aspects of geographic space*, pages 361–372. Springer, 1991.

15 Christian Freksa. Temporal reasoning based on semi-intervals. *AI*, 54(1-2):199–227, 1992.

16 Christian Freksa. Using orientation information for qualitative spatial reasoning. In *TMSTRGS*, pages 162–178. Springer, 1992.

17 Alexander Klippel. Spatial information theory meets spatial thinking: is topology the rosetta stone of spatio-temporal cognition? *Annals of AAG*, 102(6):1310–1328, 2012.

18 Alexander Klippel and Rui Li. The endpoint hypothesis: A topological-cognitive assessment of geographic scale movement patterns. In *COSIT*, pages 177–194. Springer, 2009.

19 Alexander Klippel, Jinlong Yang, Jan Oliver Wallgrün, Frank Dylla, and Rui Li. Assessing similarities of qualitative spatio-temporal relations. In *ICSC*, pages 242–261. Springer, 2012.

20 Yankai Lin, Zhiyuan Liu, Maosong Sun, Yang Liu, and Xuan Zhu. Learning entity and relation embeddings for knowledge graph completion. In *AAAI*, volume 29, 2015.

21 David M Mark and Max J Egenhofer. Calibrating the meanings of spatial predicates from natural language: Line-region relations. In *Proceedings, SDH 1994*, volume 1, pages 538–553, 1994.

22 David A Randell, Zhan Cui, and Anthony G Cohn. A spatial logic based on regions and connection. *KR*, 92:165–176, 1992.

23 Jochen Renz, Debasis Mitra, et al. Qualitative direction calculi with arbitrary granularity.

24 Jochen Renz and Bernhard Nebel. Qualitative spatial reasoning using constraint calculi. In *Handbook of spatial logics*, pages 161–215. Springer, 2007.

25 Carl Schultz, Mehul Bhatt, Jakob Suchan, and Przemysław Andrzej Wałęga. Answer set programming modulo 'space-time'. In *IJCRR*, pages 318–326. Springer, 2018.

26 Jan Oliver Wallgrün, Diedrich Wolter, and Kai-Florian Richter. Qualitative matching of spatial information. In *the 18th SIGSPATIAL*, pages 300–309, 2010.

27 Quan Wang, Zhendong Mao, Bin Wang, and Li Guo. Knowledge graph embedding: A survey of approaches and applications. *IEEE TKDE*, 29(12):2724–2743, 2017.

Predicting Distance and Direction from Text Locality Descriptions for Biological Specimen Collections

Ruoxuan Liao
Massey Geoinformatics Collaboratory, Massey University, Auckland, New Zealand

Pragyan P. Das
Massey Geoinformatics Collaboratory, Massey University, Auckland, New Zealand

Christopher B. Jones
School of Computer Science and Informatics, Cardiff University, UK

Niloofar Aflaki
Massey Geoinformatics Collaboratory, Massey University, Auckland, New Zealand

Kristin Stock[1]
Massey Geoinformatics Collaboratory, Massey University, Auckland, New Zealand

Abstract

A considerable proportion of records that describe biological specimens (flora, soil, invertebrates), and especially those that were collected decades ago, are not attached to corresponding geographical coordinates, but rather have their location described only through textual descriptions (e.g. *North Canterbury, Selwyn River near bridge on Springston-Leeston Rd*). Without geographical coordinates, millions of records stored in museum collections around the world cannot be mapped. We present a method for predicting the distance and direction associated with human language location descriptions which focuses on the interpretation of geospatial prepositions and the way in which they modify the location represented by an associated reference place name (e.g. ***near*** *the Manawatu River*). We study eight distance-oriented prepositions and eight direction-oriented prepositions and use machine learning regression to predict distance or direction, relative to the reference place name, from a collection of training data. The results show that, compared with a simple baseline, our model improved distance predictions by up to 60% and direction predictions by up to 31%.

2012 ACM Subject Classification Computing methodologies → Machine learning

Keywords and phrases geospatial prepositions, biological specimen collections, georeferencing, natural language processing, locative expressions, locality descriptions, geoparsing, geocoding, geographic information retrieval, regression, machine learning

Digital Object Identifier 10.4230/LIPIcs.COSIT.2022.4

Funding This work was partly funded under the New Zealand Ministry of Business, Innovation and Enterprise Endeavour Fund BioWhere project, and used data collections owned by Manaaki Whenua – Landcare Research New Zealand.

1 Introduction

Around the world, vast collections of biological specimens (e.g. plants, fungi, invertebrates, soil samples) held by museums, libraries and government organisations are georeferenced using text locality descriptions such as *North Canterbury, Selwyn River near bridge on Springston-Leeston Rd*. While specimens collected in recent years usually have geographic coordinates (latitude and longitude) captured from GPS, the locations of many millions of specimens, sometimes going back hundreds of years, are recorded only in text format.

[1] Corresponding author

Many of the text descriptions used are complex, multi-clausal and consist of a mixture of place names/toponyms (*North Canterbury, Selwyn River, Springston-Leeston Rd* in the above example), generic geographic feature types that appear in the landscape but are not sufficiently notable to have toponyms (*bridge* in the above example), and location terms such as prepositions (*near, on* in the above example) and multi-word phrases (e.g. *4km north-east of, 30 miles along the road from*). Furthermore, many of the descriptions include abbreviations, and while attempts are being made to enforce standard approaches to the description of location to support automated georeferencing [2], many historical records do not follow these standards.

Descriptions of location such as the examples above are referred to as locative expressions [16] and are typically expected to include a located object (or locatum), a spatial relational term or phrase, and a reference object or relatum [16]. In biological records the locatum is sometimes implicit (being the sample that was collected), but in more complex phrases, sub-clauses may include a locatum, such as the word *bridge* in the previous example. When georeferencing textual descriptions of locations, Named Entity Recognition methods which recognise specific types of entities in text, including locations [1], have commonly been used to recognise named places, while gazetteers are used to attach coordinates to (i.e. geocode) the names [10, 18]. However, this only provides part of the picture, as location terms (geospatial prepositions and other modifiers) provide offsets relative to place names (e.g. *next to the Manawatu River*), and thus are key to achieving the level of precision needed to make use of this vast repository of biological data for species mapping and monitoring over time. A number of works have identified common forms of location descriptions, such as distance and cardinal direction (e.g. *4km north-east of <place name>*), defining rule-based [13] and probabilistic [12] models for the interpretation of these common forms, and the errors associated with them. However, the range of possible forms of location descriptions is vast, and many do not conform to these common structures. Furthermore, the interpretation of location descriptions is often context dependent, and the way a description is interpreted many depend on the topography and other physical characteristics, or may vary depending on the methods used for data collection (in part a function of date) or the collector.

In this paper, we present a method for determining the distances and directions associated with a set of location terms (prepositions/prepositional phrases and cardinal directions) using machine learning regression. We predict the offset distance and direction from a reference object associated with the location term using semantic and contextual features of the locality description and the reference object, compare the results to a baseline and evaluate the importance of the features in the model. We demonstrate that regression is a useful tool for predicting the distance or direction associated with location descriptions (depending on the type of description), and that there is scope to further expand this method with additional features and machine learning models. Our method is evaluated using 15311 locality descriptions from the biological collections held by Manaaki Whenua - Landcare Research (MWLR), New Zealand. We identified the most frequently occurring prepositions or nouns/adverb + preposition pairs (e.g. *base of* and *north of* respectively), resulting in eight distance-oriented (*near, above, below, at, head of, end of, mouth of, tributary of*) and eight direction prepositions (*north of, south of, east of, west of, north-east of, north-west of, south-east of, south-west of*). We confined our attention to the last place name in the location description that is preceded by one of those terms. In future work, multiple clauses and place names should be parsed to make use of all the information in the descriptions and further improve results.

The paper is structured as follows: Section Two describes previous work; Section Three presents the method used for extracting relevant terms from the descriptions and calculating the dependent variables (the values we wish to predict, in this case distance and direction) and the features used in our model. Section Four presents the results and evaluates feature importance, and Section Five contains conclusions.

2 Previous work

One of the most common methods for georeferencing text location descriptions is the use of Named Entity Recognition to identify place names, referred to as geoparsing, before extracting their coordinates from a gazetteer, known as toponym disambiguation or resolution [10, 17]. However, this simply georeferences the place names mentioned, but ignores the impact of spatial relation terms that describe a location offset from the place name (e.g. *near the Manawatu River; north-west of Lincoln; outside Auckland*). Such descriptions are known as relative location descriptions, because they describe a location relative to a reference object, and rely on a spatial relation term to do this. Many spatial relation terms are prepositions (e.g. *at, on, near, beside*), though they may be other parts of speech including verbs and adverbs [6].

A number of works have developed models of spatial relation terms, including mapping of terms to spatial relations formally modelled with qualitative spatial reasoning methods (QSR) addressing topological, proximity, orientation and projective relations, e.g. [7, 9, 24]. Such models have only had quite limited application to quantification of distances or angles associated with specific natural language spatial relational terms due to the challenges of interpeting the vagueness inherent in human language.

Several studies have proposed fuzzy logic models of proximity relations and conducted human subjects experiments, including for geographical contexts [26, 31, 8, 11], but such models do not appear to have been applied to the interpretation of natural language texts. A regression model of various forms of nearness and farness that considers several contextual factors, again in a geographical context, was presented in [32] but its application there was to predict the linguistic description for given metric measures. A quantitative analysis of the use of *near* within n-grams was conducted in [5] based on text sources mined from the web. Triples of the located object, spatial relation and reference object were used to examine distances between points of interest within three cities, and between populated places and each of the cities. They found that distances were smaller for *near* in New York compared to San Francisco and Los Angeles, but did not study context specific differences in distances relating to the different feature types. Another study in a similar context analysed the proximity of *close, near* and *next to* spatial relation terms [30] to derive reference object locations. The study discusses the importance of contextual variables like geometry, size and travel distance in deriving coordinates. However, factors like the cardinal direction and angle of the reference object to the located object were not taken into consideration in this study.

Another approach defines spatial templates [21], also known as applicability models or probabilistic density fields, for particular spatial relation terms, that describe the areas in which a term may apply, and depict the variation in applicability for example in the form of a density surface. Thus proximity relation terms such as *near* may be highly applicable at distances close to an object, but gradually become less so as the distance increases. Individual templates can be constructed using multiple examples of observations of the location of a located object relative to a reference object. While most applications of spatial templates have been in table-top space [27, 19, 28], they have also been applied in the geo-spatial

domain [14], including for purposes of interpreting location descriptions [13], with models of several proximal and projective relations being instantiated with data from the Geograph photo-sharing web site and human subjects experiments. Various forms of applicability models were used in the study of [3] to infer distances implied by individual spatial relations between places in text describing city locations. They used the gazetteer coordinates of known locations to derive the coordinates of the non-gazetteered places, based on applying their models of the respective spatial relations. The approach was rule based and depended upon the prior existence of a graph ('place graph') representation of the respective locations.

In the context of image analysis and retrieval, spatial templates have been used with machine learning models to predict applicable spatial relations and the locations at which spatial relation terms apply, allowing the context of particular situations to be taken into account. In [22] deep learning methods are used with spatial templates (constructed from multiple examples of, mostly projective, spatial relations) to infer the spatial relations between objects in images. In a related study [4], deep learning is used with spatial templates to infer the coordinates of objects in images that have been described with verbal action relations to a given subject. Their input includes the word embeddings of the subject, the relation and the object, along with the location and size of the subject. Notably the latter study [4] provides an example of using regression methods to infer coordinates relative to a reference object. These studies were both conducted in the context of image analysis and retrieval without reference to geographic space.

This paper differs from the previous work in developing a predictive, machine learning model of a selection of distance and direction oriented terms, and incorporating novel contextual factors in the model. We do not go as far as georeferencing since we only predict either distance or direction, but this does enable the area to which a description refers to be narrowed down to a more precise location relative to the reference place name.

3 Method

3.1 Data

The biological specimen dataset that we use in this paper comes from Manaaki Whenua - Landcare Research (MWLR), New Zealand, and consists of four separate collections, the details of which are listed in Table 1. The MWLR database is constantly updated and maintained, and the MWLR version we used was extracted on July 30, 2021. Table 2 shows some examples of the kinds of locality descriptions that appear in the database. While these collections in combination contain many millions of records, only a small proportion are digitised and have geographic coordinates, and we use a subset of these to train and test our model.

A notable characteristic of the data set is that the coordinates are highly variable in spatial accuracy. The biological specimens in the collections range in age from those collected during Cook's voyages of New Zealand (1769-1779) to the present day. Older specimens may rely on place names that did not continue to be used and whose location has been lost, and textual location descriptions were sometimes very imprecise. Specimens collected in the last few years have been coordinated with GPS, but before that a range of practices were used to derive the coordinates that we used in this work. Some were heavily manual processes involving examination of maps, aerial photos and records to allocate coordinates, but also several automated processes were applied, one example being the use of map sheets recorded as part of the collection record to derive coordinates, using either the centre or a specified corner of the map sheet as an approximation of the location. The result of this is

Table 1 The composition of the MWLR database.

Dataset	Original number of records	Number of records after dropping null values
Allan Herbarium (plants)	321891	13692
International Collection of Micro-organisms from Plants	22345	909
NZ Fungarium	106945	426
NZ Arthropod Collection	202676	14
Total	**633857**	**15311**

Table 2 Example locality descriptions.

Locality Description	Latitude	Longitude
Buller, Paparoa Mountains, north flank of Mt Euclid, c. 1-1.5km east of Morgan Tarn.	-41.9562	171.6032
Auckland Island, lower slopes about Musgrave Inlet	-50.6469	166.1533
Nelson, about l km SE of Lake Peel, in the track to Balloon Hut	-41.1316	172.6001
Marlborough, hills about Queen Charlotte Sound	-41.3859	173.7136
Lake Ellesmere Spit = Kaitorete Spit - About Midway along length.	-43.874	172.2679

that the accuracy of individual records in the data set is unknown. To illustrate this point, Appendix A provides scatter plots of eight cardinal direction prepositional phrases studied in this paper, and Table 3 provides figures to indicate the mean bearing (angle from north in a clockwise direction, ranging from 0° to 360°) and standard deviation of each direction. The circular nature of bearing values, where 0° is the same as 360° causes problems for predictive models, and thus following [15], we represent bearings as the ($\sin\theta$, $\cos\theta$), and the standard deviation figures are given using this representation.

Table 3 Comparison of Cardinal Directions.

Direction	Mean bearing (°)	Standard deviation of cosine of bearing	Standard deviation of sine of bearing	Mean standard deviation
north of	11.8	0.57	0.57	0.57
south of	192.6	0.54	0.44	0.49
east of	102.7	0.57	0.54	0.56
west of	263.8	0.47	0.51	0.49
north-east of	38.7	0.47	0.37	0.42
north-west of	307.8	0.50	0.46	0.48
south-east of	142.4	0.51	0.29	0.40
south-west of	223.7	0.44	0.27	0.36

The vague use of cardinal directions in natural language is well documented [14], and demonstrated by the range of locations clustered around the specified direction. However, the scatter plots demonstrate that in this data set, the presence of multiple extreme outliers is much greater than for data analysed in other work, such as [14]. This is likely to be due to inaccuracies in the data set, particularly resulting from methods used to georeference older historical data. Future work will derive and incorporate quality measures into approaches to

georeference the collection, but here we work with the data as it is, accepting inaccuracies, as well as some evident gross errors, as part of the challenge that we address. Appendix A also illustrates the tendency for the main four cardinal directions (*north, south, east, west*) to be used for a wider range of directions than the other four (more specific) directions, and this can also been in the lower mean standard deviations for the more specific directions than for (*north, south, east, west*) in Table 3.

3.2 Pre-Processing

Spatial relations are commonly described with prepositions, and for the purposes of this paper, we focus on creating distance or direction models for a set of common spatial prepositions. Having replaced common abbreviations in the descriptions with their expanded versions (e.g. SE -> south-east), we used the spaCy[2] python library part of speech (POS) tagger to identify prepositional phrases as those that were tagged either as prepositions alone (e.g. *near*), or as nouns or adverbs followed by prepositions (e.g. *base of* and *north of* respectively). We then counted the frequency of each unique prepositional phrase to identify the most frequent. We first selected the most frequently appearing eight prepositions, and since some cardinal directions were among this set, we expanded the set to include all eight cardinal directions, and to also include the next most frequent non-directional prepositions to create a balanced set of eight prepositions that describe cardinal directions (which we describe as directional terms), and eight others for which distance is often an important defining characteristic. The final data set consisting of locality descriptions that use one of these sixteen terms formed 78.60% of the 15311 descriptions mentioned in Table 1.

The final set of spatial relation terms were:

- eight directional terms: *north, south, east, west, north-west, north-east, south-west, south-east* and
- eight other spatial relation terms for which distance may be an important component: *near, at, above, below, head of, mouth of, end of, and tributary of.*

While not all of the members of the second set of the terms are primarily distance-related, incorporating elements of elevation (*above, below*) or parthood (*head of, mouth of, end of, tributary of*), in this work we attempt to predict the distances associated with them. Even though *above* and *below* describe elevation, some distance association is implied, as they would not be used with locations that were a large distance from the reference object (e.g. *the hut above Lake Wakatipu*). Similarly, while the parthood terms refer to some specific component of an object (e.g. a river), the parthood relation also implies spatial coincidence. We acknowledge that there are many other semantic aspects of these terms than just the distance, but delay those aspects for future work.

We next used Named Entity Recognition (NER) to identify place names in the locality descriptions, testing several state of the art tools and selecting spaCy's NER tool as the most accurate after testing on a sample of 200 descriptions. We attempted to retrieve coordinates for all place names within a bounding box for New Zealand using three gazetteers: GeoNames[3], the New Zealand Geographic Board Place Names Gazetteer[4] and Nominatim[5], and selected the best result amongst multiple matches (disambiguated) as the place name

[2] https://spacy.io/
[3] http://www.geonames.org/
[4] https://gazetteer.linz.govt.nz/
[5] https://nominatim.org/

that was closest to the known location for the specimen. In addition to coordinates, the feature type (e.g. lake, river) was retrieved to support extraction of features for the machine learning model (see Section 3.3).

We identified all instances of the 16 prepositions that were immediately followed by a tagged place name for which we could retrieve coordinates. We then calculated the distance and direction between the coordinates of the place name following the preposition and the ground-truth coordinates contained in the data set. These values represent the offset that describes the location of the specimen relative to the reference place name for simple preposition-place name pairs. For example, for the locality description *near Karangahake Gorge*, the distance between the coordinates of Karangahake Gorge retrieved from the gazetteer and the coordinates of the specimen contained in the collection reflects the quantitative meaning of the *near* preposition in this particular context. These, and their associated directions where prepositions are more direction-related, are the figures we aim to predict with our model.

For the eight distance-related prepositions, we use the geodetic distance between the reference object and the ground truth specimen coordinates as the dependent variable (the value we aim to predict) in our model. For the direction-related prepositions, we follow [15] and use the sine and cosine of the bearing ($\sin\theta$, $\cos\theta$) as the dependent variables for model training (and our regression model for directional prepositions thus has two dependent variables) and convert them back to bearings at the end. This approach is used to avoid problems caused by the circular nature of bearing measurements, in which 0° is the same as 360°.

3.3 Regression Model

In order to predict the distance or direction corresponding to the prepositions in our locality descriptions, we incorporate a number of features in a machine learning regression model. The features included were as follows:

- The GloVe embedding of the **feature type of the reference object**. GloVe (Global Vector for Word Representation) generates multi-dimensional vector representations of words and was first introduced by a team at Stanford University to study the similarity index between the words. It is derived from word-word occurrences in a textual description by only considering the non-zero elements, which are used to calculate the embeddings based on probabilities. We used 200 dimension GloVe embeddings pre-trained on Wikipedia + Gigaword 5 [25]. The feature type for each place name (reference object) was retrieved from the relevant gazetteer along with the coordinates.
- The vector created by averaging the GloVe embeddings for the **feature types of all place names in the locality description** (excluding the reference object) using 200 dimension embeddings pre-trained on Wikipedia + Gigaword 5 [25]. The feature type for each place name was retrieved from the relevant gazetteer alongside the coordinates.
- One hot-encoding of the **geometry type** (point, line, polygon, volume) of the reference object feature type. The geometry type was retrieved from the Linguistically Augmented Geospatial Ontology (LAGO) [29] using WordNet [23] to match feature types retrieved from the gazetteer for our reference object to feature types contained in the LAGO if they did not already appear. This feature is included because geometry type has been shown to influence the use of geospatial prepositions (e.g. *the house beside the church* vs. *the road beside the river* - the latter implies alignment as well as proximity) [29].
- One hot-encoding of the **scale** of the reference object feature type (district scale, neighbourhood scale, immediate scale). The scale was retrieved from the LAGO as for geometry type, and is included because the influence of scale on the use of geospatial prepositions

has been demonstrated [29, 20]. Although not identical, this may be considered an approximate indicator of object size (for example, district scale may refer to objects such as mountain ranges, while immediate scale may refer to smaller objects such as houses).
- One hot-encoding of the census **Territorial Authority** of the reference object, out of a total of 85 districts that cover New Zealand. This set of features indicates whether two instances are in the same geographic area.
- The **area, population, population density and length (at the longest extent) of the meshblock**[6] that the reference object is in. These features indicate how urban/rural an area is, and are included to test whether this aspect influences the use of geospatial prepositions.
- The **area, population, population density and length (at the longest extent) of the Territorial Authority** that the reference object is in.
- The **year** that the specimen was collected. This is an approximate indicator of the accuracy of the coordinates, as recent records have GPS-level accuracy, while coordinates from 200 years ago may be very approximate (e.g. derived from map sheet or description).
- A boolean value indicating whether the location is **cultivated** from the collections data.
- The **altitude** of the specimen, providing an indication of the environment type (e.g. alpine).

We used ten-fold cross validation to test a number of different regression models including Support Vector Machine with polynomial (SVM-polynomial kernel) and Radial Basis Function kernel (SVM-rbf kernel), k-nearest neighbour, gradient boosting, support vector regression and decision tree. A number of other models including linear regression and multi-layer perceptrons were tested but did not perform well so were not pursued further.

4 Results

4.1 Distance Prediction

We evaluate the results of our methods against a simple baseline that relies only on the place name immediately following the preposition (the relatum), and like most current approaches ignores the spatial relation term. Hence it assumes that the distance and direction between the place name and the predicted location are zero. Table 4 shows the mean absolute error for the baseline and our best-performing machine learning model, together with the percentage improvement that our method provides over the baseline.

Overall, we see better performance (larger percentage increase) for cardinal directions than for the distance-related prepositions. While this is unexpected in that we would expect more consistency in distance for distance-oriented prepositions than direction-oriented, this may be explained by the larger average distances between relatum and locatum for the direction-oriented prepositions, so the baseline, which assumes a distance of zero, is a poorer estimate than for prepositions that are used for smaller distances between relatum and locatum. However, following this reasoning, we might expect that the poorer result for the *at* preposition could be equivalently explained by typically shorter distances between relatum and locatum, which are better predicted by our zero-distance baseline, but this is not supported by the data. Our goal is to predict the distance between relatum and locatum and, as the baseline predicts this distance to be zero, each baseline prediction is equal to the actual distance between relatum and locatum. Thus the mean absolute error (MAE)

[6] The smallest geographic unit for which New Zealand census data is recorded.

Table 4 Results of Machine Learning Regression - Distance Prediction.

Preposition	Best-performing Model	Count	Baseline	Regression	
			MAE (m)	MAE (m)	% improv
near	svm-rbf kernel	3478	6412	4491	30%
above	svm-rbf kernel	695	3581	2634	26%
head of	svm-rbf kernel	388	5298	3465	35%
below	svm-rbf kernel	278	4256	3462	19%
at	svm-polykernel	208	6075	5140	15%
end of	svm-rbf kernel	164	5678	4512	21%
mouth of	svm-rbf kernel	115	1630	967	41%
tributary of	svm-rbf kernel	112	5347	3934	26%
north of	svm-rbf kernel	1309	7277	4343	40%
south of	svm-rbf kernel	1211	8538	4509	47%
east of	svm-rbf kernel	959	7458	3862	48%
west of	svm-rbf kernel	879	8533	5054	41%
north-east of	svm-rbf kernel	187	10139	4701	54%
south-west of	svm-rbf kernel	169	9756	3892	60%
north-west of	svm-rbf kernel	147	6116	3212	47%
south-east of	svm-rbf kernel	185	9802	4523	54%

for the baseline is equal to the average distance between relatum and locatum across all instances of a particular preposition. The baseline MAE (and therefore the average distance between relatum and locatum) for the *at* preposition is in fact higher than for all other distance-related prepositions except *near*. Furthermore, the *mouth of* preposition has the shortest average distance (baseline MAE) between relatum and locatum, but is the best predicted of the distance-oriented prepositions using our method. The *mouth of* preposition describes a wide range of distances between 85 and 13200 metres.

Although the regression models show improvement relative to the baseline across all of the prepositions, and in many cases these are substantial, we consider that the MAE values are inflated by outliers that result from the low accuracy of some of the coordinates, and in some cases challenges in identifying accurate coordinates for the relatum place names due to their absence from, or duplication in, the gazetteers. For example, Table 5 shows that 80% of the error values (absolute value of predicted - actual distance) for the *near* preposition are below 5514.45m. Thus filtering out of the worst 20% of errors results in a MAE of 1672.09m.

Table 5 Errors for each Percentile for the *near* preposition.

Percentile	Error
10th	162.56
20th	422.68
30th	768.50
40th	1208.30
50th	1794.13
60th	2612.76
70th	3806.46
80th	5514.45
90th	10038.38
100th (all values)	108772.75

Table 6 Results of Machine Learning Regression – Direction Prediction.

Preposition	Best-performing Model	Count	Baseline	Regression	
			MAE (°)	MAE (°)	% improv
north of	gradient boosting regressor	1309	48.0	47.7	0.6%
south of	gradient boosting regressor	1211	35.3	30.4	13.8%
east of	k-nn	959	46.1	42.3	8.1%
west of	k-nn	879	43.4	41.8	3.5%
north-east of	decision tree	187	35.0	26.7	23.8%
south-west of	gradient boosting regressor	169	43.5	30.2	30.6%
north-west of	support vector regression	147	46.2	48.9	-5.9%
south-east of	support vector regression	185	48.3	40.0	17.2%

4.2 Direction Prediction

For the eight direction-oriented prepositions, we evaluated the ability of our machine learning model to predict direction using the features listed in Section 3.3.

While the cardinal directions technically describe precise directions (e.g. *east of* specifies 90° using north as 0° and measuring angle in a clockwise direction, an angular measurement known as the *bearing*), research has shown that these directions are frequently used vaguely in natural language [14] to refer to a range of directions that are more or less in the direction. As a result of this tendency to use direction terms vaguely, rather than defining our baseline as the precise direction that corresponds to each term, we instead use the average deviation from the precise direction specified by the direction term. We thus use the average difference between the bearing of the actual line between relatum and locatum and 90° as the baseline for *east*, and evaluate the ability of our regression model to predict that difference.

As explained in Section 3.2, we represent these angles as two numbers: $\sin\theta$, $\cos\theta$, and perform a multivariate regression to predict both values simultaneously. Table 6 presents the results for the direction prediction, with the $\sin\theta$, $\cos\theta$ values converted back into errors in degrees and compared to the baseline. This means, for example, that if we simply assumed that the preposition *south of* means a bearing of 180°, we would get a MAE of 35.3° using our dataset. However, the use of our regression model reduces this MAE to 30.4°, giving a 13.4% improvement. It must be acknowledged that the MAE for both the baseline and the regression model are relatively high. The high MAE for the baseline is an indication of the large spread of directions for which a given cardinal direction term is used, in some cases deviating substantially from the precise direction indicated by the term (e.g. 90° for *east*), as indicated in Appendix A, and while the regression model improves on the baseline by up to 30%, we anticipate that improvements could be achieved by the inclusion of additional features that focus on directional semantics. It is also of note that the regression model for *north-west of* predicts direction less well than if the precise direction were used (315°). This is most likely because the spread of data points for *north-west of* is relatively narrow compared to the other directions, with few outliers, and thus the regression models ability to model contextual variations in the use of the direction term is less effective.

4.3 Feature Importance

We analyse the importance of different features in the model by calculating the correlation coefficient between each feature and the three dependent variables (distance, $\sin\theta$ and $\cos\theta$). Figure 1 shows the 25 features with the highest correlations. The light grey area indicates the

importance of the GloVe embeddings for the relatum feature type across all of the distance and direction predictions. The dark grey squares represent the average embedding of the feature types of all other place names, as well as feature types mentioned explicitly in the descriptions (e.g. *pit in back paddock*), and are also important.

Geometry type is among the most highly correlated features with both distance and direction. For example, the boolean point geometry feature is negatively correlated with distance for *near* while the line geometry is negatively correlated with distance for *tributary of*. This means that expressions with point reference objects are more likely to be used for short distances than for other geometry types. The most common point reference object (which is also classified as a polygon reference object) is a small populated place or locality, and it is not unexpected that these would be referenced when closer to the specimen collection location than when further away, in contrast to non-point objects, which are likely to be larger in scale. Although they also appear among the 25 most correlated features with the two direction-related dependent variables, geometry type features do not exhibit a consistent pattern across multiple cardinal directions, with the line geometry boolean feature being positively correlated for *west of* and *south of*, negatively for *north of* and *south-east* of and not for the others.

The scale features are not strongly correlated with distance, but do appear in several of the cardinal directions, although the direction of the correlation (positive or negative) varies. The importance of territorial authorities (being the most highly correlated feature for distance for *head of* and *at* and for $\cos\theta$ for *north-west* and *south-east*) indicates a geographic pattern in the way that geospatial prepositions are interpreted. There are 85 territorial authorities throughout New Zealand, with areas ranging from 19 to 29,552 square kilometres, and while some are very small and urban in nature, many cover widely varying terrains and environments including a single authority covering all of Fiordland and much of Southland. The meshblock geometry characteristics (length, area and to a lesser extent population) shown in yellow, and those of the territorial authorities are also important for some of the prepositions, as is altitude (shown in brown, along with year and cultivation).

5 Conclusion

In this paper we used regression to predict the distance and direction associated with 16 prepositions. We demonstrated that regression is a useful tool for predicting the distance associated with location descriptions, with improvements for distance-oriented prepositions of up to 41%, and for direction-oriented prepositions of up to 60%. We also showed the significant impact of outliers in this data set, highlighting the need to consider accuracy in these kinds of biological collections data sets that contain historical records. Results for prediction of direction (bearing) were less promising, with the best result showing an improvement of 31% for *south-west of* (the preposition that yielded the best direction prediction results). We also evaluated the importance of the features used in the model through correlation with the dependent variables, showing that relatum feature type is very important, but a range of other features also contribute, such as territorial authority, geometry type and scale.

In order to further improve the results of these models, future work will derive and incorporate spatial data accuracy measures so that greater weight is given to the coordinate data that is known to be accurate. In addition, we will explore more advanced methods for identifying place names that relate to specific prepositions, and for disambiguating place names. In future work we also plan to add further contextual features to the models, and to apply transformer-based neural network approaches such as BERT to the challenge.

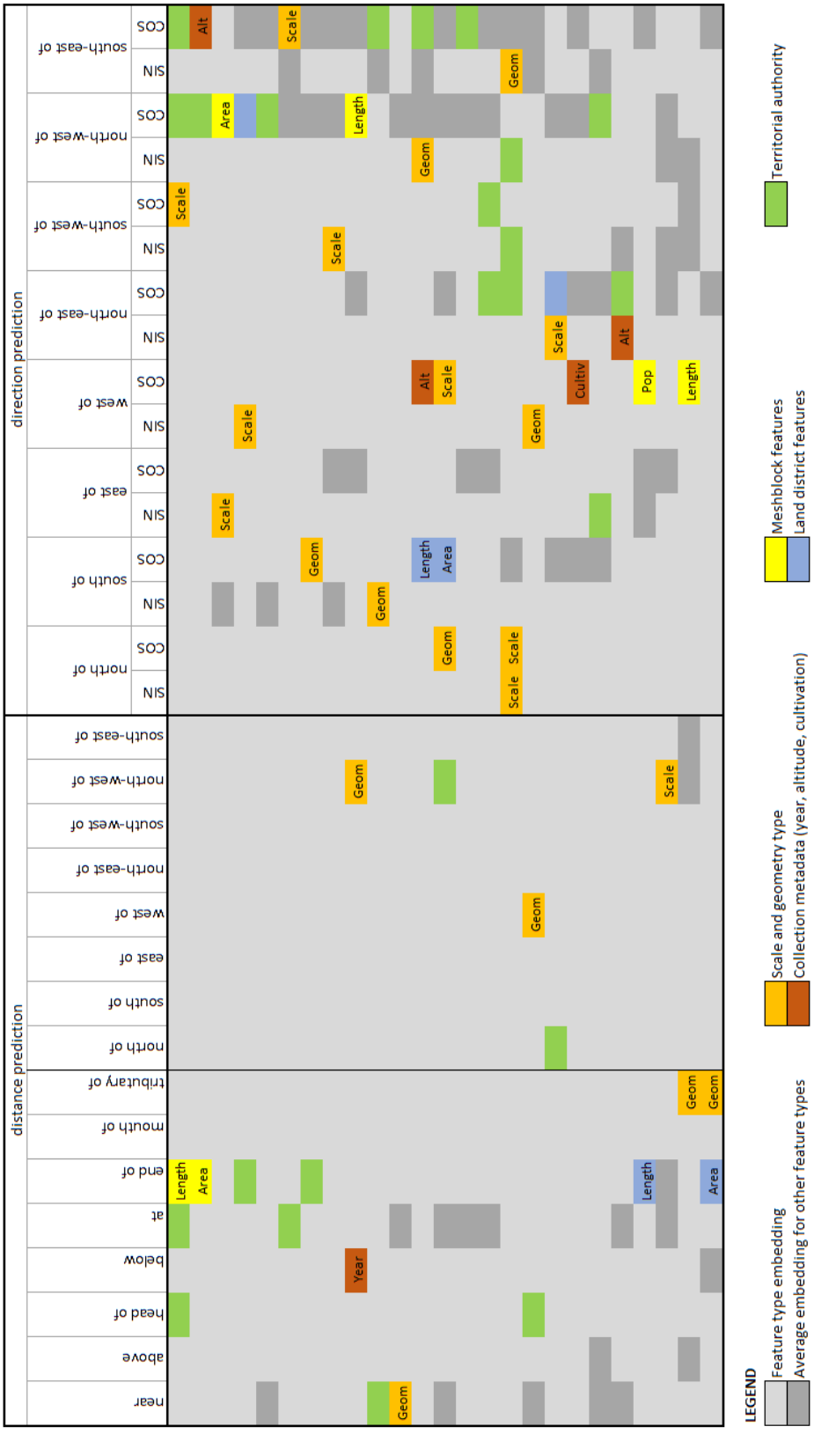


Figure 1 Feature Correlation with Dependent Variables.

References

1. S. Atdağ and V. Labatut. A comparison of named entity recognition tools applied to biographical texts. In *2nd International Conference on Systems and Computer Science*, pages 228–233, August 2013.
2. Arthur D Chapman and John R Wieczorek. *Georeferencing Best Practices*. GBIF Secretariat, Copenhagen, 2020. doi:10.15468/doc-gg7h-s853.
3. Hao Chen, Stephan Winter, and Maria Vasardani. Georeferencing places from collective human descriptions using place graphs. *Journal of Spatial Information Science*, 0(17):31–62, 2018.
4. Guillem Collell, Luc Van Gool, and Marie-Francine Moens. Acquiring common sense spatial knowledge through implicit spatial templates. In *Thirty-second AAAI conference on artificial intelligence*, 2018.
5. Curdin Derungs and Ross Purves. Mining nearness relations from an n-grams web corpus in geographical space. *Spatial Cognition and Computation*, 16, October 2016.
6. André Dittrich, Maria Vasardani, Stephan Winter, Timothy Baldwin, and Fei Liu. A classification schema for fast disambiguation of spatial prepositions. In *Proceedings of the 6th ACM SIGSPATIAL International Workshop on GeoStreaming*, pages 78–86. ACM, 2015.
7. M.J. Egenhofer. Reasoning about binary topological relations. In *Second Symposium on Large Spatial Databases*, volume 525 of *Lecture Notes in Computer Science*, pages 143–160. Springer-Verlag, 1991.
8. M. Gahegan. Proximity operators for qualitative spatial reasoning. In *Spatial Information Theory A Theoretical Basis for GIS*, pages 31–44. Springer Berlin / Heidelberg, 1995.
9. K.P. Gapp. Angle, distance, shape and their relationship to projective relations. In *Proceedings of the 17th Annual Conference of the Cognitive Science Society*, pages 112–117, 1995.
10. Milan Gritta, Mohammad Taher Pilevar, and Nigel Collier. A pragmatic guide to geoparsing evaluation: Toponyms, named entity recognition and pragmatics. *Language Resources and Evaluation*, 54, September 2019.
11. Hans W. Guesgen. Reasoning about distance based on fuzzy sets. *Applied Intelligence*, 17:265–270, 2002.
12. Q. Guo, Y. Liu, and J. Wieczorek. Georeferencing locality descriptions and computing associated uncertainty using a probabilistic approach. *International Journal of Geographical Information Science*, 22(10):1067–1090, 2008.
13. Mark Hall, Philip Smart, and Christopher B. Jones. Interpreting spatial language in image captions. *Cognitive processing*, 12(1):67–94, 2011.
14. Mark M. Hall and Christopher B. Jones. Generating geographical location descriptions with spatial templates: a salient toponym driven approach. *International Journal of Geographical Information Science*, 36(1):55–85, 2021.
15. Kota Hara, Raviteja Vemulapalli, and Rama Chellappa. Designing deep convolutional neural networks for continuous object orientation estimation. *arXiv preprint*, 2017. arXiv:1702.01499.
16. Annette Herskovits. Semantics and pragmatics of locative expressions. *Cognitive science*, 9(3):341–378, 1985.
17. Morteza Karimzadeh. Performance evaluation measures for toponym resolution. In *Proceedings of the 10th workshop on geographic information retrieval*, pages 1–2, 2016.
18. Morteza Karimzadeh, Scott Pezanowski, Alan MacEachren, and Jan Oliver Wallgrün. Geotxt: A scalable geoparsing system for unstructured text geolocation: Geotxt: A scalable geoparsing system. *Transactions in GIS*, 23, January 2019.
19. J.D. Kelleher and F.J. Costello. Applying computational models of spatial prepositions to visually situated dialog. *Computational Linguistics*, 35(2):271–306, 2009.
20. Anna-Katharina Lautenschütz, Clare Davies, Martin Raubal, Angela Schwering, and Eric Pederson. The influence of scale, context and spatial preposition in linguistic topology. In *International Conference on Spatial Cognition*, pages 439–452. Springer, 2006.
21. G.D. Logan and D.D. Sadler. A computational analysis of the apprehension of spatial relations. *Language and space*, pages 493–529, 1996.

22 Mateusz Malinowski and Mario Fritz. A pooling approach to modelling spatial relations for image retrieval and annotation. *arXiv preprint*, 2014. arXiv:1411.5190.

23 George A Miller. *WordNet: An electronic lexical database*. MIT press, 1998.

24 Reinhard Moratz and Thora Tenbrink. Spatial reference in linguistic human-robot interaction: Iterative, empirically supported development of a model of projective relations. *Spatial cognition and computation*, 6(1):63–107, 2006.

25 Jeffrey Pennington, Richard Socher, and Christopher D Manning. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 1532–1543, 2014.

26 V.B. Robinson. Individual and multipersonal fuzzy spatial relations acquired using human–machine interaction. *Fuzzy Sets and Systems*, 113(1):133–145, 2000.

27 J.R.J. Schirra. A contribution to reference semantics of spatial prepositions: The visualization problem and its solution in VITRA. *The Semantics of prepositions: from mental processing to natural language processing*, page 471, 1993.

28 Michael Spranger and Luc Steels. Co-acquisition of syntax and semantics: An investigation in spatial language. In *Proceedings of the 24th International Conference on Artificial Intelligence*, IJCAI'15, pages 1909–1915. AAAI Press, 2015.

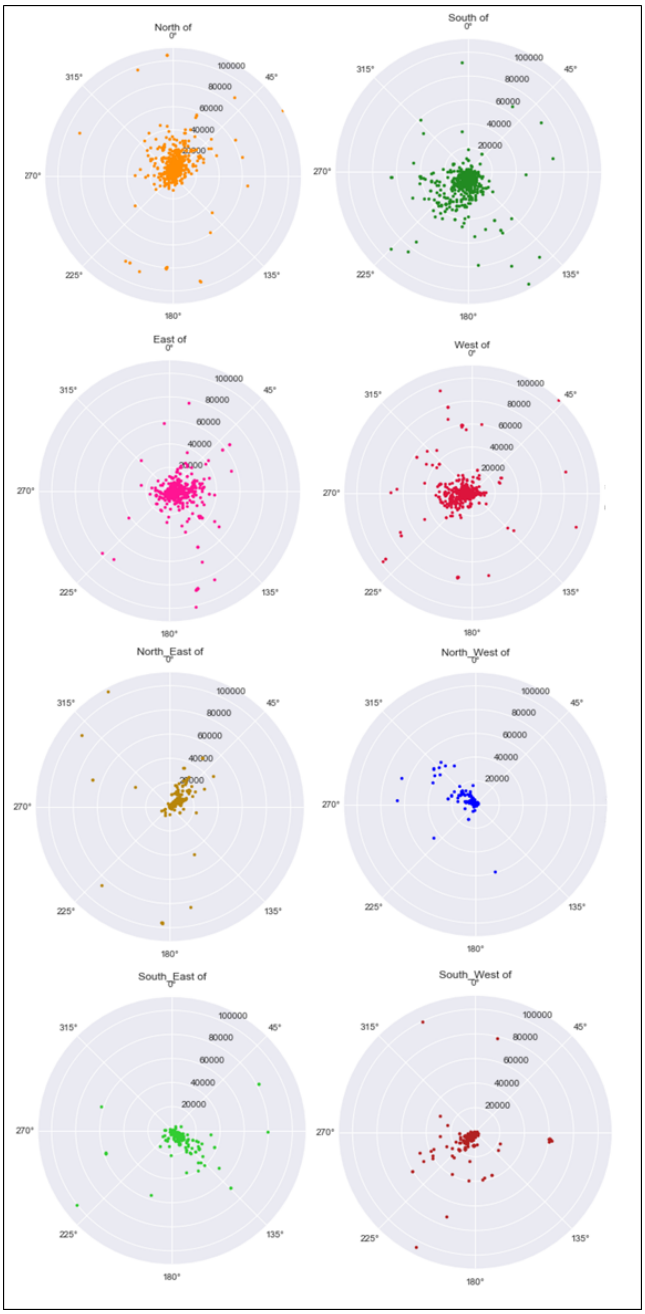
29 Kristin Stock and Javid Yousaf. Context-aware automated interpretation of elaborate natural language descriptions of location through learning from empirical data. *International Journal of Geographical Information Science*, 32(6):1087–1116, 2018.

30 Jan Oliver Wallgrün, Alexander Klippel, and Timothy Baldwin. Building a corpus of spatial relational expressions extracted from web documents. In *Proceedings of the 8th workshop on geographic information retrieval*, GIR '14, New York, NY, USA, 2014. Association for Computing Machinery. doi:10.1145/2675354.2675702.

31 M. Worboys. Nearness relations in environmental space. *International Journal of Geographic Information Science*, 15(7):633–651, 2001.

32 Xiaobai Yao and Jean-Claude Thill. How far is too far? – A statistical approach to context-contingent proximity modeling. *Transactions in GIS*, 9(2):157–178, 2005.

A Radial scatter plots for cardinal direction prepositional phrases



An Incremental Algorithm for Handling Qualitative Spatio-Temporal Information

Zhiguo Long
School of Computing and Artificial Intelligence, Southwest Jiaotong University, Chengdu, China

Qiyuan Hu
School of Computing and Artificial Intelligence, Southwest Jiaotong University, Chengdu, China

Hua Meng[1]
School of Mathematics, Southwest Jiaotong University, Chengdu, China

Michael Sioutis
Faculty of Information Systems and Applied Computer Sciences, Universität Bamberg, Germany

Abstract

In this paper, we present an online (incremental) algorithm for checking the satisfiability of qualitative spatio-temporal data, with direct implications to other fundamental knowledge representation and reasoning problems for such data, like the problems of deductive closure and redundancy removal. In particular, qualitative data come in the form of human-like, symbolic, descriptions such as "region x *contains* or *overlaps* region y", which are abundant in the Web of Data. Our approach is also able to maintain, to some extent, any sparse graph structure that may be inherent in the data, i.e., it acts parsimoniously and only tries to infer new information when needed for soundness and completeness. To this end, we complement our practical algorithm with certain theoretical results to assert its correctness and efficiency. A subsequent evaluation with publicly available large-scale real-world and random datasets against the state of the art, shows the interest and promise of our method.

2012 ACM Subject Classification Theory of computation → Constraint and logic programming; Computing methodologies → Temporal reasoning; Computing methodologies → Spatial and physical reasoning

Keywords and phrases Online algorithm, qualitative data, spatio-temporal reasoning, satisfiability checking, knowledge representation and reasoning

Digital Object Identifier 10.4230/LIPIcs.COSIT.2022.5

Supplementary Material *Software (Source Code)*: `https://github.com/ZhiguoLong/DPCI_code`

Funding This work was supported by the National Natural Science Foundation of China (61806170), the Humanities and Social Sciences Fund of Ministry of Education (18XJC72040001), and the National Key Research and Development Program of China (2019YFB1706104).

1 Introduction

Real-world spatially or temporally annotated Big Data, largely due to their huge size, naturally pose significant technical challenges in terms of implementing efficient tools for associated practical tasks such as verification, repair, and visualization.[2][3] An Artificial Intelligence area that aims to handle such challenges relating to spatio-temporal data is Qualitative Spatio-Temporal Reasoning (QSTR) [19]. QSTR abstracts away the exact metric information that may be associated with such data, e.g., geometries or time points, and instead uses some natural language, qualitative, descriptions, e.g., *is right of*, *before*, or *inside*, to represent and reason about the data [19, 27]. This qualitative approach has applications

[1] Corresponding author
[2] `https://catalog.ldc.upenn.edu/LDC2006T08`
[3] `http://gadm.geovocab.org/`

in a plethora of areas and domains that include cognitive robotics [9], spatio-temporal design [28], qualitative model generation from video [6], ambient intelligence [3], visual sensemaking [29], and qualitative case-based reasoning and learning [14], to name a few. A survey of representation languages, called qualitative constraint languages (or calculi) in QSTR, for various aspects of space and time, e.g., orientation or intervals, appears in [7].

Motivation

We are particularly interested in offering a parsimonious and incremental (online) algorithm for verifying, or, more formally, *checking the admissibility/satisfiability* of big qualitative spatio-temporal data of dynamic nature, which is an NP-hard task in the general case for most spatio-temporal calculi [7]. As satisfiability checking lies at the heart of most (if not all) reasoning tasks for such data, this algorithm also has direct implications to other fundamental knowledge representation and reasoning problems for the data, such as the problems of *deductive closure (minimal labeling)* and *redundancy removal* [25]; these are in fact polynomial-time Turing reducible to the satisfiability checking problem [12]. In sum, deductive closure concerns identifying all the information that either is or can be inferred to be true/valid and subsequently building a *minimal* knowledge base (KB), and redundancy removal concerns removing any information that can be inferred from the rest of the data and and subsequently building a non-redundant or *prime* KB. Clearly, these problems also relate to one another [18]. As noted in [13], a minimal KB is a quite useful knowledge compilation, since it can allow one to answer some queries in polynomial time that would otherwise be (at least) NP-hard. Indeed, in the context of QSTR, for instance, one could exploit minimality of a qualitative spatio-temporal KB to immediately deduce whether some task a should be scheduled before another task b. A non-redundant KB is particularly important in cases where we want to have our knowledge to be as concise as possible, yet without sacrificing essential information, like in the cases of pattern discovery or search in data mining [15, 16]; for instance, in [15] qualitative temporal data signatures are used for sepsis prediction, and explanation thereof, in intensive care medicine, and the authors identify the need to *optimize* such data with redundancy removal.

Related Work

We build upon the work of [26], where a technique for checking the satisfiability of a particular, so called *distributive* [21], class of qualitative spatio-temporal data is proposed, that performs a single pass over the spatial or temporal entities in the data and hence results in a dramatic performance boost with respect to the state of the art. That technique is based on a notion of weak local consistency, called $\overleftarrow{\diamond}_G$-consistency (directional partial path consistency with *weak composition* [20], to be detailed in Section 2), and was subsequently used as the backbone of a stronger consistency, called $\diamond_G$-consistency, in a later work [22]. However, the aforementioned technique is static in nature, i.e., it solely operates on fixed input data, which dramatically limits its applicability in real-world, dynamic, and time-critical situations; clearly, this limitation extends to other methods that rely on it, such as the one in [22] mentioned earlier. Here, we address this limitation and propose an online variant of the technique in [26] that is able to perform on-demand satisfiability checking of dynamic, incrementally-available, qualitative spatio-temporal data. Our approach mirrors to some extent that of [4], where an online algorithm is proposed for satisfiability checking of *quantitative* temporal data. Our own method is firstly distinguished by the fact that the involved consistency notions and operations are tailored to handle infinite domains and qualitative relations. Further, in our

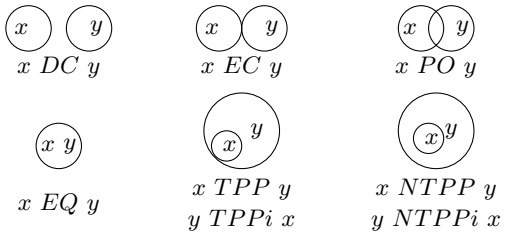


Figure 1 A $2D$ representation of the 8 base relations of RCC8, each one relating two potential regions x and y as in $x\ b\ y$; here, bi denotes the converse of b (formally b^{-1}).

work we handle more generic updates to existing (past) information, i.e., not just refinements of such information as in [4], but information that may be completely different to (or even contradicting) what exists in the KB. In our discussion, we focus on the spatial calculus of RCC8 [23] (to be detailed in Section 2) as far as examples, datasets, and evaluations are concerned, but it must be noted that the approach presented here is generic and applies to most widely adopted qualitative constraint languages, as they are listed in [7] (we explicitly state the required properties for a calculus in Section 2).

Contributions

In this paper, we make the following contributions: (i) we present an online and parsimonious algorithm for checking the satisfiability of qualitative spatio-temporal data, and recall certain implications to other fundamental knowledge representation and reasoning problems for such data, viz., deductive closure and redundancy removal, (ii) we establish the correctness of our algorithm with some theoretical results that relate to both its decision (output) and the triangulation technique that it uses to maintain any sparsity that may be inherent in the data, and (iii) we implement the algorithm and compare it against the state of the art with real-world and random datasets that scale up to thousands of variables, and subsequently demonstrate the efficiency of our approach.

The rest of the paper is organized as follows. In Section 2 we introduce some necessary terminology and notations that are followed thoughout the paper. In Section 3 we present our algorithm along with the results that establish its soundness and completeness. In Section 4 we evaluate an implementation of our algorithm against the state of the art. Finally, in Section 5 we conclude with a discussion and some directions for future work.

2 Preliminaries

A *binary* qualitative constraint language is based on a finite set B of *jointly exhaustive and pairwise disjoint* relations, called the set of *base relations* (*atoms*), that is defined over an infinite domain D (e.g., some topological space) [20]. These base relations represent definite knowledge between two entities of D; indefinite knowledge can be specified by a union of possible base relations, and is represented by the set containing them. The set B contains the identity relation Id, and is closed under the *converse* operation $(^{-1})$. The entire set of relations 2^{B} is equipped with the set-theoretic operations of union and intersection, the converse operation, and the *weak composition* operation denoted by $\diamond$ [20]. The weak composition $(\diamond)$ of two base relations $b, b' \in \mathsf{B}$ is the smallest relation $r \in 2^{\mathsf{B}}$ that includes $b \circ b'$; formally, $b \diamond b' = \{b'' \in \mathsf{B} : b'' \cap (b \circ b') \neq \emptyset\}$, where $b \circ b' = \{(x, y) \in \mathsf{D} \times \mathsf{D} : \exists z \in \mathsf{D}$ such that $(x, z) \in b \wedge (z, y) \in b'\}$. Finally, for all $r \in 2^{\mathsf{B}}$, $r^{-1} = \bigcup\{b^{-1} : b \in r\}$, and for all $r, r' \in 2^{\mathsf{B}}$, $r \diamond r' = \bigcup\{b \diamond b' : b \in r, b' \in r'\}$.

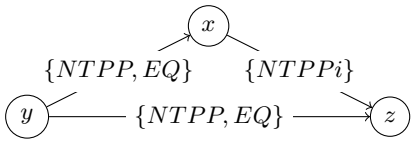


Figure 2 Illustration of a QCN $\mathcal{N}$ and $\overleftarrow{\diamond}_G$-consistency: $\mathcal{N}$ is $\overleftarrow{\diamond}_G$-consistent w.r.t. the ordering (z, y, x), i.e., x is "eliminated" first and z last, but not w.r.t. the ordering (x, y, z) (the base relation EQ between variables x and y would have to be removed).

As an illustration, consider the Region Connection Calculus (RCC), which is a first-order theory for representing and reasoning about mereotopological information [23]. The domain D of RCC comprises all possible non-empty regular closed subsets of some topological space [24]; these subsets serve as regions in RCC. A fragment of the Region Connection Calculus (RCC), denoted by RCC8, is the dominant qualitative spatial constraint language for representing and reasoning about qualitative spatial information [23]. In particular, RCC8 makes use of the mereotopological relations *disconnected* (DC), *externally connected* (EC), *equal* (EQ), *partially overlapping* (PO), *tangential proper part* (TPP) and its inverse ($TPPi$), and *non-tangential proper part* ($NTPP$) and its inverse ($NTPPi$) to encode knowledge about the spatial relations between two potential regions, as depicted in Figure 1. Relation EQ is the identity relation Id of RCC8.

The problem of representing and reasoning about qualitative spatial (or temporal) information may be tackled via the use of a *Qualitative Constraint Network*, defined in the following manner:

▶ **Definition 1** (QCN). *A* qualitative constraint network (QCN) *is a tuple* (V, C) *where:*

- $V = \{v_1, \ldots, v_n\}$ *is a non-empty finite set of variables, each representing an entity of an infinite domain* D*;*
- *and* C *is a mapping* $C : V \times V \to 2^{\mathsf{B}}$ *such that* $C(v, v) = \{\mathsf{Id}\}$ *for all* $v \in V$ *and* $C(v, v') = C(v', v)^{-1}$ *for all* $v, v' \in V$*.*

An example of a QCN is shown in Figure 2; for clarity, neither converse relations nor Id loops are shown in the figure, but they are part of any QCN.

▶ **Definition 2.** *Let* $\mathcal{N} = (V, C)$ *be a* QCN*, then:*

- *a* solution *of* $\mathcal{N}$ *is a mapping* $\sigma : V \to \mathsf{D}$ *such that* $\forall v, v' \in V$*,* $\exists b \in C(v, v')$ *such that* $(\sigma(v), \sigma(v')) \in b$*, and* $\mathcal{N}$ *is* satisfiable *iff it admits a solution;*
- $\mathcal{N}$ *is* trivially inconsistent*, denoted by* $\emptyset \in \mathcal{N}$*, iff* $\exists v, v' \in V$ *such that* $C(v, v') = \emptyset$*;*
- *a* sub-QCN $\mathcal{N}'$ *of* $\mathcal{N}$*, denoted by* $\mathcal{N}' \subseteq \mathcal{N}$*, is a* QCN (V, C') *such that* $C'(v, v') \subseteq C(v, v')$ $\forall v, v' \in V$*;*
- $\mathcal{N}$ *is* atomic *iff* $\forall v, v' \in V$*,* $C(v, v') = \{b\}$ *with* $b \in \mathsf{B}$*;*
- *the* constraint graph *of* $\mathcal{N}$*, denoted by* $\mathsf{G}(\mathcal{N})$*, is the graph* (V, E) *where* $\{v, v'\} \in E$ *iff* $C(v, v') \neq \mathsf{B}$ *and* $v \neq v'$*;*

Given a QCN $\mathcal{N} = (V, C)$, a relation $C(u, v) = R$ of $\mathcal{N}$ can be denoted by $\mathcal{N}[u, v]$ and $(u\ R\ v)$ too, if it facilitates presentation.

We recall the definition of $\overleftarrow{\diamond}_G$-consistency [26], which is a fundamental local consistency for reasoning with QCNs. For simplicity, in what follows, a set of variables $V = \{v_1, v_2, \ldots, v_n\}$ and an ordering $(v_1, v_2, \ldots, v_n)$ is implied whenever a bijection $\alpha : V \to \{1, 2, \ldots, n\}$ is defined.

▶ **Definition 3** ($\overleftarrow{\diamond}_G$-consistency). *A* QCN $\mathcal{N} = (V, C)$ *is* $\overleftarrow{\diamond}_G$-consistent *with respect to a graph* $G = (V, E)$ *and an ordering* $(\alpha^{-1}(1),\alpha^{-1}(2),\ldots,\alpha^{-1}(n))$ *defined by a bijection* $\alpha : V \to \{1, 2, \ldots, n\}$ *iff for all* $v_i, v_j, v_k \in V$ *such that* $\{v_k, v_i\}, \{v_k, v_j\}, \{v_i, v_j\} \in E$, $\alpha(v_i) < \alpha(v_k)$, *and* $\alpha(v_j) < \alpha(v_k)$ *we have that* $C(v_i, v_j) \subseteq C(v_i, v_k) \diamond C(v_k, v_j)$.

In sum, $\overleftarrow{\diamond}_G$-consistency entails consistency for all *ordered* triples of variables of a QCN that correspond to triangles of a given graph G. This ordering can be specified by a bijection between the set of the variables of a QCN and a set of integers, and can be chosen randomly, or via an algorithm or some heuristic [26]. Figure 2 shows an example of how $\overleftarrow{\diamond}_G$-consistency may relate to a QCN.

In what follows, we assume that the following two properties for a given qualitative constraint language $\mathcal{L}$ hold; these hold for most well-known calculi [7, 8]:

$\mathcal{L}$ is a relation algebra. (1)

Every atomic $\diamond$-consistent QCN of $\mathcal{L}$ is satisfiable. (2)

These properties are only needed to simplify our algorithms and establish certain syntactic satisfiability conditions that allow us to be consistent with our claims. However, they can be further relaxed to accommodate some more "exotic" calculi, subject to complicating the algorithms and related operations and conditions of course.

3 Approach

In Algorithm 1 we present our online method for checking the satisfiability of a spatial or temporal QCN by means of $\overleftarrow{\diamond}_G$-consistency; we call this method DPCI. First, we briefly describe how DPCI works, and later we establish the assumptions under which it is sound and complete.

The basic idea of DPCI is that, when a constraint $C(v_i, v_k)$ in a $\overleftarrow{\diamond}_G$-consistent QCN is updated, we often do not need to run a full pass of the static algorithm for enforcing $\overleftarrow{\diamond}_G$-consistency as in [26]. Instead, an updated constraint $C(v_i, v_k)$ ($\alpha(v_i) < \alpha(v_k)$) may affect the constraints $C(v_i, v_j)$ where $\{v_k, v_i\}, \{v_k, v_j\}, \{v_i, v_j\} \in E$ and $\alpha(v_i) < \alpha(v_j) < \alpha(v_k)$, and any further updates might propagate in this way. In addition, if a constraint $C(v_i, v_j)$ in a $\overleftarrow{\diamond}_G$-consistent QCN is updated, then it might invalidate the previously established relation $C(v_i, v_j) \subseteq C(v_i, v_k) \diamond C(v_k, v_j)$ for some k s.t. $\{v_k, v_i\}, \{v_k, v_j\}, \{v_i, v_j\} \in E$ and $\alpha(v_i) < \alpha(v_j) < \alpha(v_k)$; this must be checked before propagating the updates.

As DPCI is an online algorithm, it assumes that some input has already been processed piece-by-piece in a serial fashion, and specifically that a $\overleftarrow{\diamond}_G$-consistent QCN has been formed. Clearly, in the base case, the initial QCN would not contain any constraints (or, equivalently, every constraint would be defined by a universal relation), which would make the QCN by default $\overleftarrow{\diamond}_G$-consistent. This base case of constructing an $\overleftarrow{\diamond}_G$-consistent QCN from scratch is presented in Algorithm 2. Specifically, a QCN $\mathcal{N}$ with no constraints is initialized in line 1 of the algorithm, which is then continuously updated with new constraints via calls to DPCI (lines 3–6); it is assumed that the new constraints are taken from some QCN $\mathcal{M}$ that is defined on the same set of variables as $\mathcal{N}$. The number m of m unprocessed constraints in line 4 may vary from 1 to the total number of (non-universal) constraints in $\mathcal{N}$. When $m = 1$, the simulation acts in a fully online manner, since it adds constraints 1 by 1, and when $m =$ *total number of constraints*, the simulation acts in a fully static manner, since it degenerates into a single application of DPCI. Next, we continue to describe how a call to DPCI behaves.

DPCI receives as input a $\overleftarrow{\diamond}_G$-consistent QCN $\mathcal{N} = (V, C)$ w.r.t. a graph $G = (V, E)$ and an ordering $(\alpha^{-1}(1),\alpha^{-1}(2),\ldots,\alpha^{-1}(n))$ defined by a bijection $\alpha : V \to \{1, 2, \ldots, n\}$, and a set of additional constraints C'. In the first two lines of the algorithm we introduce a dictionary

Algorithm 1 DPCI($\mathcal{N}, G, \alpha, C'$): Incremental Directional Path Consistency Algorithm.

Input: A $\overleftarrow{\diamond}_G$-consistent QCN $\mathcal{N} = (V, C)$ w.r.t. a graph $G = (V, E)$ and an ordering $(\alpha^{-1}(1),\alpha^{-1}(2),\ldots,\alpha^{-1}(n))$ defined by a bijection $\alpha : V \to \{1, 2, \ldots, n\}$; and a set of constraints C' of the form $(v_{i_0}\ R'_{i_0,j_0}\ v_{j_0})$, where $\{v_{i_0}, v_{j_0}\} \in E' \subseteq V \times V$.
Output: True or False, the updated graph G, and the updated QCN $\mathcal{N}$.

1 **foreach** $v_i \in V$ **do**
2 $P[v_i] \leftarrow \emptyset$;
3 $Q \leftarrow \emptyset$;
4 **foreach** $(v_{i_0}\ R'_{i_0,j_0}\ v_{j_0}) \in C'$ **do**
5 **foreach** v_k *s.t.* $\{v_{i_0}, v_k\}, \{v_k, v_{j_0}\} \in E$, $\alpha(v_k) > \alpha(v_{j_0}) > \alpha(v_{i_0})$ **do**
6 $R'_{i_0,j_0} \leftarrow R'_{i_0,j_0} \cap (\mathcal{N}[v_{i_0}, v_k] \circ \mathcal{N}[v_k, v_{j_0}])$;
7 **if** $R'_{i_0,j_0} = \emptyset$ **then**
8 **return** (False, $\emptyset, \emptyset$)
9 **if** $R'_{i_0,j_0} \neq \mathcal{N}[v_{i_0}, v_{j_0}]$ **then**
10 $\mathcal{N}[v_{i_0}, v_{j_0}] = R'_{i_0,j_0}$;
11 $\mathcal{N}[v_{j_0}, v_{i_0}] = R'_{j_0,i_0}$;
12 $P[v_{j_0}] \leftarrow P[v_{j_0}] \cup \{v_{i_0}\}$;
13 $Q \leftarrow Q \cup \{\alpha(v_{j_0})\}$;
14 **while** $Q \neq \emptyset$ **do**
15 $p \leftarrow \max(Q)$;
16 $v_k \leftarrow \alpha^{-1}(p)$;
17 $Q \leftarrow Q \setminus \{p\}$;
18 **foreach** $\{v_i, v_j\}$ *s.t.* $\alpha(v_i) < \alpha(v_j) < \alpha(v_k) \wedge \{v_i, v_k\}, \{v_j, v_k\} \in E \wedge (v_i \in P[v_k]$ *or* $v_j \in P[v_k])$ **do**
19 $T_{ij} \leftarrow \mathcal{N}[v_i, v_j] \cap (\mathcal{N}[v_i, v_k] \diamond \mathcal{N}[v_k, v_j])$;
20 $E \leftarrow E \cup \{\{v_i, v_j\}\}$;
21 **if** $T_{ij} = \emptyset$ **then**
22 **return** (False, $\emptyset, \emptyset$);
23 **if** $T_{ij} \neq \mathcal{N}[v_i, v_j]$ **then**
24 $\mathcal{N}[v_i, v_j] \leftarrow T_{ij}$;
25 $\mathcal{N}[v_j, v_i] \leftarrow T_{ij}^{-1}$;
26 $Q \leftarrow Q \cup \{\alpha(v_j)\}$;
27 $P[v_j] \leftarrow P[v_j] \cup \{v_i\}$;
28 **return** (True, $G, \mathcal{N}$).

of lists P to keep track of affected edges (corresponding to constraints) that need to be processed. Every key in the dictionary corresponds to a vertex $v \in V$ and maps to a list, and the edges that are incident on v *and* need to be processed are stored in the form of adjacent vertices to v in the respective list. In lines 4–13 the new constraints are added to the QCN. Specifically, in lines 5–8 a simple preprocessing takes place to ensure that no inconsistency is introduced, and in lines 9–13, in lack of such inconsistency, we move on and perform the constraint updates on the QCN and store all affected edges to be processed later on. Here, we also keep track of where the update occured with respect to our ordering, by pushing the respective index information into a queue Q (line 13). The main functionality of the algorithm occurs in lines 14–27. We start from the highest index in our ordering where an earlier corresponding constraint update occurred (line 15), and propagate the constrainedness of that update to constraints that are associated with variables that are indexed earlier in

Algorithm 2 FullSimulation$(\mathcal{M}, G, \alpha)$: Simulating the run of Algorithm 1 from scratch.

Input: A QCN $\mathcal{M} = (V, C)$, a graph $G = (V, E)$, and an ordering $(\alpha^{-1}(1), \alpha^{-1}(2), \ldots, \alpha^{-1}(n))$ defined by a bijection $\alpha : V \to \{1, 2, \ldots, n\}$.
Output: True or False, the updated graph G, and a new QCN $\mathcal{N}$.

1 $\mathcal{N} \leftarrow$ a QCN (V, C^*) s.t. $C^*(v, v') = \mathsf{B}$ for all $v, v' \in V$ with $v \neq v'$;
2 decision $\leftarrow$ True;
3 **while** $\exists$ *unprocessed constraints* $(v_i\ R_{i,j}\ v_j) \in C \wedge$ *decision* $\neq$ False **do**
4 Choose a subset C' of m unprocessed constraints $(v_i\ R_{i,j}\ v_j) \in C$;
5 Mark the chosen constraints as processed;
6 (decision, G, $\mathcal{N}$) $\leftarrow$ DPCI$(\mathcal{N}, G, \alpha, C')$;
7 **return** $(decision, G, \mathcal{N})$.

the ordering; as a reminder, this is the definition of $\overleftarrow{\diamond}_G$-consistency (see again Definition 3). It is important to note that when an edge $\{v_i, v_k\}$ or $\{v_j, v_k\}$ of G has been affected and we have that $\alpha(v_i) < \alpha(v_j) < \alpha(v_k)$, i.e., v_i is indexed earlier than v_j and v_j is indexed earlier than v_k in the ordering, we *must* also consider the edge $\{v_i, v_j\}$ in order for the constraints to propagate soundly with respect to enforcing $\overleftarrow{\diamond}_G$-consistency (line 20). Clearly, whenever further constraint updates occur, and no inconsistency is reported (lines 21–22), the information about the affected edges and the respective indices are maintained in the way that we have explained so far. This process is repeated until no more indices that correspond to constraint updates exist (Q empties), and consequently no more affected edges exist either.

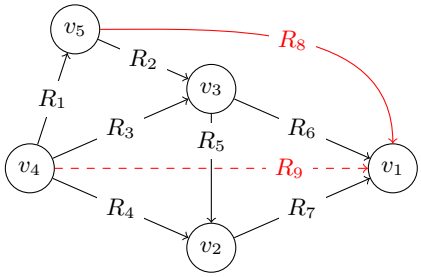


Figure 3 DPCI illustration.

▶ **Example 4.** Figure 3 illustrates how DPCI works. Initially, the QCN $\mathcal{N}$ contains the constraints shown as labelled black arrows in the figure, e.g., $(v_4\ R_1\ v_5)$, and it is $\overleftarrow{\diamond}_G$-consistent w.r.t. the ordering $(v_1, v_2, v_3, v_4, v_5)$. A new and different constraint $(v_5\ R_8\ v_1)$ is then added to the QCN, which is the same as updating the constraint between v_5 and v_1. Since there is no v_k satisfying the conditions in line 5, the loop in lines 5–8 will be skipped. The condition in line 9 is then satisfied by $R_8 \neq \mathcal{N}[v_{i_0}, v_{j_0}]$. Therefore, $\mathcal{N}[v_{i_0}, v_{j_0}]$ is updated to R_8, $P[v_5]$ is set to $\{v_1\}$, and $Q = \{5\}$. The process then moves to the loop in lines 14–27. In Q there is currently only one element, and $v_k = v_5$. The edges $\{v_i, v_j\}$ satisfying the conditions in line 18 are $\{v_1, v_3\}$ and $\{v_1, v_4\}$, because $P[v_5] = \{v_1\}$. Note that the edge $\{v_3, v_4\}$ will not be processed here because $v_3, v_4 \notin P[v_1]$, which is one of the benefits of our online algorithm. The contraints for the edges $\{v_1, v_3\}$ and $\{v_1, v_4\}$ may then get updated, and the edge $\{v_1, v_4\}$ is added to the graph (see the dashed edge). Suppose that only the constraint for $\{v_1, v_4\}$ is changed, then $\alpha(v_4) = 4$ is added to Q, and $P[v_4]$ is changed to $\{v_1\}$. In the next iteration of the loop, v_k would be v_4 and a similar process would be executed. On the other hand, if the initial QCN includes all the constraints shown

in Fig. 3, and the constraint $(v_4\ R_9\ v_1)$ is updated by another new constraint $(v_4\ R'_9\ v_1)$, then $R'_9 \leftarrow R'_9 \cap (R_1 \circ R_8)$ needs to be calculated. This is because we need to ensure that the updated $\mathsf{QCN}\ \mathcal{N}$ is $\overleftarrow{\diamond}_G$-consistent w.r.t. the ordering $(v_1, v_2, v_3, v_4, v_5)$, and thus for v_4, v_1, v_5 we should have $\mathcal{N}[v_4, v_1] \subseteq \mathcal{N}[v_4, v_5] \circ \mathcal{N}[v_5, v_1]$ (see lines 5–8).

Next, we recall and prove some results to establish the correctness of our algorithm. First, we introduce some necessary graph theoretic and other necessary concepts.

▶ **Definition 5** (PEO). *Given an undirected graph $G = (V, E)$ and an ordering $(\alpha^{-1}(1), \alpha^{-1}(2),\ldots,\alpha^{-1}(n))$ defined by a bijection $\alpha : V \to \{1, 2, \ldots, n\}$, let $F_k = \{v_j \in \mathsf{adj}(v_k) : \alpha(j) < \alpha(k)\}$; the ordering is a* perfect elimination ordering (PEO) *if and only if F_k induces a complete subgraph of G for every k.*

We can now recall the following result:

▶ **Lemma 6** ([11]). *A graph G is chordal iff G admits a PEO.*

A recall of *distributive subclasses* of relations is also required [21].

▶ **Definition 7.** *A* subclass *of relations is a subset $\mathcal{A} \subseteq 2^{\mathsf{B}}$ that contains the singleton relations of 2^{B} and is closed under converse, intersection, and weak composition.*

Given three relations r, r', $r'' \in 2^{\mathsf{B}}$, we say that weak composition distributes over intersection if we have that $r \diamond (r' \cap r'') = (r \diamond r') \cap (r \diamond r'')$ and $(r' \cap r'') \diamond r = (r' \diamond r) \cap (r'' \diamond r)$.

▶ **Definition 8** (distributive subclass). *A subclass $\mathcal{A} \subseteq 2^{\mathsf{B}}$ is* distributive *iff weak composition distributes over non-empty intersections for all relations r, r', $r'' \in \mathcal{A}$.*

We will use the following result to link $\overleftarrow{\diamond}_G$-consistency to chordal graphs, which are graphs that may allow for retaining the sparsity of a QCN [26]:

▶ **Proposition 9** ([26]). *Let $\mathcal{N} = (V, C)$ be a QCN defined over a distributive subclass of relations of a qualitative constraint language that satisfies properties 1 and 2, and $G = (V, E)$ a chordal graph s.t. $\mathsf{G}(\mathcal{N}) \subseteq G$ and $(\alpha^{-1}(1),\alpha^{-1}(2),\ldots, \alpha^{-1}(n))$ defined by a bijection $\alpha : V \to \{1, 2, \ldots, n\}$ is a PEO of G. If $\mathcal{N}$ is $\overleftarrow{\diamond}_G$-consistent w.r.t the PEO and $\emptyset \notin \mathcal{N}$, then $\mathcal{N}$ is satisfiable.*

We are ready to introduce our novel results.

▶ **Lemma 10.** *Let $G = (V, E)$ be a chordal graph, $(\alpha^{-1}(1),\alpha^{-1}(2),\ldots,\alpha^{-1}(n))$ defined by a bijection $\alpha : V \to \{1, 2, \ldots, n\}$ a PEO of G, and $|V| = n$. If E' is a set of new edges on V, and $G^{(n+1)} = (V, E^{(n+1)})$, where $E^{(n+1)} = E \cup E'$, and $E^{(k)} = E^{(k+1)} \cup E_k$, where $E_k = \{\{v_i, v_j\} : \{v_i, v_k\}, \{v_k, v_j\} \in E^{(k+1)}$ and $\alpha(v_i) < \alpha(v_j) < \alpha(v_k)\}$, then $G^{(1)} = (V, E^{(1)})$ is also a chordal graph with a same PEO.*

Proof. By Lemma 6, we only need to show that for any $v_k (k = n, \ldots, 1)$, the induced subgraph of $G^{(1)}$ on the set $F_k = \{v_i : \{v_i, v_k\} \in E^{(n)}$ and $\alpha(v_k) > \alpha(v_i)\}$ is a complete graph. In fact, by the definition of E_k, for any two vertices $v_i, v_j \in F_k$ s.t. $\alpha(v_i) < \alpha(v_j)$, there will be an edge $\{v_i, v_j\}$ in E_k. Since $E_k \subseteq E^{(1)}$, we know that the subgraph of $G^{(1)}$ on F_k is complete. Therefore, $G^{(1)} = (V, E^{(1)})$ is chordal and admits a same PEO. ◀

The previous lemma tells us that after adding new edges to a chordal graph G, we can easily construct a new chordal graph $G^{(1)}$ containing all the edges of G and maintain its PEO, by simply making sure that the induced subgraph on every F_k is complete. The following theorem shows the correctness of Algorithm 1.

▶ **Theorem 11.** *Let $\mathcal{N} = (V, C)$ be a* QCN *defined over a distributive subclass of relations of a qualitative constraint language that satisfies properties 1 and 2, $G = (V, E)$ a chordal graph s.t.* $\mathsf{G}(\mathcal{N}) \subseteq G$ *and* $(\alpha^{-1}(1), \alpha^{-1}(2), \ldots, \alpha^{-1}(n))$ *defined by a bijection $\alpha : V \to \{1, 2, \ldots, n\}$ is a PEO of G, and C' a set of new constraints of the form $(v_{i_0}\ R'_{i_0,j_0}\ v_{j_0})$, where $\{v_{i_0}, v_{j_0}\} \in E' \subseteq V \times V$ and $\alpha(v_{i_0}) < \alpha(v_{j_0})$. If $\mathcal{N}$ is $\overleftarrow{\diamond}_G$-consistent w.r.t the PEO and $\emptyset \notin \mathcal{N}$, then algorithm* DPCI *terminates and returns* $(\mathsf{True}, G', \mathcal{N}')$*, where $G' \supseteq G$ is still chordal and admits a same PEO, and $\mathcal{N}' \subseteq \mathcal{N}$ is $\overleftarrow{\diamond}_G$-consistent w.r.t. the PEO, if and only if $\mathcal{N} \setminus \{(v_{i_0} R_{i_0,j_0} v_{j_0}) : \{v_{i_0}, v_{j_0}\} \in E'\} \cup C'$ is satisfiable.*

Proof. The "only if" part is obvious by Proposition 9. For the "if" part we reason as follows. The algorithm returns False only when there is some v_i, v_j, v_k, where $\alpha(v_i) < \alpha(v_k)$, $\alpha(v_j) < \alpha(v_k)$, such that $\mathcal{N}[v_i, v_j] \cap (\mathcal{N}[v_i, v_k] \circ \mathcal{N}[v_k, v_j]) = \emptyset$. If $\mathcal{N} \setminus \{(v_{i_0} R_{i_0,j_0} v_{j_0}) : \{v_{i_0}, v_{j_0}\} \in E'\} \cup C'$ is satisfiable, then by Proposition 9 this cannot happen. Next, we assume that the algorithm returns True.

First, we show that the returned graph $G' \supseteq G$ is chordal. If we update G in the manner exactly as in Lemma 10 after adding E' and obtain G', then by Lemma 10 we know that G' is also chordal with a same PEO. In the sequel, we show that G' is exactly $G^{(1)}$ as defined in Lemma 10. Let $M^{(n+1)} = E \cup E'$ and $M^{(k)} = M^{(k+1)} \cup M_k$, where M_k is the set of edges added in line 20 in the iteration for v_k. We want to prove $M^{(k)} = E^{(k)}$ (with $E^{(k)}$ as in Lemma 10) for any k. Suppose by induction that $M^{(k)} = E^{(k)}$ holds for $u + 1 \leq k \leq n$; we will show $M^{(u)} = E^{(u)}$. To this end, we show that $\{v_i, v_u\} \in M^{(u)}$ for any $v_i, v_j \in V$ s.t. $\{v_i, v_u\}, \{v_u, v_j\} \in E^{(u+1)}$ and $\alpha(v_i) < \alpha(v_j) < \alpha(v_u)$. In fact, if the condition in line 18 of the algorithm is satisfied, i.e., $v_i \in P[v_u]$ or $v_j \in P[v_u]$, then $\{v_i, v_u\} \in M_u \subseteq M^{(u)}$. If $v_i \notin P[v_u]$ or $v_j \notin P[v_u]$, by the definition of $P[v_u]$ in the algorithm, we can see that $\{v_i, v_u\}$ and $\{v_j, v_u\}$ must have been in $M^{(u+1)} \subseteq M^{(u)}$ already. Therefore, in either case, $\{v_i, v_u\} \in M^{(u)}$ and thus $E^{(u)} \subseteq M^{(u)}$. Note that it is easy to see $M^{(u)} \subseteq E^{(u)}$, because the condition to add edges to M_u is stronger than to include an edge in E_u and thus $M_u \subseteq E_u$. In this way, we showed that the chordal graph $G^{(1)} = (V, E^{(1)})$ as defined in Lemma 10 is the same as the graph $G' = (V, M^{(1)})$ returned by the algorithm. Therefore, G' is also chordal and with a same PEO as G.

To show that the returned $\mathcal{N}'$ is $\overleftarrow{\diamond}_G$-consistent w.r.t. the PEO, suppose v_i, v_j, $v_k \in V$, $\{v_i, v_k\}, \{v_k, v_j\} \in E^{(n)}$, and $\alpha(v_i) < \alpha(v_j) < \alpha(v_k)$; we will show $\mathcal{N}'[v_i, v_j] \subseteq \mathcal{N}'[v_i, v_k] \circ \mathcal{N}'[v_k, v_j]$. If $v_i \in P[v_u]$ or $v_j \in P[v_u]$, then by the operations in lines 19, 24, and 25 of the algorithm, it is easy to see $\mathcal{N}'[v_i, v_j] \subseteq \mathcal{N}'[v_i, v_k] \circ \mathcal{N}'[v_k, v_j]$. If $v_i \notin P[v_u]$ or $v_j \notin P[v_u]$, then $\mathcal{N}[v_i, v_j] \subseteq \mathcal{N}'[v_i, v_k] \circ \mathcal{N}'[v_k, v_j]$ already holds from the beginning, because $\mathcal{N}'[v_i, v_k]$ and $\mathcal{N}'[v_i, v_k]$ are not changed by the algorithm, and $\mathcal{N}'[v_i, v_j]$ can only be refined (w.r.t. $\subset$) by the algorithm. Therefore, the returned $\mathcal{N}'$ is $\overleftarrow{\diamond}_G$-consistent w.r.t. to the PEO. ◀

We invite the reader to view Corollaries 3.2 and 3.3 in [26] for implications of the satisfiability checking result of Theorem 11 here to the fundamental knowledge representation and reasoning problems of *deductive closure (minimal labeling)* and *redundancy removal* respectively for spatial or temporal QCNs. The implications are direct, since we already mentioned in the introduction that these problems are polynomial-time Turing reducible to the satisfiability checking problem [12].

Complexity

Algorithm DPCI has a runtime of $O((|C'| + \Delta^2)|V|)$ for a given QCN $\mathcal{N} = (V, C)$ and a set of (new) constraints C', where Δ is the maximum vertex degree of the graph G returned through its output. In the worst case, the constraint updates may trigger a full-pass application of

$\overleftarrow{\diamond}_G$-consistency on the entire QCN, thus emulating the static algorithm of [26] with a runtime of $O(\Delta^2|V|)$. Specifically, if $\max(Q) = n$, i.e., the variable with order n is in Q, then a full-pass application of $\overleftarrow{\diamond}_G$-consistency will be triggered. Our algorithm also has some overhead costs for preprocessing the new constraints in C' (lines 4–13). In the worst case, for each of the new constraints, all vertices $v_k \in V$ satisfy the conditions in line 5, which in total will take $O(|C'||V|)$ time. Thus, the overall runtime of DPCI is $O((|C'| + \Delta^2)|V|)$.

4 Experimental Evaluation

We make use of the publicly available datasets from previous studies in QSTR, including real-world datasets from [18], i.e., `Footprint-`k (`F-`k, $k = 1, \ldots, 6$) and `StatArea-`k (`SA-`k, $k = 1, \ldots, 6$), and synthetic datasets from [26]. The datasets `F-1` to `F-6` have 108, 217, 434, 867, 1 736, and 3 470 variables, respectively, and `SA-1` to `SA-6` have 51, 100, 196, 376, 659, and 1 562 variables, respectively; all of them are satisfiable atomic QCNs of RCC8. The synthetic datasets range from 10^3 to 10^4 variables with a step size of 10^3. For each size n, we have 10 satisfiable QCNs of RCC8 over a distributive subclass of relations, generated using the model BA(n, m = 2) [1] for scale-free networks; networks of this type imitate real-world ones [1, 26]. All evaluations were performed on a PC with Ubuntu 18.04, CPU i7-8700 3.2GHz, RAM 64GB, and Python 3.8, and the time for each n is averaged over 10 tests.

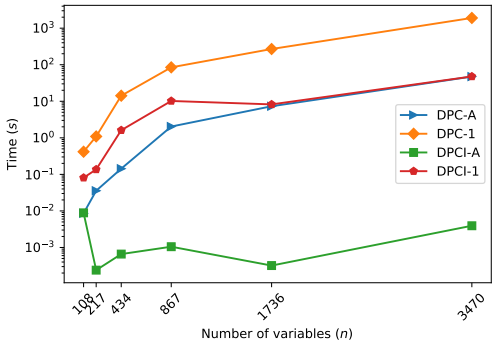
We compare the proposed online algorithm, viz., DPCI, against the state-of-the-art static algorithm of [26] for applying $\overleftarrow{\diamond}_G$-consistency on QCNs, viz., DPC. We note that we consider two ways of processing the additional constraints in C': (i) 1 *by* 1 and (ii) *altogether at once*. Regarding 1 *by* 1, each time we call DPCI we feed it a single new constraint from C', and we repeat this process until no constraints are left in C'; this method of using DPCI is denoted by DPCI-1. Regarding *altogether at once*, we feed all the additional constraints in C' at once to DPCI; this method is denoted by DPCI-A. Likewise, DPC-1 and DPC-A denote the 1 *by* 1 and *altogether at once* ways of processing constraints for the static algorithm, respectively. All of these approaches are analysed in our evaluations.

We want to measure the efficiency of enforcing $\overleftarrow{\diamond}_G$-consistency when new constraints arrive. Thus, we use the runtime as our performance metric. Specifically, we randomly generate 20 new constraints for each QCN, which can correspond to either existing or non-existing edges of each accompanying graph G. As a reminder, when edges do not exist in G, they have to be added to G *and* establish/maintain its chordality.

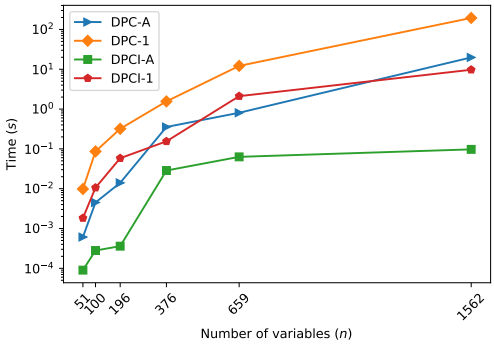
Results and Analysis

The results are shown in Figures 4 and 5, and it is clear that DPCI-A dominates all other approaches for all datasets. This is expected, as DPCI-A only updates part of a network, and processes the new constraints in bulk to avoid repeated updates triggered by individual new constraints. Notably, DPCI-1 is more efficient than DPC-1, by at least about an order of magnitude, which shows that our algorithm can indeed save a lot of time on calculations when single updates are considered. In fact, due to the time saved on calculations even when *repeated* single updates are considered, i.e., constraints are added in 1 *by* 1 fashion, DPCI-1 is also comparable to DPC-A on real-world datasets and more efficient on synthetic ones.

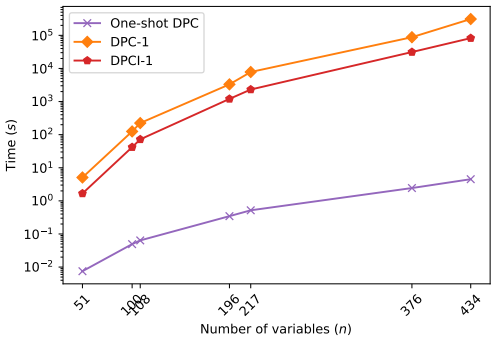
As the number of variables (and hence the number of constraints) increases, the runtime for all methods increases too, except for DPCI-A on the Footprint datasets due (to quick detection of inconsistency). This behaviour verifies that the runtimes of these methods are strongly related to the number of variables. Regarding the performance of DPCI-A on the Footprint datasets, it is due to the fact that these datasets are atomic networks, and are



(a) `Footprint`: Adding new constraints.

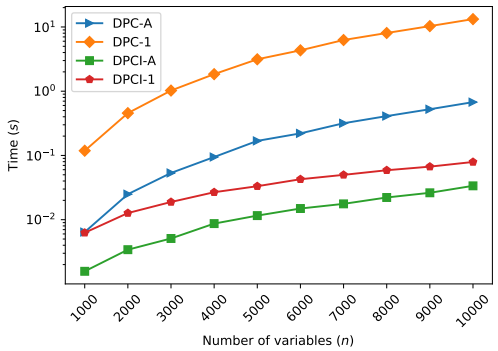


(b) `StatArea`: Adding new constraints.

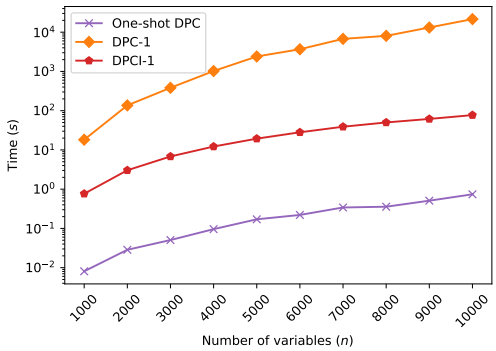


(c) `Footprint` (F-$\{1,\ldots,3\}$) + `StatArea` (SA-$\{1,\ldots,4\}$): Adding original constraints 1 *by* 1.

Figure 4 Results on real-world datasets, where 20 random new constraints are added in 4a and 4b and all original constraints are added in 1 *by* 1 fashion in 4c, i.e., QCNs are built from scratch (basically, Algorithm 2 with $m = 1$ in line 4). One-shot DPC applies $\overleftrightarrow{G}$-consistency on the entire original QCN in a single step (our baseline, which serves as a practical lower bound).



(a) Adding new constraints.



(b) Adding original constraints 1 *by* 1.

Figure 5 Results on synthetic datasets, where 20 random new constraints are added in 5a, and all original constraints are added in 1 *by* 1 fashion in 5b, i.e., QCNs are built from scratch.

hence more likely to become inconsistent when some constraints are altered; DPCI-A can quickly detect this (due to its bulk constraint processing). Such inconsistencies appear very often in real-world data. For example, with the ever-increasing enrichment of the Semantic Web with geospatial data [10, 17], it is often the case that the geometries of geographical

objects are not captured correctly due to contradictory data of different sources. Thus, we can obtain inconsistent topological information when extracting topological relations from such geometries (e.g., two overlapping regions may be stated to be identical to a third region, which is impossible as they would also have to be identical to each other if that was the case); see also [5] in this respect.

Figure 5(b) shows the results of building $\overleftarrow{\diamond}_G$-consistent QCNs from scratch. For each QCN, we manually set all of its constraints to be the universal relation (viz., B), keep the graph G as is, and then add the original constraints back in 1 *by* 1 fashion. From this experiment, we can see how the proposed algorithm scales compared to the static algorithm for enforcing $\overleftarrow{\diamond}_G$-consistency, and subsequently how much better our online algorithm is compared to the static algorithm when new information arrives in a serial manner. It can be seen that, although the online algorithm takes more time to process new information serially (than in bulk, see baseline), it is much faster than using the static one in a serial manner.

In summary, we can conclude that the proposed online algorithm for enforcing $\overleftarrow{\diamond}_G$-consistency improves the static algorithm in processing dynamic information.

▶ Remark 12. To keep our evaluation concise, the number m of new constraints to add was fixed to 20 in Figures 4a, 4b, between 1 and n, the performance difference between DPCI-1 and DPC-1 largely remains qualitatively similar. On the other hand, the performance difference between DPCI-A and DPC-A might be affected by m, because the preprocessing step in lines 4–13 of Algorithm 1 might be more costly than the overhead of enforcing $\overleftarrow{\diamond}_G$-consistency in a static manner when m is close to n; in such cases, where almost the entire network is updated, DPC-A is by design the better choice.

5 Conclusion and Future Work

In this paper, we proposed an incremental (online) algorithm for checking the satisfiability of qualitative spatio-temporal data, which has important implications to other fundamental knowledge representation and reasoning problems for such data too, such as the problems of deductive closure and redundancy removal. Contrary to the state of the art, our approach acts parsimoniously and only infers new information when needed, subsequently maintaining soundness and completeness. An evaluation with publicly available large-scale real-world and random datasets against the state of the art, showed the interest and efficiency of our method. For future work, we would like to extend the current method with vertex-incremental capabilities [2], i.e., handle also the cases where new variables are incrementally made available, and not just new constraints among established variables.

References

1 Albert-László Barabási and Réka Albert. Emergence of Scaling in Random Networks. *Science*, 286(5439):509–512, 1999.

2 Anne Berry, Pinar Heggernes, and Yngve Villanger. A vertex incremental approach for maintaining chordality. *Discrete Mathematics*, 306(3):318–336, 2006.

3 Mehul Bhatt, Frank Dylla, and Joana Hois. Spatio-terminological Inference for the Design of Ambient Environments. In *COSIT*, pages 371–391, 2009.

4 Nicolas Chleq. Efficient Algorithms for Networks of Quantitative Temporal Constraints. In *CONSTRAINT@FLAIRS Conference*, pages 901–907, 1995.

5 Jean-François Condotta, Issam Nouaouri, and Michael Sioutis. A SAT Approach for Maximizing Satisfiability in Qualitative Spatial and Temporal Constraint Networks. In *KR*, pages 432–442, 2016.

6 Krishna Sandeep Reddy Dubba, Anthony G. Cohn, David C. Hogg, Mehul Bhatt, and Frank Dylla. Learning Relational Event Models from Video. *J. Artif. Intell. Res.*, 53:41–90, 2015.

7 Frank Dylla, Jae Hee Lee, Till Mossakowski, Thomas Schneider, André van Delden, Jasper van de Ven, and Diedrich Wolter. A Survey of Qualitative Spatial and Temporal Calculi: Algebraic and Computational Properties. *ACM Comput. Surv.*, 50(1):7:1–7:39, 2017.

8 Frank Dylla, Till Mossakowski, Thomas Schneider, and Diedrich Wolter. Algebraic Properties of Qualitative Spatio-temporal Calculi. In *COSIT*, pages 516–536, 2013.

9 Frank Dylla and Jan Oliver Wallgrün. Qualitative Spatial Reasoning with Conceptual Neighborhoods for Agent Control. *J. Intell. Robotic Syst.*, 48(1):55–78, 2007.

10 Max J. Egenhofer. Toward the semantic geospatial web. In *ACM-GIS*, pages 1–4, 2002.

11 Delbert Ray Fulkerson and Oliver Alfred Gross. Incidence matrices and interval graphs. *Pacific J. Math.*, 15(3):835–855, 1965.

12 Martin Charles Golumbic and Ron Shamir. Complexity and Algorithms for Reasoning about Time: A Graph-Theoretic Approach. *J. ACM*, 40(5):1108–1133, 1993.

13 Georg Gottlob. On minimal constraint networks. *Artif. Intell.*, 191-192:42–60, 2012.

14 Thiago Pedro Donadon Homem, Paulo Eduardo Santos, Anna Helena Reali Costa, Reinaldo Augusto da Costa Bianchi, and Ramón López de Mántaras. Qualitative case-based reasoning & learning. *Artif. Intell.*, 283:103258, 2020.

15 Zina Ibrahim, Honghan Wu, and Richard Dobson. Modeling Rare Interactions in Time Series Data Through Qualitative Change: Application to Outcome Prediction in Intensive Care Units. In *ECAI*, pages 1826–1833, 2020.

16 Orestis Kostakis, Nikolaj Tatti, and Aristides Gionis. Discovering recurring activity in temporal networks. *Data Min. Knowl. Discov.*, 31(6):1840–1871, 2017.

17 Manolis Koubarakis, Manos Karpathiotakis, Kostis Kyzirakos, Charalampos Nikolaou, and Michael Sioutis. Data Models and Query Languages for Linked Geospatial Data. In *Reasoning Web*, pages 290–328, 2012.

18 Sanjiang Li, Zhiguo Long, Weiming Liu, Matt Duckham, and Alan Both. On redundant topological constraints. *Artif. Intell.*, 225:51–76, 2015.

19 Gérard Ligozat. *Qualitative Spatial and Temporal Reasoning*. ISTE Series. Wiley, 2011.

20 Gérard Ligozat and Jochen Renz. What Is a Qualitative Calculus? A General Framework. In *PRICAI*, pages 53–64, 2004.

21 Zhiguo Long and Sanjiang Li. On Distributive Subalgebras of Qualitative Spatial and Temporal Calculi. In *COSIT*, pages 354–374, 2015.

22 Zhiguo Long, Michael Sioutis, and Sanjiang Li. Efficient Path Consistency Algorithm for Large Qualitative Constraint Networks. In *IJCAI*, pages 1202–1208, 2016.

23 David A. Randell, Zhan Cui, and Anthony Cohn. A Spatial Logic Based on Regions and Connection. In *KR*, pages 165–176, 1992.

24 Jochen Renz. A Canonical Model of the Region Connection Calculus. *J. Appl. Non Class. Logics*, 12(3-4):469–494, 2002.

25 Jochen Renz and Bernhard Nebel. Qualitative Spatial Reasoning Using Constraint Calculi. In *Handbook of Spatial Logics*, pages 161–215. Springer, 2007.

26 Michael Sioutis, Zhiguo Long, and Sanjiang Li. Leveraging Variable Elimination for Efficiently Reasoning about Qualitative Constraints. *Int. J. Artif. Intell. Tools*, 27(4):1–37, 2018.

27 Michael Sioutis and Diedrich Wolter. Qualitative Spatial and Temporal Reasoning: Current Status and Future Challenges. In *IJCAI*, pages 4594–4601, 2021.

28 Philip A. Story and Michael F. Worboys. A Design Support Environment for Spatio-Temporal Database Applications. In *COSIT*, pages 413–430, 1995.

29 Jakob Suchan, Mehul Bhatt, and Srikrishna Varadarajan. Out of Sight But Not Out of Mind: An Answer Set Programming Based Online Abduction Framework for Visual Sensemaking in Autonomous Driving. In *IJCAI*, pages 1879–1885, 2019.

Rethinking Route Choices! On the Importance of Route Selection in Wayfinding Experiments

Bartosz Mazurkiewicz
Geoinformation, TU Wien, Austria

Markus Kattenbeck
Geoinformation, TU Wien, Austria

Ioannis Giannopoulos
Geoinformation, TU Wien, Austria

Abstract

Route selection for a wayfinding experiment is not a trivial task and is often made in an undocumented way. Only recently (2021), a systematic, reproducible and score-based approach for route selection for wayfinding experiments was published. However, it is still unclear how robust study results are across all potential routes in a particular experimental area. An important share of routes might lead to different conclusions than most routes. This share would distort and/or invert the study outcome. If so, the question of selecting routes that are unlikely to distort the results of our wayfinding experiments remains unanswered. In order to answer these questions, an agent-based simulation study with four different sample sizes (N = 15, 25, 50, 3000 agents) comparing Turn-by-Turn and Free Choice Navigation approaches (between-subject design) regarding their arrival rates on more than 11000 routes in the city center of Vienna, Austria, was run. The results of our study indicate that with decreasing sample size, there is an increase in the share of routes which lead to contradictory results regarding the arrival rate, i.e., the results become less robust. Therefore, based on simulation results, we present an approach for selecting suitable routes even for small-scale in-situ studies.

2012 ACM Subject Classification Information systems → Decision support systems; Computing methodologies → Agent / discrete models; Information systems → Location based services

Keywords and phrases Route Selection, Route Features, Human Wayfinding, Navigation, Experiments, Experimental Design

Digital Object Identifier 10.4230/LIPIcs.COSIT.2022.6

1 Introduction

Novel navigation system paradigms for wayfinders are still the subject of ongoing research. Regardless of the target group, i.e., whether it would be pedestrians [5, 7, 12], cyclists [20, 16] or car drivers [11] many decisions during experimental design must be made. While these decisions may impact the study results, this impact is often neither evident nor easy to estimate. One of these decisions relates to the selection of a route suitable for a particular wayfinding study. Given a potential experimental area of non-trivial size, there are at least thousands of potential routes researchers can select from (see Section 3). The potential influence of different routes on study results, however, has not been scrutinized systematically. Given the myriad of potential routes and the different characteristics they come with, there might be an important share of routes that lead to study results deviating from the mean calculated over all possible routes (population mean). By means of an agent-based simulation study comparing two different navigation approaches for all potential routes in a selected experimental area, we will provide evidence that with decreasing sample size, the share of routes which lead to contradicting results increases. Given these differences in results, we

propose a selection process of appropriate routes, i.e., routes that provide stable results across sample sizes and lead to results congruent with the population mean. Hence, our approach is useful for route selection in comparative wayfinding studies, even for smaller sample sizes.

We will provide evidence that route selection is a crucial step in experimental design, as it shows the potential to turn around the study results. Therefore, more attention should be given to this phase of experimental design. Our approach can be combined (see Section 5.3) with route selection methods for wayfinding experiments (see e.g., [15]), which have been proposed so far.

2 Related Work

In this section, two branches of related work will be discussed. First, we review systematic approaches for route selection during experimental design and route justification. Second, we will discuss comparative wayfinding studies which involve at least two routes and examine whether the route itself was treated as an independent variable in the analysis.

2.1 Systematic Route Selection for Wayfinding Studies

In our previous work, we did an exhaustive search of 'six major venues (conferences and journals) in the broader area of geographic information science and related fields' [15, p. 2] between 2010 and early 2020 regarding route descriptions and/or justifications in studies involving wayfinding tasks with a predefined route. In total, 32 papers fell into this category. The conclusion was that, in general, route choice was poorly justified and that only half of the selected publications mentioned the route length, which was considered a basic feature. This leaves the impression that route selection in wayfinding experiments tends to lack appropriate justification, given the potential impact a route may have on results. In very recent studies (i.e. from 2020 onwards), examples of both missing and explicit route selection justification can be found. Dong and colleagues [3] compared augmented reality (AR) and 2D navigation electronic maps in pedestrian wayfinding. The selection of three experiment routes was not explicitly justified. There are as well examples of explicit and elaborated route justifications. Benelia compared paper maps with audiovisual Turn-by-Turn (TBT) instructions in the context of spatial learning for car drivers [2]. While selecting the route, Benelia tried to maximize personal safety, to avoid high levels of stress in participants and to have sufficient stimuli along the route. Another example of explicit route justification can be found in [20] comparing TBT and ACTF (As-The-Crow-Flies) navigation approaches for cyclists. Both routes used in this publication were designed to contain a segment on which the participant had to cycle contrary to the compass direction pointing to the destination. This feature was crucial to the experimental design.

Although both examples present an explicit justification for route selection, they are not necessarily reproducible because several routes with those characteristics are possible and they might lead to different results. In order to tackle this problem, we previously proposed a methodological average-based framework for systematic and reproducible route selection [15]. All possible routes are ranked according to criteria and corresponding weights, which the researcher must set. This flexibility allows finding routes that exactly fit the requirements of the study. However, our framework does not provide any information on how the routes may impact the study results.

2.2 Comparative Wayfinding Studies and the Importance of Route as Independent Variable

This section will review comparative wayfinding studies and verify whether the route was used as an independent variable, thereby providing examples for both cases.

It is not new to consider the route itself an important variable in comparative wayfinding studies. Savino and colleagues [20] considered the potential influence of the two selected routes in their comparative wayfinding study for cyclists and, in consequence, analyzed the data for each route separately. For both routes, the authors came to the same conclusion regarding differences in route length, task load and orientation. However, the number and the type of errors committed differed. Dong et al. compared two navigation systems on three different routes [3]. In their ANOVA analysis, the route was treated as a factor. For none of the compared eye-tracking metrics, `route` yielded a significant effect. It was only significant for the metric wayfinding duration, which is expected as route lengths differed and were not normalized. Moreover, without justification, the authors do not include route as a predictor (logistic regression) when analyzing the sketch maps. Richter et al. compared consistent and inconsistent navigation instructions on eight routes in a desktop virtual environment [18]. The selected routes had a similar number of turns and a landmark at every intersection. In the analysis, the potential influence of the route was not considered. Kuo and colleagues compared four different navigation systems on four different roads in a virtual reality (VR) environment [9]. Here, the route was also stated to be used as one of four predictors (linear regression). However, this variable, as well as two further ones, were not mentioned in the analysis. Therefore, it remains unclear if the route had an effect on the results, although this expectation was made explicit. Another study conducted in VR compared three AR-based navigation interfaces on three different routes [21]. The routes were designed to have the same length, number of turning points and street crossings. To each interface, exactly one route was assigned. The route was not treated as a factor, and in the end, it is unclear whether the observed effects come from the navigation system, the route or a combination of both.

Generally speaking, only a few routes are compared within a single wayfinding study, which seems reasonable from a research economics perspective. Simulation studies, however, are a notable exception. Amores and colleagues [1] proposed a novel navigation paradigm *most recoverable path*. Their approach was tested by means of a simulation study in Quito, Paris and Melbourne in which 13500 routes per city were selected. However, they analyzed the influence of network topology on their approach but did not analyze the data on a route level. Another example of a simulation study in which a novel navigation paradigm was proposed is our previously published work [14]. We tested our approach with 100 routes in Vienna, Djibouti City and Mexico City, respectively. Differences between those cities were found, but route-wise differences were not analyzed.

Taken together, these examples give the impression that route selection is not always given sufficient relevance, even though it might have an impact on study results. There is no systematic approach, first, to show that different routes may lead to different results, and second, how to select routes for a wayfinding study congruent with the population mean of all routes. This paper aims to fill these gaps.

3 Experimental Setup

In this section, the agent-based simulation study is described in detail. We will elaborate on the experimental area and all potential routes with pre-defined features, such as route length. Furthermore, the sample sizes and both navigation approaches, namely Turn-by-Turn (TBT) and Free Choice Navigation (FCN), will be described. The study follows a between-subject design comparing two navigation systems.

3.1 Experimental Area and Potential Routes

As the experimental area, the city center (surface area 2.5 km^2) of Vienna, Austria is chosen (see Figure 1). According to the classification by Thompson et al. [22], the network layout is of type *high transit*. For this area, the raw network data were downloaded from OpenStreetMap (OSM)[1]. The intersections and their characteristics were calculated using the Intersections Framework [4], whereas street segments were extracted with a custom script. Taking these pieces together, the city center is represented as a networkx graph having 1848 nodes and 2722 edges.

For every experimental design, several decisions regarding route choice have to be made (e.g., route length, sequence of left, right and non-turns, number of decision points and experimental area). In order to reduce the search space of potential routes, we will consider only shortest path routes (see e.g., [19]) with 12 decision points [15] and a length between 550 m and 1000 m (see e.g., [17, 19]) in order to avoid trivial route length on the one hand and, on the other hand, to ensure a reasonable duration for an in-situ study (1000 m would result in a duration of 12.5 minutes based on an average walking speed of 4.85 km/h [10]). In order to find all possible routes sharing these characteristics and comprising no loops, SageMath 9.1 with its SubgraphSearch function[2] was used, as in our previous work about the route selection framework [15]. The resulting $N_r = 11373$ routes are the whole population of routes being shortest paths and matching the mentioned lengths and number of decision points and were used for the simulation, which was implemented in Python 3.6.

3.2 Sample Size

The simulation is run with four different samples sizes ($n = 15, 25, 50, 3000$ agents) following a between-subject design. The first three sample sizes can be considered realistic in wayfinding studies [21, 18, 20, 9, 6]. The largest sample size ($n = 3000$) is considered to be representative for the whole population of participants of such studies. Different sample sizes are tested in order to investigate whether the sample size impacts the results for both a single route and the whole route population. Each group navigates each of the 11373 routes.

3.3 Navigation Systems

The presented simulation approach will work for any two navigation systems, as we want to demonstrate that the comparison results may vary depending on the route choice. However, we continue our previous simulation study [14] and compare Free Choice Navigation (FCN) and Turn-by-Turn (TBT). While the particular figures will likely change for other navigation approaches, the proposed route selection process (see Section 5.2) based on the results remains unchanged.

---

[1] `https://www.openstreetmap.org`, last access February 4th, 2022

[2] `https://doc.sagemath.org/html/en/reference/graphs/sage/graphs/generic_graph_pyx.html`, last access January 30th, 2022

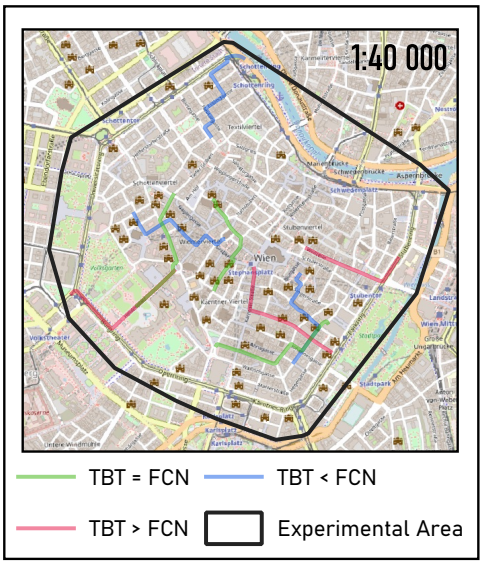


Figure 1 The experimental area in Vienna, Austria and 9 sample routes on which one navigation system performed better or they performed equally well across sample sizes. Basemap OpenStreetMap.

As the primary purpose of navigation systems is to assist wayfinders in reaching the destination, we choose arrival rate as the success metric. However, any other suitable success metric can be chosen by the researcher. As in our previous work [14], an agent is considered successful if it reaches the destination within 150% of the shortest path length. In the same work, we compared these two navigation systems in three different cities [14]. TBT lead between 5% and 10% more agents to their destination in all cities. Now, the main features and mechanics of both navigation approaches will be described.

3.3.1 Turn-by-Turn (TBT)

By analogy with commercial wayfinding assistance systems for pedestrians, the agent is supposed to follow the shortest path between origin and destination and receives navigation instructions at turning points only. If agents have to go straight ahead at a junction, then no instruction is issued and the agent continues straight.Going straight ahead is considered walking in a direction that does not deviate by more than 10 degrees to either side from the current one. Every agent has a probability to interpret a generic navigation instruction correctly. If an instruction is issued, the agent interprets it based on a weighted random choice: The branch to follow, indicated in the instruction and following the shortest path from the current junction, is assigned a weight equal to the agent's probability to interpret generic navigation instructions correctly. The remaining probability is split equally over all remaining branches (excluding the one indicated in the navigation instruction and the one the agent has come from). The trial ends when the agent reaches the destination.

3.3.2 Free Choice Navigation (FCN)

Free Choice Navigation is a novel navigation paradigm aiming for more freedom of choice during navigation [14]. The system allows the agent for some exploration but, on the other hand, tries to avoid costly mistakes by weighing the number of free choices, the number of

given instructions and a maximum allowed route length. The working mechanism can be seen in the following example: Alice, a good wayfinder, is navigating to a museum. Before the navigation starts, the system gives her information about the beeline direction and distance to the museum. At the first two junctions, the system does not issue any instructions because it is assumed that the beeline direction is still clear to the user. In consequence, Alice decides on her own which branch to take. The upcoming junction, however, is rather complex as it has five branches. Alice is quite sure about the beeline direction, but there are two branches that seem equally well suited to her. The system detects this difficulty based on internal computations that take the environmental structure and spatial abilities of the user into account and issues an instruction because one of the branches leads to a considerable deviation from the allowed maximum route length. Alice interprets it correctly and continues her walk.

This example shows that the navigation system issues an instruction based on environmental spatial abilities of a user, the characteristics of the current junction and the already walked route. If an instruction is issued, then the same procedure as above applies with the difference that the branch the agent has come from is not excluded but has a lower probability of being taken. Again, the probabilities of available branches to be taken depend on the agent's probability of interpreting generic navigation instructions correctly, which in turn depends linearly on its environmental spatial abilities. Furthermore, FCN has six parameters that steer when an instruction is given. We used the best parameter set for Vienna, which is a trade-off between the percentage of successful trials and the number of given instructions [14].

For every agent, regardless of the condition, the ability to interpret navigation instructions correctly ranges between 0.8 and 1 and is fixed before the experiment. Please refer to our previous work for further modeling details regarding the agents and their decision mechanism [14].

4 Simulation Results

In this section, we, first, present descriptive statistics for each of the systems separately and, second, discuss the differences originating from different routes. Differences between both conditions are calculated using bootstrapping ($B = 10000$ runs) and 95% confidence intervals (CIs) are reported in square brackets. As mentioned above, the arrival rate (each agent walked each route) for both systems is compared (see Section 3.3). In order to ensure that the common ability of agents to interpret navigation instructions correctly (co-domain [.8; 1] [14]) did not influence the results, a Wilcoxon Signed-Rank Test was performed for every sample size. No significant ($\alpha = .05$) differences between both conditions were found ($n = 15$ ($Z = 1.14$, $p = .25$, $r = 0.29$), $n = 25$ ($Z = 1.17$, $p = .24$, $r = 0.23$), $n = 50$ ($Z = .05$, $p = .96$, $r = 0.01$), $n = 3000$ ($Z = .00$, $p = .99$, $r = 0.00$)). Furthermore, this ability defines good and weak wayfinders. We assured that agents from both groups are present in every sample size, which is a realistic scenario for real-world wayfinding studies. The presented figures are computed based on all potential routes, which were walked by all agents of a given sample size. There are 11 373 potential routes in the experimental area. This is an exhaustive sample considering the selected route features (see Section 3.1).

4.1 Turn-by-Turn

In the case of the Turn-by-Turn condition, the simulation for all four sample sizes yields similar results regarding the arrival rate (co-domain [0; 1])(see Table 1). The mean arrival rate is around 0.97 across all four sample sizes. This is in contrast to the minimum arrival

Table 1 Descriptive statistics of the arrival rate [0; 1] for the TBT condition for all four sample sizes tested in the simulation. The figures are rounded to 3 decimals.

Sample Size	Mean	SD	Median	Min	Max
15	.976 [.975; .977]	.048 [.047; .05]	1 [1;1]	.66 [.66; .66]	1 [1; 1]
25	.975 [.974; .976]	.043 [.042; .044]	1 [1;1]	.68 [.68; .72]	1 [1; 1]
50	.973[.972; .973]	.038 [.037; .039]	.98 [.98; .98]	.72 [.72; .74]	1 [1; 1]
3000	.97 [.97; .97]	.033 [.032; .034]	.982 [.982; .983]	.788 [.788; .798]	1 [1; 1]

rate, which shows considerable variation between sample sizes: For sample size $n = 15$, the minimum arrival rate for a route is 0.66 (only $\frac{2}{3}$ of the agents reached the destination), whereas for sample size $n = 3000$ it is 0.788. The range of the arrival rate decreases with increasing sample size. In order to see whether there are route-wise differences between sample sizes, the range ($max - min$) for every route is calculated (see Figure 2). Over 20% of the routes have a range greater than or equal to 0.05. There is no difference across sample sizes for 716 routes (6.2%), whereas the biggest difference encountered for a single route across sample sizes is 0.25.

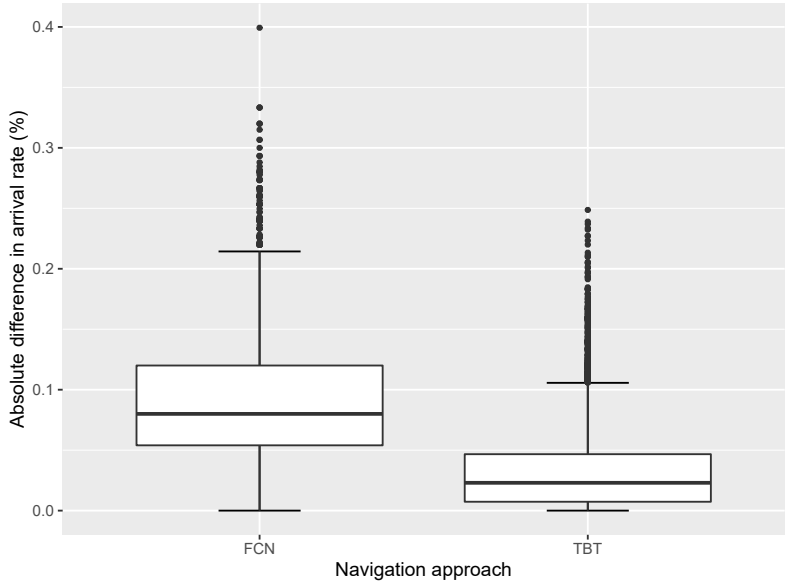


Figure 2 Route-wise ranges ($max - min$) across all sample sizes for the condition FCN (left) and TBT (right).

4.2 Free Choice Navigation

Analyzing all routes together, the four sample sizes yield, again, similar results (see Table 2). The mean arrival rate is approx. 0.90; in contrast to the TBT condition, the range remains almost identical across sample sizes. Route-wise ranges ($max - min$), however, reveal a higher variance in the FCN condition (see Figure 2): More than 80% of the routes show a difference greater than or equal to 0.05 and the biggest difference for a route across sample sizes is 0.4. For 34 routes (0.3%), there is no difference across sample sizes.

Table 2 Descriptive statistics of the arrival rate [0; 1] for the FCN condition for all four sample sizes tested in the simulation. The figures are rounded to 3 decimals.

Sample Size	Mean	SD	Median	Min	Max
15	.919 [.917; .921]	.104 [.099; .109]	.933 [.933; .933]	0 [0; 0]	1 [1; 1]
25	.908 [.906; .91]	.096 [.091; .101]	.92 [.92; .92]	0 [0; 0]	1 [1; 1]
50	.896 [.894; .897]	.091 [.086; .096]	.92 [.92; .92]	0 [0; 0]	1 [1; 1]
3000	.902 [.901; .904]	.08 [.074; .086]	.918 [.917; .919]	0 [0; 0]	.996 [.994; .996]

4.3 Differences within both Systems

In both approaches, of course, the ability to interpret navigation instructions plays a role, but as it is constant for all routes, it is not mentioned as a factor. As indicated by the figures in tables 1 and 2, arrival rates differ between both navigation systems. These differences may stem from navigation system mechanics and street layout. In the TBT condition, routes with less turning points likely lead to a higher arrival rate, as the agent has to make fewer decisions and, in consequence, has lower chances to commit an error. On the other hand, in the FCN approach, route features like junction complexity or junction skewness [4] are likely to play a role. A detailed analysis of route features leading to differences is beyond the scope of this paper (see Section 6).

4.4 Differences between both Systems

Based on the within-system results, both navigation systems will be compared regarding the arrival rate. Again, first, the whole population is analyzed, and second, route-wise differences will be inspected in order to investigate whether sample size impacts the share of routes that lead to contradicting results. Across all sample sizes, the TBT approach leads, on average, more agents to the destination (see Table 3). The sample size with the highest mean difference in arrival rate across routes is $n = 50$, whereas the lowest value can be observed for $n = 15$. Mean, standard deviation, median and maximum values are similar in all simulation runs; however, there are differences in the minimum: All minimum values are negative, meaning that there is at least one route on which the FCN approach performed better than TBT. Therefore, we will inspect per route differences between both conditions by subtracting FCN from TBT arrival rates for the respective sample size.

For every sample size, we count the number of routes which lead to a congruent result with the population mean (TBT performs better), as well as routes on which FCN performed better than or as good as TBT (see Table 4). For the sample size $n = 3000$, which is the most representative one, there are around 8% of routes on which FCN performed better or as good as TBT. For smaller sample sizes, this figure increases, reaching around 47% for $n = 15$. Contrary to the within-system results (see Sections 4.1 and 4.2), here, considerable differences between sample sizes can be observed.

5 Discussion and Limitations

This section will discuss the results, which suggest that route selection is an important part of experimental design and should be given more importance. Furthermore, a methodology that supports informed route selection is proposed. Finally, limitations that apply to our work are addressed.

Table 3 Descriptive statistics for the route-wise difference ($TBT - FCN$) in arrival rate [0; 1] for all four sample sizes tested in the simulation. Positive values mean that the TBT condition performed better and negative values indicate a better performance of the FCN navigation approach. The figures are rounded to 3 decimals.

Sample Size	Mean	SD	Median	Min	Max
15	.057 [.055; .059]	.111 [.107; .116]	.067 [.067; .067]	- .333 [- .333; - .333]	1 [1; 1]
25	.067 [.065; .069]	.101 [.096; .105]	.04 [.04; .04]	- .28 [- .28; - .24]	1 [1; 1]
50	.077 [.075; .079]	.094 [.089; .098]	.06 [.06; .06]	- .24 [- .24; - .18]	1 [1; 1]
3000	.067 [.066; .069]	.081 [.075; .086]	.057 [.056; .058]	- .119 [- .119; - .104]	1 [.983; 1]

Table 4 Shares of routes on which TBT performed better than, as good as and worse than FCN regarding the arrival rate. The figures are rounded to 1 decimal.

Sample Size	TBT Better	TBT = FCN	FCN Better
15	53 %	35.6 %	11.5 %
25	70.4 %	19.3 %	10.3 %
50	84 %	7.5 %	8.5 %
3000	92.1 %	0.1 %	7.8 %

5.1 Discussion

Regardless of the sample size, the simulation, which considers all potential routes in the experimental area, yields similar results, indicating the superiority of TBT over FCN regarding the arrival rate (see Table 3). Looking at the results for the whole population, one might think that route selection is not so critical because, independently of the sample size, the big picture is preserved. This picture is, however, somewhat misleading as wayfinding studies, of course, are conducted with a small-sized subsample of the whole population, considering both routes and participants. By means of keeping the population of routes constant across sample sizes, our simulation results indicate that different routes can lead to contradicting results (see Table 4). In consequence, ad-hoc decisions on route selection can lead to contrary results compared to the whole population of routes. This situation worsens with decreasing sample size as the chance of selecting such a route increases (see Table 4). The results, therefore, suggest that selecting a route is all the more important in the case of small numbers of participants. For samples sizes ($n = 15, 25, 50$), which can be considered realistic for comparative wayfinding experiments (see e.g. [21, 18, 20, 9, 6]), the probability to select a route that will yield results incongruent with the population mean varies between 16% and 47%. This means, if we planned an experiment with two groups, with 15 participants each, and we randomly picked a route from our experimental area, we would have a 47% chance to conclude that TBT is not superior regarding the arrival rate, although it actually is (see Table 3). Almost every second route would lead to the contrary conclusion in the case of $n = 15$, whereas, for sample sizes $n = 25$ and $n = 50$, it would be every third and sixth route, respectively. Given this high share, we want to draw attention to the importance of the route selection process as it can influence study results, in particular, given the relatively small number of participants, which is quite common in the wayfinding domain.

Taken together, in the selected experimental area, the lower the number of agents, the higher the probability of choosing a route which leads to results that are contradictory to the population mean, i.e., the route becomes more crucial with decreasing sample size. This is a

problem as routes often seem to be selected in an ad-hoc manner during experimental design in wayfinding studies (see Section 2.1). Given that wayfinding studies are not conducted with 3000 participants, a method to select those routes which are likely to lead to a conclusion corresponding to the whole population is proposed.

5.2 Route Selection

Depending on the sample size, the chance of selecting a route that leads to conclusions that are not in line with the whole route population may be considerable. This section suggests an approach that allows for an informed route selection. This process is based on the simulation results, i.e., the simulation for the compared navigation systems needs to be run beforehand. It is a two-step approach. First, routes without great variance across all sample sizes are selected, and second, those which lead to results congruent with the population mean (based on the simulation) are chosen. In both steps, the researcher needs to select a filtering threshold depending on the selected performance metric and observed differences. The underlying idea is to select routes that lead to similar results across all sample sizes and are compatible with the population mean. In our example, two navigation systems are compared. Therefore, their differences in arrival rate are used in the presented filtering process. The same approach can be applied with one navigation system only by using the arrival rates directly instead of the differences or any other success metric chosen by the researcher.

5.2.1 Consistent Routes

In this step, routes will be selected which are *consistent* regarding differences in arrival rates, i.e., they do not vary considerably in arrival rates across sample sizes. Given that there are four values (one per sample size) for each route to consider, we refrain from calculating the standard deviation and will consider the range as the measure of variability. The applied measure with a corresponding threshold can be adapted according to the number of tested sample sizes and the researcher's needs. For every route, we calculate the range across all four sample sizes and select those routes whose range is not greater than 0.03, which means that the biggest allowed difference across sample sizes is 3%. This value can be set according to the simulated data. The smaller the value, the more restrictive this filtering step will be. In this case study, 618 (5.4%) routes have a range smaller than or equal to 0.03. By this filtering step, routes with high variance across sample sizes are excluded. However, routes that are not *close* to the population mean are still possible, or even routes on which the drawn conclusion is contrary to the population mean. Therefore, a second filtering step is necessary.

5.2.2 Routes in Concordance with the Population Mean

In order to find routes that are congruent with the most representative sample size ($n = 3000$), they are filtered by their mean across sample sizes. Routes whose means do not differ considerably from the population mean (see Table 3) are selected for being considered suitable routes. For this step, another threshold needs to be selected by the researcher. In consequence, routes are selected whose means do not deviate by more than the selected threshold from the population mean. Given that the population mean difference is 0.067, we set this threshold to 0.02. Therefore, routes with an average between 0.047 and 0.087 are considered in our case study as acceptable. With this second filtering step, 304 routes are

left. This is 2.67% of the whole population. For this proof of concept, the exact threshold values are of less relevance. The smaller both thresholds are set, the more restrictive the filtering process is, i.e., less routes are considered suitable. This has to be decided based on the simulation results at hand and the researcher's needs.

5.3 Route Ranking

Our approach delivers a list of suitable routes with regard to the most representative sample size but does not state explicitly which one to choose. Our approach can, however, be combined with our route selection framework [15]. In doing so, potential route biases can be further mitigated. First, the routes are ranked according to features selected by the researcher [15], e.g., mean segment length, traffic, average number of branches or number of left, right and non-turns along the route. Second, the ranked routes are filtered according to our proposed approach. This results in routes that satisfy both the researcher's needs regarding route characteristics and being close to the global mean across sample sizes.

5.4 Limitations

Running a simulation implies simplifying certain aspects of the real world. In our simulation, the street network and the agent's spatial abilities are used to model the agent's behavior. Compared to our previous work [14], the agents, their reasoning mechanism and the environment could have been adapted regarding complexity (see e.g., [13, 8]). In addition to that, there may be relationships that have not been yet discovered and, therefore, are not considered in the simulation process. Given that randomness plays a role in our simulation, running the simulation once is a limitation. However, several seeds were tested with a subset of routes during a pretest and the results were quite consistent.

6 Conclusion and Future Work

By means of an agent-based simulation study, which was run on all potential routes in a selected experimental area, it was shown that depending on the route selection, the study results can be contradictory. Although the results for the whole population lead, on average, to the same conclusion, there is an important share of routes that lead to contrary results. Given that the route selection process usually does not receive much attention in wayfinding studies, with this simulation, we direct researchers' attention to the potentially harmful effects of ad-hoc route selections. Therefore, we propose a selection method based on running the same simulation with different sample sizes. The resulting selection of routes should lead to results that are congruent with the population mean.

Furthermore, our proposed simulation approach with different sample sizes allows for detecting weak points of a given navigation system. Researchers will find routes on which their proposed navigation system does not perform as good as expected and their examinations will lead to further improvements. Moreover, our simulation approach makes it possible to identify spatial configurations (routes and their neighborhood) favorable or adverse to the navigation system at hand by analyzing route features that cause differences in the selected performance metric. This analysis would provide valuable feedback in order to improve the tested navigation approach. The in-depth analysis of route properties and their influence on the success metric is part of our future work. One could improve the navigation system until it is robust on all routes, i.e., it performs equally well on the whole population of routes.

A series of simulation studies in different geographic areas is planned in order to see whether different network types [22] lead to the same results. *Motor cities* might be less vulnerable to ad-hoc route selection. In addition, the route properties which caused differences in arrival rates will be examined in depth. Moreover, we plan as well to increase the complexity of the models to increase the validity of the simulation. Adding more complexity would expand the search space for possible explanations because the differences in the selected success metric could be explained by additionally modeled features like points of interest, buildings or terrain slope.

We are aware that implementing a simulation is not a trivial task and not every researcher has the resources to do it. Therefore, another research direction could be the prediction of route suitability based on route features and the characteristics of the navigation approach without running a simulation study.

Our approach still needs to be verified in real-world environments. Therefore, a series of human subject experiments will be conducted. In these experiments, the results of several routes selected with our approach will be compared with the population mean resulting from a simulation study. Following our selection approach, we expect that those routes considered suitable will lead to consistent and congruent with the population mean results and the routes considered non-suitable will more likely lead to contrary conclusions. However, this hypothesis needs to be verified in a real-world setting as the routes are selected based on simulation results.

References

1 David Amores, Egemen Tanin, and Maria Vasardani. A proactive route planning approach to navigation errors. *International Journal of Geographical Information Science*, 35(6):1094–1130, 2021. doi:10.1080/13658816.2020.1820508.

2 Eran Ben-Elia. An exploratory real-world wayfinding experiment: A comparison of drivers' spatial learning with a paper map vs. turn-by-turn audiovisual route guidance. *Transportation Research Interdisciplinary Perspectives*, 9:100280, 2021. doi:10.1016/j.trip.2020.100280.

3 Weihua Dong, Yulin Wu, Tong Qin, Xinran Bian, Yan Zhao, Yanrou He, Yawei Xu, and Cheng Yu. What is the difference between augmented reality and 2d navigation electronic maps in pedestrian wayfinding? *Cartography and Geographic Information Science*, 48(3):225–240, 2021. doi:10.1080/15230406.2021.1871646.

4 Paolo Fogliaroni, Dominik Bucher, Nikola Jankovic, and Ioannis Giannopoulos. Intersections of Our World. In Stephan Winter, Amy Griffin, and Monika Sester, editors, *10th Int. Conf. on Geographic Information Science (GIScience 2018)*, volume 114 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 3:1–3:15, Dagstuhl, Germany, 2018. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik. doi:10.4230/LIPIcs.GISCIENCE.2018.3.

5 Ioannis Giannopoulos, Peter Kiefer, and Martin Raubal. Gazenav: Gaze-based pedestrian navigation. In *Proc of the 17th Int Conf on Human-Computer Interaction with Mobile Devices and Services*, MobileHCI '15, pages 337–346. ACM, 2015.

6 Antonia Golab, Markus Kattenbeck, Georgios Sarlas, and Ioannis Giannopoulos. It's also about timing! when do pedestrians want to receive navigation instructions. *Spatial Cognition & Computation*, 0(0):1–33, 2021. doi:10.1080/13875868.2021.1942474.

7 Haosheng Huang, Thomas Mathis, and Robert Weibel. Choose your own route – supporting pedestrian navigation without restricting the user to a predefined route. *Cartography and Geographic Information Science*, 0(0):1–20, 2021. doi:10.1080/15230406.2021.1983731.

8 Peter M. Kielar, Daniel H. Biedermann, Angelika Kneidl, and André Borrmann. A unified pedestrian routing model for graph-based wayfinding built on cognitive principles. *Transportmetrica A: Transport Science*, 14(5-6):406–432, 2018. doi:10.1080/23249935.2017.1309472.

9 Ting-Yu Kuo, Hung-Kuo Chu, and Yung-Ju Chang. Comparing the effects of reference-based, orientation-based, and turn-by-turn navigation guidance on users' independent navigation. In *Adjunct Proceedings of the 2020 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2020 ACM International Symposium on Wearable Computers*, UbiComp-ISWC '20, pages 63–66, New York, NY, USA, 2020. Association for Computing Machinery. doi:10.1145/3410530.3414424.

10 R. V Levine and A. Norenzayan. The pace of life in 31 countries. *J.of Cross-Cultural Psychology*, 30(2):178–205, 1999.

11 Xiangdong Ma, Mengting Jia, Zhicong Hong, Alex Pak Ki Kwok, and Mian Yan. Does augmented-reality head-up display help? a preliminary study on driving performance through a vr-simulated eye movement analysis. *IEEE Access*, 9:129951–129964, 2021. doi:10.1109/ACCESS.2021.3112240.

12 Charlotte Magnusson, Kirsten Rassmus-Gröhn, and Delphine Szymczak. Navigation by pointing to GPS locations. *Personal and Ubiquitous Computing*, 16(8):959–971, 2012.

13 Tsubasa Maruyama, Satoshi Kanai, Hiroaki Date, and Mitsunori Tada. Simulation-based evaluation of ease of wayfinding using digital human and as-is environment models. *ISPRS International Journal of Geo-Information*, 6(9), 2017. doi:10.3390/ijgi6090267.

14 Bartosz Mazurkiewicz, Markus Kattenbeck, and Ioannis Giannopoulos. Navigating Your Way! Increasing the Freedom of Choice During Wayfinding. In K. Janowicz and J. Verstegen, editors, *11th Int. Conf. on Geographic Information Science (GIScience 2021) - Part II*, volume 208 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 9:1–9:16. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2021. doi:10.4230/LIPIcs.GIScience.2021.II.9.

15 Bartosz Mazurkiewicz, Markus Kattenbeck, Peter Kiefer, and Ioannis Giannopoulos. Not Arbitrary, Systematic! Average-Based Route Selection for Navigation Experiments. In K. Janowicz and J. Verstegen, editors, *11th Int. Conf. on Geographic Information Science (GIScience 2021) - Part I*, volume 177 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 8:1–8:16. Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 2020. doi:10.4230/LIPIcs.GIScience.2021.I.8.

16 Martin Pielot, Benjamin Poppinga, Wilko Heuten, and Susanne Boll. Tacticycle: Supporting exploratory bicycle trips. In *Proceedings of the 14th International Conference on Human-Computer Interaction with Mobile Devices and Services*, MobileHCI '12, pages 369–378, New York, NY, USA, 2012. Association for Computing Machinery. doi:10.1145/2371574.2371631.

17 Karl Rehrl, Elisabeth Häusler, Sven Leitinger, and Daniel Bell. Pedestrian navigation with augmented reality, voice and digital map: final results from an in situ field study assessing performance and user experience. *Journal of Location Based Services*, 8(2):75–96, 2014. doi:10.1080/17489725.2014.946975.

18 Kai-Florian Richter, Róisín Devlin, and Filippo La Greca. Investigating wayfinding under inconsistent information. In Jurģis Šķilters, Nora S. Newcombe, and David Uttal, editors, *Spatial Cognition XII*, pages 191–195, Cham, 2020. Springer International Publishing.

19 S. Robinson, M. Jones, P. Eslambolchilar, R. Murray-Smith, and M. Lindborg. "I Did It My Way": Moving Away from the Tyranny of Turn-by-Turn Pedestrian Navigation. In *Proc. of MobileHCI '10*, pages 341–344, 2010.

20 Gian-Luca Savino, Laura Meyer, Eve Emily Sophie Schade, Thora Tenbrink, and Johannes Schöning. Point me in the right direction: Understanding user behaviour with as-the-crow-flies navigation. In *22nd International Conference on Human-Computer Interaction with Mobile Devices and Services*, pages 1–11, 2020.

21 Tram Thi Minh Tran and Callum Parker. Designing exocentric pedestrian navigation for ar head mounted displays. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI EA '20, pages 1–8, New York, NY, USA, 2020. Association for Computing Machinery. doi:10.1145/3334480.3382868.

22 J. Thompson, M. Stevenson, and J. S. Wijnands et al. (8). A global analysis of urban design types and road transport injury: an image processing study. *The Lancet Planetary Health*, 4(1):e32–e42, 2020.

Empirical Evidence for Concepts of Spatial Information as Cognitive Means for Interpreting and Using Maps

Enkhbold Nyamsuren[1]
Department of Human Geography and Spatial Planning, Utrecht University, Utrecht, The Netherlands

Eric J. Top
Department of Human Geography and Spatial Planning, Utrecht University, Utrecht, The Netherlands

Haiqi Xu
Department of Human Geography and Spatial Planning, Utrecht University, Utrecht, The Netherlands

Niels Steenbergen
Department of Human Geography and Spatial Planning, Utrecht University, Utrecht, The Netherlands

Simon Scheider
Department of Human Geography and Spatial Planning, Utrecht University, Utrecht, The Netherlands

Abstract

Due to the increasing prevalence and relevance of geo-spatial data in the age of data science, Geographic Information Systems are enjoying wider interdisciplinary adoption by communities outside of GIScience. However, properly interpreting and analysing geo-spatial information is not a trivial task due to knowledge barriers. There is a need for a trans-disciplinary framework for sharing specialized geographical knowledge and expertise to overcome these barriers. The *core concepts of spatial information* were proposed as such a conceptual framework. These concepts, such as *object* and *field*, were proposed as cognitive lenses that can simplify understanding of and guide the processing of spatial information. However, there is a distinct lack of empirical evidence for the existence of such concepts in the human mind or whether such concepts can be indeed useful. In this study, we have explored for such empirical evidence using behavioral experiments with human participants. The experiment adopted a contrast model to investigate whether the participants can semantically distinguish between the *object* and *field* core concepts visualized as maps. The statistically significant positive results offer evidence supporting the existence of the two concepts or cognitive concepts closely resembling them. This gives credibility to the core concepts of spatial information as tools for sharing, teaching, or even automating the process of geographical information processing.

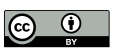
2012 ACM Subject Classification Information systems → Geographic information systems; General and reference → Empirical studies

Keywords and phrases core concepts, cognition, map interpretation, spatial analysis

Digital Object Identifier 10.4230/LIPIcs.COSIT.2022.7

Supplementary Material
Dataset: `https://github.com/quangis/Core-Concept-Study---Contrasting-Maps`
Dataset (Dataset backup at DataverseNL): `https://doi.org/10.34894/IOGWDP`
Dataset (Dataset backup at DANS EASY): `https://doi.org/10.17026/dans-xg5-cw67`

[1] The corresponding author



Funding This work was supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 803498).

1 Introduction

Going beyond manipulating map layouts and data formats, concepts of spatial information enable us to effectively handle maps by accounting for a map's semantics. This, in turn, allows us to decide on analytic methods, answer questions, and make decisions with maps. For example, we can use a map of height contour lines to assess the slope and aspect of the terrain in order to assess the potential for solar energy. Yet, the way that maps represent these concepts is rather indirect. Despite what one might think, maps do not wear their content on their sleeves. In the solar energy example, the contour map may be encoded as vector polygons and might be visualized in terms of graded colors, just like a choropleth map. Yet, in contrast to a choropleth map, human map interpreters need to conceive of the contour lines not as boundaries of objects, but rather as boundaries of height intervals. Thus, the polygon map really represents a spatial height field broken down into intervals, and not a collection of objects. This *conceptualization* of the map's information content does not follow from the way it is encoded, and thus requires a skilled human interpreter.

Since the concepts represented by maps are usually not (fully) explicit in a map, empirical investigation is needed to find out which conceptual distinctions are used and which role they play in map interpretation and map usage. Spatial concepts have been found to play a role early in the development of young children [5, 8, 9]. One example is the concept of (relative) location, which is a primary concept in spatial cognition and orientation [22, 10] and is also underlying spatial reference systems. Yet for interpreting maps, further concepts are required. Research in human cognition found that space is only one out of four main systems of core knowledge acquired early in life, including also objects, actions and numbers [27]. Naive Geography set out to study cognitive models of the common-sense geographic world, including topology, metrics, as well as discrete and continuous spatial entities [6]. However, potential concepts underlying Geography abound [10], and it remained unclear which ones should be regarded as essential for geographic information. More recently, the *core concepts of spatial information* were suggested as a concise model of different conceptualizations of the environment in this context [20], forming a basis for trans-disciplinary spatial thinking[2]. They include *objects* (e.g. buildings or administrative units), *fields* (e.g. temperature), *events* (e.g. earthquakes), and *networks* (e.g. commuter flows). Related concepts of measurement (such as extensive and intensive amounts and measurement levels) have been suggested earlier in theories about Geographic Information Systems (GIS) [3], and were recently used together with core concepts to describe spatial data models on a conceptual level (e.g. [25]) to automate the answering of geographic questions and the synthesis of workflows [19].

Although this provides a kind of indirect evidence for the importance of concepts in handling spatial information, there is still a lack of primary *empirical evidence* for such concepts as cognitive tools for interpreting and using maps[3]. Hence, it is still unclear which concepts precisely should serve as a transdisciplinary framework [20] for sharing geographic

[2] Concepts are regarded as trans-disciplinary because the underlying (GIS) methods are used across many disciplines, just like in Statistics. For a justification of core concepts in this respect, cf. [20].

[3] The fact that a concept is part of documented knowledge does not yet mean it is used effectively. Cognitive research is required to determine this [27].

knowledge. Furthermore, it is also unclear whether information systems that utilize such concepts (e.g. [26, 21]) reflect distinctions that mirror human cognition. Therefore, the goal of this study is to collect such evidence. Our main research question is:

Are there mental skills that allow users to distinguish maps on a conceptual level that, at least partially, corresponds to the core concepts of spatial information?

Knowing the definition of concepts does not imply the ability to effectively use them for spatial analysis. For example, human users may still have difficulty differentiating between the field and object interpretations despite knowing the conceptual difference. This leads to the sub-question *R1* below. Next, the visual perception of maps can influence the interpretation of spatial data. Spatial datasets representing the same phenomenon may still invoke different interpretations due to differences in symbology or geometry (e.g. point vs lines). Sub-question *R2* should account for this possible visual interference in our study. Finally, the ability to distinguish concepts and to effectively use them for interpreting maps may be an acquired skill that develops with the level of experience. This concern leads to sub-question *R3*.

- RQ1: *To what extent can users effectively distinguish maps that are attributed to different core concepts?*
- RQ2: *To what extent is the ability to distinguish conceptually different maps dependent on visual geometric properties?*
- RQ3: *To what extent is the ability to distinguish conceptually different maps dependent on a user's level of expertise?*

2 Related work

To better understand geographic information (GI) and its applications, authors have suggested conceptual frameworks to categorize GI-tools and operations [11, 12, 1, 2], syntactic data types [11, 12] and representation models [20, 21, 25, 13, 18]. Kuhn's core concepts of spatial information [20, 21] are an example of the last group and the focus of this study. The core concepts were used in the development of web-based ontologies [25], for pedagogical purposes [7, 17] and for automatic question-answering [30]. Multiple scholars, among them Kuhn himself, argue for the usefulness of the core concepts for the transdisciplinary communication and teaching of GI-knowledge. However, only Ishikawa [17] so far used the core concepts for the evaluation of empirical data.

Table 1 Semantically distinct dataset types based on the core content concept *object* and *field* and geometry combinations. The terms in brackets are abbreviations.

	Point	Line	Non-tessellated polygon	Tessellated polygon		
Object	Point object (PO)	Line object (LO)	Region object (RO)	Lattice (LA)	Lattice (LA)	Square lattice (SL)
Field	Point measure (PM)	Line measure (LM)	Region patch (RP)	Contour (CO)	Coverage (COV)	Raster field (RF)
Group	PO-PM	LO-LM	RO-RP	LA-CO	LA-COV	SL-RF

In the latest iteration [21], the core concepts framework includes five content concepts (*location*, *field*, *object*, *network*, and *event*) and two quality concepts (*granularity* and *accuracy*). The quality concepts are of less relevance to this study and, therefore, ignored. *Location* denotes the relation of some spatial phenomenon with its space or 'grounds'. A *Field* measures time-varying spatial phenomena "*...that have a scalar or vector attribute everywhere in a space of interest, for example, air temperatures on the Earth's surface*" [20, p. 2272]. *Objects*

describe spatially bounded individuals with identity and spatial and thematic properties that vary in time. *Events* have temporal boundaries (e.g. an earthquake). Finally, a *network* is a binary relation between objects, such as a road network or import/export trade flows.

The content concepts are generally not mutually exclusive because they are dependent on each other and the interpretation of a map therefore can be ambiguous. For example, *location* is an orthogonal concept used to represent the spatial aspect of every other core concept. It is also a relational concept often defined relative to *object*. *Network* is a relation between pairs of *objects* [24], while *event* is a temporal phenomenon in which *objects* or *fields* can participate. This naturally leads to variations in the interpretation of a given map. Furthermore, in which way the important concepts of amount and measurement relate to core concepts is still an open question [28]. However, one and the same represented phenomenon is usually not interpreted both as *object* and *field* at the same time. That is, the interpretation of an entity as *object* usually excludes parallel interpretation of the same as *field*. As mentioned later, this independence is important and makes *object* and *field* the main focus of our study.

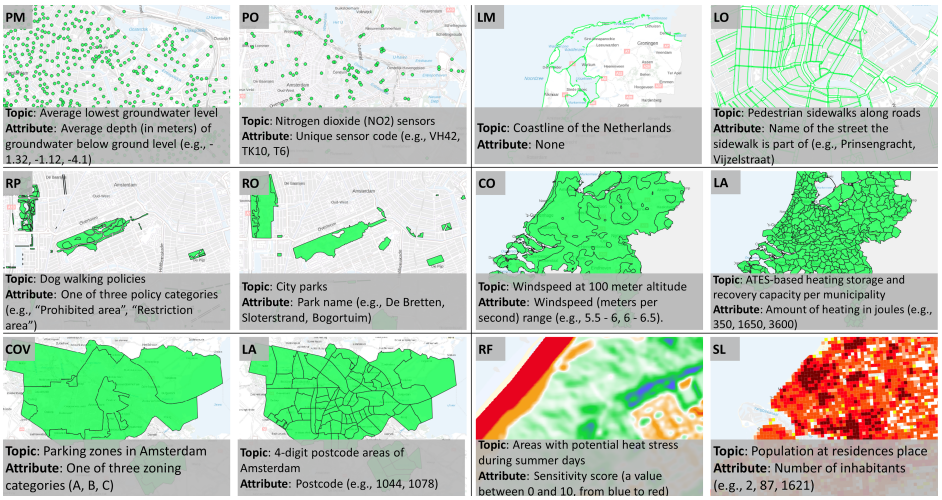


Figure 1 Examples for each dataset type listed in Table 1. The letters in the top-left corners are the abbreviations for the dataset types.

In a map, a given concept can be represented by different types of geometries: points, lines, and polygons (vector) or tessellated squares (raster). When polygon datasets are *tessellated*, they are covering the entire extent of the dataset without any gaps. Furthermore, different conceptualizations of the same geometry express different spatial semantics [25]. In Table 1, we suggested a way to capture the resulting diversity of dataset interpretations based on the combinations of geometries and the *object* and *field* concepts. Different semantic interpretations of the same geometry type are summarized into groups (see table columns). Example maps for these dataset types are shown in Fig. 1. Points and lines can represent fields (e.g. pointwise temperature measurements; contour lines) as well as objects (intersections; roads). We furthermore distinguish tessellated datasets that represent objects, called *lattices* (e.g. administrative units), from ones representing fields. An example of the latter is *contour polygon maps* which depict a strictly ordered value gradient of a field [14]. Following [25], we also gave *coverages* (e.g. landcover data) and *patches* (non-tessellated regions of homogeneous landuse) a field interpretation. This allowed us to distinguish landcover maps that are self-similar and thus can be dissected arbitrarily for spatial analysis from objects with spatial

unity[4]. Raster datasets predominantly represent fields. However, we consider a square lattice as a particular form of object tessellation where each polygon is squarely shaped and measures some amount, e.g., of population.

3 Methodology

3.1 Theoretical framework

As discussed earlier, the core concepts can be used to interpret questions, datasets, and analytical tools. This study focuses on dataset interpretation as the simplest form of the three. Spatial datasets can be visualized as maps, which are easier to interpret even for participants who do not have formal GIS training. This choice makes the experiment accessible to a wide range of participants with varying GIS experiences. Furthermore, the scope of the experiment is focused on *object* and *field* core concepts only. The two content concepts are often mutually exclusive, limiting the possibility of concurrent interpretation of the same map. They are also the most prevalent core concepts occurring in spatial datasets. Hence, these two concepts are most likely to result in observable effects during the experiment.

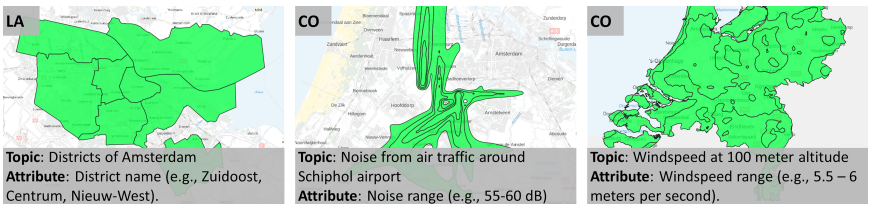


Figure 2 An example triplet of maps (LA-CO). The left-most map is city districts as tessellated objects. The other two maps visualize noise and wind-speed fields as contour maps.

As an experiment framework, we adopt a contrast model for exploring semantic similarity [15]. The basic experiment setup is to display to a participant a set of three maps (see Fig. 2). Two maps (the contrast maps) represent the same concept while the remaining map (the odd map) represents a different concept in our interpretation. For example, three maps can visualize two object datasets and one field dataset, or vice versa. Each map is accompanied by a short description of the topic and the attribute of the dataset. The participants are asked to compare the three maps and identify the one that is semantically at odds with the two others. We hypothesize that (1) participants need to rely on some form of spatial cognitive concepts to identify the odd dataset, and (2) these cognitive concepts correspond, to some degree, to our distinction of *field* and *object* core concepts.

Since core concepts are, as explained above, largely agnostic to the geometry types, participants should be able to differentiate between object and field equally well for all geometry types. In our experiment, a triplet contrasts the types from the same column of Table 1 but not the types across the columns. For example, a point object dataset is contrasted against point measure datasets only and not against any other dataset types. Similarly, a square lattice is contrasted against a raster field due to geometrical similarities when visualized as maps. This restriction limits geometric differences that may interfere

[4] A more nuanced alternative would be to interpret coverage regions and patches as particular *amount objects*, and distinguish them from objects with unity. However, since the amount concept is still under development, we remained with the simpler interpretation here. For the purpose of this paper, the relevant conceptual distinctions can still be drawn.

with the interpretation of semantic differences. Overall, Table 1 lists six possible contrasts of field and object, further referred to as contrast groups. The last row lists the names of the contrast groups. Each contrast group has two possible combinations for a triplet (two objects and one field or vice versa) resulting in 12 possible triplets in total.

3.2 Dataset compilation and annotation

Three sources provided geo-spatial datasets. *PDOK* (`https://www.pdok.nl/datasets`) is an open platform for accessing the geodata of the Dutch government. *Nationaal Georegister* (`https://nationaalgeoregister.nl`) is the data portal of the Dutch National Geo-registry. *Maps Amsterdam* (`https://maps.amsterdam.nl/open_geodat`) is the open geodata portal of Amsterdam municipality. These sources follow the same regulations for sharing open data thereby ensuring the comparable quality of datasets. All collected datasets were manually annotated by the authors of this study. This preliminary annotation involved assigning to each dataset one of the types from Table 1. The preliminary annotations were then finalized by discussing and resolving any annotation disagreements between the annotators. The main source of disagreement was the datasets with multiple attributes corresponding to different core concepts. In such cases, only one attribute was picked as being representative of the dataset. Subsequently, the selected attribute was included in the description of the corresponding map (see Fig. 1 and Fig. 2 for example descriptions). From the collected pool, we selected 36 datasets to be used in our experiment. Except for *Lattice*, three distinct datasets were selected for each dataset type. Six distinct datasets were selected for *Lattice* since it is contrasted against two other *field* types.

3.3 Survey design

We have used two online survey platforms to collect responses, Google Surveys (`https://surveys.google.com`) and Qualtrics (`https://www.qualtrics.com`). Google Surveys was used to survey random participants from the general public, while Qualtrics was used to survey a controlled selection of participants who work or study in Geography and GIScience domains. Each survey mainly consisted of a set of questions. In each question, a participant had to select an odd map when presented with a distinct triplet of maps.

Table 2 Combining Region Object (RO) and Region Patch (RP) datasets into distinct triplets.

Question Id	Odd dataset	Contrast dataset 1	Contrast dataset 2
Q13	RO1	RP1	RP2
Q14	RO2	RP2	RP3
Q15	RO3	RP3	RP1
Q16	RP1	RO2	RO1
Q17	RP2	RO3	RO2
Q18	RP3	RO1	RO3

The questions were generated with the 36 annotated datasets. Each contrast group has six datasets (three for *field* and *object* each), which were used to generate six questions each with a unique triplet combination. As an example, Table 2 demonstrates how the questions were generated for the RO-RP group. Three rules were used to assign the datasets to the triplets. First, each dataset was used as the odd one in one triplet only. Second, each dataset was used as a contrasting dataset in exactly two triplets. Third, the same combination of two contrasting datasets occurred in one triplet only. These three rules ensure that all six

datasets occur equally often in different roles while preventing the repetition of the same combination. This design ensures that no bias based on presentation frequency is introduced to participants. This design is used to generated the questions for the other contrast groups.

As discussed earlier (Fig. 2), the datasets were visualized as maps in the questions. The visual style was homogeneous across all maps except for *Raster Field* and *Square Lattice*. The polygon datasets were visualized with the same green shading and black border. The line datasets were visualized with lines of the same green color and width. Similarly, the point datasets were depicted with circles of the same size, green shading color, and black border color. *Raster Field* and *Square Lattice* datasets were visualized with color gradients that did not repeat between the datasets.

3.3.1 Design for the survey on Google Surveys

The free version of Google Surveys allows 10 questions per survey. Hence, we used only one contrast group in the survey. The survey started with a single-choice question: "*Categorize your expertise with Geographic Information Systems (GIS)*". The options were "*Laymen: never used GIS*", "*Beginner: can use basic GIS functions*", "*Trained: formally trained by a GIS course*", and "*Expert: used GIS for 5 years or more*". Except for Laymen, the three expertise categories were reused from an existing validated questionnaire on GIS [31]. Next, the survey presented the six questions from the RO-RP contrast group. The order of these questions was randomized for the survey but not per participant (not supported by the platform). Each of the six questions was accompanied with the instruction text: "*Which one of the three spatial datasets is more different from the two others in terms of spatial analyses that can be done on it.*" The final 8th question asked if participants were familiar with the core concepts of spatial information. Google Surveys was arranged to collect responses from 100 people. In total, 1205 random people from the United States were screened for the survey. 1055 participants reported as being Laymen and were screened out. Of the remaining 150 participants, 101 participants completed the survey. Another 12 participants responded as being familiar with the core concepts and were also filtered out. The responses from the remaining 89 participants (further referred to as the *general* cohort) were analyzed. The focus on non-laymen participants increases the chances of finding a positive effect and maximizes the information gain in this uncharted territory. In case of absence of a positive effect, we can safely assume that a laymen group will also not perform well.

3.3.2 Design for the survey on Qualtrics

The survey on Qualtrics started with informed consent, an agreement to which was necessary for further progression. The consent was followed by questions about the age, gender, and GIS expertise level of the participants. The expertise question used the same four options as in Google Surveys. Next, instruction on how to answer the contrast questions was shown to the participants. Finally, the participants were shown 18 questions from three randomly selected contrasts groups. The order of the 18 questions and the order of three maps within each question were also randomized per participant. We targeted two cohorts of participants differing in level of GIS expertise. The first cohort, the *student* cohort, included 61 students who were either Bachelor students in a Geo-Information minor program or attending our Applied Data Science MSc course focusing on spatial data analysis. As part of their study, the students were taught the core concepts. We selected the students who were not yet introduced to the core concepts. The second cohort, the *skilled* cohort, included participants who were manually evaluated by the investigators to have sufficient skills in Geography

and GIScience. These participants had to have a completed Master's degree in a relevant domain and actively practice in either academia or industry. Of the 40 invited participants, 18 participants completed the survey in the *skilled* cohort.

4 Results

4.1 Comparing responses in the RO-RP contrast group

Due to randomization, 29 and 9 participants from the *student* and *skilled* cohorts respectively answered the six questions from the RO-RP contrast group. These responses were compared with the responses from the *general* cohort with the 89 participants.

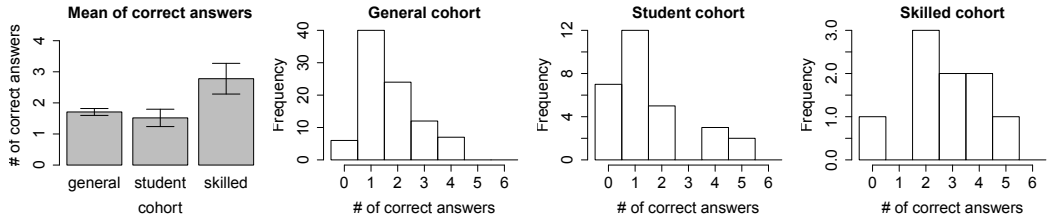


Figure 3 The left-most graph depicts means with standard errors of correct questions answered by the participants in each cohort. The remaining three graphs depict distributions of participants in three cohort according to the number of correct responses.

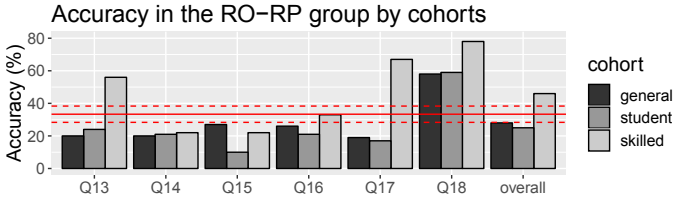


Figure 4 Accuracy for RO-RP individual questions and overall accuracy for the three cohorts. The solid red line depict expected accuracy based on a random choice. The dashed red lines depict 5% thresholds around the random choice probability.

The mean accuracies and the three accuracy distributions in Fig. 3 suggest the general and student cohorts have similar performances and the skilled cohort shows a better performance. The three distributions were analyzed for identicality with a Kruskal Wallis Test. The test is an alternative to one-way ANOVA for cases with non-normal distributions and uneven sample sizes. The test indicates a significant difference between the three distributions: $H(2) = 8.25, p = .02$. We did a follow-up pairwise comparison of the distributions with the Dunn's test with the Holm–Bonferroni correction for multiple testing. As suspected, the general and student cohorts are not significantly different ($p = 0.15$, *adjusted* $p = 0.46$). The skilled cohort is significantly more accurate than the student cohort ($p < 0.01$, *adjusted* $p = 0.01$). The skilled cohort is not significantly different from the general cohort with the Holm–Bonferroni correction (*adjusted* $p = 0.07$) but is significantly more accurate without the adjustment ($p = 0.02$). Therefore, we suspect there may have been a significant difference between the two cohorts if the sample size for the skilled cohort was bigger.

Table 3 *p-value* from two-sided exact binomial tests of accuracies in Fig. 4 for significant difference from the expected probability of 0.33 from random choice. The *correct* rows list the numbers of correct responses. The colored cells indicate significant or near significant p-values.

		Q13	Q14	Q15	Q16	Q17	Q18
general ($N = 89$)	correct	18	18	24	23	17	52
	p-value	0.01	0.01	0.26	0.18	0.01	< 0.01
student ($N = 29$)	correct	7	6	3	6	5	17
	p-value	0.43	0.17	0.01	0.17	0.08	0.01
skilled ($N = 9$)	correct	5	2	2	3	6	7
	p-value	0.17	0.73	0.73	1	0.07	0.01

We have calculated overall accuracies from the pooled responses of all participants for all questions and applied exact binomial tests against the success rate of 33% from the random choice strategy. The results are 28% ($N = 534$, $p = 0.03$), 25% ($N = 174$, $p = 0.03$), and 46% ($N = 54$, $p = 0.04$) for the *general*, *student*, and *skilled* cohorts respectively. The significant results indicate that the participants use specific strategies instead of random guesses. However, both *general* and *student* cohorts use ineffective strategies with their performance being below the chance threshold, and only the *skilled* cohort uses strategies that are more effective than random guessing. Finally, Fig. 4 depicts accuracies for individual questions and by cohorts. Table 3 lists the results of exact binomial tests of these accuracies against the chance probabilities. For the *general* and *student* cohorts, the accuracies are either below or at the chance level except for question Q18. The accuracies are exceptionally and significantly high for Q18 (RO-RP) in all three cohorts. We explore potential explanations for these results in the Discussion section.

4.2 Comparing performance across the six contrast groups

Out of 18 questions, on average, 8.9 ($SD = 2.6$) and 10.9 ($SD = 2.7$) questions are answered correctly in the *student* and *skilled* cohorts respectively (Fig. 5). These constitute 49% and 61% success rates respectively, which are considerably higher than the 33% success rate expected with the random choice strategy. The Kruskal Wallis test indicates that the *skilled* cohort is significantly more accurate than the *student* cohort ($H(1) = 8.79, p < .01$). A follow-up test is not necessary since there are only two cohorts.

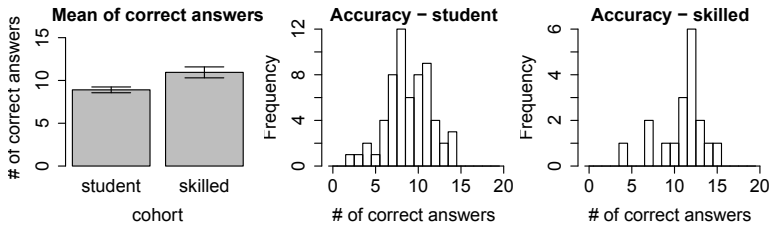


Figure 5 The left-most graph depicts means with standard errors of correct questions answered by the participants in each cohort. The remaining two graphs depict distributions of participants in the *student* and *skilled* cohorts according to the number of correct responses.

Next, we have calculated the averages of participants' accuracies for each contrast group. The results are shown in Fig. 6. Interestingly, compared to the RO-RP contrast group, the participants performed considerably better in the five other contrast groups. According to the

results of the two-sided exact binomial tests (Table 4) both cohorts demonstrate above chance accuracies in these five contrast groups. The only exception is M=43% ($SE = 10.4\%$) average accuracy of the *skilled* cohort in the SL-RF group. However, considering the magnitude of the standard error, we suspect the test would have been significant with a bigger sample size for the *skilled* cohort.

The overall results suggest that participants can distinguish well between the maps depicting *object* and *field* across most data representations. Tessellated polygon and line representations achieve higher accuracies across both student and skilled cohorts, whereas PM-PO and RO-RP were mastered significantly better by skilled users. However, it is also interesting that overall, participants' performance in the RO-RP contrast group is significantly different than in the other contrast groups. A probable explanation for this result is discussed in the next section.

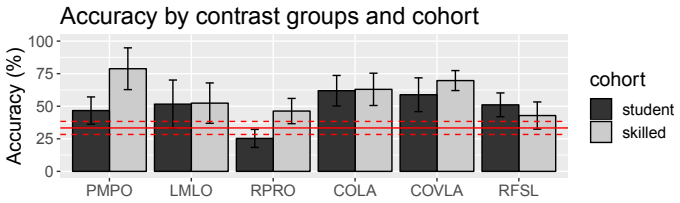


Figure 6 Mean accuracies for the six contrast groups calculated separately for the *student* and *skilled* cohorts. The mean is calculated from the proportions of the correct answers per a participant. The interval bars depict standard error intervals.

Table 4 *p-value* from two-sided exact binomial tests of accuracies in Fig. 6 for significant difference from the expected probability of 0.33 from the random choice strategy. The rows *correct/total* list correct and total responses respectively.

		PO-PM	LO-LM	RO-RP	LA-CO	LA-COV	SL-RF
student	correct/total	84/180	96/186	44/174	104/168	120/204	95/186
	p-value	0	0	0.03	0	0	0
skilled	correct/total	52/66	22/42	25/54	34/54	46/66	18/42
	p-value	0	0.01	0.04	0	0	0.19

5 Discussion

The responses in the RO-RP contrast group suggest two distinct dominant strategies used by the participants. The basic strategy is to select the option that is the most visually contrasting from the two other maps. For example, the correct option in Q18 (Fig. 7) is also the most visually contrasting map, which results in the majority of correct responses in all three cohorts (Fig. 4). The participants in the general and student cohorts seem to prefer this strategy in most questions. However, this visual strategy fails to consider semantics leading to mostly incorrect responses and low accuracy (Fig. 4) for these two cohorts. The existence of this strategy also explains why the two cohorts perform worse than if they simply would have guessed. For example, Fig. 7 depicts how this visual strategy leads the general and student cohorts into incorrect responses in questions Q17 and Q15. For Q17, 54% and 69% of all responses selected the middle map in the general and student cohorts respectively. Similarly, the visually distinct right map in Q14 is selected in 58% and 69% of all responses.

The second more advanced strategy is to compare the maps based on the attribute descriptions. More specifically, the participants try to distinguish between the maps with categorical and non-categorical attributes. Only the skilled cohort seems to efficiently employ this strategy, which explains why its response pattern diverges from the other two cohorts' patterns. This strategy is also more likely to lead to correct responses, which explains the higher accuracy of the skilled cohort (Fig. 4). For example, in question Q17 (Fig. 7), only the correct option mentions categories in its attribute description, thus, resulting in a correct response. In question Q14, however, the skilled participants fail to identify that the right dataset is also categorical since the same place name can be associated with multiple polygons and does not identify a distinct object on a map. Therefore, the strategy leads to the incorrect map that explicitly mentions category, which accounts for 56% of all responses.

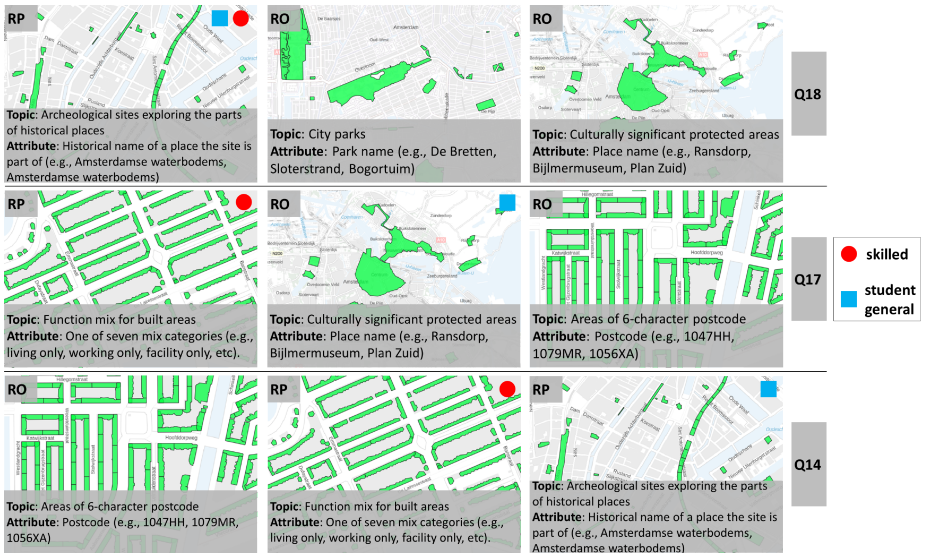


Figure 7 The questions Q18, Q17, and Q14 from the the RO-RP contrast group. In each question, the left-most map is the odd one (the correct option). The blue rectangles and the red circles in the top-right corners mark the most frequently selected options among the student/general and skilled cohorts respectively.

The two strategies together explain the most frequent responses by cohorts in each question of the RO-RP contrast group. Visual contrast-based strategy is commonly observed in many decision-making and visual search tasks [23]. The strategy often relies on the well-known bottom-up visual pop-out effect [29] where an object with an odd color (e.g. a red dot among blue dots), shape, or orientation automatically stands out and attracts priority attention. The low-effort bottom-up (automatic) nature of the pop-out effect makes the visual strategy more preferred to the less trained participants in the general and student cohorts. The second strategy of comparing attributes requires a more thoughtful approach and, more importantly, recognition that attributes play an important role in the conceptual interpretation of the datasets [27, 25]. Such a strategy requires an ability to distinguish between categorized and named datasets, which is not easy but can be improved with experience. The more experienced skilled cohort is more willing to apply the second strategy. Such experience-based transition from a simple bottom-up visual strategy to a top-down mental strategy is documented in other tasks [4, 16]. Hence, we can reasonably assume that such a transition is also happening from the basic to the advanced strategies.

Fig. 6 shows that the student cohort is more accurate in the five other contrast groups than in the RO-RP contrast group. These five contrast groups are geometrically represented by points, lines, and tessellations. Unlike the RO-RP maps, the maps within these group are visually more homogeneous, giving less room to apply the basic visual strategy. Hence, we assume that the student cohort is incentivised to use the advanced strategy that is more likely to lead to the correct responses and overall higher accuracies than in the RO-RP group. This implies that the participants in the student cohort know the advanced strategy but prefer to use the basic strategy when possible. This preference likely stems from the fact the advanced strategy requires more effort that can be minimized with training and more experience. Nevertheless, there is also unexplained variance in performance between the contrast groups suggesting other decision making factors that should be investigated further.

6 Conclusion

In most of the contrast groups, the student and skilled cohorts demonstrated significantly high accuracies (Fig. 6) thereby supporting the thesis that people can effectively distinguish maps with different concept-based interpretations (Research Question 1). An exception is the RO-RP contrast group where the general and student cohorts showed lower accuracies than would have been achieved with the naive guessing strategy. The pattern of responses suggests that a visual presentation of geometric shapes significantly interferes with a participant's ability to conceptually interpret the maps (Research Question 2). However, it should be noted that this interference can be an artifact of the experimental design based on the contrasting three maps, and requires further investigation. Finally, the skilled cohort consistently demonstrated better performance than the general and student cohorts suggesting that experience plays an important role (Research Question 3). However, higher experience seem to result in better utilization of the existing concepts necessary for interpretation of the maps rather than the development of new concepts. This result suggests that even people untrained with spatial data may have certain conceptual notions similar to the core concepts of spatial information. Overall, the study provides evidence supporting the existence of mental analytical skills that are of comparable use for distinguishing maps in terms of concepts of spatial information. The future studies should focus on replicating the current findings and verifying whether they can be generalized to laymen population. We should further explore if the core concept distinction applies to other aspects of geo-analytics such as analytical tools and questions.

References

1 Jochen Albrecht. Universal analytical gis operations: A task-oriented systematization of data structure-independent gis functionality. *Geographic information research: Transatlantic perspectives*, pages 577–591, 1998.

2 Johannes Brauner. *Formalizations for geooperators-geoprocessing in Spatial Data Infrastructures*. PhD thesis, Technische Universität Dresden, 2015.

3 Nicholas R Chrisman. *Exploring geographic information systems*. Wiley New York, 2002.

4 Marc Destefano, John K Lindstedt, and Wayne D Gray. Use of complementary actions decreases with expertise. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, volume 33, 2011.

5 Roger M Downs. The representation of space: Its development in children and in cartography. In *The development of spatial cognition*, pages 349–372. Psychology Press, 2013.

6 Max J Egenhofer and David M Mark. Naive geography. In *International Conference on Spatial Information Theory*, pages 1–15. Springer, 1995.

7 Thomas R Etherington. Teaching introductory gis programming to geographers using an open source python approach. *Journal of Geography in Higher Education*, 40(1):117–130, 2016.

8 Philip J Gersmehl and Carol A Gersmehl. Spatial thinking by young children: Neurologic evidence for early development and "educability". *Journal of Geography*, 106(5):181–191, 2007.

9 Philip J Gersmehl and Carol A Gersmehl. Spatial thinking: Where pedagogy meets neuroscience. *Problems of Education in the 21st Century*, 27:48, 2011.

10 Reginald G Golledge. Geographical perspectives on spatial cognition. In *Advances in psychology*, volume 96, pages 16–46. Elsevier, 1993.

11 Michael F Goodchild. A spatial analytical perspective on geographical information systems. *International journal of geographical information system*, 1(4):327–334, 1987.

12 Michael F Goodchild. Towards an enumeration and classification of gis functions. In *International Geographical Information Systems Symposium*, 1988.

13 Michael F Goodchild, May Yuan, and Thomas J Cova. Towards a general theory of geographic representation in gis. *International journal of geographical information science*, 21(3):239–260, 2007.

14 Torsten Hahmann and E Lynn Usery. What is in a contour map? In *International Conference on Spatial Information Theory*, pages 375–399. Springer, 2015.

15 U. Hahn and E. Heit. Semantic similarity, cognitive psychology of. In Neil J. Smelser and Paul B. Baltes, editors, *International Encyclopedia of the Social and Behavioral Sciences*, pages 13878–13881. Pergamon, Oxford, 2001. doi:10.1016/B0-08-043076-7/01548-5.

16 Diane F Halpern and Jonathan Wai. The world of competitive scrabble: Novice and expert differences in visuopatial and verbal abilities. *Journal of Experimental Psychology: Applied*, 13(2):79, 2007.

17 Toru Ishikawa. Spatial thinking in geographic information science: Students' geospatial conceptions, map-based reasoning, and spatial visualization ability. *Annals of the American Association of Geographers*, 106(1):76–95, 2016.

18 Donald G Janelle and Michael F Goodchild. Concepts, principles, tools, and challenges in spatially integrated social science. *The SAGE handbook of GIS and society*, pages 27–45, 2011.

19 Johannes F Kruiger, Vedran Kasalica, Rogier Meerlo, Anna-Lena Lamprecht, Enkhbold Nyamsuren, and Simon Scheider. Loose programming of gis workflows with geo-analytical concepts. *Transactions in GIS*, 25(1):424–449, 2021.

20 Werner Kuhn. Core concepts of spatial information for transdisciplinary research. *International Journal of Geographical Information Science*, 26(12):2267–2276, 2012.

21 Werner Kuhn and Andrea Ballatore. Designing a language for spatial computing. In *AGILE 2015*, pages 309–326. Springer, 2015.

22 David M Mark. Human spatial cognition. *Human factors in geographical information systems*, pages 51–60, 1993.

23 Enkhbold Nyamsuren and Niels A Taatgen. Pre-attentive and attentive vision module. *Cognitive systems research*, 24:62–71, 2013.

24 Simon Scheider and Tom de Jong. A conceptual model for automating spatial network analysis. *Transactions in GIS*, 2021.

25 Simon Scheider, Rogier Meerlo, Vedran Kasalica, and Anna-Lena Lamprecht. Ontology of core concept data types for answering geo-analytical questions. *Journal of Spatial Information Science*, 2020(20):167–201, 2020.

26 Simon Scheider, Enkhbold Nyamsuren, Han Kruiger, and Haiqi Xu. Geo-analytical question-answering with gis. *International Journal of Digital Earth*, 14(1):1–14, 2021.

27 Elizabeth S Spelke and Katherine D Kinzler. Core knowledge. *Developmental science*, 10(1):89–96, 2007.

28 Eric J Top. The semantics of extensive quantities in geographical information. Master's thesis, Utrecht University, 2021.

29 Jeremy M Wolfe and Todd S Horowitz. What attributes guide the deployment of visual attention and how do they do it? *Nature reviews neuroscience*, 5(6):495–501, 2004.

30 Haiqi Xu, Ehsan Hamzei, Enkhbold Nyamsuren, Han Kruiger, Stephan Winter, Martin Tomko, and Simon Scheider. Extracting interrogative intents and concepts from geo-analytic questions. *AGILE: GIScience Series*, 1:1–21, 2020.

31 Hao Ye, Michael Brown, and Jenny Harding. Gis for all: exploring the barriers and opportunities for underexploited gis applications. *OSGeo Journal*, 13(1):19–28, 2014.

Generalized, Inaccurate, Incomplete: How to Comprehensively Analyze Sketch Maps Beyond Their Metric Correctness

Angela Schwering[1]
Institute for Geoinformatics, Universität Münster, Germany

Jakub Krukar
Institute for Geoinformatics, Universität Münster, Germany

Charu Manivannan
Institute for Geoinformatics, Universität Münster, Germany

Malumbo Chipofya
ITC, University of Twente, Enschede, The Netherlands

Sahib Jan
Independent Researcher, Munich, Germany

Abstract

Sketch mapping is a method to investigate a person's spatial perception and knowledge about the surrounding environment. While cartographic maps can be easily evaluated with respect to the represented features, map scale, and spatial accuracy, there still does not exist a comprehensive method to evaluate sketch maps. This paper aims to overcome this gap and proposes a sketch map analysis method that allows for analyzing the completeness, generalization and (qualitative) spatial accuracy of the sketched information in a three-step process. After describing the method, we illustrate how our computer-supported method performs in a use case with three sketch maps. Our approach may assist researchers in geography, psychology, and education to evaluate spatial knowledge in a systematic way independent of specific research questions and experimental scenarios.

2012 ACM Subject Classification Applied computing → Psychology

Keywords and phrases sketch map analysis, spatial knowledge evaluation, cognitive map

Digital Object Identifier 10.4230/LIPIcs.COSIT.2022.8

Funding This work has been supported by the German Research Foundation grant SCHW1372/7-3.

Acknowledgements We thank all supporters of the Sketchmapia project over the last ten years.

1 Introduction

Studying spatial knowledge of humans is a challenge that we face in many different settings in spatial cognition, e.g. when studying the participant's performance in spatial tasks, when studying the nature of cognitive processes, or when studying the effectiveness of different spatial representations or wayfinding assistance systems. We aim to explore how much of the presented information participants recall. Did they capture every detail? Did they focus on the important spatial aspects? What type of information did they consider important? What kind of knowledge did they acquire and how accurate is this knowledge?

Sketch mapping is a method with a long tradition to explore spatial knowledge and people's mental maps of their surroundings. Researchers have studied the elements of spatial knowledge [11, 26, 3], studied the distortions of size, distance and directions [19, 26], and

[1] Corresponding author

studied factors influencing the sketch map quality [7, 2]. Sketching is commonly used to evaluate the spatial memorization performance in experiments [16]. Despite the fact that sketch mapping is a great method to capture the configuration of features in a two-dimensional map that reveals distance and directional relations between objects sketched, they suffer from the fact that to date there exists no comprehensive method to analyze sketch maps. In 2016, Montello claims that "Analyzing sketch maps is something of a notorious problem in research" [15, p. 174] and not much has changed since then. Researchers have been counting the number of sketched features, the existence of particular landmarks, and analyzing the properties of such features (distance, direction, shape). A successful approach is bidimensional regression [6, 9], which analyzes the spatial distortion between selected sets of points (e.g. a set of landmarks) ignoring the map layout (street network etc.) between these points.

A core problem in analyzing sketch maps in comparison to cartographic maps is the fact that human spatial knowledge is incomplete, generalized and schematic. So are sketch maps. To evaluate a sketch map, we need a method that can handle incompleteness, generalization and schematization while comparing sketched information to a base map which is considered as the "correct" ground truth. In this paper we present a comprehensive sketch map analysis method that analyzes the information content with respect to the

- degree of completeness: How many of the features in the base map are covered by a corresponding feature in the participant's sketch map? A participant with a more complete sketch map is considered to have a better memorization performance than participants with a less complete map.
- degree of generalization: How many features in the sketch map are represented at the same level of generalization as the base map? Do participants recall all details or do they recall the information at a more abstract level?
- qualitative spatial accuracy: How correct is the spatial configuration of skeched features? E.g. do participants place landmarks along the correct street segments which themselves are connected in the correct way?

It is necessary to analyze these three aspects within a single method for two reasons. First, it is impossible to determine one of them independently. The completeness analysis informs us about which elements should be considered in the generalization analysis. Qualitative correctness can only be analyzed if the alignment of generalized objects is done. Second, even small sketch maps easily become too complicated for humans to analyze as the number of spatial relations among the drawn features grows exponentially. Visual inspection of the accuracy of spatial relations is nearly impossible for human rater when generalization (example in Figure 1) and incompleteness (example in Figure 2) is involved. A consistent, computer-supported method is necessary to ensure a systematic evaluation.

2 Background

A cognitive map is a mental model that encompasses the internal processes that enable people to acquire and operate information about the physical environment [5]. Information in cognitive maps is not as it is in two-dimensional cartographic maps. "Instead, cognitive maps are complex, highly selective, abstract, generalized representations in various forms" [5, p. 18]. Human spatial knowledge in cognitive maps is incomplete and fragmented [8]. This is not only due to our limited memory capacity and the natural process of spatial knowledge fading out over time, but rather the result of our cognitive processes, using objects to establish a frame of reference for other objects to localize, relate and provide orientation. Capturing information at different levels of detail is not seen as a sign of memory failure, but as a

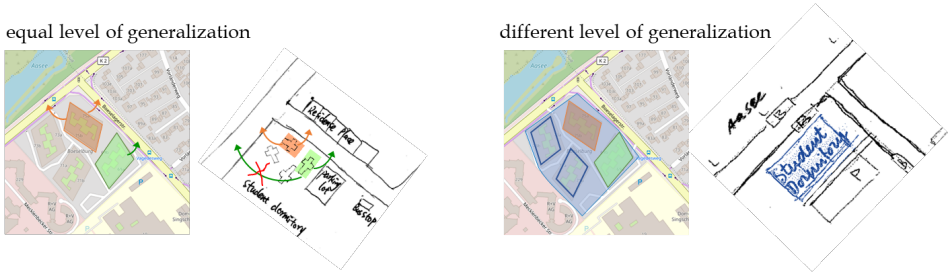


Figure 1 Challenge: Evaluating the spatial accuracy when generalization is involved. The student dormitory highlighted in orange is adjacent to two streets, while the student dormitory highlighted in green is not. To compare the spatial relations with a generalized visualization in the sketch map, we need to change the level of generalization in the base map. Similarly, the student dormitory highlighted in green has an adjacency relation to the parking lot, which is not present for the student dormitory highlighted in orange, but cannot be distinguished at a generalized level.

consequence of normal information processing. Spatial memory is organized in hierarchies and categories [26]. This leads to various effects of distortions well studied in Psychology since decades [24].

Sketch maps, as externalizations of cognitive maps, reflect distortions and errors that are originated in cognitive maps. For example, distances between near spatial objects are considered relatively longer than distances between far away ones [9]. Ordinary buildings are judged closer to landmarks than the other way around [20, 14]. Routes with more turns and intersections [20] or more landmarks [25] are judged longer. Spatial information is also simplified in cognitive maps. For instance, angles tend to be perceived more rectangular, and curved features are perceived straighter [28]. There are other typical cognitive impacts found in sketch maps such as errors of quantities, shape, size, and inconsistent scales. Another phenomenon of human spatial knowledge is its generalization. The process of generalization can be found in cartography to represent the same spatial information at different levels of detail. In cognitive maps – or following Tversky better called cognitive collages [27] - this process of generalization happens in an inconsistent way. When externalizing such knowledge, this leads to different generalization levels integrated within a single sketch map produced by a single participant. Generalization of spatial information is a well-researched problem in cartography – although their generalization is consistently applied across the whole map. The generalization types suggested below are inspired by cartographic generalization, but are however differ since generalization is rather conceptual in sketch maps and applied inconsistently across the map.

Regarding the characteristics of sketch maps, there are two principles being followed in this paper: first, sketch maps contain invariant spatial information as a necessity for people to conduct any spatial behaviour in the physical environment; second, cognitive impacts should be taken into account when sketch maps are under analysis, because they cause inaccuracy in sketch maps.

Methods to evaluate Spatial Knowledge Acquisition. Experiments in wayfinding research may pursue very different targets – e.g. investigate human wayfinding strategies, the influence of different wayfinding instructions, or the effect of digital wayfinding assistance on wayfinding performance or spatial knowledge acquisition. While some studies focus on landmark knowledge acquisition, others study route or route network knowledge acquisition, configural or survey knowledge acquisition.

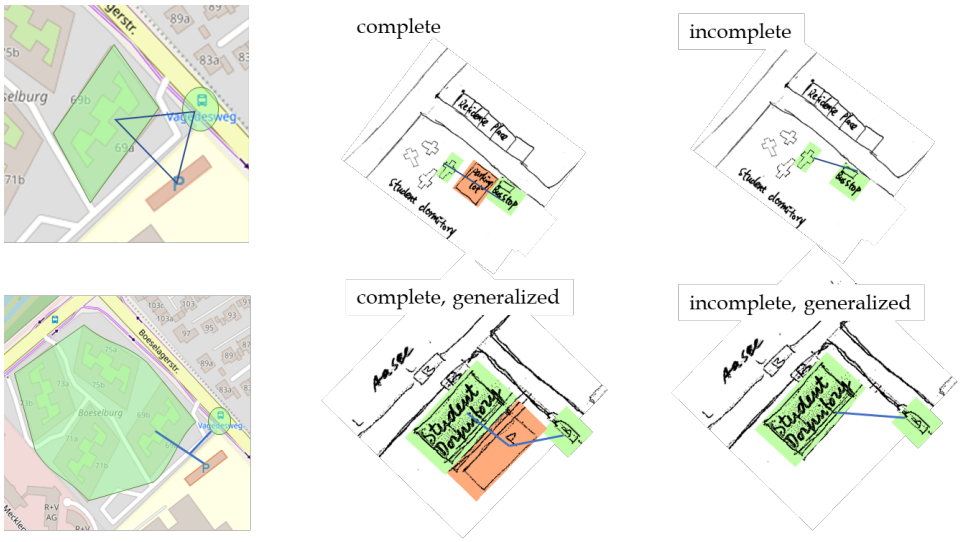


Figure 2 Challenge: Evaluating the spatial accuracy when incompleteness is involved. In the base map, the bus stop is in-between the dormitory and the parking lot. In the complete sketch maps, the spatial arrangement of bus stop and parking lot is wrong, while in incomplete sketch maps, the arrangement of dormitory, bus stop and street is (qualitatively) correct.

The probably simplest method to analyze spatial knowledge acquisition are recall tasks. Participants can be asked to mention verbally all features they recall or to select landmarks that they recall from a set of landmark pictures. To test their route knowledge, participants may be asked to order landmarks on the route correctly or to place them correctly on locations on a base map. Distance estimates are oftentimes combined with direction estimates where participants have to point in the direction of the landmark assuming that they are standing at a particular location. Ranking landmarks according to their distance and direction as well as placing them onto a two-dimensional map involve certain survey knowledge. To elicit spatial distance information between places but to avoid sketch mapping, experiments oftentimes apply the method of multidimensional regression.

Methods to analyze the information content in sketch maps. Approaches to analyze sketch maps include simple counting approaches (e.g. counting streets, landmarks, other particular features) or verifying the existence of features with a particular interest for the experiment. This method has been used in many studies, however Billinghurst and Weghorst (1995) was the first to name this method as "completeness". There are different variations in measuring completeness such as counting nodes, paths drawn in sketch map etc. Since, this method does not have any spatial relevance, it is usually accompanied by "map goodness". "Map goodness" [1] is measured by asking experts to rate sketch maps based on certain criteria such as "how useful the sketch maps are for navigation purpose?" etc.

The best known quantitative approach is bi-dimensional regression [6, 9]: It analyzes the degree of shape and scale distortions between reference points in the map. These are collectively called quantitative approaches as they only capture the metric aspect of sketch map such as scale, angle etc.

Qualitative approaches give some guidelines how to analyze qualitative aspects of sketch maps. The Qualitative Matching approach developed in [4] represents a sketch map as a set of qualitative constraint networks (QCN) one for each aspect of space. This approach supports only limited number of qualitative relations and was never extensively tested in the context of sketch map evaluation analysis.

Spatial Scene Similarity by Nedas [17, 18] proposed a similarity measure to compare two spatial scenes that takes into account (i) the similarity between objects in the two scenes; (ii) the similarity between the binary relations among objects in the two scenes; and, (iii) the ratio of the total number of objects in both scenes to the number of objects that have been matched – or equivalently, not matched. This approach goes into a promising direction, although it was never applied and tested in the context of sketch map evaluation.

Last but not least, there is our SketchMapia approach [22], analyzing different sketch aspects to align sketch maps and metric maps. A set of sketch aspects were identified that are used to compare sketch maps with a base map. So far, the approach was only applied to sketch map alignment, not sketch map evaluation. However, the set of spatial sketch aspects identified in [28] is useful for this approach, because it describes spatial relations between drawn objects, which - if the relations in the sketch map are not identical to the relations from a topographic map - should be considered as erroneous in the sketch map. The invariant sketch aspects are further described below.

3 The Research Gap

The related work in section 2 demonstrates that different approaches have been applied to separately detect completeness and spatial accuracy. We believe, that spatial accuracy cannot be analyzed without a systematic approach to capture generalization and generalization cannot be analyzed without having determined missing objects. A comprehensive sketch map analysis method therefore has to address all three aspects jointly. However, completeness, generalization, and spatial accuracy are not always easily separable. Sometimes missing a feature leads to a generalization of other objects which would have been separated if the missing feature was drawn. Generalization might affect the spatial relations that can be determined in a map. For example, grouping several objects into one such as in Figure 1 changes their spatial relation to the streets: while the dormitory area is adjacent to both streets, the single buildings are not.

Thus, the first challenge we address is the integration of the three aspects into one formalized and structured procedure applicable to different sketch maps. While simple counting methods will suffice to measure completeness, we aim to build upon our previous work on generalization types in sketch maps [12, 13] and apply it to the context of spatial knowledge evaluation. For determining the spatial accuracy, we will build upon the sketch aspects we identified for sketch map alignment [22, 28] to detect spatial configurations which should be considered as erroneous. Note that within this approach a geographically inaccurate location along the correct street segment is not considered an error.

4 A comprehensive sketch map analysis method

Our comprehensive sketch map analysis method can be useful for sketch maps in various experimental set-ups, e.g. for sketch maps drawn directly from memory, after prior exploration of the scene with or without assisted wayfinding, or for sketching as a recall task after a learning phase. We first describe how data for the sketch map analysis method might be acquired and afterwards explain, how each aspect is captured by our comprehensive method.

Data Acquisition in an Experiment. Experiments in wayfinding research may pursue very different goals. We may investigate wayfinding strategies, the influence of different wayfinding instructions, or the effect of digital wayfinding assistance on wayfinding performance. In

spatial learning tasks, the spatial knowledge acquired might tell us about the memorability of different communication formats or the nature of different cognitive processes. Sketch mapping is commonly applied to learn more about a participant's spatial knowledge, e.g.:

- what features did the participant remember,
- at which level of abstraction did the participant recall objects,
- was the participant able to recall them in the correct spatial configuration.

Based on the research question that shall be answered, the experimenter has to decide what information they want to extract from sketch maps, i.e. which landmarks and streets the participants are expected to recall in a 'perfect' (i.e., best possible) sketch map. Next, the experimenter digitizes the base map of the experimental area and the sketch maps collected throughout the experiment.

Sketch maps may also include additional features. These may be additional features from the real world that are not captured in the base map or additional features made up by the participant. As our base line for comparison are only features digitized in the base map, we do not consider such additional features in the sketch map analysis.

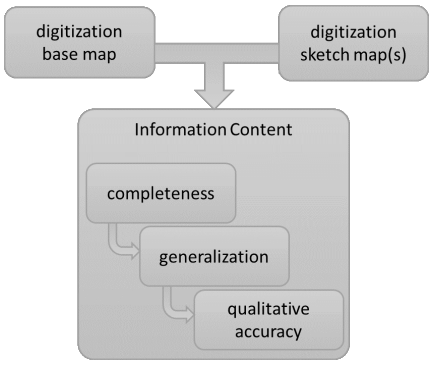


Figure 3 Procedure of the Sketch Map Analyzer: To evaluate the information content of a sketch map, we need to compare it to a base map.

Further, we distinguish between analyzing its information content and the type of the map: The information content is analyzed in a sequential order with respect to the completeness of the sketch map, the degree of generalization of sketched features and with respect to the qualitative accuracy.

Sketch Map Analysis Step 1: Completeness. The sketch map's completeness tells about the ratio of features captured in the sketch map with respect to all base map features. Depending on the design of the experiment, the sketch map's drawn features may be compared to the (very large set of) features in the corresponding map section of a topographic map or to a subset of features selected by the experimenter. We identify all features that have not been sketched. All other features – independently of whether they are drawn at the same level of generalization or in an accurate way, are counted for the number of drawn objects. We interpret the amount of recalled features as an indicator for more or for less comprehensive spatial knowledge acquisition.

Sketch map completeness is the first step of the analysis process, because it determines the set of features in the base map that need to be aligned to (generalized or non-generalized) features in the sketch map.

Sketch Map Analysis Step 2: Generalization. The degree of abstraction, respectively the detailedness, is considered as an indicator for the level of abstraction of spatial knowledge. The degree of generalization of a drawn object says something about a person's perception of the environment and thus about the mental model[2]. Generalization is the second step of the sketch map analysis process in which we establish alignment between features: This may be a one-to-one alignment (i.e. no generalization involved). If generalization is involved, we distinguish between group-to-one alignment and group-to-group alignment. In the first case, several detailed features in the base map are aligned with one abstract feature in the sketch map. In the latter case, several objects are drawn in the sketch map to indicate a particular pattern but not concrete features. The alignment is established at a generalized level, e.g. the houses at the side of a street are represented by a set of houses in the sketch map while not every single sketched house can be aligned to a particular house in the base map.

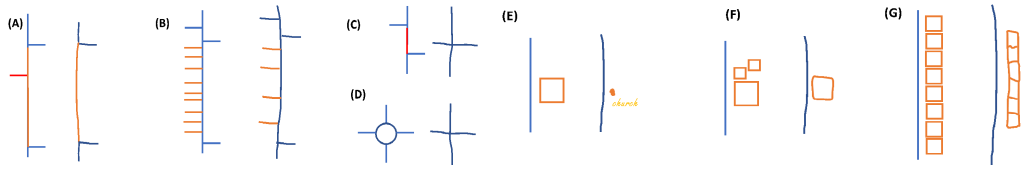


Figure 4 Seven generalization types: geometric merging for streets (A), abstraction to show existence for streets (B), junction merge (C), roundabout collapse (D), collapse for buildings (E), amalgamation for buildings (F), abstraction to show existence for buildings (G).

Figure 4 visualizes the different generalization types that we proposed in [13] based on an extensive analysis of sketch maps. Generalization type (A) and type (C) are a result of an incompletely drawn sketch map. In generalization type (A), a side street was left out and thus two street segments are merged into one. In generalization type (C), the missing street leads to a junction merge which eventually also effects the spatial accuracy of the street segments, because a new street pattern occurs. Generalization type (B) and (G) are examples of group-to-group alignments. While in the base map each feature matches one feature in the reality, the set of side streets or the set of houses do not refer to a specific streets or houses in reality; but only indicate the existence of a set of streets and houses, respectively. In generalization type (D) and (E), the extended feature in the base map is represented by a feature of lower dimensionality: The junction is collapsed into a junction, and the polygon representing the footprint of a building is collapsed into a point. Generalization type (F) amalgamates a multi-complex building to one single building.

The procedure of highlighting missing and generalized objects in a base and a sketch map was tested in [12]. Five raters received annotated the same set of 30 sketch maps which systematically differed in their degree of generalization and completeness. Out of the total of 416 features coded by the 5 participants, 53 features were generalized, 275 features were non-generalized, 82 features were not drawn (out of which 24 resulted in merging of segments due to omission of streets). Once, the color-coded data was ready, we used the Light's Kappa index [7] of the irr package in R to calculate the agreement. The overall agreement score among the participants was kappa = 0.889 with p = 0.056.

[2] Every spatial representations, not only the sketch map but also the topographic map, results from an abstraction process. Simlarly to completeness, the experimenter decides for the degree of generalization (s)he expects.

Sketch Map Analysis Step 3: spatial accuracy. The third step aims to determine spatial accuracy of a sketch map. It operates on the generalized feature alignments, i.e. in case a feature is generalized, the detailed features are replaced by the generalized one. The spatial relations are calculated also based on the generalized one. Since sketch maps are distorted and schematized due to the cognitive processes underlying the formation of a cognitive map, we strongly believe that a quantitative regression measure has only limited meaning. In our previous work [28, 22] we have investigated invariant aspects in sketch maps, i.e. (qualitative) spatial relations among sketched objects which are typically not distorted in the sketch map.

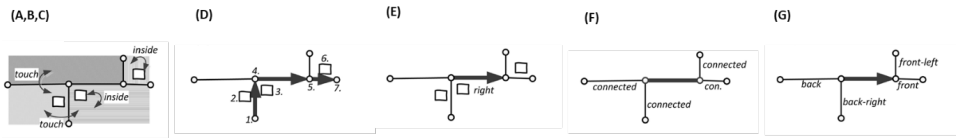


Figure 5 Seven sketch aspects to determine the qualitative correctness [22]: topological relations between landmarks (A), regions (B), and between street segments (C), linear order of features along the route (D), left/right relation of landmarks with respect to the street (E), connectivity of street segments (F), and orientation of street segments (G).

Figure 5 illustrates the six sketch aspects taken into account for calculating the qualitative accuracy. The first sketch aspects refer to topological relations: In our implementation, we calculate topological relations jointly for landmarks and regions (A, B), and separately the topological relations between street segments and landmark/regions (C). Sketch aspect (D) describes the linear ordering of landmarks and junctions along a route. A route is defined as connected street segments. Sketch aspect (E) describes whether the landmark is left or right located with respect to its nearby oriented street segments. Sketch aspect (F) describes whether two street segments are connected to each other. Sketch aspect (G) describes the binary directional relations of two street segments that coincide in at least one junction point.

In various studies [10, 21] we investigated the reliability of these sketch aspects in alignment scenario to determine at which level topological relations, ordering relations and direction relations should be distinguished to provide a reliably and accurate measurement. This challenge is highly connected to the problem of defining which factors influence a sketch map quality: Which distortions in sketch maps do we consider correct and which ones false. In our experiments we could show that the methodology of capturing qualitative relations with the abovementioned sketch aspects is feasible and demonstrates a high accuracy of >99%, but further studies with systematically varied sketch maps will be necessary to empirically validate our treshhold when a distortion is supposed to be considered as an error or not.

5 Use Case demonstration

For the demonstration of our sketch map analysis method, we simulate a wayfinding study in which participants follow a route given by wayfinding instructions which indicate which street to follow, which turns to take and refer to landmarks for better orientation. Our research aims at investigating how memorable our route instructions are. Thus, we decide to consider only features that are mentioned in our route instructions for our evaluation.

Figure 6 shows our study area. Our wayfinding instructions refer to 11 landmarks (all buildings or complex buildings) and 22 streets (12 of them form the route). Afterwards, participants are instructed to draw a sketch map on a blank piece of paper. They shall remember as many streets and landmarks mentioned in the route instructions as they can.

Figure 6 Experiment route with landmarks and streets referred to in route instructions (solid blue line), streets part of the route (dotted red), and the start and end location.

We collected a set of sketch maps out of which we will analyze three examples following the above described methodology. Figure 7 visualizes step 1 (completes) and step 2 (generalization) for sketch map 1.

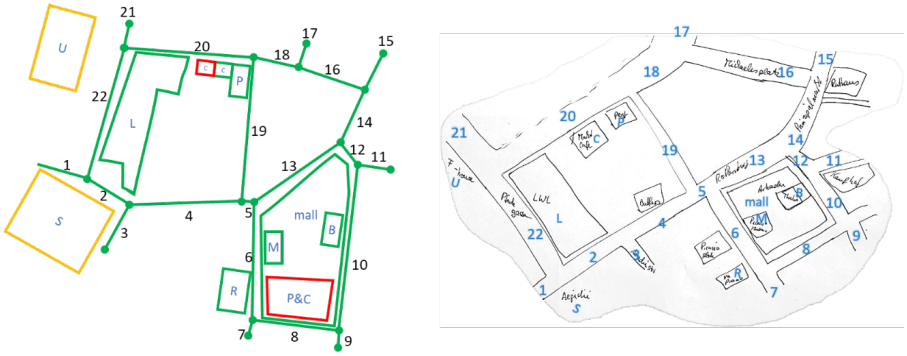


Figure 7 Sketch map 1 (right). The base map on the left indicates which features are missing (red), generalized (yellow) and which features can be one-to-one aligned with the sketch map (green).

Sketch map 1 is of very high quality: It misses only very few features: The two missing landmarks do no effect the level of generalization of the street network. In sketch map 1, nearly all objects are at the same level of abstraction as in the metric map. Only landmarks S and U were collapsed to a single label in the sketch map not specifying the footprint of the landmark (generalization type (E) in Figure 4. There are only minor qualitative errors. For example, landmark U is alongside street segment 21 and 22 in the base map, but in the sketch map landmark U is collapsed into a label and the label is only adjacent to street segment 21.

Sketch map 2 is much simpler than sketch map 1 (c.f. Figure 10 in the appendix). Many street segments are missing such that the network structure is mostly reduced to the route of the experiment. Missing street segment lead to a high generalization in the street network. Many landmarks are missing as well, however if they are drawn, they are not generalized.

This difference can be seen in the bar charts in Figure 8 (right side): while the degree of detailedness is 100% for landmarks, it is only 30% for street segments. For detailed data from the bar charts refer to Tables 1 and 2 in the appendix.

There are several qualitative errors. For example, in the base map landmark R and P&C are right/left at the opposite side of the street segment. In the sketch map, you first pass landmark R on the right side and afterwards you pass landmark P&C on the left side. This leads to wrong spatial ordering relations of landmarks on the route (Figure 9). Furthermore, street segment 22 is connected to street segments 2 in the base map, while in the sketch map segment 22 is too short (c.f. Figure 10).

In our third example (sketch map 3 is shown in Figure 11 in the appendix), less streets and landmarks are missing than in sketch map 2. This leads to a lower degree of generalization (respectively higher level of detailedness) in the street network (Figure 8). Several landmarks are placed incorrectly in the sketch map: Landmark M is supposed to be located next to street segment 6 within the mall. Landmark R is supposed to be next to street segment 6. The spatial relation between the two landmarks M and R is correct. Furthermore, landmark S is placed inccorectly. In the base map, it is adjacent to street segment 1 and 2, while in the sketch map it is adjacent to street segment 2.

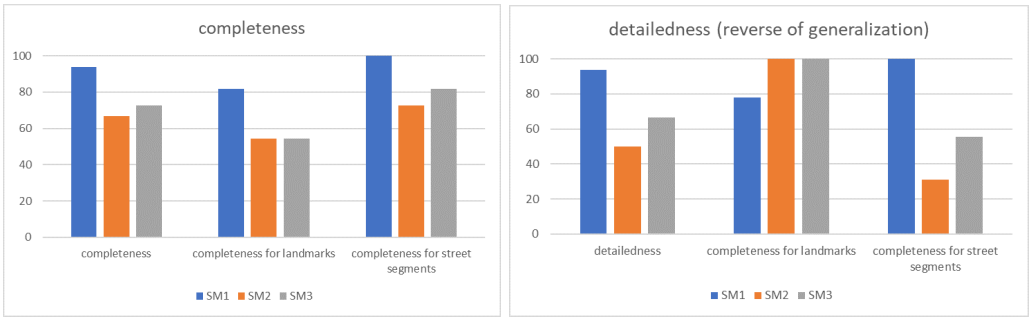


Figure 8 Results for step 1 and 2 of the sketch map analysis method: Completeness and degree of detailedness (the reverse of generalization) for all three sketch maps.

Figure 8 compares the completeness and the detailedness for all sketch maps. The missing streets lead to a high degree of generalization in sketch map 2. The charts clearly show, that completeness does not necessarily go together with a low degree of generalization. Missing streets oftentimes lead to generalization in the street network, but the degree of generalization differs.

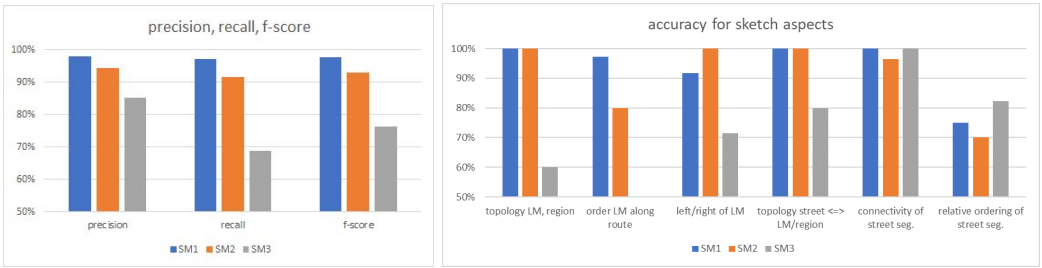


Figure 9 Results for step 3 of the sketch map analysis method: Spatial accuracy for each sketch map (precision, recall and f-score of spatial relations on the left) and the individual accuracy values for each sketch aspect (right).

For analyzing the spatial accuracy in step 3, we calculate the spatial relations for each feature and sketch aspect (c.f. Table 1 in the appendix for all relations in sketch map 1). Even for a small map such as sketch map 1, these relations easily sum up to a large number of relations which cannot be manually computed anymore. We analyse them using the precision, recall, and f-score values. Precision is the proportion of elements that are correct; recall is the proportion of information (correct or not) that was included in the sketch map; and f-score is the harmonic average of the two (their relative weight can be adjusted). Recall of spatial relations should not be confused with the completeness of features. The whole process is supported by our software implementation (screenshots in the appendix, at `http://www.sketchmapia.de` you find our open-source tool).

Sketch map 2 and 3 are missing about 50% of the landmarks, but those landmarks sketched are placed correctly in sketch map 2 and incorrectly in sketch map 3. Figure 9 shows that sketch map 1 has the highest qualitative accuracy, but also sketch map 2 – despite of the higher incompleteness, has a high precision, recall and f-score. For sketch map 3 we see a low recall, i.e. the number of correctly represented spatial relations in sketch maps compared to the number of spatial relations in the base map is relatively low. Sketch map 3 scores only 100% with respect to the connectivity of street segments, since the errors are introduced by wrong placement of landmarks.

6 Summary

Analyzing sketch maps is a common methodology to evaluate human spatial knowledge. Researchers have developed analysis methods specifically adjusted to the research question and set-up in their experiments, but to date, no methodology exists that allows for a comprehensive analysis of sketch maps. Montello outlined the key challenge: "one piece of good advice is that you should figure out what kind of information you want to get from the sketch maps, based on what research questions you want to address" [15, p. 174].

Based on previous work by the authors on analyzing sketch maps, this paper proposes to combine different analysis approaches to a three-step comprehensive sketch map analysis method accounting for the typical characteristics of cognitive maps, namely incompleteness, generalization and schematization. In step 1, completeness of the map is determined by identifying missing features in the sketch map. All remaining features will be investigated in step 2: In the generalization, we identify features that cannot be directly aligned to a single feature in the other map. We analyze the type of generalizations which end up in a one-to-group or group-to-group alignment. The third step analyzes the spatial accuracy by comparing the qualitative spatial relations among features in the base map and corresponding features in the sketch map. A software suite is implemented to support this systematic analysis.

While we are aware that the overall evaluation of the method is still to be done, we believe that – having shown the evaluations for completeness, generalization and spatial accuracy separately – we are able to demonstrate with our extensive use case that the methodology is sufficiently generic to be useful in different settings assisting researchers to analyze sketch maps in experiments in a comprehensive way. Comparisons across different experiments become possible, as the method is systematic, standardized and not research question specific.

Future Steps. Human spatial knowledge is classified into three types [23]: landmark knowledge, route knowledge, and survey knowledge. Our sketch map analysis method comprises components that evaluate different types of knowledge. For example, completeness

of landmarks, generalization of landmarks and bidimensional regression relate to landmark knowledge. Street segment completeness, generalization of street network, linear ordering of landmarks along the route as well as left/right relations along the route relate to route knowledge. Topological relations of landmarks and regions, connectivity of spatial relations leading to cycles in the graph, and the topology of street segments and regions relate to survey knowledge. As a next step, we aim to evaluate the quality of landmark, route, and survey knowledge based on the different components of our analysis method.

One open challenge is determining how individual measures obtained by our toolset correlate with each other. This way we could avoid situation in which one error in a sketch has a cascading effect on multiple measures computed by the toolset, and results in much lower scores, e.g., because it invalidated a very large number of qualitative relations. Simultaneously, other errors may have disproportionally small effect of the final scores only because they occurred in an area of a sketch map less sensitive to the problem of correlated measures or because the specific qualitative relation being wrong is less correlated with other relations.

7 References

References

1 M. Billinghurst and S. Weghorst. The use of sketch maps to measure cognitive maps of virtual environments. In *Proceedings Virtual Reality Annual International Symposium '95*, pages 40–47, 1995. doi:10.1109/VRAIS.1995.512478.

2 Mark Blades. The reliability of data collected from sketch maps. *Journal of Environmental Psychology*, 10(4):327–339, 1990.

3 Andreas D. Blaser. A study of people's sketching habits in GIS. *Spatial Cognition & Computation*, 2(4):393–419, 2000. doi:10.1023/A:1015555919781.

4 Malumbo Chipofya, Carl Schultz, and Angela Schwering. A metaheuristic approach for efficient and effective sketch-to-metric map alignment. *International Journal of Geographical Information Science*, 30(2):405–425, 2016. doi:10.1080/13658816.2015.1090000.

5 Roger M. Downs and David Stea. *Cognitive maps and spatial behavior: Process and products*, pages 8–26. Aldine Press, Chicago, USA, 1970.

6 Alinda Friedman and Bernd Kohler. Bidimensional regression: assessing the configural similarity and accuracy of cognitive maps and other two-dimensional data sets. *Psychological methods*, 8(4):468, 2003.

7 Kevin A. Hallgren. Computing Inter-Rater Reliability for Observational Data: An Overview and Tutorial. *Tutorials in quantitative methods for psychology*, 8(1):23–34, 2012.

8 Kateřina Hátlová and Martin Hanus. A systematic review into factors influencing sketch map quality. *ISPRS International Journal of Geo-Information*, 9(4), 2020. doi:10.3390/ijgi9040271.

9 Keith J Holyoak and Wesley A Mah. Cognitive reference points in judgments of symbolic magnitude. *Cognitive Psychology*, 14(3):328–352, July 1982. doi:10.1016/0010-0285(82)90013-5.

10 Sahib Jan, Angela Schwering, Carl Schultz, and Malumbo Chaka Chipofya. Cognitively plausible representations for the alignment of sketch and geo-referenced maps. *Journal of Spatial Information Science*, 14:31–59, June 2017.

11 K Lynch. *The Image of the City*. MIT Press, 1960.

12 Charu Manivannan, Jakub Krukar, and Angela Schwering. Generalisation in Sketch Maps – A Systematic Classification of Different Generalization Types for Sketch Map Analysis. *Journal of Environmental Psychology*, under review.

13 Charu Manivannan, Angela Schwering, Malumbo Chipofya, and Sahib Jan. Categorization of generalization in sketch maps. In *Poster at the 14th International Conference on Spatial Information Theory*, Regensburg, Germany, 2019.

14 Timothy P. Mcnamara and Vaibhav A. Diwadkar. Symmetry and Asymmetry of Human Spatial Memory. *Cognitive Psychology*, 34(2):160–190, November 1997. doi:10.1006/cogp.1997.0669.

15 Daniel R. Montello. Behavioral Methods for Spatial Cognition Research. In *Research Methods for Environmental Psychology*, chapter 9, pages 161–181. John Wiley & Sons, Ltd, 2016. doi:10.1002/9781119162124.ch9.

16 Stefan Münzer, Hubert D. Zimmer, and Jörg Baus. Navigation assistance: A trade-off between wayfinding support and configural learning support. *Journal of Experimental Psychology: Applied*, 18(1):18–37, 2012. doi:10.1037/a0026553.

17 Konstantinos A. Nedas. *Semantic Similarity of Spatial Scenes*. PhD thesis, The University of Maine, 2006.

18 Konstantinos A Nedas and Max J Egenhofer. Spatial-Scene Similarity Queries. *Transactions in GIS*, 12(6):661–681, 2008. doi:10.1111/j.1467-9671.2008.01127.x.

19 Thomas Saarinen, Michael Parton, and Roy Billberg. Relative Size of Continents on World Sketch Maps. *Cartographica: The International Journal for Geographic Information and Geovisualization*, 33(2):37–48, June 1996. doi:10.3138/F981-783N-123M-446R.

20 Edward K. Sadalla and Stephen G. Magel. The Perception of Traversed Distance. *Environment and Behavior*, 12(1):65–79, March 1980. doi:10.1177/0013916580121005.

21 Angela Schwering, Sahib Jan, Jakub Krukar, and Malumbo Chipofya. Evaluating Sketch Maps Qualitatively: A new Software-Supported Method. In *Poster at the International Conference on Spatial Cognition, September 2019*, Freiburg, Germany., 2019.

22 Angela Schwering, Jia Wang, Malumbo Chipofya, Sahib Jan, Rui Li, and Klaus Broelemann. SketchMapia: Qualitative Representations for the Alignment of Sketch and Metric Maps. *Spatial Cognition & Computation*, 14(3):220–254, July 2014. doi:10.1080/13875868.2014.917378.

23 Alexander W. Siegel and Sheldon H. White. The Development of Spatial Representations of Large-Scale Environments. In Hayne W. Reese, editor, *Advances in Child Development and Behavior*, volume 10, pages 9–55. JAI, January 1975. doi:10.1016/S0065-2407(08)60007-5.

24 Albert Stevens and Patty Coupe. Distortions in judged spatial relations. *Cognitive Psychology*, 10(4):422–437, October 1978. doi:10.1016/0010-0285(78)90006-3.

25 Perry W Thorndyke. Distance estimation from cognitive maps. *Cognitive psychology*, 13(4):526–550, 1981.

26 Barbara Tversky. Distortions in memory for maps. *Cognitive Psychology*, 13(3):407–433, July 1981. doi:10.1016/0010-0285(81)90016-5.

27 Barbara Tversky. Cognitive maps, cognitive collages, and spatial mental models. In Andrew U. Frank and Irene Campari, editors, *Spatial Information Theory A Theoretical Basis for GIS*, Lecture Notes in Computer Science, pages 14–24, Berlin, Heidelberg, 1993. Springer. doi:10.1007/3-540-57207-4_2.

28 Jia Wang and Angela Schwering. Invariant spatial information in sketch maps – A study of survey sketch maps of urban areas. *Journal of Spatial Information Science*, 11, 2015. doi:10.5311/JOSIS.2015.11.225.

A Appendix

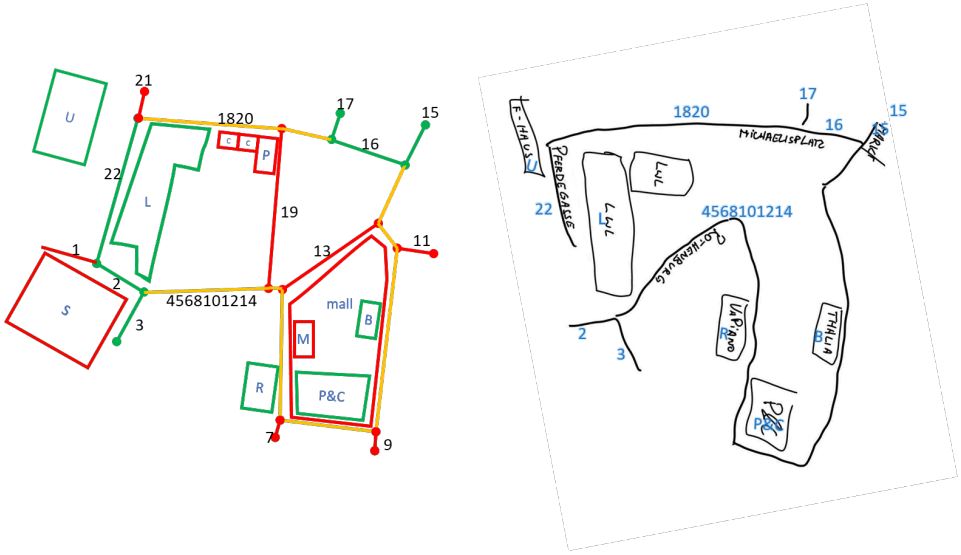


Figure 10 Sketch map 2 (right). The base map on the right indicates which features are missing (red), generalized (yellow) and which features can be one-to-one aligned with the sketch map (green).

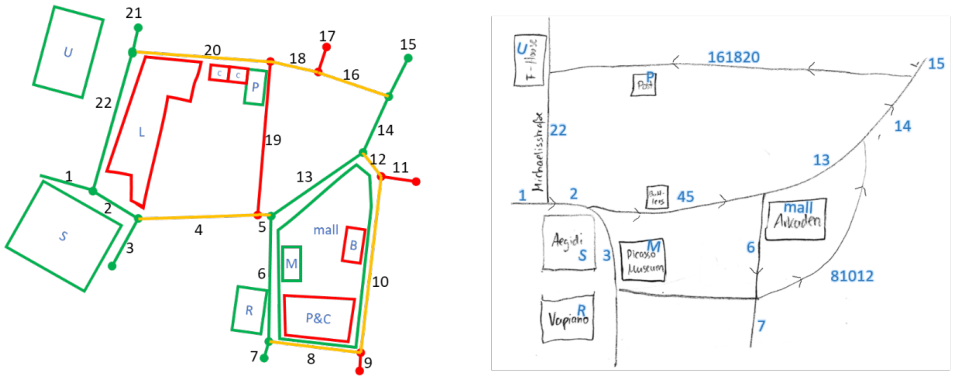


Figure 11 Sketch map 3 (right). The base map on the right indicates which features are missing (red), generalized (yellow) and which features can be one-to-one aligned with the sketch map (green).

Table 1 Sketch map 1: accuracy for each type of qualitative relation.

	base map	sketch map	correct	wrong	missing	accuracy
topology LM, region	36	36	36	0	0	100%
order LM along route	36	36	35	1	0	97%
left/right of LM	28	24	22	1	5	92%
topology street <=> LM/region	198	198	198	0	0	100%
connectivity of streets	231	231	231	0	0	100%
relative ordering of streets	36	36	27	9	0	75%

Table 2 Sketch map 2 and its completeness and generalization values for each feature.

ID	com	gen	ID	com	gen	ID	com	gen
B			1	√		12	√	merge
C1			2	√		13	√	
C2			3	√		14	√	
L			4	√	merge	15	√	
M	√		5	√	merge	16	√	merge
mall	√		6	√		17		
P	√		7	√		18	√	merge
P&C			8	√	merge	19		
R	√		9			20	√	merge
S	√	collapse	10	√	merge	21	√	
U	√		11			22	√	

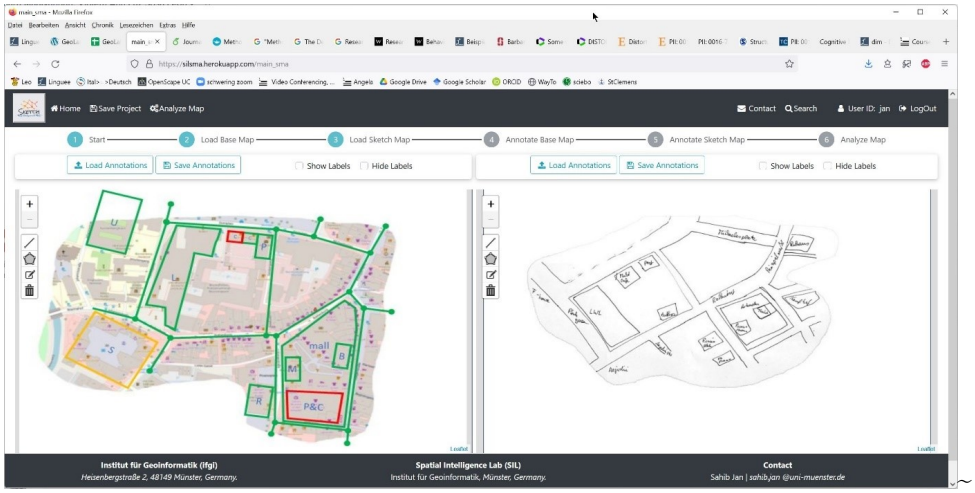


Figure 12 Screenshot 1: Annotation of incomplete, generalized and non-generalized features.

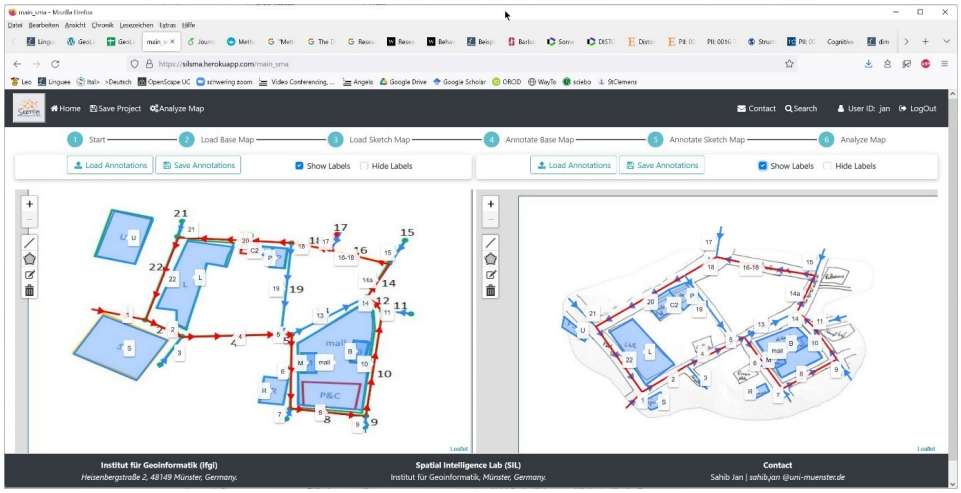


Figure 13 Screenshot 2: Digitization of base and sketch maps, indicating the route (street segment in red), and alignment via labels.

Perceptions of Qualitative Spatial Arrangements of Three Objects

Ningran Xu
Department of Infrastructure Engineering, The University of Melbourne, Parkville, Australia

Ivan Majic
Institute of Geodesy, RG Geoinformation, TU Graz, Austria

Martin Tomko[1]
Department of Infrastructure Engineering, The University of Melbourne, Parkville, Australia

Abstract

Cognitive grounding of formal models of qualitative spatial relations is important to bridge between spatial data and human perceptions of spatial arrangements. Here, we report on an experimental verification of the cognitive alignment of the recently proposed Ray Intersection Model (RIM) capturing qualitative relationships between three spatial objects, and human perceptions of spatial arrangements through a grouping task. Further, we explore arrangements with an object positioned "between" two other objects. We show that RIM has sufficient expressive power and aligns well with human perceptions of ternary spatial relationships.

2012 ACM Subject Classification Information systems → Geographic information systems; Human-centered computing → User studies

Keywords and phrases Spatial Perception, Qualitative Spatial Relationships, Betweenness, Evaluation, Ternary Relationships

Digital Object Identifier 10.4230/LIPIcs.COSIT.2022.9

Funding Work on this research has been supported by the Australian Research Council project DP170100153 Self-Healing Maps.

Acknowledgements We acknowledge the assistance of Xiaolin Zhang with the design of the back-end data structure for the survey and Yan Zhang and Maggie Zhuang for testing the survey before the main data collection.

1 Introduction

Computational representations of qualitative spatial relations need to be computationally tractable and formalised, but they also should be cognitively valid, capturing perceptual distinctions important to people. People's perception of spatial arrangements are often subjectively impacted by their perception of form, of arrangement, and other contextual and possibly dynamic aspects [16]. One could argue that subjective differences may be even more apparent when the complexity of spatial relations increases, i.e., by increasing the number of objects in the arrangement.

Take the spatial relation "between" as an example. Some may consider the object O in Figure 1 to be positioned between objects A and B, while others may disagree, because O is not colinear with A and B but offset above their line of colinearity. Common topological relation models are binary (i.e., consider two objects), and they cannot adequately distinguish the ternary relation *"between"* from other, e.g., *disjoint* cases. Considering the complexity of

[1] Correspondences regarding this article should be sent to Martin Tomko, Department of Infrastructure Engineering, The University of Melbourne, Parkville, VIC 3010, Australia. Email: tomkom@unimelb.edu.au

human perception then, is it then feasible and practical to capture consensus expressions for ternary qualitative spatial relations? In this paper, we answer the research question of how do people distinguish between the relations between objects A, B, and O such as shown in Figure 1?

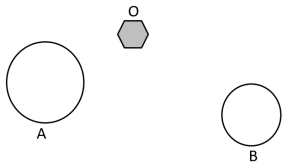


Figure 1 How do people distinguish the relations between the object O and objects A and B?

This paper investigates the human perception of different scenarios of three spatial objects that are distinguished by the recently proposed Ray Intersection Model (RIM) [12]. In an experiment, participants have been presented with image stimuli of spatial scenarios and tasked with grouping stimuli perceived as similar. Groups of stimuli common across participants may be considered as perceptually identical ternary spatial relations. The results of this study indicate that the expressive power of RIM aligns with human perception, and is therefore sufficient to facilitate reasoning about ternary spatial relations in a cognitively ergonomic manner. We further provide a breakdown of results by participant's age group, and provide insights in the verbal descriptions of the identified groupings, in particular with respect to the perception of the arrangements as "between".

The remainder of this paper is structured as follows: Section 2 briefly reviews related work on spatial relation models and ensuing experimental spatial cognition studies. Section 3 demonstrates the interactive survey designed with basic RIM scenarios. Section 4 shows the survey results and their analysis. The discussion of this study, its limitations, and concluding remarks are given in Sections 5 and 6.

2 Related works

2.1 Topological relation models

Qualitative models of spatial relations mostly focus on topological relations between objects which may be defined as relations that are invariant to topological transformations of the reference objects [5]. Most of the existing topological relation models for objects in a 2-dimensional plane are binary and based either on the intersection of point sets, such as the 4 and 9 Intersection Models (4IM and 9IM) [5], or the Region Connection Calculus [17] such as the well-known RCC-8 model [4]. However, binary relations are limited when describing complex relationships that need to consider more objects simultaneously.

Bloch et al. [2] discussed the ternary topological relations and the challenges in modelling them, but did not present a formal model. Clementini and Billen [3] proposed a ternary model for projective relations between regions that divides space around two regions into five parts (i.e., Before, Between, After, Leftside, Rightside). The third object is then described relative to the first two by its intersections with the four space divisions. Their model was named 5-intersection model and it distinguishes 31 projective relations between three regions.

2.2 Ray Intersection Model (RIM)

In this study we test the recently proposed Ray Intersection Model (RIM) introduced by Majic et al. [12] to validate its ergonomic and perceptual ability to distinguish situations differentiated by human subjects. RIM defines rays between two peripheral objects in a 2-dimensional plane (e.g., A and B in Figure 2) as straight lines that share exactly one end point with each of the peripheral objects. There may be infinite number of rays and the area covered by all possible rays is called the *ray area*, and the rays that coincide with the borders of the ray area are called *extreme rays*. The position of the third – core – object (O) in relation to peripheral objects is then represented by its topological relations with the rays, and since there may be many rays that have the same topological relation with the core object, only the distinct relations are considered (e.g., rays r_{e_1}/r_{e_2}, r_2, and r_3 in Figure 2 represent all distinct topological relations any ray between A and B can have with O as any other ray drawn between A and B will have an identical relation to O as one of these rays).

Figure 2 shows peripheral objects A and B, the core object O, and their representation with RIM. The first three matrices (3x3) show distinct 9IM relations between extreme (r_{e1}, r_{e2}) and two distinct non-extreme rays (r_2, r_3) and the core object O. These relations are then combined into a RIM matrix (4x3) that shows whether none (□), some (◪), or all (■) rays' boundaries (∂) and interiors (°) intersect the core object's interior, boundary, and exterior ($^-$). In the RIM matrix, R stands for the ray set which consists of all rays between A and B.

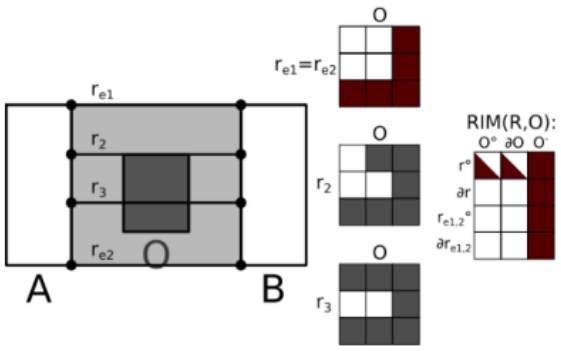


Figure 2 Basic RIM example with peripheral objects A and B, core object O, extreme rays r_{e_1} and r_{e_2}, and non-extreme rays r_2 and r_3 [12].

In [12], 23 basic RIM arrangement of three rectangles have been discussed, followed by examples of more complex scenarios discovered in a case study of betweenness of campus buildings. In theory, the expressiveness of a RIM matrix can capture 2070 different arrangements of three simple planar geometries. In [13], RIM was applied to detect missing data in OpenStreetMap, by identifying buildings with obstacles between buildings and the nearest road.

2.3 Experimental testing of perceptual categorisations

Spatial cognition research has a long history of experimental testing of the perceptual grounding of formal models of qualitative spatial relationships. Our research applies the card sorting technique [18, 7], a grouping exercise developed in human-computer interaction and usability studies where users sort cards (i.e., items) into groups that they perceive

as similar in some way (e.g., design of information architecture of Websites). While all individuals categorise objects or situations somewhat subjectively, the success of our day-to-day communication and interactions affirms a level of shared perception of membership in groups. This is indeed what the card sorting technique enables to identify, based on synthesizing outcomes of individual groupings.

The card sorting technique has been previously adapted and applied in an influential series of research into spatial and linguistic conceptualizations and their interplays by Klippel and co-authors [11, 10, 9, 8]. Beyond spatial conceptualizations and their linguistic reflections, Bianchetti et al. [1] applied this technique to map symbol representations for similar concepts, but across two distinct national cartographic symbol standards for emergency maps.

Here, we equally apply a hierarchical clustering technique to understand how the coarse-grained commonalities in groups manifest across individuals. We also elicit descriptions of the groups by participants, to understand the verbal reflections of these perceptual groupings.

3 Experiment

We present the results of a survey designed to collect answers from a perceptual test of similarity in spatial arrangements. Participants were shown different spatial configurations of two peripheral and one core object. Their task was to group the depictions and provide textual descriptions for the groups they have created, based on their own judgment of similarity of arrangements. Participants' grouping patterns were analysed to gain insights about how people perceive ternary spatial relations, with particular interest if they have recognized and grouped the spatial relation "between".

A key criterion for the grouping task design was to let participants group depictions according to their perception of spatial relations between objects instead of other object properties, such as size and color. Survey instructions were designed to be clear, but excluded any spatial relation terms or hints to avoid influencing participants' answers. Thus, any spatial relation terms used in group descriptions come solely from participants and indicate their perception of spatial relations.

All depictions used in the experiment were generated from the 23 basic RIM cases shown in [12] by varying the rotation and placement of the objects as these parameters change the alignment of objects which could affect participants' perception of spatial relations. The shape and colour of the two peripheral objects remain constant in all depictions because they do not affect the spatial relations but could tempt participants to use them as criteria for grouping. The core object is colored red to stand out. Size was controlled too, however, for some RIM scenarios, it is impossible to keep the core object exactly the same. Figure 3(a), (b) and (c) shows examples of three such cases, where the core objects have different sizes to be able to express the RIM scenarios. The rotation of the depictions is another issue we do not want the participant to concentrate on. We chose random directions of the objects from a set of 8 options in the range of $0 - 360^o$ with 45^o increments.

Figure 4 shows the web interface of the grouping task, design of which was inspired by experiments of Klippel and Li [9]. Participants start the experiment with 77 ungrouped, depictions shown on the left side of the interface in a random order to mitigate the ordering effect [14]. The number on top of the interface indicates how many depictions remain ungrouped. On the right side of the interface, participants can add an unlimited number of groups, or delete any groups they have made. The main task is to sort each depiction from the left into one of the groups on the right by drag and drop. In a second step, participants label each group and describe their reasons for grouping the depictions together. A final

(a) (b) (c)

Figure 3 Examples of RIM scenarios.

submission is only possible when all depictions are sorted and all groups have been labeled and described. Groups, labels, and descriptions can be modified at any time until participants submit their answers.

To ensure that participants are familiar with the interface operations before the task starts, the survey first presents a warm-up trial task where they use the interface to sort random pictures.

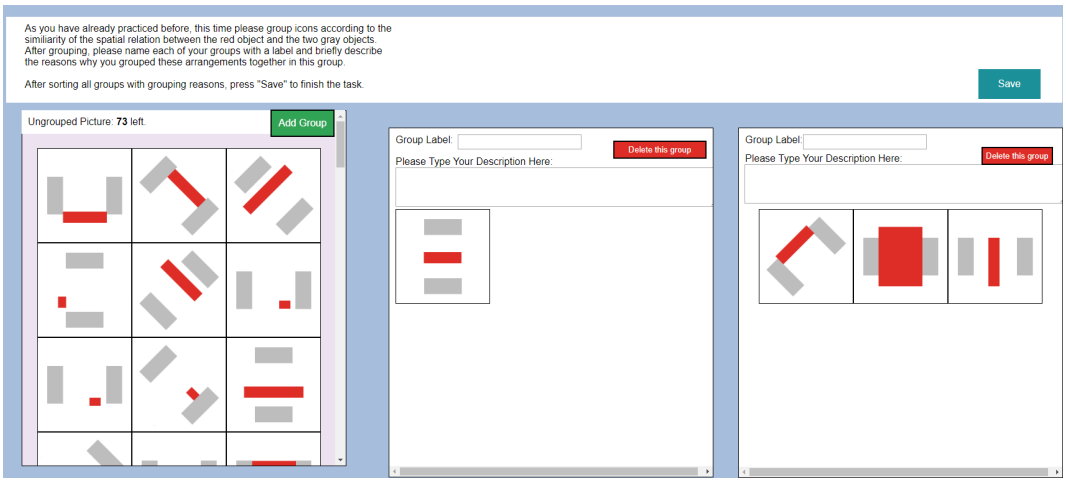


Figure 4 Grouping task interface.

The survey was administered through Amazon Mechanical Turk (Mturk) [2] and controlled the distribution of participants with regards to their primary language and age. We have divided participants into five age groups, with at least 25 participants per group: *18-25, 26-35, 36-45, 46-55, above 55* years old. In addition to their age, participants were asked to state their first language and country of residence which would provide insight into the geographical and linguistic distributions of participants. Figure 5 shows the overall flow of the survey.

[2] `https://www.mturk.com/`

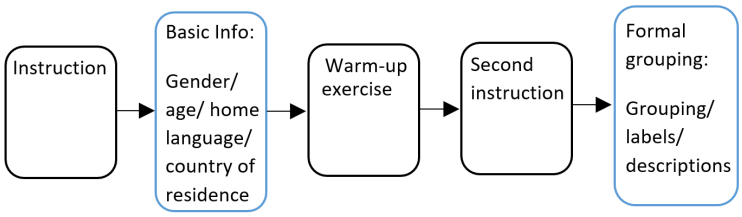


Figure 5 Overall flow of the survey.

Table 1 Summary of grouping by age group.

Participant group	Participants	Number of groups set by each participant					
		Mean	Median	Mode	Min	Max	Std. Dev.
Overall	75 (100.0%)	4.72	4	3	3	13	2.17
Age 18-25	12 (16.00%)	4.58	3.5	3	3	13	2.88
Age 25-35	13 (17.33%)	4.77	4	4	3	13	2.62
Age 35-45	25 (33.33%)	4.56	4	4	3	11	1.80
Age 45-55	10 (13.33%)	4.70	5	5	3	7	1.42
Age above 55	15 (20.00%)	5.07	4	3	3	11	2.34

4 Results

4.1 Participants overview

In a two-day period, the survey received 106 submissions from Mturk workers who were relatively evenly spread across the five age categories we have defined (Table 1). In the data cleaning process, the depiction grouping results were filtered into three classes:

1. Grouping based on spatial relations with the spatial relation terms (e.g. touching, in the middle of, between) mentioned in the group label or description.
2. Grouping based on spatial relations but no exact spatial relation terms mentioned (e.g. table-like, sandwich, pong game).
3. Grouping based solely on the directions of objects (e.g. 12 o'clock, 2 o'clock).

Although the grouping reasoning of the third class belongs to the area of spatial cognition, it is beyond the scope of this study, as the core research question is aimed at ternary spatial relations using RIM as a reference and not directional relations. Therefore, the results belonging to the third class were discarded from further analysis. After data cleaning, 75 answer sets remained. The final distributions of the age groups are shown in Tables 1 and 2.

4.2 Similarity analysis

Similarity matrices were generated to observe patterns of depictions that were frequently grouped together. Firstly, for every group from each participant, a square binary similarity matrix S_i of dimensions $m \times m$ ($m = 77$) contains entries of 1 when two depictions are grouped together (i.e., judged similar) and 0 for depictions that belong into different groups. By aggregating across matrices for all participants $S_1, S_2, S_3, \ldots, S_N$ ($N = 75$), we obtain an overall similarity matrix with element values ranging between 0 and 75. A higher number

Table 2 Descriptive statistics of all similarity matrices.

Participant group	Participants	Similarity parameter summary					
		Mean	Median	Mode	Min	Max	Std. Dev.
Overall	75 (100.0%)	22.01	13	4	1	75	21.14
Age 18-25	12 (16.00%)	3.37	3	1	1	12	3.11
Age 25-35	13 (17.33%)	3.91	3	1	1	13	3.73
Age 35-45	25 (33.33%)	7.91	3	1	1	25	7.92
Age 45-55	10 (13.33%)	2.67	2	1	1	10	2.96
Age above 55	15 (20.00%)	4.16	3	1	1	15	4.19

in the matrix means that the two depictions were more frequently grouped together by participants, while a smaller number means that the two depictions were more likely placed into different groups. A 0 entry means that two depictions were never grouped together.

Partial similarity matrices were created for each age group to compare their grouping behaviours. In table 2, the descriptive statistics are listed for all generated similarity matrices. Additionally, we normalise the data in each participant group to the range $[0, 1]$, so that each element is divided by the total number of participants in the group (as shown in Figure 6).

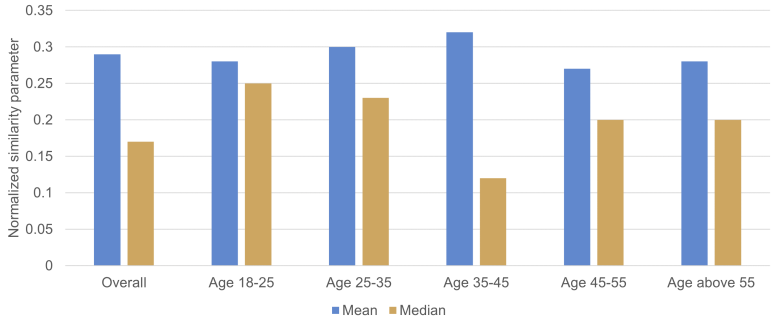


Figure 6 Normalised summary statistics for participant groups.

4.3 Hierarchical clustering

The agglomerative hierarchical cluster analysis is used to further demonstrate the patterns of participants' spatial relation groupings. All depictions are first recognized as single-object clusters, and then they may be merged into bigger clusters based on rules of the distance between each cluster [6]. Here we use Ward's method which performs better than other methods when clustering the non-multivariate data [15], which fits the requirements and the purpose of this experiment.

Therefore, we first calculate the pairwise dissimilarity matrix DS_i using the similarity matrix discussed in the previous section:

$$DS_i = 1 - \frac{S_i}{range(S_i)} \tag{1}$$

The dissimilarity matrix also contains normalised values in range [0,1] where a higher value represents less similarity between depictions. In the Ward's method, the principle of combining two clusters is based on the comparison between the squared deviations (Sq. Dev.) of all possible merges and performing the merge with minimal deviation:

$$Sq.Dev. = \sum (x_i - \bar{x}) \tag{2}$$

The result of the Ward's clustering method for all 75 selected survey answers is shown in Figure 7 as a dendogram. From the highest level of the dendogram, branches are divided in hierarchical structure and the summarized cluster levels of Figure 7 are shown in Figure 8.

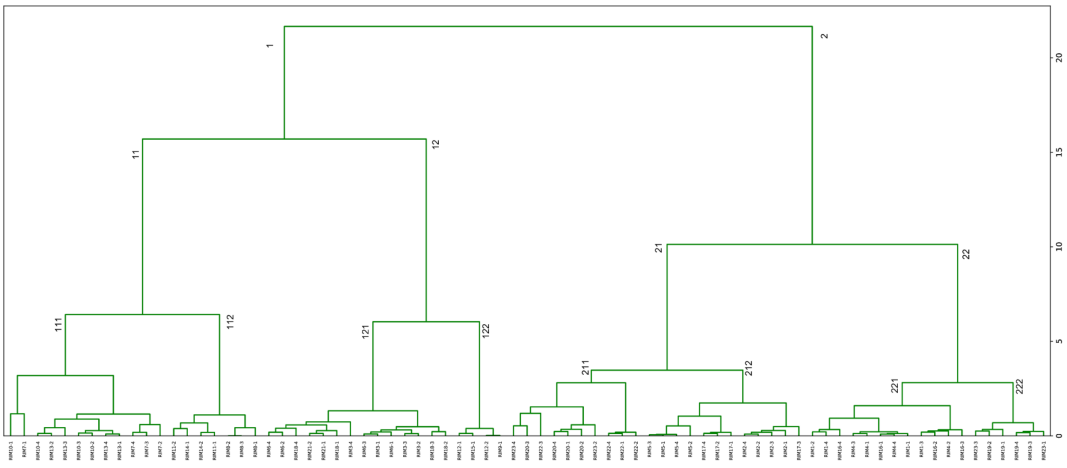


Figure 7 Dendrogram for all 75 selected answers using Ward's method.

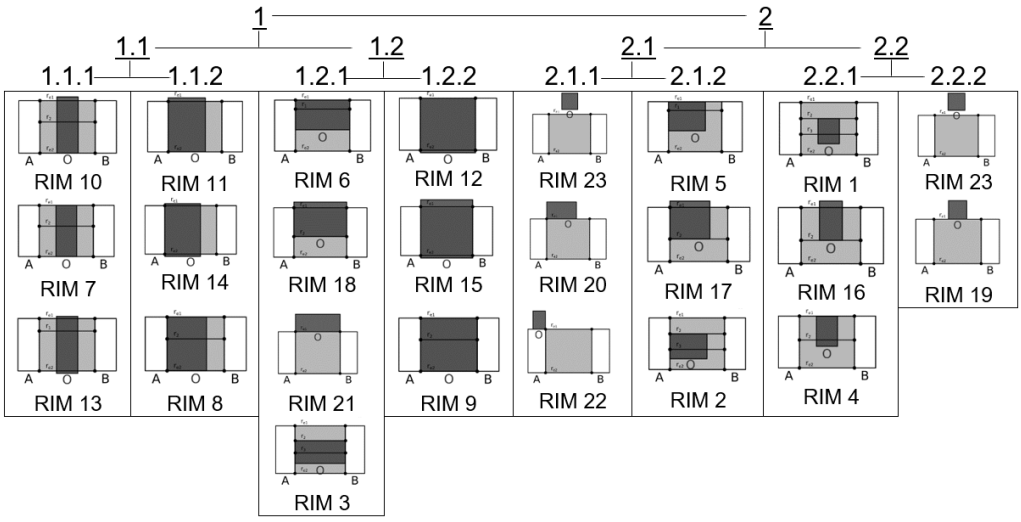


Figure 8 Hierarchical cluster structure of the overall selected answers.

5 Discussion and Limitations

5.1 Survey depiction design evaluation

We aimed to generate four depictions for each RIM scenario (where possible/meaningful) by randomly varying the rotation of objects and translating the core object. To see whether participants perceive the depictions generated from the same RIM scenario as similar, we calculate pairwise similarities between scenarios. Table 3 shows the summary statistics for the pairwise similarities of depictions generated from the same RIM scenarios. Most

Table 3 Summary statistics of the pairwise similarity of depictions within the same RIM scenario.

Mean	Median	Mode	Maximum	Minimum	Std. Dev.
0.86	0.88	0.89	1	0.51	0.09

depictions belonging to the same RIM scenario are grouped together, hence judged similar. The mean, median, and mode similarities are all above 86%. The maximum similarity of 100% is achieved for two depictions generated from the RIM 8 (Figure 8) scenario which may be due to the limited options for the placement of the core object. The lowest similarity of 51% is achieved between two depictions generated from RIM 23. A possible explanation for this is that because the core object is not touching any rays between peripheral objects, there are more possibilities for the placement of the core object than perhaps in other RIM scenarios, as shown in Figure 9.

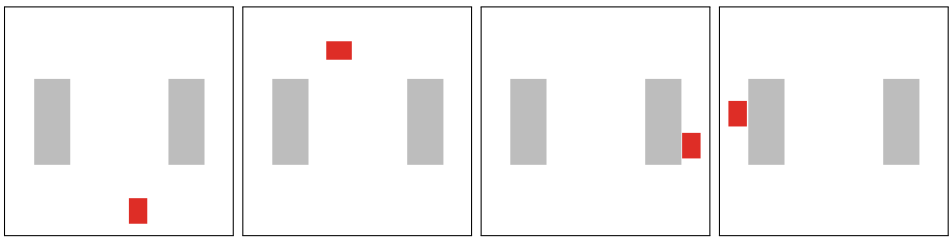


Figure 9 Four depictions randomly generated from the RIM 23 scenario.

5.2 Spatial relation clustering analysis

Figure 8 shows how the 23 RIM scenarios have been clustered in the survey results. Clusters indicate that participants differentiate depictions in a manner that can be explained through analysis of the relationship between the core object and the peripheral objects, and the relationship between the core object and the rays cast between the peripheral objects. Table 4 shows the patterns in these parameters that can be observed in each cluster.

Table 4 Relationship clustering summary (in the third root level).

Cluster	Core object - peripheral objects	Core object - rays
1.1.1	Disjoint	Intersects all rays
1.1.2	Touches one boundary	Intersects all rays
1.2.1	Touches boundaries on both sides	Intersects some rays (or extreme rays)
1.2.2	Touches boundaries on both sides	Covers all rays
2.1.1	Touches one boundary	Disjoint or only touches one extreme ray
2.1.2	Touches one boundary	Intersects some rays (or extreme rays)
2.2.1	Disjoint	Intersects some rays (or extreme rays)
2.2.2	Disjoint	Disjoint or only touches one extreme ray

Table 5 Examples of group descriptions.

Group type	Examples
Explicit spatial relation terms	• This is a small red stripe of varying orientations between two grey identical stripes. • Little red block attached to the inside of grey blocks. • There is a long red block in the middle of two grey blocks.
Metaphors for spatial relations	• In all of these, the red bar is in between the two grey ones forming an almost bench-like shape. • When two grey blocks play pong with a red ball. • Shapes that remind me of a big, red table with some grey seats on the sides.

From Table 4, three distinct perceptual situations capturing the core object's relationship with the extreme rays can be extracted:

1. The core object touches/intersects all extreme rays
2. The core object touches/intersects only one extreme ray and some other rays
3. The core object touches/intersects one extreme ray but is disjoint from other rays

The results indicate that there is no difference between cases 2 and 3 in participants' perceptions of ternary spatial relations. When the core object intersects both extreme rays (case 1), then this happens in all depictions in the cluster (e.g., clusters 1.1.1, 1.1.2, and 1.2.2). But when this is not the case (cases 2 and 3), then the intersection of the extreme ray is not the determining factor as it only happens in some depictions in the cluster (e.g., clusters 1.2.1, 2.1.2, and 2.2.1). In these cases the relationship between the core object and the peripheral objects is the determining factor and consistent in all depictions within a cluster. For example, in cluster 1.2.1 the core object touches both peripheral objects, in cluster 2.1.2 one, and in cluster 2.2.1 none.

RIM 23 is the only scenario that is grouped into two different clusters (cluster 2.1.1 and cluster 2.2.2). As discussed in Section 5.1, this may be due to the large variation in placement options for the core object in situations describable as RIM 23.

5.3 Overview of group descriptions

As presented in Section 4.1, participants' descriptions of spatial relations can be divided into those that explicitly mention spatial relation terms, and those that use metaphors to express spatial relations. Typical examples are shown in Table 5. Spatial relation terms like "between", "outside", and "in the middle of" are used to describe the RIM scenarios based on whether the core object is perceived to be "between" peripheral objects or not. Metaphoric expressions like "table", "pong" and "sandwich" indicate not only the perception of spatial relations, but also other aspects of spatial cognition such as the size and shape of the core object compared to the peripheral objects. For example, participants tend to describe a relation as "sandwich" if all three objects in the depiction are parallel, mostly aligned, and have equal length.

5.4 Spatial relation "between"

In the 75 accepted survey answer sets, 35 mention the word "between" as a spatial relation in at least one of the created groups. One example of such survey answer is shown in Figure 10.

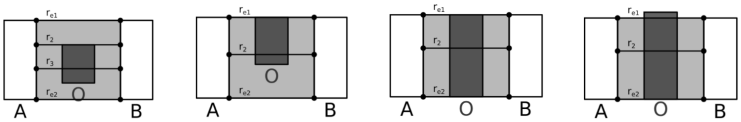
Participant ID: *removed*; **Gender:** Female; **Age:** 35-45
Home Language: English; **Country of residence:** United States of America
Group Label 1: Between touching both
Group Description 1: The red bar is between the grey bars and they are all touching.
Group Label 2: Between touching one
Group Description 2: The red bar is between the grey bars and is only touching one.
Group Label 3: Between no touching
Group Description 3: The red bar is between the grey bars but none of the bars are touching.
Group Label 4: Outside no touching
Group Description 4: The red bar is outside of the grey bars and none of the bars are touching.
Group Label 5: Outside touching
Group Description 5: The red bar is outside the grey bars but is touching one of them.

Figure 10 One participant's group labels and descriptions using the term *between*.

Different people group in different levels of detail. In the example above, the participant clearly separates the relation "between" into three classes depending on whether the core object touches the boundaries of the two peripheral objects. Other participants sometimes create the same groups that would correspond to "between", but use different parameters in their descriptions such as the size of the core object, the direction of the depiction, and whether the core object lies completely within the ray area.

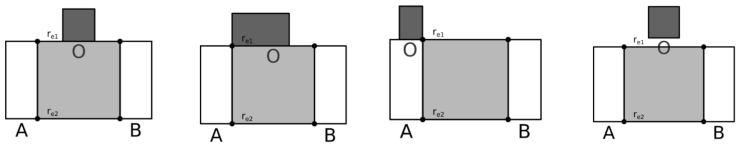
The first 4 RIM scenarios that are most frequently recognized as a relation of "between" are displayed in Figure 11, while the 4 RIM cases that are least likely to be identified as "between" are shown in Figure 12. All core objects in Figure 11 have some intersections with the rays cast between peripheral objects. Furthermore, the results indicate that these core objects should intersect some non-extreme rays. This ensures that the intersection happens mostly inside the ray area, regardless whether the core object extends outside (i.e., RIM 10) or not.

On the other hand, all core objects in Figure 12 are disjoint from the non-extreme rays. In these cases it does not matter whether they intersect one of the extreme rays as in RIM 19, RIM 20, and RIM 22, or if they are disjoint from all rays as in RIM 23. All of these core object will mostly be perceived as "not between" the peripheral objects by participants.



(a) RIM1. **(b)** RIM4. **(c)** RIM7. **(d)** RIM10.

Figure 11 RIM scenarios that can be grouped as a spatial relation "between".



(a) RIM19. **(b)** RIM20. **(c)** RIM22. **(d)** RIM23.

Figure 12 RIM scenarios that cannot be grouped as a relation "between".

5.5 Limitations of the experiment

There are several limitations regarding the human perception of ternary spatial relations that are not fully controlled in this study. Firstly, the country of origin of the majority of our participants may bias the results. 73% of the participants were from the US and 86% of the participants declared English as their first language. These biases may affect the diversity of samples collected from the survey, as language and cultural differences may affect the perception of spatial relations.

The second limitation is that only the spatial configurations that are generated from the 23 basic RIM scenarios were tested in this experiment. This covers only scenarios limited to spatial configurations of three rectangles, while perceptions of arrangements of other (circular, linear, point-like, 3D objects, complex polygons such as campus buildings in [12]) could be different.

Thirdly, in some scenes properties of spatial arrangements other than spatial relations have influenced participants answers, e.g, shape, color, size, and orientation of objects, as is visible from the depiction clusters and group descriptions. Although each cluster in Figure 8 can be justified with spatial relations, the core objects in each cluster are very similar in size. In this case, further validation may be needed to investigate whether the size of the object is tied to human perception of spatial relations.

6 Conclusion

This study has assessed the human perception of ternary spatial relations by conducting a survey with 75 participants. Their task was to sort 77 depictions showing different configuration of three spatial objects (i.e., two peripheral and one core object) into groups based on their perception of the spatial relations between these objects. These depictions were generated from 23 basic RIM scenarios by randomly translating and rotating the objects. The participants were also asked to describe each group with a label and a description. Survey answers were then analyzed with hierarchical clustering and similarity analysis to quantify participants' qualitative reasoning.

Two types of group descriptions were found in survey answers: those that explicitly use spatial relation terms and those that use metaphors to express spatial relations. We found that the participants who used metaphors were more likely to be influenced by size, shape, and rotation of objects, contrary to the survey instructions. There are no other noticeable differences between participants' grouping patterns. This is true for different age groups as well, where among five different age groups analyzed hardly any differences are found.

We note a number of patterns in the collected data. Firstly, participants perceive depictions generated from the same RIM scenario similarly, as reflected in the groupings (Table 3). This indicates that each distinct scenario captured by the RIM model [12] indeed captures ternary spatial relations that humans uniquely recognize and distinguish qualitatively,

even when challenged with changes in objects' rotation, position, and size. Secondly, participants group different RIM scenarios highly similarly and consistently (Figures 7 and 8). The RIM scenarios within a cluster are conceptually closer to each other and can be hierarchically aggregated if a cognitively aligned generalization based on spatial relations is needed. The hierarchical cluster dendrogram shows how conceptually more similar scenarios group together, revealing this cognitive hierarchy in the arrangements.

RIM distinguishes scenarios that align with linguistic descriptions. The four RIM scenarios that were most frequently described with the term "between" are RIM 1, RIM 4, RIM 7, and RIM 10 (Figure 11). The core object in all of these scenarios intersects non-extreme rays, and interestingly, does not touch any of the peripheral objects. It also does not seem to matter whether the core object extends outside of the ray area or not. The four RIM scenarios that were the least associated with the term "between" all have the core object that is outside of the ray area (i.e., has no intersection with non-extreme rays) and may intersect extreme rays (Figure 12). This shows that RIM can differentiate ternary relations that people perceive as "between" or "not between".

The main limitation of this study and the motivation for the future work is the lack of cultural and language diversity in participants. This is something that can be controlled in future experiments to also investigate if there are differences in the human perception of spatial relations based on culture and language. Another possible improvement is to better communicate the goal of the study to minimize the influence of aspects other than spatial relations such as size, rotation, and color of objects on their answers. Lastly, this experiment was based on the RIM model which is currently limited to polygons [12]. It would be interesting to test human perception with objects of a different type such as points and lines, or dimensionality such as 3D objects.

References

1 Raechel A Bianchetti, Jan Oliver Wallgrün, Jinlong Yang, Justine Blanford, Anthony C Robinson, and Alexander Klippel. Free classification of canadian and american emergency management map symbol standards. *The Cartographic Journal*, 49(4):350–360, 2012.

2 Isabelle Bloch, Olivier Colliot, and Roberto M Cesar. On the ternary spatial relation" between". *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 36(2):312–327, 2006.

3 Eliseo Clementini and Roland Billen. Modeling and computing ternary projective relations between regions. *IEEE Transactions on Knowledge and Data Engineering*, 18(6):799–814, 2006.

4 Anthony G Cohn, Brandon Bennett, John Gooday, and Nicholas Mark Gotts. Qualitative spatial representation and reasoning with the region connection calculus. *GeoInformatica*, 1(3):275–316, 1997.

5 Max J Egenhofer and Robert D Franzosa. Point-set topological spatial relations. *International Journal of Geographical Information System*, 5(2):161–174, 1991.

6 Laura Ferreira and David B Hitchcock. A comparison of hierarchical methods for clustering functional data. *Communications in Statistics-Simulation and Computation*, 38(9):1925–1949, 2009.

7 Sally Fincher and Josh Tenenberg. Making sense of card sorting data. *Expert Systems*, 22(3):89–93, 2005.

8 Alexander Klippel. Topologically characterized movement patterns: A cognitive assessment. *Spatial Cognition & Computation*, 9(4):233–261, 2009.

9 Alexander Klippel and Rui Li. The endpoint hypothesis: A topological-cognitive assessment of geographic scale movement patterns. In *International Conference on Spatial Information Theory*, pages 177–194. Springer, 2009.

10 Alexander Klippel and Daniel R Montello. Linguistic and nonlinguistic turn direction concepts. In *International Conference on Spatial Information Theory*, pages 354–372. Springer, 2007.

11 Alexander Klippel, Michael Worboys, and Matt Duckham. Identifying factors of geographic event conceptualisation. *International Journal of Geographical Information Science*, 22(2):183–204, 2008.

12 Ivan Majic, Elham Naghizade, Stephan Winter, and Martin Tomko. Rim: a ray intersection model for the analysis of the between relationship of spatial objects in a 2D plane. *International Journal of Geographical Information Science*, 35:893–918, May 2021. doi:10.1080/13658816.2020.1778002.

13 Ivan Majic, Elham Naghizade, Stephan Winter, and Martin Tomko. There is no way! Ternary qualitative spatial reasoning for error detection in map data. *Transactions in GIS*, 25(4):2048–2073, 2021. doi:10.1111/tgis.12765.

14 Paul Martin, Paul Patrick Gordon Bateson, and Patrick Bateson. *Measuring behaviour: an introductory guide*. Cambridge University Press, 1993.

15 Glenn W Milligan and Martha C Cooper. Methodology review: Clustering methods. *Applied psychological measurement*, 11(4):329–354, 1987.

16 David R Olson and Ellen Bialystok. *Spatial cognition: The structure and development of mental representations of spatial relations*. Psychology Press, 2014.

17 D A Randell, Z Cui, and A Cohn. A spatial logic based on regions and connection. In B Nebel, C Rich, and W Swartout, editors, *KR'92. Principles of Knowledge Representation and Reasoning: Proceedings of the Third International Conference*, pages 165–176, San Mateo, California, 1992. Morgan Kaufmann. URL: http://citeseer.ist.psu.edu/randell92spatial.html.

18 Jed R Wood and Larry E Wood. Card sorting: current practices and beyond. *Journal of Usability Studies*, 4(1):1–6, 2008.

Are Psychological Variables Relevant to Evaluating Geoinformatics Applications? The Case of Landmarks (Vision Paper)

Jakub Krukar[1]
Institute for Geoinformatics, Universität Münster, Germany

Angela Schwering
Institute for Geoinformatics, Universität Münster, Germany

Abstract

Interdisciplinary integration of spatial cognition and spatial computation promises to create better spatial technology based on findings from cognitive psychology experiments. Using the example of psychological studies and computational modelling of landmarks, this paper argues that core evaluation criteria of both disciplines are not well aligned with the goal of evaluating landmark-enhanced navigation support systems that support users in everyday wayfinding.

The paper raises two points. First, it reviews evaluation criteria used in the interdisciplinary field of landmark research. It is argued that when to consider the role of landmark-enhanced navigation support systems in everyday life of their users, different evaluation criteria are needed. If strictly-psychological or strictly-computational criteria continue being prioritised by the community, we risk undervaluing significant technological contributions. Second, it proposes one such potential criterion: testing whether the cognitive task has changed due to equipping users with the new technology.

This goal might be achieved at the expense of criteria typical to strictly-psychological studies (such as spatial memory of landmarks along the travelled route) or strictly-computational studies (such as efficiency and accuracy of a landmark-selection algorithm). Thus, promoting and implementing alternative evaluation criteria comes with methodological risks. In order to mitigate them we propose a process based on pre-registration of "postdiction" studies and hope to stimulate a further debate on a consensus-based approach in the community.

2012 ACM Subject Classification Human-centered computing → Empirical studies in ubiquitous and mobile computing; Applied computing → Psychology

Keywords and phrases wayfinding, navigation support systems, cognitive geoengineering, landmarks

Digital Object Identifier 10.4230/LIPIcs.COSIT.2022.10

Category Vision Paper

Supplementary Material *Dataset (Literature Review)*: `https://osf.io/ru8yb/`

Funding This work has been supported by the German Research Foundation grant SCHW1372/7-3.

Acknowledgements We wish to thank Eva Nuhn and Sabine Timpf for organising the workshop on "Modelling of landmark dimensions" held at the 8th International Conference on Spatial Cognition 2021, as well as all its participant for an inspiring discussion that lead us to writing this paper.

1 Introduction

Research on landmarks can be classified into three main streams:

(a) *psychological research* aimed at understanding how the mind works (e.g., understanding the role of landmarks in spatial cognition);

[1] Corresponding author

(b) *computational research* aimed at solving the problem of landmark identification based on human-specific criteria (e.g., modelling how the human mind would identify salient landmarks); and

(c) *navigation support research* that aims to support the user in everyday wayfinding by enhancing existing technology with landmarks.

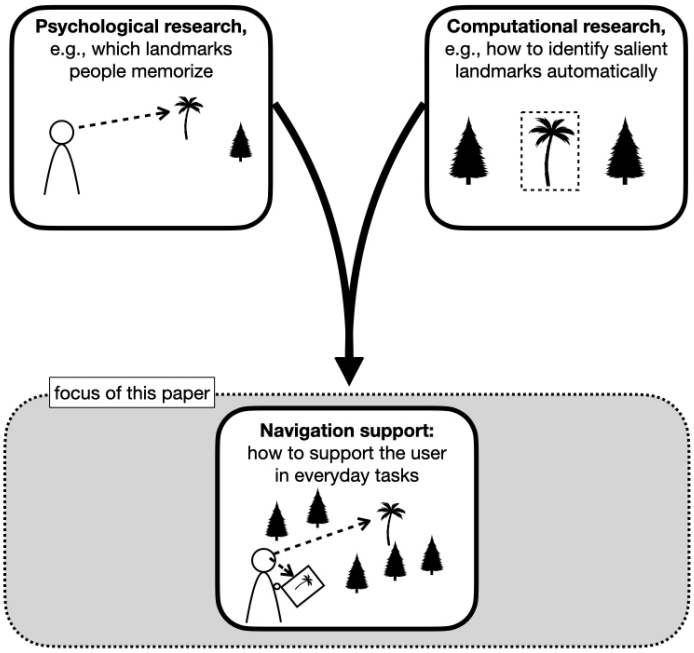


Figure 1 The practical goal of interdisciplinary collaboration in landmark research is the focus of this paper.

The practical aim of the community's interdisciplinary effort is to create (c) (Figure 1). Yet, this effort seems to have reached an impasse. Despite fundamental landmark studies spanning back at least six decades [13] it remains impossible to predict which specific landmark should be presented or highlighted for each individual user in each possible context of use; and it is uncertain whether it should be. The field's momentum seems to be directed at modelling more factors contributing to the context of use, identifying more dimensions contributing to personalised preferences, and extracting more types of potentially relevant digital data. Yet, there has been criticism of the entire concept of landmarks underlying this line of research. Montello [15] argued that landmarks are exaggerated in both basic and applied spatial cognition research, reviewing two issues in particular: that the term "landmark" is a label for several concepts; and that – even when communicated precisely – the concept of landmarks has an overestimated role in human navigation, compared to other components of spatial knowledge, memory, and reasoning.

The current paper focuses on another possibility underlying the impasse: that landmark-enhanced navigation support systems are evaluated based on criteria relevant to basic (be it psychological or computational) research, but with no clear agreement on which criteria should be employed for evaluating their applications. This paper reviews two pathways for moving beyond the impasse:

(1) re-considering the role of landmark-enhanced navigation support systems in everyday life of their users; and
(2) standardising the evaluation of a phenomenon difficult to capture with traditional measures – whether the technology changed the cognitive task performed by its user.

In order to do so, the paper presents a re-analysis of a recently published review of papers on landmark research from the domains of spatial cognition and spatial computation. As demonstrated, these domains use a diverse set of variables when studying landmarks but it is questionable how useful these variables are within the domain of navigation support research. We review what role such technology is intended to play in users' everyday life and argue that in the most popular scenarios of use the dominant variables are not critical for evaluating navigation support systems. However, increasing the use of alternative – not yet crystallised – evaluation criteria is associated with methodological risks. One proposed way to mitigate them is an approach used in psychological science for pre-registering "postdiction" analyses. The paper aims to stimulate a debate on a corresponding consensus-based approach in our community.

2 Variables and evaluation criteria within the landmark literature

In a recently published literature review Yesiltepe et al. [27] aggregated 59 papers on the visibility of landmarks, selection of landmarks, location of landmarks, personal and emotional landmarks, gaze behaviour with relation to landmarks, and landmark saliency. We extend this literature review by classifying these papers into two main categories: (a) psychological research aimed at understanding how the mind works; and (b) computational research aimed at solving the problem of landmark identification based on human-specific criteria. This classification has been performed manually by the authors, based on the content of each paper. It is available in full at: `https://osf.io/ru8yb/`.

Next, we checked which dependent variables were reported in those papers' "Results" or "Evaluation" sections (when available) and classified those into higher-level groups of *evaluation criteria*. Table 1 presents evaluation criteria across the psychological and computational paper types.

Table 1 Evaluation criteria used across two types of papers treating on the problem of landmarks. Numbers in the brackets indicate how many papers of the given category applied the given criterion.

	Psychological papers	Computational papers
number of papers	39	18
evaluation criteria	spatial memory performance (21) navigational performance (12) properties of route descriptions (8) subjective salience judgement (5) gaze behaviour (distribution) (4) quality of route descriptions (4) self-localization performance (4) gaze behaviour (fixations) (2) variance of landmark placement (2) head/body alignment (1) navigational strategy (1) stress (1)	subjective salience judgement (5) head/body alignment (1) navigational strategy (1) properties of route descriptions (1)

We identified 39 psychological papers, and 18 computational papers. Four papers remained unclassified: one literature review paper, one containing too little methodological detail, and two that focused on developing and evaluating a new navigation system but not on contributing to the understanding of psychological processes.

2.1 Psychological research

Psychological papers used 12 different types of evaluation criteria. The most popular one was spatial memory performance – this includes tests of spatial memory (survey, route, or landmark memory) where the reported variable is the accuracy or speed of response in a test. Spatial memory performance was studied in more than half of the reviewed papers. The second most-popular criterion was navigational performance – this includes measures of errors in, or optimality of, physical or virtual-reality-based wayfinding tasks. Navigational performance measurement was applied by almost 1/3 of all reviewed papers. The third most-popular evaluation criterion were properties of route descriptions – these are mainly analyses of sketchmaps produced in response to a task such as "describe the route to a stranger" (or similar), but *not* measuring the accuracy or quality of these descriptions. For instance, Peters [18] counted how many participants did vs did not include landmarks in their sketch map.

This analysis demonstrates that psychological papers seem to have converged on a set of measures: spatial memory performance and navigational performance are of interest to the highest number of papers.

2.2 Computational research

Computational papers used standardised evaluation criteria less often – out of 18 papers in this category we noted only 6 papers that applied any at all (one paper applied 3 of them). Of those 6 papers, 5 papers used participants' declared subjective judgement of landmark's salience (or landmark's "relevance") as their evaluation criterion. In the majority of cases, however, the most important evaluation criterion of the newly proposed computational solution is simply the fact that it works. Many papers demonstrated this by applying their method in a use case involving a real-world dataset.

This analysis demonstrates that research has investigated the landmark problem from diverse perspectives and with different motivations. Computational papers have diverse aims (e.g., mathematical–, engineering–, or data processing–oriented) and an empirical evaluation is not their main contribution. They maintain a connection to psychological literature but remain focused on own goals. Individual motivations of computational work might be useful, for example, to help develop methods for obtaining or processing previously underutilised datasets.

2.3 Navigation support research

However, there is another group of studies not well captured by this particular literature review: navigation support systems, also referred to as cognitive geoengineering applications [19]. These studies focus on supporting users in everyday task of navigation with the use of landmark-enhanced technology. Often, navigation support system research makes use solely of the same variables as employed by the strictly-psychological papers. For example, Wunderlich et al. [26] evaluated two different types of landmark-enhanced navigational instructions in comparison to standard no-landmark instructions. They measured spatial memory, navigational performance, and subjective mental load.

In some cases, navigation support systems research has considered different evaluation criteria as an add-on to spatial memory performance and navigational performance variables. For instance, Schwering et al. [20] designed a wayfinding-support system enhanced with on-screen and off-screen landmarks, with the aim of supporting user's orientation. In order to evaluate the approach they used measures of spatial memory performance. In addition, they discussed spatial memory variance – in particular the fact that that their wayfinding system seemed to have increased the consistency of pointing errors, regardless of their accuracy. Another example of such an approach is a paper by Smith et al. [21]. The authors implemented and tested navigation systems with off-screen landmarks inspired by those found in video games, and compared it to a standard Google Maps application in an in-situ experiment. In primary analyses they evaluated spatial memory performance through the accuracy and speed of pointing, as well as a map mark-up task. In addition, researchers analysed video recordings taken when participants performed their tasks in order to identify behavioural patterns such as: focusing on the environment vs on the phone, reorienting the device, frequency of glancing between the device and the environment, and signs of confusion. Yet another type of add-on variables used in evaluating navigation support research are measures of human-computer interaction. For instance, Li et al. [12] recorded the number of pans and zooms performed on the device during an in-situ navigational task.

This analysis demonstrates that spatial memory performance and navigational performance measures are prioritised and other measures are reported as secondary. A paper evaluating a new approach to landmark-enhanced navigation support is expected to report them, and ideally to show improvement in comparison to standard technological solutions. As our personal experience shows, when the dominant measures are not recorded in the experiment, or are reported with a downplayed importance, the paper raises concerns in the peer-review process. However, the goal of a new navigation support system technology may indeed be different from improving spatial memory or navigational performance of its user. Simply not decreasing spatial memory and navigational performance might be entirely satisfactory if the technology demonstrates other benefits or novelty.

In the publication process, however, defending such a null hypothesis is problematic. First, it requires deciding on one of few available (but conceptually different) Bayesian formulations of null-hypothesis testing [11]. Second, not highlighting measures accepted as standard in the field bears the tell-tale signs of potential "HARKing" (Hypothesizing After Results are Known) and "fishing expeditions" (performing many tests on variables of secondary importance until something significant is found) [1]. A paper that reports no significant improvement in spatial memory and navigational performance, but focuses on less popular variables (often measured with custom-built tasks), justifiably rises red flags of questionable research practices. One interpretation seems very likely: researchers had attempted to improve spatial memory or navigational performance with the proposed technology, failed to do so, and abused their secondary analyses long enough to finally find something significant that they now present as an achievement.

It is the position of this paper that the motivation standing behind developing navigation support systems is different from the categories described in the two previous sections, and that this type of research need not directly inherit evaluation criteria central to psychology and spatial computation. The open problems in the evaluation of navigation support systems are what the new criteria should be and how to give them due diligence in the process of investigation and peer review.

3 How to evaluate navigation support systems

We propose a vision of shifting the focus of navigation support research away from spatial memory and navigational performance; and instead moving towards standardising and promoting an alternative evaluation criterion. As such a criterion, we propose investigating whether the *cognitive task at hand has changed due to equipping the user with the new technology*. Below, we first justify why spatial memory and navigational performance should not be prioritised by new technology. Second, we justify why a change in the cognitive task is a relevant criterion for the scientific goal of developing novel technology. Third, we propose the means towards standardising and promoting it within the field inspired by the still scarcely used approach of pre-registering "postdiction" studies in the field of psychological science.

3.1 Counterarguments to prioritising spatial memory and navigational performance

Spatial memory performance and other criteria from strictly-psychological research are the de-facto seal of approval for navigation support systems. One line of argument behind this is a concern that spatial technologies lead to spatial infantilization [23, 14], i.e., to reducing our capabilities of performing given tasks without the technology. Thus, creating technology that increases (as opposed to diminishing) spatial knowledge seems to be a logical counterstrategy. Not only is it a direct evidence that a user learnt something new thanks to the technology; but it can also be an indirect evidence for intensified engagement with the environment, or deeper cognitive processing of spatial information.

A counterargument to this is that humans have always created tools to amplify the limitations of their own mind. These have resulted in some cognitive infantilization but also in societal benefits [5]. Ford et al. [5] describe the example of a calculator: we could argue that its introduction removed the need for learning how to count. But at the same time it allowed a vast group of untrained users to engage with activities requiring reliable counting; while those who want, need, or like to count mentally, can still do so. Evaluating the usefulness of a newly proposed navigation support systems with spatial memory or navigational performance is similar to evaluating the usefulness of a calculator by studying whether the user memorised the outcome, learnt how to perform similar calculations on their own, or how often they obtained the correct solution. The first two criteria are irrelevant – the presence of the calculator makes them obsolete. This is the reason to buy one. The latter is a given – we expect any calculator to provide the correct response at all times. Potential problems might arise due to human-computer interaction (e.g. errors in input) – those would be relevant and can be fixed but pertain to the interface design, not to the role or usefulness of this technology to the society. If the goal of the calculator is to empower its users, and not to teach them (we create other tools for that), then variables related to learning are of secondary importance.

Another popular line of argument for evaluating spatial technologies with strictly-psychological variables is the cognitive geoengineering argument that the spatial concepts used by the systems must *match* spatial concepts used by its user [19]. As Table 1 demonstrates, finding a *match* between the algorithm and human preference is the most popular evaluation criterion for strictly-technological research. This is certainly valuable, as it demonstrates a success of interdisciplinary spatial cognition community: It is routine for computer scientists to understand, target, and evaluate human-centred issues.

But it is worth considering how close a *match* we aspire to? Let us differentiate between two levels of such a *match* in the domain of automated landmark identification: a *soft* and a *hard match*. We would consider a *soft match* all instances when the human understands the system referring to some object as a "landmark" – in a similar way we understand when a stranger (with different personal background, spatial knowledge, and preferences) refers to something that they consider a "landmark". We would consider a *hard match* such instances of automatically identified landmarks that would be identical to what the user themselves would identify (in the given context, emotional state, and within the given task).

A soft match is certainty sufficient for sustaining everyday communication. Simply referring to some perceivable object in navigational instructions makes it a landmark [15], even if the object does not exist, like the famous "etak" islands [8]. Thus, the goal of landmark-enhanced navigation support need not be to mimic the human mind by detecting the same landmarks as the user would. In fact, given well-defined criteria and a large database, a computer algorithm will outperform humans in consistently detecting landmarks or classifying their salience (within the definition formalised by the researcher). When no landmark is available in the immediate environment, the system has the means to compensate for this, e.g., by visualising information that is not directly perceivable in the environment, or by imposing schematised order and salience onto a chaotic or uncharacteristic surrounding [6]. This potential of landmark-detection algorithms seems to be underutilised in designing landmark-enhanced navigation support systems. Instead, researchers focused on the challenge of "biasing" those consistent results in order to match human biases, for instance on the base of emotional judgements in landmark preference.

3.2 Evaluating the human-computer cognitive system

It is the position of this paper that the goal of navigation support systems research is not to bridge the gap between solutions of the landmark computation performed by the algorithms and the ways in which humans do it. This may be useful for other research goals: a system achieving a "hard" match between digital data and human cognition would be a theoretical achievement and may have some practical use. But within navigation support systems, constructing technology mimicking how the mind works might be as difficult, as it is unnecessary. Instead, we propose to evaluate the cognitive system consisting of the technology and its user, in line with the distributed cognition framework of cognitive engineering [4]. Within this theoretical approach, humans and the tools they use are viewed as a separate kind of a cognitive system: one that has different properties than any of its parts. Such a newly established cognitive system should not be evaluated based on criteria specific to its sub-parts. A human equipped with a navigation support system does not need to identify or remember landmarks because the system can do it for them. So what should we evaluate instead?

A first step could be deciding on (and declaring) the intended role of the landmark-enhanced technology in human everyday life. Hermann [7] presented a framework of cognitive technology, classifying technological applications based on their role in supporting or substituting human cognitive processes. Eight types of technologies are listed, for instance a "cognitive prosthesis". The role of such technology is to externalise a cognitive function, similarly to how a reminder in a smartphone externalises the need for using memory [7]. This category in Hermann's framework was further subdivided by others [2] to highlight an important difference between a "cognitive prosthesis" and a "cognitive amplifier".

The difference is that the role of a prosthesis is to bridge the gap between some characteristic or ability of an individual and an accepted average, or a standard previously achievable by experts. For instance, a pair glasses "brings" the eye-sight back to an agreed standard [2].

This could be one role of landmark-based systems: a subset of the population that cannot easily navigate with other existing tools could make good use of a "cognitive prosthesis". One such group might be visually impaired [17].

Compared to a cognitive prosthesis, a cognitive amplifier acts by allowing people to complete more advanced tasks, regardless of the agreed standard. Importantly, this happens not because tools increase computational powers of the human mind, but rather because they change the cognitive task that is being performed [2, 4]. One strategy for changing the nature of a cognitive task is to delegate what used to be an expensive mental calculation onto a simpler perceptual ("pre-cognitive") system [2]. The success of the turn-by-turn navigation systems stems from the fact that they substituted complicated cognitive operations (e.g., planning and remembering actions based on an allocentric perspective of a map) with simple perceptual ones: "take action when you see or hear a notification". This is not detectable by measuring variables specific to the user-part or the computer-part of the joint human-computer system.

In fact, human cognitive performance in spatial learning could decrease because it is being delegated to the technology. Navigational performance may lose relevance, because the human-computer system can achieve satisfactory performance in a wast majority of cases. These measures are not indicative of the entire human-computer systems because the distribution of cognitive tasks between the human and the computer changes in various human-computer systems. This phenomenon is schematically depicted in Figure 2. If to focus on performance measures of the entire human-computer system, a human with no computer (Fig. 2a) has potentially the lowest navigational efficiency but the highest share of cognitive tasks. A human with a map (Fig. 2b) can rely on the technology instead of own spatial memory, thus the navigational efficiency of the total system is higher, but the share of the human in the cognitive work is lower within the system. What is not captured clearly by any standard measure is that the *cognitive task* has changed: the user can perform new cognitive tasks (e.g., plan a route on a symbolic representation of the map) and still reach the destination. A human with a turn-by-turn wayfinding system has a higher navigational efficiency, but the share of the human-part in the computational work of the system is even lower compared to a map. The cognitive task of the human has changed because one does not need to plan turning actions (Fig. 2c). A system detecting and communicating personalised, context-specific landmarks instantly understandable by the user (Fig. 2d) is likely to reduce the human's demand for cognition because one will be able to perform the task with even less effort spent on identifying turning points. Yet, it is yet unclear whether this will change the *cognitive task* performed by the user. The potential usefulness and acceptance of this technology is also unclear. Future technological applications (Fig. 2e) are unlikely to improve the navigational efficiency any further. The involvement of the human in the cognitive work of the system might change depending on the purpose of the technology, but should not be the key variable based on which we evaluate its potential, novelty, or success.

What is needed is a way of evaluating the success (or novelty, or potential) of newly proposed navigation support systems that would depart from strictly-psychological or strictly-technological variables such as those documented in Table 1. We suggest two criteria: (1) considering the role of the technology in the society (e.g., within Hermann's framework) with one potential aim being to adapt some technological solution to a new societal role; and (2) documenting a change in the type of the cognitive task required from the human-part of the human-computer system; this would make it possible to draw qualitative distinctions between two systems where the human involvement in computation is similar or decreases.

With regard to (1), consider the societal role that a landmark-enhanced navigation support system might have inside autonomous vehicles. Although not required for navigation, it could still serve to provide a feeling of comfort or security during a trip through an unknown area, in which case it would not be a "cognitive amplifier", but a "cognitive self-care facilitator" [7]. Another example are "cognitive trainers" that are purposefully designed to train cognitive skills. It would be relevant to evaluate the performance of such system by measuring gains in spatial memory. Note, however, that the goal of most navigation support systems for everyday use is not to be "cognitive trainers".

With regard to (2), this is where taxonomies and task classifications might be particularly helpful; such as the taxonomy of wayfinding tasks proposed by [25] and extended to technology-supported wayfinding by [20]. One explicit goal of novel technology could be to change the distribution of cognitive tasks between the user and the system, i.e., to allocate the user and the newly proposed system in a different taxon. However, it is unclear how to assess whether two systems are enabling, or encouraging the user to perform different cognitive tasks. In the next section we propose a step towards building up an approach for this kind of evaluation.

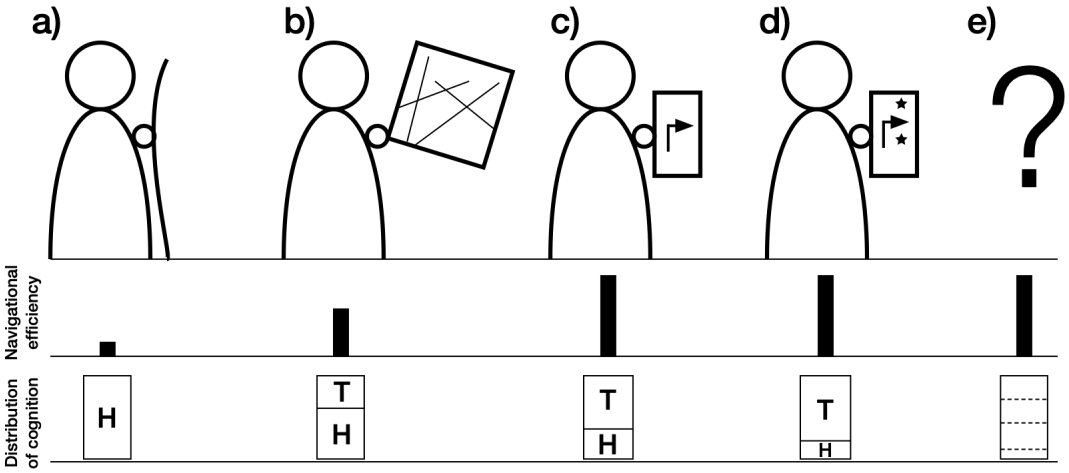


Figure 2 A schematic depiction of various human-computer cognitive systems in wayfinding with relation to standard evaluation criteria and the distribution of cognitive tasks between the human (H) and the technology (T). (a) A human with a walking stick and no computer. (b) A human with a map. (c) A human with a turn-by-turn wayfinding system. (d) A human with a system detecting and communicating personalised, context-specific landmarks instantly understandable by the user. (e) Future technological applications are unlikely to improve the navigational efficiency any further. How should they change the cognitive task performed by the user?

3.3 Pre-registration as a way to standardise and promote new criteria

Understanding how to measure the two above-mentioned research goals is an open challenge. We suggest to approach it from the perspective of exploratory research. This is different from the approach typically applied by papers introducing a new navigation support technology. The default approach is to treat it as confirmatory research, where researchers use theories of spatial cognition (e.g., a generalised assumption that salient landmarks are easier to memorise), develop some technology utilising this theory (e.g., a system that shows salient landmarks), and design an experiment to confirm if the assumption is true (e.g., if participants using the system had statistically significantly better memory performance with regard to the newly visualised salient landmarks).

This approach has two impediments. First, it binds together the test of the theory with the evaluation of the technology: if a said system would not improve participants' memory, it is impossible to conclude whether this is due to the theory being wrong (i.e., in this particular context salient landmarks are not easier to remember), or whether the technology is not a good application of this particular theoretical insight (e.g., something about the visualisation of landmarks in this particular application hinders the memorisation effect). This would not be a problem if theories we use were strongly verified in practical context of use but this is not the case for most theories in behavioural and psychological science [9]. Second, this approach creates an "all-or-nothing" narrative about the success of the proposed technology: if the new technology does not improve the memorisation of landmarks, the approach has failed, and other measures are reported at most as an interesting side-effect in secondary analyses. Reviewers' and readers' scepticism towards this kind of secondary analyses in confirmatory studies is justified. From the statistical viewpoint, one dataset cannot be used to both generate and validate a hypothesis because a dataset used to *find* a phenomenon is not going to falsify it [24].

An alternative approach is to explicitly treat validation of new navigation support systems tested with participants in naturalistic scenarios as exploratory research. The goal of exploratory research is not to validate a hypothesis but to describe interesting properties of the data, determine which (tentative) findings are interesting, and to propose steps forward for future research [24]. This means that what we are used to call an "evaluation" of new technological propositions, would in fact be an exploratory data analysis project. From the standpoint of cumulative science, a declared exploratory study has a higher value than an undisclosed exploratory practice looking like a confirmatory test [16, 24].

This is a particularly challenging issue for our discipline. Creating a novel navigation support system and testing it in a naturalistic scenario is proceeded by decades of research in psychology and spatial computation; it is based on theories confirmed in numerous experiments; it utilises advanced computational methods that must have matured before they became usable in practical contexts. To call such research "exploratory" does not give it justice. Yet, is seems methodologically more appropriate to treat it as exploratory than as confirmatory analysis. For this reason, for evaluating navigation support systems we propose to adopt the term "postdiction" – defined as research undertaken after the data is known with the goal of explaining why it occurred (also referred to as "research in the context of discovery" [16]).

Dirnagl [3] compares doing such exploratory studies to the endeavour undertaken by sailing expeditions in the times when maps were still incomplete. These attempts were pre-registered, in the sense that the goal had to be specified. Yet, the retrospective value of the expeditions is not only whether they did or did not reach the goal. Partial discoveries of predecessors were valid – even if only fragmentary – reasons to pursue or avoid particular directions in the future. But what made the cumulative learning from this expeditions possible were logs kept from the journeys. Pre-registering research as exploratory, publishing study protocols, and keeping a public log of exploratory analyses is the modern science equivalent of such expeditions.

For landmark-enhanced navigation support systems we suggest the following ways for enhancing the cumulative effort of discovering novel and useful technology:

1. A declaration of the intended role of the newly proposed technology (e.g., a cognitive amplifier or a cognitive trainer, within the above described framework [7]).
2. A pre-registration of the claim that the newly proposed system *will not significantly decrease navigational performance*. Since navigation support is already sufficiently supporting performance, expecting that new systems will further improve on this criterion

seems unjustified. In some cases, a similar claim might be made for spatial memory measures – not all contexts require that spatial memory of the travelled route increases. Pre-registering this claim makes it explicit, open, and traceable, that researchers were not focused on increasing measures accepted as dominant in the field. This can assist authors and readers in shifting attention to other aspects tested in the study.

3. A pre-registration of intended postdiction analyses aimed at detecting whether the cognitive task performed by the human in the joint human-computer navigation system has been changed by the newly tested technology. Such analyses are sensitive to the specificity of the dataset, therefore the highest value lies in detecting a change by using an approach developed by other researchers on another dataset. These are not unknown in the domain of landmark research but a set of standardised methods that would be appropriate for this research goal has yet to crystallise. Arguments have been made for *detecting strategies and variance* [10] in navigational behaviour and problem-solving, instead of methods for quantifying performance.

4 Conclusion

Landmark research spans decades, but the use of landmarks in everyday navigation support technology to date has neither been realised in common applications, nor proven necessary. Hermann [7] argued that demonstrating a successful application can be an ultimate check for basic research – a check that landmark-enhanced navigation support systems have not yet passed. The claim presented in this paper is that the reason is not a technological limitation but rather the way in which navigation support research is evaluated – with an over-reliance on evaluation criteria specific to strictly-psychological research.

Two suggestions have been made for moving beyond the impasse: (1) re-considering the role of landmark-enhanced navigation support systems in everyday life of their users; and (2) standardising the evaluation of a phenomenon difficult to capture with traditional measures – whether the technology changed the cognitive task performed by its user. As no clear set of methods for detecting the latter are agreed upon, we suggest *pre-registering exploratory analyses* prior to collecting the data as a means of highlighting this goal. Methods for detecting change and variance in cognitive or behavioural strategies, developed by researchers on other datasets, should be prioritised.

This turn to exploratory analyses is not an attempt to diminish or criticise the quality of navigation support research but an attempt to set it apart from priorities central to strictly-psychological or strictly-computational research. Moving towards this goal requires the development of evaluation criteria better reflecting the societal potential (or scientific novelty) of the new technology. In the spirit of "use-inspired basic research" [22] we are strongly convinced about the potential that such technology can also feed back into basic science; for instance by shifting research focus onto cognitive processes and strategies that gain importance in everyday life inevitably assisted by technology.

References

1 Chittaranjan Andrade. HARKing, Cherry-Picking, P-Hacking, Fishing Expeditions, and Data Dredging and Mining as Questionable Research Practices. *The Journal of Clinical Psychiatry*, 82(1), 2021. doi:10.4088/jcp.20f13804.

2 Richard Arias-Hernandez, Tera M Green, and Brian Fisher. From Cognitive Amplifiers to Cognitive Prostheses: Understandings of the Material Basis of Cognition in Visual Analytics. *Interdisciplinary Science Reviews*, 37(1):4–18, 2013. doi:10.1179/0308018812z.0000000001.

3 Ulrich Dirnagl. Preregistration of exploratory research: Learning from the golden age of discovery. *PLOS Biology*, 18(3):e3000690, 2020. doi:10.1371/journal.pbio.3000690.

4 Itiel E Dror and Stevan Harnad. Offloading cognition onto cognitive technology. In *Cognition Distributed: How Cognitive Technology Extends Our Minds*, volume 16, pages 1–23. John Benjamins Publishing, 2008. arXiv:0808.3569.

5 Kenneth M. Ford, Clark Glymour, and Patrick J. Hayes. On the Other Hand ... Cognitive Prostheses. *AI Magazine*, 18(3):104, 1997. doi:10.1609/aimag.v18i3.1317.

6 Marcelo De Lima Galvão, Jakub Krukar, Martin Nöllenburg, and Angela Schwering. Route schematization with landmarks. *Journal of Spatial Information Science*, 21:99–136, December 2020. doi:10.5311/JOSIS.2020.21.589.

7 Douglas J. Hermann. The potential of cognitive technology. In W. Richard Walker and Douglas J. Herrmann, editors, *Cognitive Technology: Essays on the Transformation of Thought and Society*, pages 5–19. McFarland & Co, Jefferson, N.C, 2005.

8 Edwin Hutchins. *Cognition in the Wild*. The MIT Press, Cambridge, MA, US, 1995.

9 Hans IJzerman, Neil A. Lewis, Andrew K. Przybylski, Netta Weinstein, Lisa DeBruine, Stuart J. Ritchie, Simine Vazire, Patrick S. Forscher, Richard D. Morey, James D. Ivory, and Farid Anvari. Use caution when applying behavioural science to policy. *Nature Human Behaviour*, 4(11):1092–1094, 2020. doi:10.1038/s41562-020-00990-w.

10 Alan Kingstone, Daniel Smilek, and John D. Eastwood. Cognitive Ethology: A new approach for studying human cognition. *British Journal of Psychology*, 99(3):317–340, 2008. doi:10.1348/000712607x251243.

11 John K. Kruschke. Bayesian Assessment of Null Values Via Parameter Estimation and Model Comparison. *Perspectives on Psychological Science*, 6(3):299–312, 2011. doi:10.1177/1745691611406925.

12 Rui Li, Amichi Korda, Maurin Radtke, and Angela Schwering. Visualising distant off-screen landmarks on mobile devices to support spatial orientation. *Journal of Location Based Services*, 8(3):166–178, 2014. doi:10.1080/17489725.2014.978825.

13 K Lynch. *The Image of the City*. MIT Press, 1960.

14 Daniel R Montello. Cognitive Research in GIScience: Recent Achievements and Future Prospects. *Geography Compass*, 3(5):1824–1840, 2009. doi:10.1111/j.1749-8198.2009.00273.x.

15 Daniel R. Montello. Landmarks are Exaggerated. *KI - Künstliche Intelligenz*, 31(2):193–197, 2017. doi:10.1007/s13218-016-0473-5.

16 Brian A. Nosek, Charles R. Ebersole, Alexander C. DeHaven, and David T. Mellor. The preregistration revolution. *Proceedings of the National Academy of Sciences*, 115(11):201708274, 2018. doi:10.1073/pnas.1708274114.

17 Rajchandar Padmanaban and Jakub Krukar. Increasing the density of local landmarks in wayfinding instructions for the visually impaired. In Georg Gartner and Haosheng Huang, editors, *Progress in Location-Based Services 2016*, pages 131–150, Cham, 2017. Springer International Publishing.

18 Denise Peters, Yunhui Wu, and Stephan Winter. Testing Landmark Identification Theories in Virtual Environments. In Christoph Hölscher, Thomas F. Shipley, Marta Olivetti Belardinelli, John A. Bateman, and Nora S. Newcombe, editors, *Spatial Cognition VII*, Lecture Notes in Computer Science, pages 54–69, Berlin, Heidelberg, 2010. Springer. doi:10.1007/978-3-642-14749-4_8.

19 Martin Raubal. Cognitive engineering for geographic information science. *Geography Compass*, 3(3):1087–1104, 2009. doi:10.1111/j.1749-8198.2009.00224.x.

20 Angela Schwering, Jakub Krukar, Rui Li, Vanessa Joy Anacta, and Stefan Fuest. Wayfinding Through Orientation. *Spatial Cognition & Computation*, 17(4):273–303, 2017. doi:10.1080/13875868.2017.1322597.

21 Alastair D Smith, Gary Priestnall, and Juliette Cross. Supporting spatial orientation during route following through dynamic maps with off-screen landmark persistence. *Spatial Cognition & Computation*, pages 1–28, 2021. doi:10.1080/13875868.2021.1985122.

22 Donald E Stokes. *Pasteur's Quadrant: Basic Science and Technological Innovation*. Brookings Institution Press, 1997.

23 Tyler Thrash, Sara Lanini-Maggi, Sara I. Fabrikant, Sven Bertel, Annina Brügger, Sascha Credé, Cao Tri Do, Georg Gartner, Haosheng Huang, Stefan Münzer, and Kai-Florian Richter. The Future of Geographic Information Displays from GIScience, Cartographic, and Cognitive Science Perspectives (Vision Paper). In Sabine Timpf, Christoph Schlieder, Markus Kattenbeck, Bernd Ludwig, and Kathleen Stewart, editors, *14th International Conference on Spatial Information Theory (COSIT 2019)*, volume 142 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 19:1–19:11, Dagstuhl, Germany, 2019. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik. doi:10.4230/LIPIcs.COSIT.2019.19.

24 Eric-Jan Wagenmakers, Ruud Wetzels, Denny Borsboom, Han L. J. van der Maas, and Rogier A. Kievit. An Agenda for Purely Confirmatory Research. *Perspectives on Psychological Science*, 7(6):632–638, 2012. doi:10.1177/1745691612463078.

25 Jan M. Wiener, Simon J. Büchner, and Christoph Hölscher. Taxonomy of Human Wayfinding Tasks: A Knowledge-Based Approach. *Spatial Cognition & Computation*, 9(2):152–165, 2009. doi:10.1080/13875860902906496.

26 Anna Wunderlich, Sabine Grieger, and Klaus Gramann. Landmark information included in turn-by-turn instructions induce incidental acquisition of lasting route knowledge. *Spatial Cognition & Computation*, pages 1–26, 2022. doi:10.1080/13875868.2021.2022681.

27 Demet Yesiltepe, Ruth Conroy Dalton, and Ayse Ozbil Torun. Landmarks in wayfinding: a review of the existing literature. *Cognitive Processing*, 22(3):369–410, 2021. doi:10.1007/s10339-021-01012-x.

New Human Dynamics in the Emerging Metaverse: Towards a Quantum Phygital Approach by Integrating Space and Place

Daniel Sui
Department of Geography, Virginia Tech, Blacksburg, VA, USA

Shih-Lung Shaw[1]
Department of Geography, University of Tennessee, Knoxville, TN, USA

Abstract

With the convergence of mirror worlds, virtual worlds, lifelogging, and augmented/virtual reality, the emerging metaverse is rapidly becoming a major platform where humans work, shop, entertain themselves, and socialize with others. Human dynamics, which refers to all forms of human activities and interactions, will undergo profound transformations in the coming years with the advent of the metaverse. The new human dynamics will be neither **phy**sical nor di**gital** but a seamless integration of both – **phygital**[2]. The goal of this vision paper is to develop a phygital approach to support human dynamics research in the spirit of GIScience as a convergence. Built on our earlier work in human dynamics research, we argue that the current discussions on human dynamics are conceptually constrained by their physical and digital silos. The new phygital approach we are envisioning aims to transcend the simplistic dichotomy by integrating both space and place perspectives. This paper also draws on basic concepts in quantum physics and earlier discussions on their potential applications in geography and GIScience to espouse a quantum turn in exploring the human dynamics in the emerging metaverse. It explores how concepts, methods, and understandings from quantum physics and emerging quantum computing and communication technologies can be translated into addressing fundamental geographical analyses for this phygital world.

2012 ACM Subject Classification Theory of computation → Quantum information theory; Human-centered computing; Computing methodologies

Keywords and phrases metaverse, human dynamics, phygital, space-place, quantum, GIScience theory

Digital Object Identifier 10.4230/LIPIcs.COSIT.2022.11

Category Vision Paper

Acknowledgements Comments on an earlier draft by Dr. Laurel Miner and research assistance by Dr. Scarlett Jin are gratefully acknowledged. Constructive comments from five anonymous reviewers have also significantly clarified our thinking about the critical issues discussed in our paper. We are solely responsible for any remaining errors.

1 Introduction – The brave new metaverse

When Aldous Huxley penned the historical *Brave New World* back in 1932, he would never have imagined that his brave new world evolved into a brave new metaverse 90 years later in 2022 as evidenced by all the attentions metaverse, which is a hybrid world in which the virtual world based upon digital bits is increasingly linked to the atom-based physical world (`http://www.metaverseroadmap.org`), received by the media, business/industry, and

[1] Corresponding author
[2] According to Awabot, an Australian agency Momentum invented the term "phygital" and claimed copyright in 2013. A definition and advantages of phygital are provided at Cyberclick.

the research community during the past year. Indeed, those with values different from the prevailing social vision may believe the metaverse will save them from Huxley's Brave New World [8], which some are convinced is already here/inevitable [2].

Human dynamics research, which studies all forms of human activities and interactions in both physical and virtual worlds, currently faces a computing environment that has drastically changed during the past two decades. Instead of the traditional distinction of hardware and software, we have witnessed the emergence of 'everyware' [14]. The future scenario of everyware (sometimes used interchangeably as ubicomp or ambient computing) when people and objects are connected via distributed computing and unconstrained by geographical contexts has arrived faster than expected. Concomitant with the growth of ubicomp/everyware, we are also rapidly entering a new age of the metaverse.

First coined by Neal Stephenson's [28] science fiction novel Snow Crash, metaverse refers to a fictional virtual world where humans, as avatars, interact with each other and software agents in a three-dimensional space that uses the metaphor of the real world. The rapidly evolving metaverse is a result of several converging technologies. According to the metaverse road map report, the browser for engaging this metaverse will be based upon a 3-D Web that brings together the following four technologies:

- Mirror worlds – digital representations of the atom-based physical world, such as Google Earth, Microsoft Virtual Earth, NASA World Winds, ESRI ArcGlobe, USGS National Map, and the massive georeferenced GIS databases developed during the past fifty years, virtual geographical environment (VGE), and a variety of digital twins of the physical world at different scales.
- Virtual worlds – digital extensions of the physical world (e.g., amazon.com) and/or digital representations of imagined worlds (e.g., Second Life, World of Warcraft).
- Lifelogging - the digital capture of information about people and objects in the real or digital worlds (e.g., Twitter, Instagram, YouTube, Facebook/Meta, and TikTok).
- Augmented and virtual reality– sensory overlays of digital information on the real and virtual worlds using a head-up display (HUD) or other mobile/wearable devices such as cell phones or sensors via participatory sensing.

With the rebranding of Facebook as the new Meta [35], 2021 will go down in history as a watershed year for the development of the metaverse. When we now think about GIScience research in general and human dynamics in particular, we cannot separate either of them from the emerging metaverse. Viewed from a metaverse perspective, our discussions of human dynamics within the geospatial community have focused almost exclusively on the perspectives of mirror worlds with growing interests in social media/lifelogging in recent years.

With the convergence of mirror worlds, virtual worlds, lifelogging, and augmented/virtual reality, the emerging metaverse is becoming a platform where humans work, shop, entertain themselves, and socialize with others. Human dynamics will undergo profound transformations in this emerging metaverse. Some scholars also call it synthetic or reality media [6]. The new human dynamics will be neither **phy**sical nor di**gital** but a seamless integration of both – **phygital**. The goal of this vision paper is to develop a phygital approach to support human dynamics research from a broader GIScience perspective. Extending the earlier idea about GIS as media, GIS is increasingly becoming indistinguishable from the reality media created by the emerging metaverse [30, 32].

The rest of this vision paper is organized as follows. After a brief introduction on metaverse in this section, section 2 presents a synoptic overview of the current state of human dynamics research and its limitations. To address these limitations, section 3 outlines key features of a

quantum phygital approach for conducting human dynamics research in the metaverse by integrating space and place. Section 4 further elaborates the theoretical, methodological, and legal/ethical issues of conducting human dynamics and GIScience research in the age of metaverse. Summary and conclusions are contained in the last section.

2 Human Dynamics Research at a Cross Road

Human beings carry out various activities and interactions to meet their needs. According to Maslow's hierarchy of needs, human needs are arranged in hierarchies of predominance that consist of (1) physiological needs (e.g., food, water, sleep), (2) safety needs (e.g., health, employment), (3) love needs (e.g., family, friend), (4) esteem needs (e.g., accomplishment, confidence), and (5) self-actualization needs (e.g., creativity, meaning) [17]. The activities and interactions performed by people collectively become the foundation of the economic, social, cultural, political, and other systems in human societies. In the meantime, the dynamics of these human activities and interactions evolve with the changing environments and technologies over time [25, 23].

Most human activities and interactions were performed in physical space through in-person contacts before the modern technologies made it feasible for us to interact and accomplish certain tasks remotely. As the modern technologies such as personal computers, the internet, and mobile phones became available in the late twentieth century, they enabled an increasing number of human activities and interactions taking place in the so-called virtual space. For example, online shopping and online social networks have replaced some shopping trips and social gatherings in physical space. Mobile phones further relaxed the constraint of staying connected to the network at fixed locations. During the COVID-19 pandemic, teleworking, online education, and many other virtual activities also surged to an unprecedented level. These changes during the recent decades have created an increasingly hybrid physical-virtual world. We now use transportation to move among different places in physical space while we navigate among different places in virtual space via information and communications technology (ICT) [26]. It is important to note that what happens in virtual space often influence and are influenced by what takes place in physical space and vice versa. For example, online orders at amazon.com trigger specific activities and shipments in physical space. It therefore is critical to treat today's world as a hybrid physical-virtual world rather than two independent physical and virtual worlds. In this paper, we use the term human dynamics to cover all forms of human activities and interactions in today's hybrid physical-virtual world (or phygital world).

Batty [4] argues that "the future subject matter and method of geography will be very different as place and space and time itself become virtual in an age where the digital permeates all human activity" (p.351). There is no doubt that we now live in a phygital world envisioned by Michael Batty twenty-five years ago. Taylor [34] also discusses emerging geographies of virtual worlds enabled by virtual reality (VR) technology. Nevertheless, most conventional geographic information systems (GIS) methods have focused on human dynamics in physical space with the following assumptions: (1) objectivism which assumes that objects exist independent of the subjects who observe them, (2) materialism which assumes that the elementary units of reality are physical objects, (3) reductionism which assumes that larger objects can be reduced to smaller ones, (4) determinism which assumes that objects behave in law-like ways; and (5) mechanism which assumes that causation is mechanical and local. Even when virtual activities are considered in a study [3], the above assumptions often are implied in the study by treating human dynamics in physical space and

human dynamics in virtual space as two independent and parallel worlds. Some studies have attempted to associate human activities in virtual space with human activities in physical space through data such as geotagged tweets [27]. However, the location where a tweet is sent may have nothing to do with the content of a tweet which could generate misleading analysis results. As we move into a hybrid phygital world, it is imperative to pursue human dynamics research with approaches beyond the limits set by the above assumptions and develop approaches that can better integrate human dynamics in a hybrid phygital world.

Taking online shopping at amazon.com as an example, there exist various challenges that we must address. For example, what is the location of amazon.com? How should we represent amazon.com in a GIS environment? In practice, we could use the location of Amazon headquarters office to represent its location in GIS. However, this location may be irrelevant to most transactions at amazon.com. An alternative is to use the street address of the specific vendor's location to represent where an item will be shipped out. But, the vendor may have multiple warehouses from which the ordered item could be shipped. In reality, most people who place orders at amazon.com do not care much about the vendor's location. Instead, they may pay more attention to online reviews or delivery date. In this case, the identity of amazon.com and/or the vendor in virtual space is more critical than their locations in physical space when users place an order at amazon.com. Such human dynamics can be better handled as a relational space, which represents the relations among different entities such as a social network graph, than as an absolute space that assumes an infinite and immovable space which exists independent of other things [24]. Furthermore, online reviews, comments made by friends, and our own experiences with different vendors and online shopping websites also influence our perceptions and attitudes in mental space which in turn will affect our behaviors. These examples illustrate why we need to develop a new framework for studying human dynamics in today's hybrid phygital world. The recent development of Amazon Go and Amazon Fresh, built upon a combination of their "just walk-out" technologies, is the latest quintessential example of the phygital shopping we all will soon experience in the metaverse.

3 Human dynamics and the emerging metaverse: Towards a quantum phygital approach by integrating space and place

3.1 Human dynamics will be increasingly phygital

The new buzzword phygital went viral during the past two years since the beginning of the global pandemic. Technically speaking, "phygital" refers to the seamless integration of both physical and digital universes that captures the essential features of both virtual world and physical world so that we can maximize and optimize our experiences in both.

Operationally, a phygital strategy is closely related to other business strategies such as immersive marketing, omnichannel, or O2O (online-to-offline), but with the accelerated maturing and advances in virtual reality, augmented reality, social media, digital twins and other mirror world technologies, the emerging metaverse has made the following three characteristics a reality for the phygital world: (1) Immediacy: It works to ensure things happen at an exact moment in time; (2) Immersion: The user is an integral part of the experience; and (3) Interaction: Communication is constant and activates the more physical and emotional part of the experience.

There have been many successful phygital implementation examples lately, such as Amazon's Go store, Pokemon Go game, Magik Book, smart tourism, and the emerging phygital banking. This trend is certainly not confined to business and retail, but is also

rapidly diffusing to other sectors such as the government operations, higher education, worship activities, and other non-profit operations. This is indeed a phygital age and this neologism captures a pivotal moment in human history. For the geospatial world, we have become increasingly phygital as well since the early 1960s as we gradually move away from the world of analog paper maps to digital geospatial information. From the early adoption of Global Positioning System (GPS) technologies in civilian applications to the growing popularity of Uber and more broadly to the emerging spatial computing paradigm, what we are dealing with is neither physical nor digital, but increasingly phygital. Concomitantly with this transformation, we are not just dealing with physical/absolute space but also a variety of other spaces and places. This mandates that we need to broaden our conceptual framework to move beyond the Newtonian physical world we are accustomed to in order to better understand the new reality we are in.

3.2 Integrating space and place to understand phygital human dynamics

Space, place, and human are three fundamental elements in geography. Geographers traditionally focus on the spaces that are relevant to human life and the places that are created by human activities. There exist different approaches to conceptualizing space and place. In cartography and GIS, space is often represented according to Newton's concept of *absolute space* and operationalized through Cartesian coordinates and Euclidean geometry. Under this representation, objects can be placed and events can take place at various locations in an empty and objective absolute space. The concept of absolute space can be transformed into the concept of *relative space* by relaxing the assumption of a fixed origin point in absolute space. If we allow the origin point to move with the observer, it becomes a relative space that represents the spatial separation between an observer and other objects based on their relative locations.

We also can create schematic maps that focus on the connections among a set of places such as a map showing the subway network in London, United Kingdom. On a schematic map, the actual locations of network nodes and network links in physical space are not critical as far as the network links represent the correct topological connections among the network nodes. Such schematic maps are examples of representing the relationships in a *relational space*. Furthermore, we can create mental maps to represent maps in human mind that reflect our understanding of the world around us based on a mixture of objective observations and subject perceptions. Locations on a mental map are usually distorted from their physical locations in absolute space. Mental maps, therefore, are examples of representing objects in a mental space.

In addition to the concepts of space, there also exist different concepts of place. Tuan [36] indicates that human beings create meanings to an area in space that becomes a place. Places, therefore, are social constructs that can have different meanings to different people and can evolve over time [22]. Agnew [1] further suggests that the concept of place covers three different dimensions, which are location, locale, and sense of place. When a place is viewed as a location, it is considered as a site in space where an object or an event is located. In this case, we can use (x,y,z) coordinates to define a location in absolute space. When a place is viewed as a locale, it is considered as a setting where activities take place. Locale, therefore, refers to the physical, socioeconomic, and cultural context within which activities occur. The concept of relative space that focuses on the surrounding environments around the observer fits well with the concept of locale. When a place is viewed as sense of place, it is associated with identification with a place such as a sense of belonging to a place. Such human subjective perceptions or attachments to a particular location or locale can be

associated with the concept of mental space. Furthermore, place identity is the most critical element under the concept of relational space to identify the relationships among different entities.

Based on the concepts of absolute space, relative space, relational space, and mental space as well as the concepts of location, locale, place identity, and sense of place, Shaw and Sui [24] propose a space-place (or splatial) framework for a better understanding of human dynamics in a hybrid phygital world (Figure 1). This framework puts humans at the center since human dynamics is created by human activities and interactions to fulfill various needs. In addition, humans are not static objects at fixed locations. They navigate among different places in both physical space and virtual space to carry out different activities and interactions. Humans therefore are treated as dynamic objects in this framework. Humans, who have experienced a "digitization" of every facet in their lives, have discovered new ways to fulfill their needs. Just as Amazon digitized written knowledge with eBooks and YouTube digitized verbal/visual knowledge, blockchain now has digitized property rights for transactions in virtual space. Those who recognize and believe this shift are willing to give this new world an ontological status, thus investing heavily into entities in this new virtual space [11].

The space-place couple of *absolute space/location* in this framework addresses questions such as "Where are the different objects?" that have been implemented in conventional GIS and many other spatial analysis methods. The space-place couple of *relative space/locale* addresses questions such as "What is around us?" that places an emphasis on the context and surrounding environments. The space-place couple of *relational space/place identity* addresses questions such as "What is related to us?" that focuses on the connections among different people and/or entities. Finally, the space-place couple of *mental space/sense of place* addresses questions such as "What do people have in mind?" that refers to the cognitive and mental aspects of human dynamics. The lines connecting these four space-place couples and humans in this framework indicate that they are not independent of each other. Instead, a particular human activity or interaction could be represented by multiple space-place couples that are linked with each other. This framework, therefore, offers a robust yet flexible design to integrate human dynamics in a hybrid phygital world according to different application needs. For example, if an application only requires representations of the absolute space/location couple in physical space and the relational space/place identity couple in virtual space, it does not need to create representations of the relative space/locale couple or the mental space/sense of place couple.

3.3 Rethinking the phygital human dynamics in the metaverse: Towards a quantum leap

If indeed we need to seamlessly integrate space and place to better capture the phygital process that defines the new human dynamics in the metaverse as we move into a post-pandemic world, the assumptions based upon the Newtonian classic world view seem to be unwarranted. Instead we need to reframe human dynamics research by espousing an explicit quantum turn [33].

Quantum theory, initially developed in the early twentieth century, disrupts all the five assumptions in the traditional research outlined in section 2, reflecting the world view of classic Newtonian physics. Viewed from quantum perspectives at the sub-atomic level, systems are not independent of observers; physical objects dissolve into ghost-like processes; and the whole cannot be reduced to parts. Quantum physics recognizes that elements exist as both waves and particles, and an object's state is a wave function that collapses upon measurement. Furthermore, the world does not behave deterministically, and causation is non-local – a phenomenon also known as quantum entanglement.

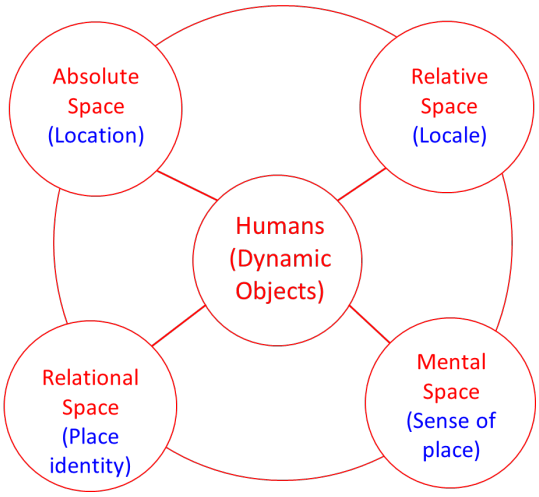


Figure 1 A space-place (splatial) framework for human dynamics research (Adapted from [24]).

Importantly, these findings do not necessarily invalidate the classical worldview at the macro level, since quantum states normally 'decohere' into classical ones above the molecular level, which is why the everyday world appears to us as conforming to the classical worldview. Decoherence has been a barrier to developing a unified quantum theory encompassing both micro and macro levels, and is a fundamental obstacle to the quantum consciousness hypothesis in particular [39]. Nevertheless, at the nano-level, the quantum revolution has decisively overturned the claim of the classical worldview to provide a complete description of reality.

Inspired by Zohar's [40] ground-breaking work, we find the following four quantum concepts are particularly useful when we try to reframe the human dynamics research that is increasingly phygital in the emerging metaverse, often requiring integration of multiple spaces and places to capture their complexity.

1. *Complementarity (holism)*: This refers to the tenet that a complete knowledge of phenomena on atomic dimensions requires a description of both wave and particle properties, the 'wave-particle duality'. Similarly, human dynamics are no longer either physical or digital but both. The portmanteau word phygital captures this new duality of human dynamics.
2. *Entanglement (non-locality)*: Entanglement in quantum physics refers to the phenomenon that measuring one particle immediately alters the properties of the other, even when they are physically separated. Entanglement, also described as 'spooky action at a distance' by skeptical Einstein and his colleagues [12], negates the idea of local realism, in which every event has an immediate cause. The new human dynamics in the metaverse are deeply entangled, at least metaphorically (if not physically), at the global scale as everybody is connected through the Internet with the information exchange instantly.
3. *Superposition (potentiality)*: Much like waves in classical physics, any two (or more) quantum states can be added together ("superposed") and the result will be another valid quantum state; and conversely, every quantum state can be represented as a sum of two or more other distinct states. Unlike the macro physical world, quantum physics reveals that the nature or the behavior of matter at the sub-atomic scale can be actually in all possible states simultaneously [16]. We do not know what the state of any object is as long as we do not look to check. The act of measurement itself causes the object to be

limited to single possibility. The concept of superposition thus disrupts our traditional separation of object from subject. Apparently, the phygital human dynamics in the metaverse are always in a state of high potentiality, with multiple states superimposed with one another until measurement/observation is taking place.

4. *Uncertainty (non-determinism)*: The uncertainty principle asserts that our ability to completely measure of quantum states is fundamentally limited. Knowledge about two dimensions is complementary, and the more accurately we measure one property (e.g., particle speed) and less likely we can measure accurately another (e.g., particle position). The quantum uncertainty is not a limitation in measurement technology or experimental sophistication, but a fundamental attribute of the mathematical formalism that underlies quantum theory and has been supported by a century of experimentation. With the growing misinformation, anonymity online. and the rise of bots, we should be more mindful about the non-deterministic aspect of human dynamics in the metaverse. The only certainty about the phygital world is its uncertainty.

Drawing from the concept of superposition, Bittner [5] proposed a quantum theory of geographic fields that allows for the possibility of representing multiple incompatible states simultaneously at a given point of the geographic field. This quantum theory of geographic fields provides a new level of synthesis and understanding of indeterminacy and ontological vagueness in the geographic world. Indeed a quantum-inspired ontology (based upon entanglement and superposition) and epistemology (based upon the inseparability of the observer and the observed) would provide us a more robust conceptual framework to implement the splatial framework to better understand human dynamics in the phygital metaverse.

4 Further Discussions

We hope that the broad-brush road map we charted above for exploring the new human dynamics in the emerging metaverse will entice more followers in the GIScience community to hit the road to start their journey in the phygital world. We should point out that multiple of, if not all, the key elements of the metaverse have been discussed in the literature since the late 1990s [10, 13, 30, 32]. What is new and exciting is the accelerated convergence of these diverse technologies and theoretical frameworks. There are many fundamental issues that need to be addressed for the quantum approach to studying human dynamics in the phygital metaverse. We have aimed to connect the dots in this paper. The geospatial community needs to make concerted efforts to further investigate the theoretical/philosophical, methodological/technical, and legal/ethical issues related to the metaverse. It is beyond the scope and page limit of this paper to cover these issues in detail, but suffice it here we can present a synoptic overview for further discussion.

4.1 Theoretical and Philosophical Issues

We would like to place the exploration of phygital human dynamics in the metaverse in a broad theoretical/philosophical context. Even before metaverse became trendy, Oxford philosopher Nick Bostrom [7] had made the simulation argument, hypothesizing that we humans may be increasingly living in a matrix-like simulation. According to this argument, it is quite possible that multiple versions of ourselves exist in parallel universes living out their lives in different spaces and timelines. Virk [37] further expanded this argument - if we are indeed living in a simulated universe composed of information that is rendered around us, then many

of the complexities and baffling characteristics of our reality start to make more sense. In particular, Virk [37] argued that the two most popular interpretations of quantum mechanics, the Copenhagen Interpretation and the Many Worlds interpretation, which are thought to be mutually exclusive, can be unified in an information-based framework. Quantum computing, in theory, can enable us to simulate complex phenomena in parallel, allowing the simulation to explore many realities at once to find the most "optimum" path forward. This could explain not only the enigmatic Mandela Effect, but more importantly, provides us with a new understanding of time and space, consistent with the splatial framework we discussed earlier.

Is the simulation argument just another metaphor invoked by scholars based upon the dominating technologies of our time as we did in the past [31], or is it really a very profound theory that is physically and literally true that can help GIScientists better address the ontological and epistemological deficiencies critics leveled against GIS earlier [29]?

4.2 Methodological and Technical Issues

Methodologically, the use of 'quantum-like' mathematical and statistical models to study probabilistic dynamical systems has increasingly become popular. Working along this line will surely have implications for exploring the new human dynamics in the metaverse of a phygital world. In the emerging quantum social science literature, Haven and Khrennikov [15], along with Orrell [20, 21] and Project Q, have already demonstrated the potential applications of a quantum approach in studying a wide range of issues in economics, finance, psychology, sociology, and other domains of inquiry with the help of formal models and concepts used in quantum physics. In particular, recent advances in quantum cognition and psychological modeling [38, 9] may be the most relevant and even directly applicable for studying human dynamics in the metaverse of a phygital world. Quantum decision-making recognizes that judgments and decisions are influenced by context, and that entangled systems cannot, in theory, be modeled as separate systems. It draws attention to quantum theory as a statistical theory, recognizing that the interference of probabilities is a basic statistical feature of quantum theory. Quantum formalisms are merely considered a more effective way of processing incomplete information and accounting for the interference of probabilities in macroscopic quantum systems.

By exploiting collective properties of quantum states, such as superposition and entanglement, to perform computation, quantum computers have been proven to be able to solve certain computational problems substantially faster than classical computers. The demands for computing power for data processing related to human dynamics in the metaverse of a phygital world will increase exponentially. The goal of quantum supremacy or quantum advantage is to demonstrate that a programmable quantum device can solve a problem that no classical computer can solve in any feasible amount of time (irrespective of the usefulness of the problem). As of now, there are generally four ways to build quantum computers (`https://spectrum.ieee.org/4-ways-to-make-bigger-quantum-computers`), but fundamental to all is the quantum bit (qubit or qbit), which provides an exponential advantage over classical computers that are based upon digital bits in binary states of either 0 or 1. Furthermore, quantum computers are the most promising to address issues related to communication and security in the metaverse, with improvements unmatched by classical computers. Most if not all current blockchain code is reliant on SHA-256 hash functions, which is secure enough for classical computing. But preliminary work showing the power of quantum computing to break a SHA-256 code has further elevated the importance and urgency of embracing quantum computing into the future of the metaverse. Indeed, the potential impacts of quantum supremacy for human dynamics in the metaverse of a phygital world cannot be overestimated.

4.3 Legal and Ethical Issues

Although still at an early stage, core metaverse technologies have demonstrated that we are enabled to travel across space and time at an unprecedented level of granularity and a high degree of fidelity. Concomitantly, metaverse has also transported us onto a new uncharted legal and ethical territory that deserves our attention. Due to enormous financial potentials, there are mounting intellectual property issues looming, especially related to patents, contracts, and non-fungible tokens (NFTs) in the metaverse. The emergence of the metaverse challenges the core propositions in our current patent law, i.e., what is patentable and what is not? As for the content in the metaverse, most of the claims will fall into three categories - copyright, trademark, and right of publicity. Legal boundaries are currently not clearly defined regarding what can and can't be included in the metaverse content. There are also liability claims of users against metaverse companies and users against other users in the emerging metaverse.

More than simply a technological marvel and advance, we must regard the emerging metaverse as one of the greatest social experiments humanity has ever taken throughout history, one that merges our physical and digital identities and our physical and virtual presence. This phygital world allows us to create and recreate ourselves as many times as we choose across multiple spaces and places in both physical space and virtual space. Our avatars will not be limited by space and time constraints of our own physical bodies.

So, in addition to the legal challenges, our new phygital persona in metaverse also raises a profound set of ethical questions related to privacy, safety, equity, and agency. It remains to be seen how all the stakeholders can work together to create spaces and places for everyone to thrive in the metaverse, including the GIS community which must deal with the new concepts of space, place, time, and human dynamics in the metaverse. Additionally, we have to have strong identity and security solutions for the convergence of digital and physical worlds to protect the creators, the brands, and the consumers.

Scientific discourses are not a separate sphere of society. They are part of social cultures. Scientific discourses contribute to shaping ontologies and causal stories. Ontological imaginaries shape our political and research practices and how we validate them. In an entangled quantum world, we may also be able to bring about big changes through small local actions (micropolitics). Ethically and politically, quantum onto-epistemologies raise the bar for adjudicating ethical choices, and at the same time, open up possibilities for further actions and engagement. With its non-local, non-deterministic, and participatory approach to social change, the quantum perspective advocated in this paper sheds a brighter light on the new ethics in the phygital age. By taking non-local, hidden, and subjective factors seriously and explicitly, quantum ethics perhaps can guide our social practices to address complex issues facing humanity today, such as increasing global polarization, growing economic disparity, and worsening global environmental change. Indeed, uninformed populism and nationalist approaches go against an entangled worldview as espoused by the quantum turn. O'Brien [18] lays out a road map on how the quantum perspective could help us better cope with the challenges posed by global climate change through meaningful social transformation.

5 Concluding Remarks: GIScience as a convergence science

GIScience in general and human dynamics research in particular have undergone major changes during the past three decades, and yet nothing is as profound and far-reaching as we are going through right now. With the maturing of metaverse technologies, quantum

computing, AI/machine learning, and blockchains, GIScience and human dynamics research are on the cusp of another major paradigm shift that calls for new theories, methodologies, and ethics for us to better deal with the brave new phygital world we increasingly live in.

In the spirit of accelerating convergence research, this vision paper has tried to present a unified splatial framework to better understand the phygital human dynamics by integrating previous conceptualization of multiple spaces and places. Moving away from the absolute conceptualization of space and place as defined by Newtonian physics, the splatial framework embodies the core concepts of quantum physics. We hope this vision paper has charted new territory for further exploration along theoretical, methodological, and ethical fronts. More than ever, GIScience and human dynamics research need to take a convergence approach instead of the traditional siloed approach.

According to the U.S. National Academy of Sciences [19], convergence research must have two primary characteristics:

- Transdisciplinarity: As experts from different disciplines pursue common research challenges, their knowledge, theories, methods, data, research communities, and languages become increasingly intermingled or integrated. New frameworks, paradigms, or even disciplines can form sustained interactions across multiple communities.
- Stakeholder synergy: In order to have broader impacts, research should be conducted by drawing together academic researchers, policymakers, and industry partners. Convergence research is generally inspired by the need to address a specific challenge or opportunity, whether it arises from deep scientific questions or pressing societal needs.

As shown throughout this paper, no single discipline can actually claim exclusive ownership on addressing any challenges or issues outlined in this paper. Convergence GIScience and human dynamics research entails a transdisciplinary approach to seamlessly integrate theories, methods, and data from multiple disciplines. Moreover, convergence further mandates stakeholder synergy in order to create broader public impacts of the research we are conducting. Stakeholder synergy – the integration of academia, industry, and government – is critically important for the success of GIScience and human dynamics research. By default, seeking stakeholder synergy automatically mandates the synthesis of creative works by multiple teams with diverse backgrounds. Apparently, a team science approach is needed to develop GIScience and human dynamics research in the context of stakeholder synergy, which mandates academics to move beyond their comfort zone to address pressing issues facing society today. Of course, we are mindful of the future of government in light of the accelerated adoption of blockchain technology. Bitcoin (or cryptocurrency more generally) was created in the wake of the global financial crisis of 2008 because of a growing, and still quite prevalent, distrust of government, authorities/institutions, and big corporations. A key attribute of the cryptocurrency is its libertarian ideal of decentralized, permissionless, self-governing operation, not controlled by a central authority.

Over the past two decades, there has been an emerging emphasis on scientifically addressing multi-factorial problems, such as climate change, fighting global terrorism, the rise of infectious/chronic diseases, the health impacts of social stratification, and growing concerns of social disparity. This has contributed to a surge of interest and investment in team science. Increasingly, scientists across many disciplines and settings are engaging in team-based research initiatives. These include small and large teams, uni- and multi-disciplinary groups, and efforts that engage multiple stakeholders such as scientists, community members, and policymakers. Since the first road map for our phygital future in the metaverse was articulated by a team of industry leaders, it is even more critical that academic GIScience researchers should aggressively seek stakeholder synergy for our future endeavors.

References

1 John A. Agnew. *Space and place*, chapter The SAGE Handbook of Geographical Knowledge, pages 316–330. SAGE Publications, Los Angeles, 2011.

2 Tom Arant. *Welcome to Our Real Matrix*. Thomas F. Arant, 2021. URL: https://www.ebook.de/de/product/41008107/tom_arant_welcome_to_our_real_matrix.html.

3 Albert-László Barabási. The origin of bursts and heavy tails in human dynamics. *Nature*, 435(7039):207–211, 2005. doi:10.1038/nature03459.

4 Michael Batty. Virtual geography. *Futures*, 29(4-5):337–352, 1997. doi:10.1016/s0016-3287(97)00018-9.

5 Thomas Bittner. Towards a Quantum Theory of Geographic Fields. In Eliseo Clementini, Maureen Donnelly, May Yuan, Christian Kray, Paolo Fogliaroni, and Andrea Ballatore, editors, *13th International Conference on Spatial Information Theory (COSIT 2017)*, volume 86 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 5:1–5:14, Dagstuhl, Germany, 2017. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik. doi:10.4230/LIPIcs.COSIT.2017.5.

6 Jay David Bolter, Maria Engberg, and Blair MacIntyre. *Reality Media*. The MIT Press, 2021. doi:10.7551/mitpress/11708.001.0001.

7 Nick Bostrom. The simulation argument: Why the probability that you are living in a matrix is quite high. https://www.simulation-argument.com/matrix.html, 2013. Accessed: 2022-1-23. URL: https://www.simulation-argument.com/matrix.html.

8 Andrew Bosworth and Nick Clegg. Building the metaverse responsibly. https://about.fb.com/news/2021/09/building-the-metaverse-responsibly/, 2021. Accessed: 2022-1-23.

9 Jerome R. Busemeyer and Zheng Wang. Quantum cognition: Key issues and discussion. *Topics in Cognitive Science*, 6(1):43–46, 2014. doi:10.1111/tops.12074.

10 Mike Crang, Phil Crang, and Jon May. *Virtual geographies*. Routledge London, 1999.

11 Christopher DiLella and Andrea Day. Investors are paying millions for virtual land in the metaverse, 2022. Accessed: 2022-1-23. URL: https://www.cnbc.com/2022/01/12/investors-are-paying-millions-for-virtual-land-in-the-metaverse.html.

12 A. Einstein, B. Podolsky, and N. Rosen. Can quantum-mechanical description of physical reality be considered complete? *Physical Review*, 47(10):777–780, 1935. doi:10.1103/physrev.47.777.

13 Peter Fisher and David Unwin. *Virtual reality in geography*. CRC Press, 2001.

14 Adam Greenfield. *Everyware: The dawning age of ubiquitous computing*. New Riders, 2006.

15 Emmanuel Haven and Andrei Khrennikov, editors. *The Palgrave handbook of quantum models in social science*. Springer, 2017.

16 Julia Layton. Scientists prove Schrodinger's cat can be in two places at once. https://science.howstuffworks.com/science-vs-myth/everyday-myths/scientists-prove-schrodingers-cat-can-be-two-places-once.htm, 2021. Accessed: 2022-1-23.

17 Abraham H Maslow. *A dynamic theory of human motivation*, chapter Understanding Human Motivation, pages 26–47. Howard Allen Publishers, Cleveland, OH, 1958.

18 Karen O'Brien. *You Matter More Than You Think: Quantum Social Change for a Thriving World*. cChange Press, Oslo, Norway, 2021. URL: https://www.ebook.de/de/product/41827474/karen_o_brien_you_matter_more_than_you_think_quantum_social_change_for_a_thriving_world.html.

19 National Academies of Sciences (NAS). *Convergence: Facilitating Transdisciplinary Integration of Life Sciences, Physical Sciences, Engineering, and Beyond*. National Academies Press, Washington D.C., 2014.

20 David Orrell. *Economyths: 11 Ways Economics Gets It Wrong*. Icon Books Ltd, July 2017. URL: https://www.ebook.de/de/product/28540206/david_orrell_economyths.html.

21 David Orrell. *Quantum Economics: the new science of money*. Icon Books Ltd, London, 2018.

22 Stéphane Roche. Geographic information science II. *Progress in Human Geography*, 40(4):565–573, 2015. doi:10.1177/0309132515586296.

23 Shih-Lung Shaw and Daniel Sui. *Human Dynamics Research in Smart and Connected Communities*. Springer International Publishing AG, Cham, Switzerland, 2018.

24 Shih-Lung Shaw and Daniel Sui. Understanding the new human dynamics in smart spaces and places: Toward a splatial framework. *Annals of the American Association of Geographers*, 110(2):339–348, 2019. doi:10.1080/24694452.2019.1631145.

25 Shih-Lung Shaw, Ming-Hsiang Tsou, and Xinyue Ye. Editorial: human dynamics in the mobile and big data era. *International Journal of Geographical Information Science*, 30(9):1687–1693, 2016. doi:10.1080/13658816.2016.1164317.

26 Shih-Lung Shaw and Hongbo Yu. A GIS-based time-geographic approach of studying individual activities and interactions in a hybrid physical–virtual space. *Journal of Transport Geography*, 17(2):141–149, 2009. doi:10.1016/j.jtrangeo.2008.11.012.

27 Luke Sloan and Jeffrey Morgan. Who tweets with their location? understanding the relationship between demographic characteristics and the use of geoservices and geotagging on twitter. *PLOS ONE*, 10(11):e0142209, 2015. doi:10.1371/journal.pone.0142209.

28 Neal Stephenson. *Snow Crash*. Bantam Books, New York, 1992.

29 Daniel Z. Sui. Gis and urban studies: Positivism, post-positivism, and beyond. *Urban Geography*, 15(3):258–278, 1994. doi:10.2747/0272-3638.15.3.258.

30 Daniel Z. Sui. Cybergis: A perspective from the emerging metaverse. In Shaowen Wang, Nancy R. Wilkins-Diehr, and Timothy L. Nyerges, editors, *Proceedings of the NSF CyberGIS Workshop*. University Consortium of Geographic Information Science (UCGIS), 2010.

31 Daniel Z. Sui. *Rethinking progress in urban modeling: Models, metaphors, and meaning*, chapter Urban Remote Sensing: Monitoring, synthesis, and modeling in the urban environment, pages 371–382. Wiley, New York, 2010.

32 Daniel Z. Sui. From the mirror worlds to everyware in the metaverse: Or what is special about spatial (computing)? In *Proceedings of the Spatial Computing Workshop*. The U.S. National Academies of Science, 2012.

33 Daniel Z. Sui. *The quantum turn for geospatial technologies and society*, chapter The Routledge Handbook of Geospatial Technologies and Society. Routledge, London, 2021.

34 Jonathan Taylor. The emerging geographies of virtual worlds. *Geographical Review*, 87(2):172–192, 2010. doi:10.1111/j.1931-0846.1997.tb00070.x.

35 TechFacebook. Connect 2021: Our vision for the metaverse, 2021. Accessed: 2022-1-23. URL: https://tech.fb.com/connect-2021-our-vision-for-the-metaverse/.

36 Yi-Fu Tuan. *Space and Place: The Perspective of Experience*. University of Minnesota Press, Minneapolis, 1977.

37 Rizwan Virk. *The Simulated Multiverse: An MIT Computer Scientist Explores Parallel Universes, The Simulation Hypothesis, Quantum Computing and the Mandela Effect*. Bayview Books, Moutainview, CA, 2021.

38 Zheng Wang, Jerome R. Busemeyer, Harald Atmanspacher, and Emmanuel M. Pothos. The potential of using quantum theory to build models of cognition. *Topics in Cognitive Science*, 2013. doi:10.1111/tops.12043.

39 Alexander Wendt. *Quantum Mind and Social Science*. Cambridge University Press, Cambridge, 2015.

40 Danah Zohar. *The quantum society: mind, physics and a new social vision*. Morrow, New York, 1994.

Large-Scale Spatial Prediction by Scalable Geographically Weighted Regression: Comparative Study

Daisuke Murakami[1]
Institute of Statistical Mathematics, Tokyo, Japan

Narumasa Tsutsumida
Saitama University, Japan

Takahiro Yoshida
The University of Tokyo, Japan

Tomoki Nakaya
Tohoku University, Seindai, Japan

Abstract

Although the scalable geographically weighted regression (GWR) has been developed as a fast regression approach modeling non-stationarity, its potential on spatial prediction is largely unexplored. Given that, this study applies the scalable GWR technique for large-scale spatial prediction, and compares its prediction accuracy with modern geostatistical methods including the nearest-neighbor Gaussian process, and machine learning algorithms including light gradient boosting machine. The result suggests accuracy of our scalable GWR-based prediction.

2012 ACM Subject Classification Computing methodologies → Model development and analysis

Keywords and phrases Spatial prediction, Scalable geographically weighted regression, Large data, Housing price

Digital Object Identifier 10.4230/LIPIcs.COSIT.2022.12

Category Short Paper

Funding This research was funded by the Joint Support Center for Data Science Research at Research Organization of Information and Systems (ROIS-DS-JOINT) under Grant 006RP2018, 004RP2019, 003RP2020, and 005RP2021.

1 Introduction

Geostatistical Gaussian process (GP) models have been used for spatial prediction in geology, environmental science, and other fields (see [2]). Although GP-based spatial prediction is known to be accurate as demonstrated in [9], the computational complexity inflates in an order of N^3 where N is the sample size due to a matrix inversion. The classical GP model is unavailable for large samples (e.g., $N > 10,000$). To address the drawback, fast GP approximations have been developed in geostatistics (see [7]) and machine learning areas (see [11]). For example, nearest-neighbor Gaussian process (NNGP [3]) is widely accepted as a fast approximate GP in geostatistics. NNGP and other scalable GPs achieve linear-time computational complexity (i.e., the computational complexity increases in an order of N). They are available for very large samples.

[1] corresponding author

Scalable GPs usually model stationary spatial process assuming model parameters including regression coefficients as constant over space. However, [4, 13] among others have demonstrated that regression coefficients can vary over geographical space. Nevertheless, fast prediction technique considering such spatially varying coefficients (SVCs) is quite limited.

Geographically weighted regression (GWR [1]) is a popular SVC modeling technique that has been used for spatial prediction (e.g., [6, 5]). It is hard to apply the classical GWR for very large data in terms of the computational complexity and memory usage. To overcome the limitation, [10] and [12] developed algorithms for estimating the GWR model computationally efficiently. In particular, the latter developed the scalable GWR technique achieving a quasi-linear computational complexity with very small approximation error. The scalable GWR is potentially useful for spatial prediction. However, it has never been used for spatial prediction.

The objective of this study is to examine the usefulness of the scalable GWR in terms of spatial prediction for large samples through a comparison with modern prediction methods in geostatistics and machine learning areas.

2 GWR model

2.1 Basic GWR model

GWR describes the explained variable y_i at i-th sample site on a two-dimensional space using the following model:

$$y_i = \sum_{k=1}^{K} x_{i,k}\beta_{i,k} + \epsilon_i, \qquad \epsilon_i \sim N(0, \sigma^2) \tag{1}$$

where $x_{i,k}$ is the k-th explanatory variable and σ^2 is the variance parameter. $\beta_{i,k}$ is the k-th regression coefficient at the i-th location. The model is estimated by a weighted least squares (WLS) method assuming greater weights for nearby samples using a distance-decaying kernel whose decay-speed is dependent on a bandwidth parameter w. Later, we will use an exponential kernel $k(d_{i,j}; w) = \exp(d_{i,j}/w)$, where $d_{i,j}$ is the Euclidean distance between locations i and j. The bandwidth parameter w is typically estimated by a leave-one-out cross-validation (LOOCV). In each iteration of the LOOCV, regression coefficients must be estimated for all the N sample sites. This property makes GWR slow for very large samples.

2.2 Scalable GWR model

Scalable GWR estimates the same local model (Eq. 1). To lighten the computational cost, the kernel function $k(d_{i,j}; w)$ with unknown b is replaced with a linear combination of kernel functions with known b values.

$$k^*(d_{i,j}; a, b) = a + \sum_{l=1}^{L} b^l k(d_{i,j}; \tilde{w})^{4/2^l}, \tag{2}$$

where $\tilde{w}$ is a known bandwidth, which is specified based on the median of the 100-nearest neighbor distance. a and b are parameters being estimated through the LOOCV. The first term represents a global weight assigning a constant weight across samples while the second term represents a local weight assigning greater weight on nearby samples. The l-th kernel $k(d_{i,j}; \tilde{w})^{4/2^l}$ has a faster-decay for small l while slower-decay for large l. If $b > 1$, the weight b^l for faster-decay kernels are larger than those for slower-decay kernel, while the opposite is true $b < 1$. Thus, Eq. 2 estimates the decay speed of the kernel by estimating the b parameter.

Unlike $k(d_{i,j};w)$, which is used in the ordinary GWR, $k^*(d_{i,j};a,b)$ is just a linear function with respect to the parameters a and b^l. Thus, a quasi-linear time algorithm, which is explained in [12] is available for the model estimation.

2.3 Spatial prediction using the scalable GWR model

The basic GWR is readily applicable for spatial prediction by assuming a spatial kernel centered on the prediction site. Similarly, once the a and b parameters are estimated by the LOOCV, the regression coefficient at the prediction site s_0 is estimated by a WLS in which the samples are weighted by using the following kernel function:

$$k^*(d_{0,j};a,b) = a + \sum_{l=1}^{L} b^l k(d_{0,j};\tilde{w})^{4/2^l}, \tag{3}$$

where $d_{0,j}$ is the distance between the prediction site and j-th sample site. Eq. 3 assigns large weights on observations nearby the site s_0. Thus, the regression coefficients are estimated to reflect the local property nearby the prediction site. Spatial prediction at site s_0 is performed by substituting the estimated local coefficient $\hat{\beta}_{i,k}$ into the following model:

$$\hat{y}_0 = \sum_{k=1}^{K} x_{0,k}\hat{\beta}_{0,k} \tag{4}$$

Thus, the scalale GWR is easily employed for spatial prediction. Importantly, the spatial interpolation achieves a (quasi-)linear computation cost that is considered as desirable for large-scale spatial predictions in geostatistics. Nevertheless, the scalable GWR has never been applied fro spatial prediction.

3 Application

This section examines the performance of our proposed prediction method through a comparison of prediction accuracy with modern prediction techniques including (conjugate) NNGP and the light gradient boosting machine (LightGBM), which is known as an accurate and computationally efficient gradient boosting algorithm [8]. We also consider the linear regression (LM) for reference.

The Lucas housing data, which consists of the data of 25,357 single family houses sold in Lucas, Ohio in 1993-1998 ($N = 25,357$), is used for the comparison. The conventional GWR is hard to apply because of the computational burden. The explained variable is logged housing price (see Figure 1) and the explanatory variables are total living area in square feet (TLA), garage area in square feet (garagesqft), and building age (AGE).

Prediction accuracy is compared though a 5-fold cross-validation (CV). The root mean squared errors (RMSEs) are as follows: 0.462 (LM); 0.360 (LightGBM); 0.362 (NNGP); 0.309 (Scalable GWR). Surprisingly, scalable GWR has outperforms NNGP and LightGBM, which are modern geostatistical method and machine learning method respectively. Figure 2 plots the observed price in the x-axis and the price predicted during the CV in the y-axis. Based on this figure, the scalable GWR tends to have smaller prediction error relative to alternatives. The result suggests the potential of the scalable GWR as a spatial predictor. From Figure 2, it is also observed that several predicted values of the scalable GWR have large error. Further stabilization might be required to improve the prediction accuracy.

Finally, Figure 3 compares interpolated housing price map. Here, logged price in each 250 m grid covering the study area is predicted. Because explanatory variables are unavailable on the grids, we compare NNGP with constant mean and the scalable GWR with spatially varying intercept. Based on the result, the scalable GWR tends to have smoother prediction result relative to NNGP.

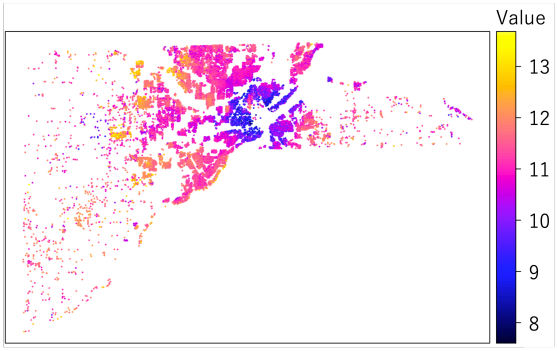


Figure 1 Logged housing price in Lucas county, Ohio.

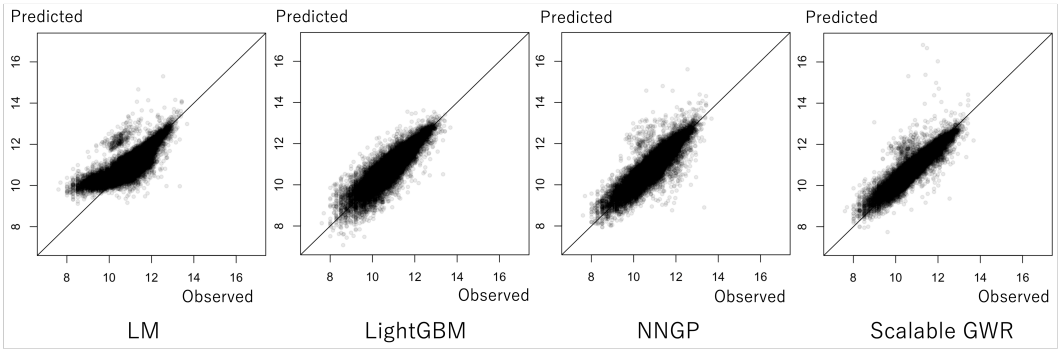


Figure 2 Comparison of observed (x-axis) and predicted housing prices (log-scale; y-axis).

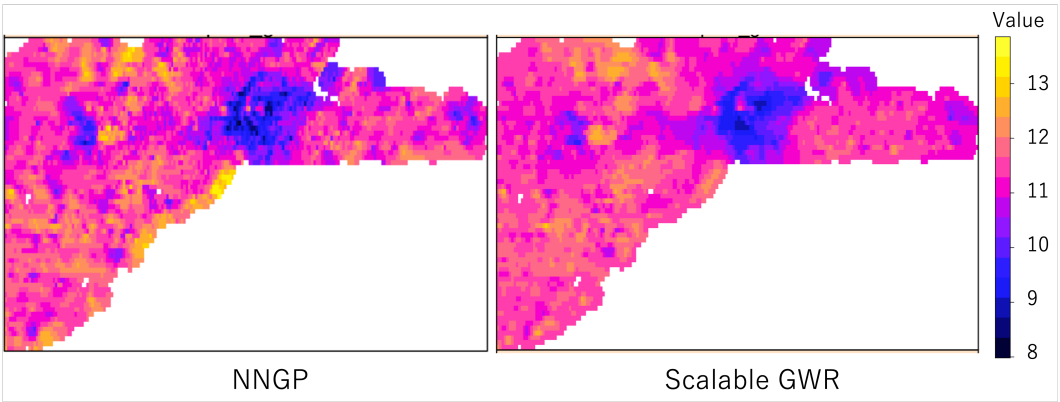


Figure 3 Comparison of observed and predicted housing prices (log-scale). Intercept-only model is used for comparison.

References

1 Chris Brunsdon, A Stewart Fotheringham, and Martin Charlton. Some notes on parametric significance tests for geographically weighted regression. *Journal of regional science*, 39(3):497–524, 1999.

2 Noel Cressie. *Statistics for spatial data*. John Wiley & Sons, 2015.

3 Abhirup Datta, Sudipto Banerjee, Andrew O Finley, and Alan E Gelfand. Hierarchical nearest-neighbor gaussian process models for large geostatistical datasets. *Journal of the American Statistical Association*, 111(514):800–812, 2016.

4 A Stewart Fotheringham, Chris Brunsdon, and Martin Charlton. *Geographically weighted regression: the analysis of spatially varying relationships*. John Wiley & Sons, 2003.

5 Melanie S Hammer, Aaron van Donkelaar, Chi Li, Alexei Lyapustin, Andrew M Sayer, N Christina Hsu, Robert C Levy, Michael J Garay, Olga V Kalashnikova, Ralph A Kahn, et al. Global estimates and long-term trends of fine particulate matter concentrations (1998–2018). *Environmental Science & Technology*, 54(13):7879–7890, 2020.

6 Paul Harris, AS Fotheringham, R Crespo, and Martin Charlton. The use of geographically weighted regression for spatial prediction: an evaluation of models using simulated data sets. *Mathematical Geosciences*, 42(6):657–680, 2010.

7 Matthew J Heaton, Abhirup Datta, Andrew Finley, Reinhard Furrer, Rajarshi Guhaniyogi, Florian Gerber, Robert B Gramacy, Dorit Hammerling, Matthias Katzfuss, Finn Lindgren, et al. Methods for analyzing large spatial data: A review and comparison. *arXiv preprint*, 22, 2017. `arXiv:1710.05013`.

8 Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. Lightgbm: A highly efficient gradient boosting decision tree. *Advances in neural information processing systems*, 30, 2017.

9 Jin Li and Andrew D Heap. A review of comparative studies of spatial interpolation methods in environmental sciences: Performance and impact factors. *Ecological Informatics*, 6(3-4):228–241, 2011.

10 Ziqi Li, A Stewart Fotheringham, Wenwen Li, and Taylor Oshan. Fast geographically weighted regression (fastgwr): a scalable algorithm to investigate spatial process heterogeneity in millions of observations. *International Journal of Geographical Information Science*, 33(1):155–175, 2019.

11 Haitao Liu, Yew-Soon Ong, Xiaobo Shen, and Jianfei Cai. When gaussian process meets big data: A review of scalable gps. *IEEE transactions on neural networks and learning systems*, 31(11):4405–4423, 2020.

12 Daisuke Murakami, Narumasa Tsutsumida, Takahiro Yoshida, Tomoki Nakaya, and Binbin Lu. Scalable gwr: A linear-time algorithm for large-scale geographically weighted regression with polynomial kernels. *Annals of the American Association of Geographers*, 111(2):459–480, 2020.

13 Tomoki Nakaya, Alexander S Fotheringham, Chris Brunsdon, and Martin Charlton. Geographically weighted poisson regression for disease association mapping. *Statistics in medicine*, 24(17):2695–2717, 2005.

Geographically Varying Coefficient Regression: GWR-Exit and GAM-On?

Alexis Comber
University of Leeds, UK

Paul Harris
Rothamsted Research, Harpenden, UK

Daisuke Murakami
Institute of Statistical Mathematics, Tokyo, Japan

Narumasa Tsutsumida
Saitama University, Japan

Chris Brunsdon
Maynooth University, Ireland

Abstract

This paper describes initial work exploring two spatially varying coefficient models: multi-scale GWR and GAM Gaussian Process spline parameterised by observation location. Both approaches accommodate process spatial heterogeneity and both generate outputs that can be mapped indicating the nature of the process heterogeneity. However the nature of the process heterogeneity they each describe are very different. This suggests that the underlying semantics of such models need to be considered in order to refine the specificity of the questions that are asked of data: simply seeking to *understand process spatial heterogeneity* may be too semantically coarse.

2012 ACM Subject Classification Information systems; Theory of computation

Keywords and phrases Geographically weighted regression, Spatial Analysis, Process Spatial Heterogeneity, Model Semantics

Digital Object Identifier 10.4230/LIPIcs.COSIT.2022.13

Category Short Paper

Funding This work was supported by ROIS-DS-JOINT (005RP2021), `https://ds.rois.ac.jp/en/`.

1 Introduction

Geographically varying regression models are those that estimate coefficients locally rather than globally. A key feature of such models is that they accommodate process spatial heterogeneity and support local understandings of how the relationship between different inputs and an outcome vary spatially. Geographically Weighted Regression (GWR) [2] is the best known method to calibrate spatially varying regression models. It uses a moving window (kernel) weighted regression centred on locations in the study area, to estimate local coefficients. The Geographically Weighted (GW) framework has been extended to other statistical methods as well as regression such as GW-PCA [10], GW Discriminant Analysis [3] and GW correspondence matrices [4]. The GWR framework has also been extended to accommodate parameter (predictor variable) specific bandwidths in Multi-Scale GWR (MS-GWR) [18, 8, 13].

Determining the kernel bandwidth (size) in any GW analysis is critical as this defines the variation in the local outputs (i.e. the degree of smoothing). Optimal bandwidths are identified through some measure of model fit. Thus calibrating GWR model bandwidth(s) in this way provides an indicator of the spatial scales over which heterogeneous processes

operate, enhancing process understanding. There are some concerns about the identification of optimal bandwidths for GWR and MS-GWR using current search heuristics, as in some cases optimisation searches may return local rather than global optima [6]. The potential for this is particularly acute when the bandwidth search space is multi-dimensional (MS-GWR). This may be because the increased complexity / dimensionality of the bandwidth search space means that the potential for local optima to be identified by search heuristics is greater.

An alternative approach to calibrating local coefficient models can be constructed using Gaussian Processes (GPs) to model terms in Generalised Additive Models (GAMs) [17, 7]. A GP is a random process over *functions* and GAMs are a general approach to calibrating regression models with unspecified functions of the predictor variables, of the form:

$$y = \alpha + f_1(z_1) + f_2(z_2) + \cdots + f_m(z_m) + \epsilon$$

where z_j may be a vector.

These can be extended so that each $f_j(z_j)$ is a linear regression coefficient on another predictor x_j:

$$y = \alpha(z_0) + x_1 f_1(z_1) + x_2 f_2(z_2) + \cdots + x_m f_m(z_m) + \epsilon$$

Finally, if $z_0 = z_1 = \cdots z_m = z$ say, and z is a vector specifying spatial *locations* then this specifies a *geographically varying regression model*:

$$y = \alpha(z) + x_1 f_1(z) + x_2 f_2(z) + \cdots + x_m f_m(z) + \epsilon$$

One way of specifying $\alpha(z) \cdots f_m(z)$ is that each function is generated from a GP and each function estimate is an *a posteriori* estimate of a GPs with a zero mean. GPs also have a covariance function:

$$\kappa_m(\delta) = \mathrm{Cov}(f_m(z), f_m(z + \delta))$$

These control the "smoothness" of $f_m(z)$ - the more rapidly $\kappa_m(\delta)$ reduces as the magnitude of δ increases, the "smoother" $f_m(z)$ tends to be. These are similar to models based on Kriging as semivariogram functions are related to covariance functions. In a similar way to MS-GWR, the covariance function for each $f_m(z)$ is individually calibrated to optimise model fit. One task of the GAM is to estimate parameters in each $\kappa_j(\delta)$ and so estimate $f_m(z)$.

Thus both MS-GWR and GAM GPs with a GP smooth construct spatially varying coefficient models: both require the degree of smoothing to be determined or specified, with this done through the optimisation of the bandwidth for each predictor variable via a back-fitting operation for MS-GWR, and in a GAM GP spline over geographic space, this is similarly determined through a smoothing parameters for each GP. Both provide a measure of the process heterogeneity specific to each predictor variable in a regression.

The aim of this paper is to explore the complementaries between MS-GWR and GAM GPs specified with observation spatial locations as different approaches for specifying geographically varying regression models in terms of the process understanding (the scale of spatial heterogeneity) they support. It will also reflect on how fit GA / GP and MS-GWR models can be optimised and issues around using them for prediction.

2 Background: GWR and GAM GP

2.1 GWR

Geographically Weighted Regression (GWR) [2] is spatially varying coefficient model, that uses a kernel based approach to create a series of local regression models, for which local coefficient estimates of the predictor variable can extracted and mapped. GWR attempts to calibrate regression models of the form:

$$y_i = \beta_0(u_i, v_i) + \beta_1(u_i, v_i)x_{1i} + \cdots + \beta_m(u_i, v_i)x_{mi} + \epsilon_i$$

where y_i and $\{x_{1i}, \cdots, x_{m_i)}\}$ for $i = 1, \cdots, n$ are a set of observations with m predictor variables and a response, $\{\beta_0(.,.), \cdots \beta_m(.,.)\}$ are functions of two variables providing a mapping from location to a regression coefficient, (u_i, v_i) for $i = 1, \cdots, n$ are locations associated with each of the n observations, and ϵ_i is a random variable, typically from a Normal distribution. The functions $\beta_j(.,.)$ are usually of most interest, and after a model is calibrated these are typically illustrated cartographically. GWR estimates these functions using data subsets falling under a weighted kernel around a point (u, v) to calibrate a local weighted least squares regression. The kernel generally takes a distance decay function such as a Gaussian decay of the form:

$$w_i = \exp\left(-\frac{d_i^2}{2h^2}\right)$$

where d_i is the distance from (u, v) to (u_i, v_i) and h is a quantity termed the "bandwidth" that determines the size of the regression window. Bandwidths can be a fixed distance or a fixed number of nearest data points (i.e., an adaptive radius depending on the local density of points. If h is large, the functions $\beta_j(.,.)$ become smoother and thus determining the size of bandwidth in any GWR analysis is critical as this defines the variation in the local outputs (i.e. the degree of smoothing). Various approaches to finding an "optimal" h for fitting a model to a given data set exist and generally these try to optimise some measure of model fit and parsimony, such as AIC.

A standard GWR operates determines a single bandwidth and thus implicitly assumes that each input variable operates over the same scale with respect to its relationship to the response variable. In reality some relationships may operate over larger scales than others and a standard GWR finds a *best-on-average* scale of relationship non-stationarity [5]. Multi-Scale GWR (MS-GWR) can be used to address this [18, 8, 13]. It determines a bandwidth for each predictor variable plus the intercept individually, thereby allowing individual response-to-predictor variable relationships to vary. Recent thinking has suggested that because of this, MS-GWR should be the default GWR [5], and a standard GWR only used if there is evidence to support a single scale of relationship (and bandwidth).

2.2 Statistical Tensions with GWR

GWR has at its core the idea that "whole map" global statistical models ignore any process heterogeneity and implicitly assume it does not exist. GWR is attractive to geographers and GIScience because it shows how and where processes vary spatially and because it explicitly reflects Tobler's First Law of Geography [15]. This tension between advocates of global models in classic statistics and local models in spatial statistics can be observed in some of the critiques of GWR and other nonstationary models [16, 14]. On one side, the main

critique is that GWR and nonstationary models provide only a collection of local models in order to model a non-stationary process (in this case the coefficients), whereas Bayesian models [9], for example, provide a full single model able to capture a non-stationary process. A further critique is that if the global model has locally clustered outliers (a key indicator of the potential suitability of a GWR model), then some explanatory variables are missing, or the process under investigation has been poorly represented by the model inputs or some theoretical understanding of the process is lacking (and not represented in the predictor variables).

On the other, advocates of GWR and local statistical analyses argue that whole map regression models may unreasonably assume stationary regression coefficients [12] and process heterogeneity (spatial variation in data relationships), whereas in reality the processes *do* vary spatially, for example the relationship between crime and unemployment levels is not the same everywhere. And potentially these arguments are more pertinent to socio-economic processes, which are more likely to be specifically concerned with *how* processes manifest themselves in different socio-economic (local) contexts local, while many environmental or physical processes have fixed global (mathematical) relationships - i.e. laws. An additional related argument is that socio-economic analyses are frequently in the situation where ideal data are never available for actual real-world analysis, as opposed to simulated data commonly used in theoretical statistics. Large swaths of GIScience and geography deal only in secondary data – data that was collected by someone else, usually for a different purpose – and frequently have to use proxies for the data they would really like to use. Very rarely are bespoke data, collected under experimental design and with full understanding, used in geographical analyses [1], especially socio-economic ones. In this sense, the use of local statistical models or global ones, and whether you believe in a global truth or local process understanding, are a bit like a religious belief: it either makes sense (conceptually or practically) or it does not, and no amount of logic will sway opinion.

2.3 GAMs, Bases and Gaussian Processes

A Generalized Additive Model (GAM) uses smooth functions of the predictor variables in which the values of y are assumed to be of an exponential distribution, such as a Gaussian one. If

$$y = f(x) + \epsilon$$

where f is the function being sought in the model, then in GAMs, rather than assuming y to be some linear function of x, a space of functions, or *basis*, is chosen of which f is some element. This allows the basic formula above to be expanded:

$$y = f(x) + \epsilon = \sum_{j=1}^{d} \beta_j(x)\gamma_j + \epsilon$$

where each β_j is a basis function of the transformed x and the γ are the corresponding regression coefficient estimates. One example of a basis is a Gaussian Process basis. If there are n distinct geographical locations in the data set, knowing the locations and the covariance function κ the variance covariance function of the values of β_j in each location can be found, giving a variance covariance matrix R. This can be translated into a set of n basis vectors $\beta_j(x)$ [11], and the GAM can be calibrated in this way. Thus, in contrast to standard linear models, the predictors in a GAM include smooth functions of some or all of the covariates, which allow for non-linear relationships between the predictors and the target variable.

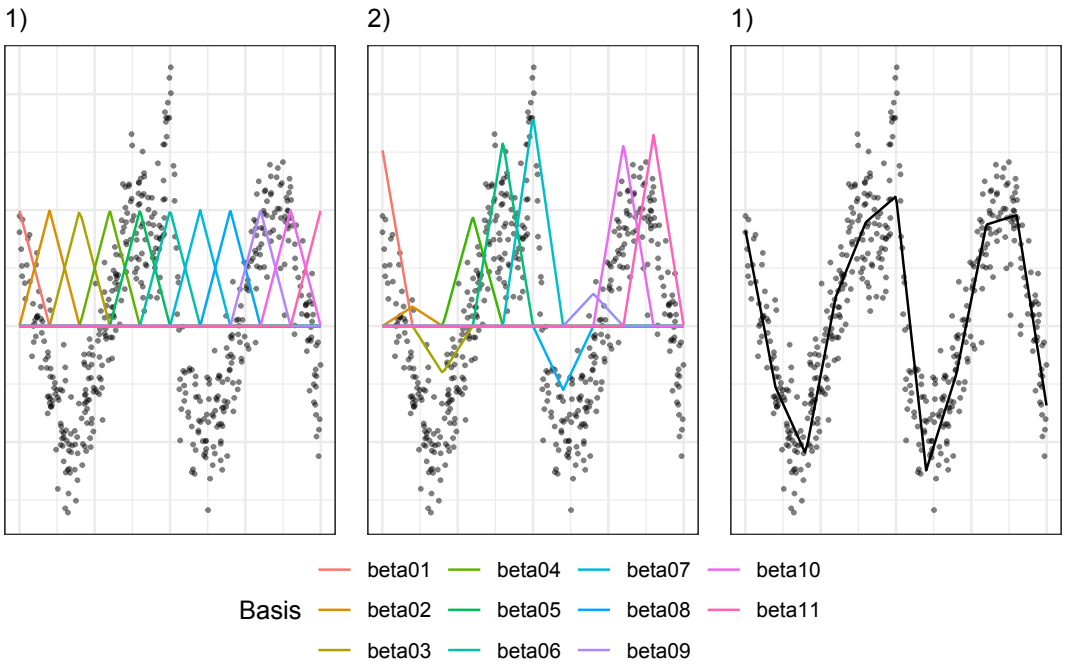


Figure 1 The working of a GAM spline, with simulated x and y data: the LHS plot shows a linear regression fitted between knots, the centre shows each basis multiplied by the corresponding piecemeal linear regression coefficient, and the RHS plot the sum of the basis functions.

GAMs have at their heart the fitting of a series of non-linear functions through the data as illustrated in Figure 1. In this, the various increasing and decreasing functions (centre panel) indicate the slope of coefficients from the data defined by a set of x-points called *knots* (LHS panel). The RHS panel shows the sum of the basis functions, and is equivalent to taking the fitted values from a regression on the basis expansion of x. Is this local fitting that starts to hint at how GAMs can be used with geographic data, where not only are the splines constructed in attribute space but over geographical space if that is included in the inputs to the spline. It suggests that GAMS have the potential to bridge between the need for local, spatial understanding such as is provided by a GWR analysis and for global models, as well as for the enhanced predictive power of non-linear statistical models.

3 Analysis

3.1 Overview and Data

Socio economic data from the `gw` R package is used to illustrate both spatial understanding from GWR and spatial prediction using GAMs. This has census data for the counties in state of Georgia in the USA from the 1990s. It has 159 observations and 6 variables of interest, median income, % of the population that is rural, % with degrees, % elderly, % foreign born and % black. The analyses below construct a MS-GWR model and a GP-derived GAM spline model of Median Income. Both generate local coefficient estimates, which can be mapped.

3.2 MSGWR vs GAM with GP splines

MS-GWR analyses require the explicit identification of the individual bandwidths for each covariate. These provide important information about spatial scales at which the relationship between the predictor variable and the target variables operate. Small bandwidths indicate a local scale and large ones a global scale. It is quite common for these to vary from highly local to highly global within a single MS-GWR analysis.

The MS-GWR model shows the covariates to have a range of adaptive bandwidths (Table 1), and the varying degrees of process heterogeneity are shown by the distributions of the coefficient estimates in Table 1 and by spatial distribution in Figure 2. In this case the bandwidths indicate that the variables for % rural (`PctRural`), % with degrees (`PctBach`) and % elderly (`PctEld`) are highly localised whereas % foreign born (`PctFB`) and % black (`PctBlack`) are global. These are also indicated by the variation in the coefficient estimates for the covariates in Figure 2.

Table 1 The MS-GWR bandwidths (BW) and distribution of the coefficient estimates.

	BW	Min	1stQ	Median	Mean	3rdQ	Max
Intercept	25	27.98	38.85	46.12	45.60	52.45	59.97
PctRural	88	0.04	0.06	0.10	0.10	0.14	0.16
PctBach	26	0.02	0.58	0.95	0.87	1.15	1.82
PctEld	31	−2.46	−1.81	−1.42	−1.48	−1.15	−0.48
PctFB	157	−1.48	−1.46	−1.41	−1.37	−1.30	−1.09
PctBlack	157	−0.24	−0.23	−0.22	−0.22	−0.21	−0.20

In a similar way, Gaussian Process splines can be used in a GAM model, parameterised with observation location, in this case projected Easting and Northing. Each spline was specified with 7 knots in order to ensure sufficient degrees of freedom across the data and the splines. Splines optimise a smoothing parameter which controls the degree of smoothing of the data and as such indicates the locally varying nature of the coefficient. The GPs modelled in the GAM function all have a mean of zero, so for each covariate an extra fixed offset term is added. The fixed terms are shown in Table 2 and the spatially smoothed terms in Table 3. Here it can be seen that of the fixed terms, the Intercept, % with degrees (`PctBach`), % elderly (`PctEld`) and % black (`PctBlack`) are globally significant, while the Intercept, % with degrees (`PctBach`), % elderly (`PctEld`) and % black (`PctBlack`) are locally significant.

It is also possible to map locally significant predictors of Median Income arising from the GAM splines as in Figure 3. The trends in smoothed coefficients broadly show East-West gradients for the Intercept, % with Degree and % Elderly, and North-South ones for % Black.

Table 2 The coefficient estimates of the GAM fixed terms.

	Estimate	Std. Error	t-value	p-value
Intercept	46.398	4.004	11.588	0.000
PctRural	−0.541	0.882	−0.613	0.541
PctBach	0.356	0.179	1.985	0.049
PctEld	−0.482	0.112	−4.313	0.000
PctFB	−0.445	0.294	−1.513	0.133
PctBlack	−0.159	0.023	−6.840	0.000

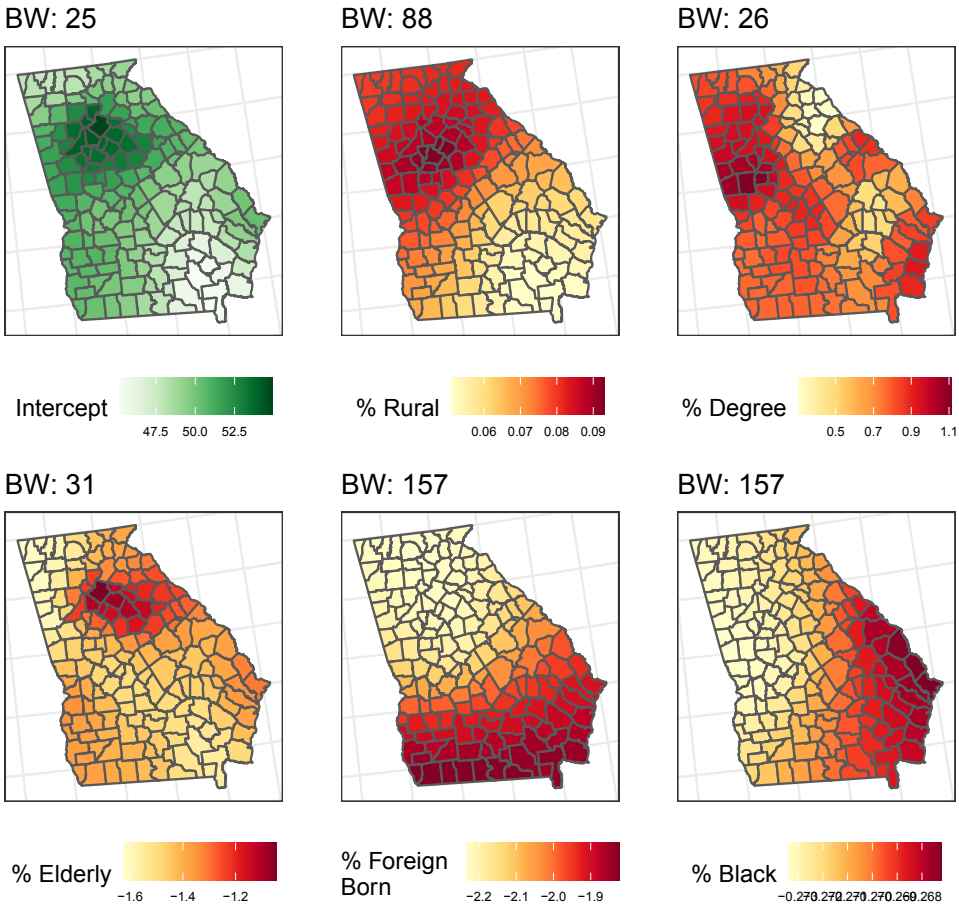


Figure 2 MS-GWR coefficient estimates.

Table 3 The GAM spatially smoothed terms from the GP splines with location.

	Effective df	Ref. df	F	p-value
s(X,Y):Intercept	15.538	18.932	2.093	0.008
s(X,Y):PctRural	5.616	6.044	2.081	0.075
s(X,Y):PctBach	2.517	2.529	2.780	0.031
s(X,Y):PctEld	2.500	2.500	4.812	0.015
s(X,Y):PctFB	2.500	2.500	0.797	0.614
s(X,Y):PctBlack	2.500	2.500	11.675	0.000

3.3 MSGWR bandwidth vs GAM Spline smoothing parameter

The MS-GWR and the smooth terms from the GAM GPs splines constructed with locations, both construct spatially varying coefficient models. They also both include some measure of the degree of local smoothing: in a MS-GWR this is specified through the optimisation of the bandwidth for each predictor variable via a back-fitting operation and in a GAP GP this is indicated through smoothing parameters for each spline. Importantly both methods provide a measure of the process heterogeneity that is specific to each predictor variable. However, it is evident from Figures 2 and 3 that the way that the spatial processes are being modelled by the 2 approaches is very different. The MS-GWR results have distinct but different locales

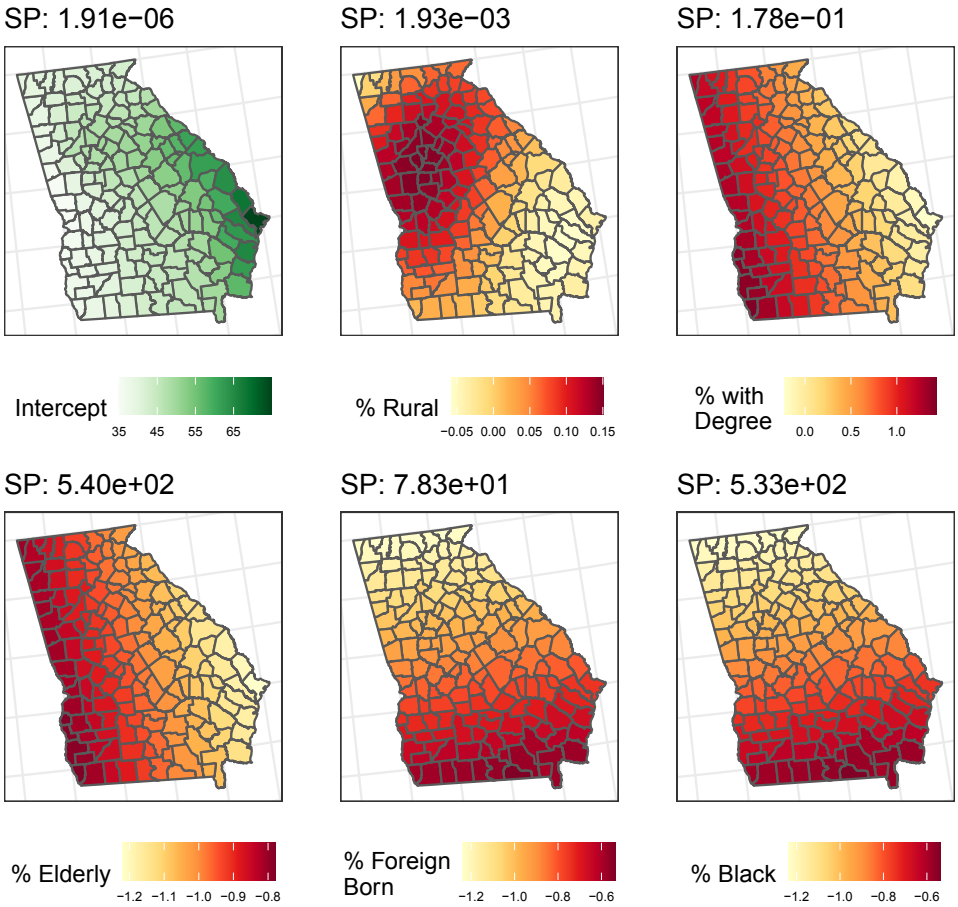
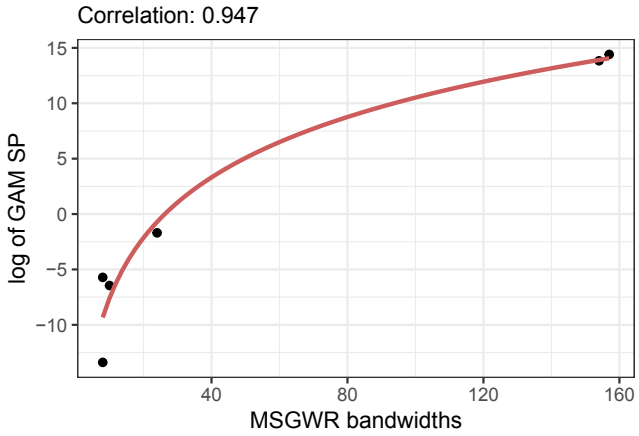


Figure 3 The local coefficient arising from the spatial GAM splines.

of high coefficient estimate value for the different covariates. For example, the association of % with Degree with Median Income is much greater in a region located in the centre of the Western part of Georgia under a MS-GWR, whereas under the GAM it is in the South West corner.

Evidently, despite using the same data, and generating spatially varying coefficient estimates, the 2 approaches are very different. It would be useful to better understand how their results relate, if at all, either to be able to link them, or to be able to infer the circumstances in which each approach may be most useful, under the assumptions that MS-GWR provides an intuitive understanding of the spatial of the relationship between target and individual predictor variables and GAMs probably have a stronger theoretical background.

To investigate how these different ways of capturing and modelling process spatial heterogeneity, relate if at all, the MSGWR bandwidths and GP spline smoothing parameters were examined. The smoothing parameters and the MSGWR coefficients were extracted from the models A correlation showed a high degree of fit for and when they were plotted as in Figure 4, there is a clear trend between them as shown in the tabular values and in the scatter plot. This shows, that at least in this example, there is some form of relationship between MSGWR bandwidth and smoothing parameters of Gaussian Process splines.



Term	MSGWR	GAM
Int	8	1.53e−06
PctBach	8	3.29e−03
PctEld	10	1.58e−03
PctRural	24	1.83e−01
PctBlack	154	1.02e+06
PctFB	157	1.81e+06

Figure 4 MSGWR bandwidths (x-axis) against the log of the GAM GP spline smoothing parameter (SP) (y-axis) with a log trend, and a table of the data included.

4 Final Comments

Spatially vary coefficient models are useful because they explicitly accommodate process spatial heterogeneity, where the relationships between an outcome and factors used to model or predict that outcome, may change with location. The locally varying coefficient estimates that are generated by such approaches provide an explicit representation of process spatial heterogeneity that can be easily communicated via maps. This features has underpinned the popularity of GWR and and other GW models.

In this work, the same process (Median Income in 159 counties of the state of Georgia) is clearly being modelled in different ways by two different spatially varying coefficient models, MS-GWR and GP splines with location in a GAM. The maps in Figures 2 and 3 show these differences, and provide an indication of the different model semantics. It suggests that the concept of "process spatial heterogeneity", which is frequently referred to in the GIScience and spatial analysis literature with a link to Tobler [15] needs to be refined. The results of the two methods used here indicate the need for more nuance in the way that we ask data for "process understanding".

There is evidently a relationship between MSGWR bandwidths and GP spline smoothing parameters. The primary role of GAMs is prediction, but here we have shown that when constructed over geographic space, they can be used to generate locally estimated coefficients. Further work will explore GAM GP splines over other spatial datasets including simulated data with known spatial properties. It may extend the splines to the temporal domain.

References

1 Chris Brunsdon and Alexis Comber. Opening practice: supporting reproducibility and critical spatial data science. *Journal of Geographical Systems*, 23(4):477–496, 2021.

2 Chris Brunsdon, A Stewart Fotheringham, and Martin E Charlton. Geographically weighted regression: a method for exploring spatial nonstationarity. *Geographical Analysis*, 28(4):281–298, 1996.

3 Chris Brunsdon, Stewart Fotheringham, and Martin Charlton. Geographically weighted discriminant analysis. *Geographical Analysis*, 39(4):376–396, 2007.

4 Alexis Comber, Chris Brunsdon, Martin Charlton, and Paul Harris. Geographically weighted correspondence matrices for local error reporting and change analyses: mapping the spatial distribution of errors and change. *Remote Sensing Letters*, 8(3):234–243, 2017.

5 Alexis Comber, Christopher Brunsdon, Martin Charlton, Guanpeng Dong, Richard Harris, Binbin Lu, Yihe Lü, Daisuke Murakami, Tomoki Nakaya, Yunqiang Wang, et al. A route map for successful applications of geographically weighted regression. *Geographical Analysis*, 0:1–24, 2022.

6 Alexis Comber, Nick Malleson, Hang Nguyen Thi Thuy, Thanh Bui Quang, Minh Kieu, Hoang Huu Phe, and Paul Harris. Multiscale geographically weighted discriminant analysis. In *GIScience 2021 Short Paper Proceedings*, pages 17–32. Springer, 2021. doi:10.25436/E2PP4F.

7 Ludwig Fahrmeir, Thomas Kneib, Stefan Lang, and Brian D Marx. Regression models. In *Regression*, pages 23–84. Springer, 2021.

8 A Stewart Fotheringham, Wenbai Yang, and Wei Kang. Multiscale geographically weighted regression (mgwr). *Annals of the American Association of Geographers*, 107(6):1247–1265, 2017.

9 Alan E Gelfand and Dipak K Dey. Bayesian model choice: asymptotics and exact calculations. *Journal of the Royal Statistical Society: Series B (Methodological)*, 56(3):501–514, 1994.

10 Paul Harris, Chris Brunsdon, and Martin Charlton. Geographically weighted principal components analysis. *International Journal of Geographical Information Science*, 25(10):1717–1736, 2011.

11 Trevor J Hefley, Kristin M Broms, Brian M Brost, Frances E Buderman, Shannon L Kay, Henry R Scharf, John R Tipton, Perry J Williams, and Mevin B Hooten. The basis function approach for modeling autocorrelation in ecological data. *Ecology*, 98(3):632–646, 2017.

12 Stan Openshaw. Developing gis-relevant zone-based spatial analysis methods. *Spatial analysis: modelling in a GIS environment*, pages 55–73, 1996.

13 Taylor M Oshan, Ziqi Li, Wei Kang, Levi J Wolf, and A Stewart Fotheringham. mgwr: A python implementation of multiscale geographically weighted regression for investigating process spatial heterogeneity and scale. *ISPRS International Journal of Geo-Information*, 8(6):269, 2019.

14 PD Sampson, D Damian, and P Guttorp. Advances in modeling and inference for environmental processes with nonstationary spatial covariance. In *GeoENV III—Geostatistics for Environmental Applications*, pages 17–32. Springer, 2001.

15 Waldo R Tobler. A computer movie simulating urban growth in the detroit region. *Economic geography*, 46(sup1):234–240, 1970.

16 Levi John Wolf, Taylor M Oshan, and A Stewart Fotheringham. Single and multiscale models of process spatial heterogeneity. *Geographical Analysis*, 50(3):223–246, 2018.

17 Simon N Wood. *Generalized additive models: an introduction with R*. chapman and hall/CRC, 2006.

18 Wenbai Yang. *An extension of geographically weighted regression with flexible bandwidths*. PhD thesis, University of St Andrews, 2014.

3D Sketch Maps: Concept, Potential Benefits, and Challenges

Kevin Gonyop Kim
Institute of Cartography and Geoinformation,
ETH Zürich, Switzerland

Jakub Krukar
Institute for Geoinformatics,
Universität Münster, Germany

Panagiotis Mavros
Future Cities Laboratory,
Singapore ETH Center, Singapore

Jiayan Zhao
Chair of Cognitive Science,
ETH Zürich, Switzerland

Peter Kiefer
Institute of Cartography and Geoinformation,
ETH Zürich, Switzerland

Angela Schwering
Institute for Geoinformatics,
Universität Münster, Germany

Christoph Hölscher
Chair of Cognitive Science,
ETH Zürich, Switzerland

Martin Raubal
Institute of Cartography and Geoinformation,
ETH Zürich, Switzerland

Abstract

Studying the 3D aspect of spatial information has become increasingly important due to changes in the way we interact with the surrounding environments as well as technological innovations. Current pen-and-paper approaches of sketch mapping have a limitation in investigating 3D spatial knowledge as they are forced to be drawn on 2D interfaces. In this paper, we propose the concept of 3D sketch mapping as a tool to study human spatial knowledge by externalizing the mental models of spatial information with 3D representations. The goal of this paper is to introduce the concept, discuss its potential importance and challenges, and share our vision for future research directions.

2012 ACM Subject Classification Human-centered computing → Visualization theory, concepts and paradigms

Keywords and phrases Sketch maps, mental representations, spatial knowledge

Digital Object Identifier 10.4230/LIPIcs.COSIT.2022.14

Category Short Paper

Funding This work has been supported by Swiss National Science Foundation (Sinergia 202284, "3D Sketch Maps") and German Research Foundation (SCHW1372/7-3, "Sketchmapia").

1 Introduction

The spaces we live in are naturally three-dimensional. Although many aspects of spatial cognition [10], such as navigation, have been predominantly studied with a focus on the horizontal plane, there are situations in which the vertical dimension is equally important. The 3D aspect of spatial information is an increasingly important issue and it is due to the changes in the environment surrounding us (e.g., more complex buildings) as well as technological advancement (e.g., 3D virtual-reality simulations or flying a drone). Thus, understanding how people perceive, navigate, and interact with spaces while accounting for the 3D aspect has become essential for various applications such as training pilots or wayfinding in multi-level buildings.

One of the popular methods to study human spatial knowledge is the use of sketch maps. Sketch maps are drawings of spatial environments typically made based on spatial memories of the person and they have been a popular research tool to study human spatial understanding and decision making [13]. However, one of the limitations regarding the use of

sketch maps in contemporary research is that they are forced to be drawn on 2D interfaces, often on a piece of paper. Studying the understanding of the 3D aspect of spatial knowledge with 2D sketch maps is challenging as drawing mental models that potentially contain 3D information on a 2D interface requires a set of mental transformations that can be highly prone to distortion and errors [14].

In this paper, we introduce the concept of sketch mapping in 3D. A variety of recent technological developments from desktop-based 3D drawing software to extended reality (XR) devices opens new possibilities to implement novel digital interfaces for 3D sketch maps. We argue that 3D sketch maps can enable studying people's understanding of 3D spatial information directly with 3D interfaces, which may seem intuitively plausible, but it can also introduce potential complexity to a simple drawing process. The aim of this paper is to propose the concept of 3D sketch mapping, with a particular focus on XR technologies as a means of implementing it, and to discuss potential research topics and challenges while comparing it with the conventional 2D method.

2 Background

2.1 Mental models of spatial information

A mental model of spaces is a mental representation of the relative locations and attributes of phenomena in spatial environments [5]. For example, it is required for a pilot to construct a mental model of a flight path together with the relative locations of weather-related spatial information in order to perform the task of flying an aircraft. Similarly, a nurse navigating between wards in a large hospital, a shopper visiting shops in a complex mall, and a commuter going through a large underground interchange, all have in common a need for a mental model of their environments.

Understanding and assessing mental models of spatial information has been an important research topic in the field of cognitive science. Especially for navigation and wayfinding domains, the majority of studies have been based on spatial objects and associated relations lying on a horizontal surface. With respect to 3D, researchers have been interested in the unique characteristics of mental representations along the vertical axis compared to the two horizontal dimensions. For example, previous research has revealed an anisotropy of vertical and horizontal spaces – people are better at navigating horizontally (e.g., within floors) than vertically (e.g., between floors) [8].

One major barrier to probing people's mental models of spatial information for the tasks in which the vertical dimension is important is the lack of appropriate tools to externalize such models. Traditional media only allow the depiction of spatial information on a 2D screen or paper. In response to this barrier, researchers have explored different methods to examine the vertical aspect of mental models such as structure mapping, 3D pointing, and object recognition [8, 4]. Despite these efforts in exploring the 3D aspect of spatial knowledge, it is still largely undiscovered what makes the difference in human cognitive processes and how people perceive and understand 3D spatial information, especially with regard to navigation. Moreover, many previous studies utilized the 2D-based approach to assessing spatial knowledge for the environments and tasks where the vertical dimension is important. This is largely due to the absence of a standardized method for studying spatial understanding of the vertical component of spatial knowledge.

2.2 Sketch maps: A tool to externalize mental models

Sketch maps have been an established tool to externalize mental models and assess spatial knowledge of an environment [2]. They have been widely used in the fields that study human spatial decision-making and information processing, such as navigation and wayfinding [12]. Sketch maps are particularly well suited to extracting survey knowledge of environments because they require combining multiple spatial relations in a single sketch. This can be a good indicator of overall spatial knowledge of larger environments [11]. Sketch maps are routinely distorted in an inconsistent way [14] but not all information in sketch maps is distorted inconsistently – some qualitative spatial relations remain invariant [15]. This is one reason why the key challenge in analyzing sketch maps is to decide on what information to extract from them [11]. The main approaches focused on analyzing sketch maps' type, their metric accuracy, the correctness of qualitative relations represented on them, and the level of generalization [13].

One of the limitations of using sketch maps is that it enforces the use of a 2D sketch for representing spatial knowledge of an environment. For example, imagine a person trying to create a 2D sketch map of a complex university building. They can try to either project all the information on a plane (i.e., flatten the vertical information), create separate sketches for each floor, or draw from an isometric point of view. These options, however, make them choose to represent only a subset of information and this is against the main advantage of sketch maps which is the ability to represent all information at once. Humans can maintain and utilize 3D spatial information, however, it is unclear how they could externalize this information in an accessible and intuitive way.

3 Sketch mapping in 3D

The limitation in the current practice of using sketch maps for 3D spatial information motivated the conceptual development of 3D sketch maps. Our hypothesis is that it will be more effective to study 3D spatial knowledge with 3D sketches as this requires less mental projection compared to 2D ones. Although the idea of representing 3D spatial knowledge with 3D interfaces may seem natural, it is necessary to carefully consider the consequences of adding one more dimension. The main research question that needs to be addressed in the study of 3D sketch maps is whether it can be an adequate method for extracting mental models of 3D spatial information from individuals. Answering this research question will require investigating 3D sketch mapping from different perspectives: (1) *conceptual* and *methodological* perspective that looks into the validity of the concept and the methodology, (2) *cognitive* perspective for understanding underlying cognitive processes, and (3) *technological* perspective on how to design an interface for 3D sketch mapping, especially using XR technologies. As shown in Table 1, we have developed research questions for each aspect which will be discussed in the following.

3.1 Conceptual and methodological perspective

The validity and effectiveness of 3D sketch mapping can vary depending on its use cases. While the main goal of 3D sketch mapping is to assess the knowledge of 3D spatial information, there exist differences among the use cases in different domains of spatial cognition [10] in terms of the purpose of using sketch mapping, their requirements, and user characteristics. For example, one use case of 3D sketch mapping can be in the field of aviation where the pilots' spatial knowledge of flight routes and weather situations needs to be assessed.

Table 1 Research questions to be addressed for 3D sketch maps.

	Research questions
Conceptual and methodological perspective	- For which use cases do 3D sketch maps allow the creation of more valid and reliable sketch maps compared to the 2D approach? - How should 3D sketch maps be analyzed?
Cognitive perspective	- What are the cognitive processes underlying the creation of 3D sketches? - How do 3D sketch maps differ from 2D ones with regard to their conveyed information? - What are the behavioral features suitable for analyzing the cognitive processes involved in 3D sketch mapping?
Technological perspective	- Where on the reality-virtuality continuum are tools that optimally enable 3D sketch mapping? - Can XR-based 3D sketch mapping tools be more valid and reliable than the 2D pen-and-paper method? If so, in which aspect?

Another use case is to measure the accuracy of mental representations of laypeople for the task of navigating buildings. In addition, some groups, such as architects, might be much more used to encoding and storing vertical aspects of spatial knowledge; a tool that makes 3D sketch maps easy to draw will allow us to study these mental processes, and their different development between professionals and laypeople. Understanding the similarities and differences among these cases will help define 3D sketch mapping as a tool.

Another important aspect from a methodological perspective is how to analyze 3D sketch maps. Interpretation of 2D sketch maps has been a topic of interest and there exist methods to evaluate them both qualitatively and quantitatively, and classify them based on their type [7]. For 3D sketch maps, it is still a question whether these methods that have been developed for 2D sketch maps are also valid. We can expect differences between the 2D and 3D sketch maps in terms of the types of spatial knowledge that are likely to be expressed in them and also how they are expressed. For example, vertical information might be distorted in a way different from horizontal information and these distortions might weigh differently on the overall quality of the sketch in the given context.

3.2 Cognitive perspective

From a cognitive perspective, the creation of 3D sketch maps can have different underlying cognitive processes compared to 2D ones. Cognitive processes may differentiate themselves from understanding to transforming onto a sketch in 3D compared to 2D. Creating 3D sketch maps might require less cognitive load regarding mental projection, but the added dimension for sketching may naturally increase the cognitive demand. Moreover, there will be additional extrinsic cognitive load related to the interaction with the technology used for enabling 3D drawing, compared to the simple pen-and-paper interface. The pen-and-paper interface is a medium that we are extremely familiar with and such familiarity may facilitate the mental projection process required in sketch mapping, whereas this is not the case with the 3D interaction tools that are currently unfamiliar to most people.

3D sketch maps can have different characteristics in terms of conveyed information compared to 2D ones. Given the additional dimensionality, people may show different behavior regarding the type of information that they tend to express in 3D. The potential difference in conveyed information will directly influence the way they are analyzed. All methods that are currently available for interpreting sketch maps focus only on a subset of all information contained in a sketch map. It is currently unknown how this subset of

information differs in 3D sketch maps, e.g., what are the new types of information that are commonly communicated by users when they are asked to represent a space in 3D. To answer this question, an experimental investigation will be necessary that compares 2D and 3D sketch maps in terms of the types of spatial information conveyed in them.

In the research studies of 3D sketch maps, the behavior analysis of how people create them will help understand the underlying cognitive processes. The behavior of users is closely related to the technological implementation of the interface since the way people interact with a tool is often shaped by the interface provided to them and the tool's capabilities. For sketch mapping, analyzing behavior measures such as hand gestures, sequences of lines drawn, or pause/hesitation patterns can help understand the underlying cognitive processes. One particular measure that can provide us with rich information would be eye-tracking data [6] as sketch mapping heavily involves visual attention and perspective changes. Although the use of gaze information is sometimes considered limited and challenging as an input modality in 3D interaction [3], it can be an interesting source of information for analyzing behavior and understanding cognitive processes.

3.3 Technological perspective

One of the reasons for utilizing the 2D approach of sketch mapping even when studying the 3D aspects of spatial knowledge is the lack of technologies that enable 3D sketching. The recent advancement of XR technologies allows us to design an interface that allows 3D inputs from users. For example, there already exist non-professional tools for 3D drawing using XR such as Google Tilt Brush[1] and they enable novel interactions between users and 3D information. As mentioned earlier, we mainly consider XR technologies as a means of enabling 3D sketch mapping in this paper.

The wide spectrum of XR technologies allows for designing different interfaces for sketching. In Figure 1, we show some possible interfaces for 3D sketch mapping using XR together with existing methods where the horizontal axis is the reality-virtuality continuum [9] and the vertical axis is the dimension of the externalized output. The area that corresponds to the research gap this paper aims to address is shown with a rectangle in the figure. To provide an example, a VR-based interface can enable 3D sketch mapping, either with a desktop or with a head-mounted display (HMD). Another possibility is to use mixed reality (MR) to combine a physical object from reality with additional digital information augmented on it. An MR-based interface can augment the vertical dimension on top of the traditional paper interface and allow users to enter and modify 3D information. With different options for implementing the interfaces for 3D sketch mapping, it requires careful consideration in designing and evaluating these technical configurations as they often define the boundaries of what is possible using the interface and affect the behavior of the users. Comparing the positions along the horizontal axis in the figure corresponds to the question of how to realize the 3D representations. This is the opportunity that technological advancements provide for us to explore. Moving vertically from 2D to 3D in the figure corresponds to the question of when it is beneficial to use 3D representations. And this relates back to the importance of considering the purpose and the use case of sketch mapping as discussed earlier.

As 3D sketch mapping is a newly proposed concept for studying human spatial knowledge, it is important to confirm its validity and reliability from a technological point of view. We need to study whether users can create more valid/correct and reliable representations of

[1] Tilt Brush by Google. `https://www.tiltbrush.com`

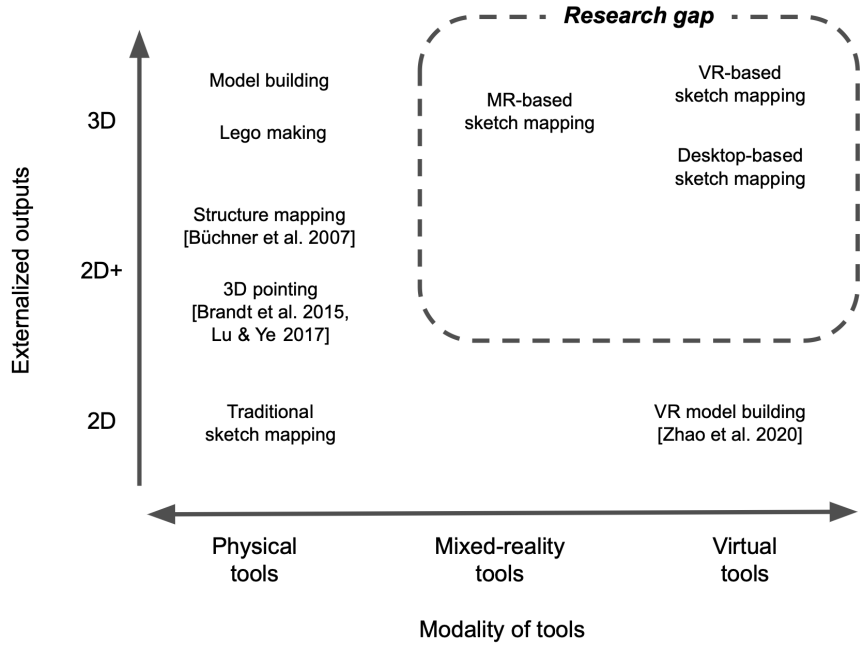


Figure 1 Tools for externalizing spatial knowledge. The x-axis represents the modality of externalizing a mental representation of space, from physical to virtual tools. The y-axis shows different types of produced outputs based on their dimensions – this does not refer to the spatial information that is encoded but to how it is externalized.

their spatial knowledge of environments using an XR-based 3D interface. Previous research shows evidence that free 3D drawing in the air is often less accurate than drawing on 2D surfaces [1]. It can be a concern because accuracy, or quantitative correctness, is one of the core criteria for evaluating sketch maps, and thus it requires empirical validations while considering the technological design of the interface that can improve the accuracy. On the other hand, from a user's point of view, there can be differences between 2D and 3D sketching in terms of perceived workload. It is important to study whether the perceived workload is higher in the 3D case, how the physical and temporal demands are related to the richness of the representation, and how it affects the user experience. Another interesting topic in this direction of research that connects the technological and cognitive aspects is to study the influence of users' spatial abilities on the validity and reliability of the 3D sketch maps that they create.

4 Concluding remarks

In this paper, we introduce the concept of 3D sketch maps while considering its potential benefits and challenges. As the importance of studying human understanding of 3D spatial knowledge has increased recently, we envision that 3D sketch maps can become an effective tool to assess it. It is an exciting time to conduct research in this direction as the recent advances in XR technologies make it possible to implement the interfaces for creating sketches in 3D. There can be different configurations of technologies to design an interface for 3D sketch maps and it is important to compare them both theoretically and empirically in order to make it an effective tool. As discussed in this paper, 3D sketch maps can potentially solve some of the limitations of using the traditional 2D approach, but there exist challenges such

as increased cognitive load and reduced accuracy that need to be addressed in future studies. In conclusion, the idea of sketch mapping in 3D proposes interesting directions for spatial sciences research and invites researchers to explore them.

References

1 Rahul Arora, Rubaiat Habib Kazi, Fraser Anderson, Tovi Grossman, Karan Singh, and George Fitzmaurice. Experimental evaluation of sketching on surfaces in vr. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 5643–5654, New York, NY, USA, 2017. Association for Computing Machinery.

2 Mark Billinghurst and Suzanne Weghorst. The use of sketch maps to measure cognitive maps of virtual environments. In *Proceedings Virtual Reality Annual International Symposium'95*, pages 40–47. IEEE, 1995.

3 Arzu Çöltekin, Ian Lochhead, Marguerite Madden, Sidonie Christophe, Alexandre Devaux, Christopher Pettit, Oliver Lock, Shashwat Shukla, Lukáš Herman, Zdeněk Stachoň, et al. Extended reality in spatial sciences: A review of research challenges and future directions. *ISPRS International Journal of Geo-Information*, 9(7):439, 2020.

4 Laurent Dollé, Jacques Droulez, Daniel Bennequin, Alain Berthoz, and Guillaume Thibault. How the learning path and the very structure of a multifloored environment influence human spatial memory. *Advances in Cognitive Psychology*, 11(4):156, 2015.

5 Roger M Downs and David Stea. Cognitive maps and spatial behaviour: Process and products, the map reader: Theories of mapping practice and cartographic representation, 2011.

6 Peter Kiefer, Ioannis Giannopoulos, Martin Raubal, and Andrew Duchowski. Eye tracking for spatial research: Cognition, computation, challenges. *Spatial Cognition & Computation*, 17(1-2):1–19, 2017.

7 Jakub Krukar, Stefan Münzer, Lucas Lörch, Vanessa Joy Anacta, Stefan Fuest, and Angela Schwering. Distinguishing sketch map types: A flexible feature-based classification. In *German Conference on Spatial Cognition*, pages 279–292. Springer, 2018.

8 Panos Mavros, Michael van Eggermond, and Christoph Hoelscher. Human navigation in a multilevel travelling salesperson problem. *PsyArXiv*, 2022.

9 Paul Milgram and Fumio Kishino. A taxonomy of mixed reality visual displays. *IEICE TRANSACTIONS on Information and Systems*, 77(12):1321–1329, 1994.

10 Daniel Montello and Martin Raubal. Functions and applications of spatial cognition. In *Handbook of spatial cognition*, pages 249–264. American Psychological Association, 2013.

11 Daniel R Montello. Behavioral methods for spatial cognition research. *Research methods for environmental psychology*, pages 161–181, 2016.

12 Daniel R Montello and Corina Sas. Human factors of wayfinding in navigation, 2006.

13 Angela Schwering, Jia Wang, Malumbo Chipofya, Sahib Jan, Rui Li, and Klaus Broelemann. Sketchmapia: Qualitative representations for the alignment of sketch and metric maps. *Spatial cognition & computation*, 14(3):220–254, 2014.

14 Barbara Tversky. Structures of mental spaces: How people think about space. *Environment and behavior*, 35(1):66–80, 2003.

15 Jia Wang and Angela Schwering. Invariant spatial information in sketch maps – A study of survey sketch maps of urban areas. *Journal of spatial information science*, 11:31–52, 2015.

16 Jiayan Zhao, Tesalee Sensibaugh, Bobby Bodenheimer, Timothy P McNamara, Alina Nazareth, Nora Newcombe, Meredith Minear, and Alexander Klippel. Desktop versus immersive virtual environments: effects on spatial learning. *Spatial Cognition & Computation*, 20(4):328–363, 2020.

The Effect of Abstract vs. Realistic 3D Visualization on Landmark and Route Knowledge Acquisition

Armand Kapaj[1]
Department of Geography, University of Zurich, Switzerland
Digital Society Initiative, University of Zurich, Switzerland

Enru Lin
Department of Geography, University of Zurich, Switzerland
Digital Society Initiative, University of Zurich, Switzerland

Sara Lanini-Maggi
Department of Geography, University of Zurich, Switzerland
Digital Society Initiative, University of Zurich, Switzerland

Abstract

Even though humans perform it daily, navigation is a cognitively challenging process. Landmarks have been shown to facilitate navigation by scaffolding humans' mental representation of space. However, how landmarks can be effectively communicated to pedestrians to support spatial learning of the traversed environment remains an open question. Therefore, we assessed how the visualization of landmarks on a mobile map (i.e., abstract 3D vs. realistic 3D symbols) influences participants' spatial learning, visual attention allocation, and cognitive load during an outdoor map-assisted navigation task. We report initial results on how exposing pedestrians to different landmark visualization styles on mobile maps while navigating along a given route in an urban environment can have differing effects on how they remember landmarks and routes. Specifically, we find that navigators better remember landmarks visualized as 3D realistic-looking symbols compared to 3D abstract landmark symbols on the mobile map. The pattern of results shows that displaying realistic 3D landmark symbols at intersections potentially helps participants to remember route directions better than with landmarks depicted as abstract 3D symbols. The presented methodological approach contributes ecologically valid insights to further understand how the design of landmarks on mobile maps could support wayfinders' spatial learning during map-assisted navigation.

2012 ACM Subject Classification Information systems → Geographic information systems; Applied computing → Cartography; Human-centered computing → Empirical studies in visualization

Keywords and phrases Abstraction, realism, 3D, landmark visualization, mobile map design, cartography, real-world navigation, spatial learning

Digital Object Identifier 10.4230/LIPIcs.COSIT.2022.15

Category Short Paper

Supplementary Material *Collection (Paper Analyses and Data)*: `https://gitlab.uzh.ch/giva/geovisense/cosit2022_analyses.git`
archived at `swh:1:dir:7f87d08639ce1ff4c00bba49173b5c07500d42f9`

Funding This work was supported by the H2020 European Research Council (ERC) Advanced Grant GeoViSense (no. 740426).

Acknowledgements The authors thank Bingjie Cheng, Ian T. Ruginski for their comments on experimental design, and Sara I. Fabrikant for her supervision on this research project.

[1] Winterthurstrasse 190, 8057 Zurich, Switzerland

1 Introduction

Imagine you are in a new town and are meeting new friends at a café in the historic downtown area. As you live in a spatially enabled society, you most likely rely on your trusted mobile navigation service to get you there [4]. Once you approach the narrow medieval streets of the downtown area, you realize it is getting harder for you to make destination-relevant navigation decisions. You are standing at an intersection and wondering which way to go. You rotate your navigation device and look around to guess the correct direction. You notice historic buildings with unique facades, but these are only indicated as plain gray footprints on your mobile map. You wished you saw some of these beautiful building facades on your mobile map to facilitate your wayfinding, especially at street intersections.

Unique buildings serve as landmarks during navigation and facilitate the challenging wayfinding process [1]. Landmarks are geographic features that help structure humans' mental representation of space, and when placed on a mobile map, they support the matching of information on the mobile map display with what is seen in the environment [12]. Hence, landmarks play a key role in humans' everyday spatial mobility [11]. Despite their acknowledged importance for navigation, landmarks are not communicated effectively on most mobile wayfinding services. They are either omitted entirely and shown only as building footprints or substituted with commercial points of interest [9]. Aside from GPS reliance for localization, the omission of relevant landmarks on GPS-enabled maps might also explain navigators' deteriorated spatial learning [4]. This is because GPS reliance increasingly shifts navigators' visual attention to the mobile map display [3, 6] without offering additional navigation-relevant information.

Past research has shown that landmarks can be depicted on a visual abstraction (geometric symbol vs. photorealistic symbol) and dimensionality continuum (2D vs. 3D symbols) on mobile maps, ranging from abstract text labels to realistic 3D building models with different levels of realism [2, 10]. However, how landmarks can be communicated on mobile map services for pedestrian navigation to help shift users' attention away from the map and back to the environment, thus improving spatial learning, has not been explored systematically.

Our research program aims to identify cartographic design solutions for landmark visualization by including first-person perspective viewing of 3D landmarks on planar mobile maps, with the goal of increasing the visual saliency of relevant landmarks while inhibiting irrelevant features in mobile maps [7]. We propose that a closer visual match between information on the mobile map display and the physical environment would increase wayfinders' engagement with their surroundings. Consequently, navigators' spatial learning would improve without increasing cognitive load. In this article, we are particularly interested in assessing how landmark visualization style (3D abstract vs. 3D realistic symbols) might influence participants' landmark and route knowledge acquisition. We hypothesize that participants can recall landmarks and route directions better when navigating with realistic 3D landmark symbols on the mobile map than with abstract 3D landmarks. Next, we detail our empirical approach that emphasizes ecological validity while assessing wayfinders' spatial learning in situ.

2 Methodology

We conducted an outdoor navigation experiment in a residential area in Zurich, Switzerland, on days without precipitation, in December 2021. We used a within-subject experimental design with landmark visualization style as the independent variable and landmark and route knowledge as the dependent variables. We designed a mobile map application containing the

route depicted as a black line and ten landmarks (five depicted as realistic 3D symbols and five as abstract 3D symbols) on a 2D basemap (Figure 1). We counterbalanced the order of the landmark symbol style, starting with landmarks visualized as realistic 3D symbols (Figure 1.A) and then as abstract 3D symbols (Figure 1.B) or vice versa. Prior to developing the mobile map, we conducted a pilot survey ($N = 9$) to select prominent buildings along the route that would serve as landmark candidates. Participants in the survey rated the most prominent buildings and the buildings they would use to give directions [9].

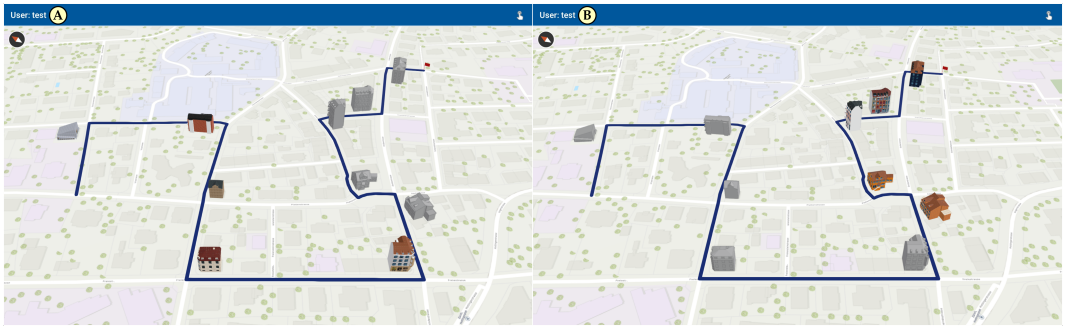


Figure 1 Mobile map stimuli with counterbalanced order of landmark visualization, either with realistic 3D (A) or abstract 3D landmark symbols (B) shown first.

2.1 Participants and procedure

Forty-six (22 females; avg. age = 27.5 yrs., range = 21–46 yrs.) healthy adults with normal or corrected-to-normal vision (i.e., contact lenses only) participated in the study. The experimental task consisted of following a given route for approximately 1 km with the help of a mobile map (Figure 1) in an urban environment (77% of participants were unfamiliar with the study area). Participants were asked to identify the ten visualized landmarks on the mobile map in the environment by raising a hand while passing the given landmarks on the predefined route on their way to the set destination. We followed participants at a safe distance and corrected their direction if a wrong navigation decision was made. To provide a naturalistic navigation experience, we asked participants to walk as they would when exploring a new environment. Additionally, they could interact (i.e., zoom, pan, rotate, and tilt) with the mobile map application as desired. To control for intentionality in learning, we informed participants that they would be tested on their newly acquired environmental knowledge at the end of the navigation task. After arriving at the destination, participants completed a questionnaire (see subsection 2.2) designed to assess their spatial learning.

2.2 Post-navigation questionnaires

The landmark and route questionnaire presented 30 images in random order, with ten images each for 1) relevant, 2) irrelevant, and 3) novel landmarks as seen from participants' perspective during navigation. Relevant landmarks (REL) refer to landmarks that participants saw along the route where a navigation decision was required and which were presented as either realistic or abstract 3D symbols on the mobile map (Figure 1). Irrelevant landmarks (IRL) refer to buildings located along the route but not depicted on the mobile map and where no navigation decision was required (i.e., participants continued straight when passing an IRL). Novel landmarks (NOL) were buildings similar in style to the buildings in the study area but that were neither along the route nor visualized on the map. We asked participants

if they remembered seeing a given building during navigation for each of the 30 building images. If they answered yes, we asked participants whether they saw the building visualized on the mobile map **and** in the environment or **only** in the environment to distinguish between REL and IRL landmarks. If participants classified the building as a REL landmark, they were asked to provide the navigation direction they took after seeing it.

3 Data analyses and results

To assess the effect of landmark visualization style on wayfinders' acquired landmark knowledge, we employed analyses according to signal detection theory (SDT) using the *psycho* package in R (v.4.1.2).

3.1 Data analyses

SDT distinguishes response accuracy (i.e., stimulus, signal, or target) from background noise or distractors [13]. SDT encodes participants' correct responses as "Hit" or "Correct Rejection", and false responses are recorded as "Miss" or "False Alarm" [13]. To statistically compare landmark recall accuracy, we used the discriminability index d' (d prime), computed as *Z(Hit) − Z(False Alarm)*. Additionally, we used criterion location (c) computed as *−0.5 * [Z(Hit) + Z(False Alarm)]* to assess participants' response bias against zero (no bias). We ran SDT analyses in two ways. Firstly, we analyzed whether participants remembered seeing landmarks that could be seen in the environment during navigation (i.e., REL and IRL) or were not present at all (i.e., NOL). Secondly, we analyzed whether participants remembered seeing landmarks visualized on the mobile map display and in the environment (REL) or only in the environment (IRL).

We additionally ran multilevel regression analyses to compare landmark and route knowledge recall performance by landmark visualization condition. Multilevel models were executed using the *lme4* package, with the significance threshold set at $p < .05$. Analyses outcomes are detailed next.

3.2 Results

3.2.1 Results on recall of landmarks seen in the environment

First, we turn to the landmark recall analyses by visualization condition using SDT (Figure 2).

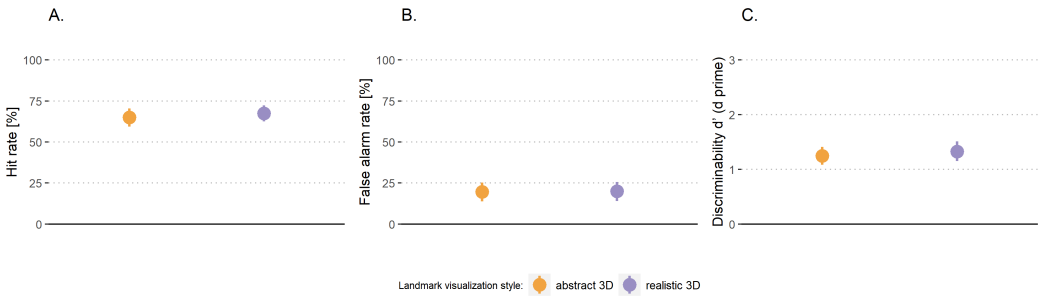


Figure 2 Participants' hit (A) and false alarm rates (B), and (C) recall performance of landmarks seen in the environment by visualization style (dots indicate means; bars indicate 95% CIs).

The average hit rate of approximately 75% together with the average false alarm rate of approximately 25% equates to a moderate discriminability between signal and noise, with a d' score of 1.23 ($SD = 0.68$) for 3D realistic landmarks compared to a d' score of 1.19 ($SD = 0.61$) for abstract 3D landmarks. This difference is not statistically significant ($p = 0.74$, $r = 0.06$). Additionally, we find that participants' response bias is significantly higher than zero for both conditions (abstract 3D: $M = 0.19$, $t(45) = 2.8$, $p = 0.007$; realistic 3D: $M = 0.17$, $t(45) = 2.7$, $p = 0.009$), indicating that participants were more likely to respond that they did not see the landmark in the environment. The difference in response bias is not significant between conditions ($t(45) = 0.29$, $p = 0.77$).

3.2.2 Results on recall of landmarks seen on the mobile map

Figure 3 illustrates the results on participants' recall of landmarks seen in the mobile map. Specifically, participants had both higher hit rates ($M = 74.3\%$, $SD = 23.7\%$) and lower false alarm rates ($M = 5.7\%$, $SD = 10.8\%$) for realistic 3D landmarks compared to abstract 3D landmarks (hit rate: $M = 66.5\%$, $SD = 25.3\%$, and false alarm rate: $M = 8.3\%$, $SD = 15.5\%$). In line with our hypothesis, the d' prime analysis reveals significantly better recall ($p = 0.03$, $r = 0.4$) for the realistic 3D landmarks ($M = 1.81$, $SD = 0.74$) compared to the abstract 3D landmarks ($M = 1.54$, $SD = 0.79$). Results revealed that participants' response bias is significantly higher than zero for both conditions (abstract 3D: $M = 0.35$, $t(45) = 6.02$, $p < 0.001$; realistic 3D: $M = 0.29$, $t(45) = 5.3$, $p < 0.001$), indicating that participants were more likely to respond that they did not see the landmark on the mobile map display. The difference for response bias between conditions is not significant ($t(45) = 0.93$, $p = 0.36$).

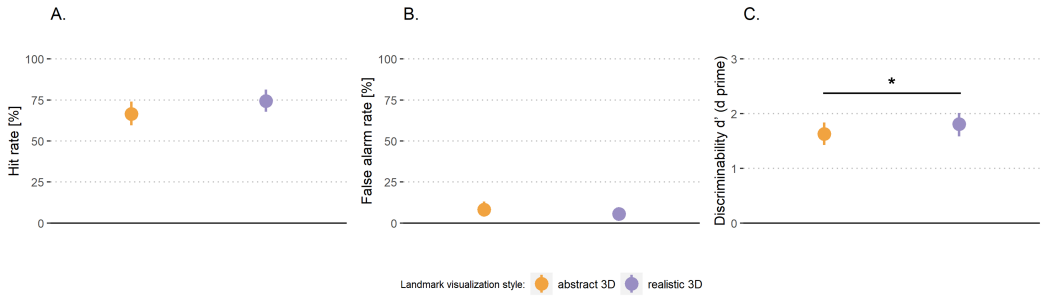


Figure 3 Participants' hit (A) and false alarm rates (B), and (C) recall performance of landmarks seen on the mobile map by visualization style (dots indicate means; bars indicate 95% CIs, $*p < .05$).

3.2.3 Results on route direction recall

Finally, the analysis of route direction recall accuracy (Figure 4) reveals that both groups performed significantly better (abstract 3D: $t(45) = 4.8$, $p < 0.001$; realistic 3D: $t(45) = 7.7$, $p < 0.001$) than chance level (green line) at 33.3% (three possible route directions: straight, left, and right). Participants, on average, seem to remember route directions better when navigating with realistic 3D landmarks ($M = 59.56\%$, $SD = 23.3\%$) on the mobile map compared to abstract 3D landmarks ($M = 52.17\%$, $SD = 26.5\%$). Contrary to our hypothesis, this difference is not statistically significant ($p = 0.09$). However, this difference shows a medium effect size ($r = 0.3$).

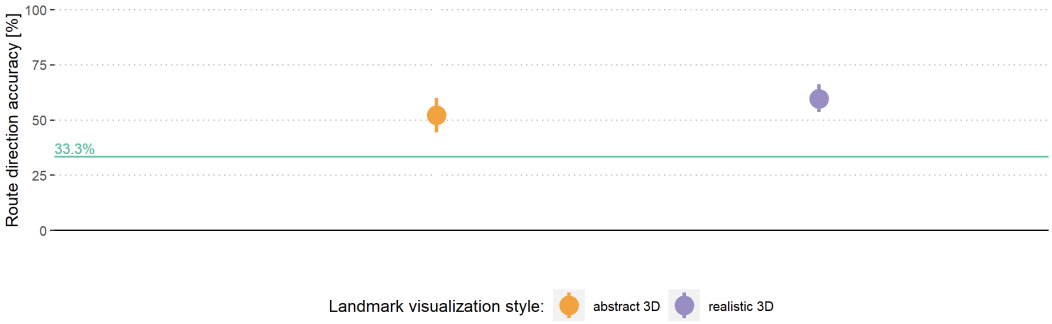


Figure 4 Accuracy of route direction recall by landmark visualization condition (dots indicate means; bars indicate 95% CIs; green line indicates performance at chance level).

4 Discussion

We conducted a map-assisted navigation experiment outdoors in an urban area to assess how landmark visualization style (i.e., 3D abstract vs. 3D realistic symbols) on a mobile map influences participants' recall of landmarks and route direction along the traversed environment. Our first results show that participants better remembered landmarks shown as 3D realistic symbols on the mobile map than 3D abstract landmark symbols (Figure 3). In support of our hypothesis, adding more details to the 3D landmarks (realistic) on the mobile map display could have made them more memorable compared to 3D landmarks with fewer details (abstract) and thus made it easier for participants to recall these symbols [7]. Our results on landmark recall are consistent with previous findings showing that 3D realistic landmarks on mobile maps facilitate the matching of navigation-relevant information with proximal features in the environment [6], and in doing so, facilitate landmark learning [7]. Although participants' response biases were significantly higher than zero, they did not differ between landmark visualization conditions and thus do not affect our interpretation of this result.

Contrary to our hypothesis, the results did not reveal any significant differences in recalling landmarks seen in the environment (Figure 2) and route direction accuracy (Figure 4). The lack of differences between landmark visualization styles could be because having 3D landmarks in general (regardless of the level of detail) could have already provided the necessary visual information to help participants recognize these buildings and the associated navigation decisions [2, 10]. This result is consistent with previous studies showing that participants do not necessarily perform better with increased realism [10, 15]. Another possible explanation could be that participants were explicitly told that their acquired knowledge of the environment would be tested [14], which might have motivated them to pay more attention to task-relevant features and directions along the navigation route, or use other spatial-learning strategies, reducing the benefits of depicting landmarks on the mobile map [8].

5 Implications and future directions

This navigation study revealed promising results on how the design of landmarks on mobile maps can support spatial learning during navigation. We conducted the study in an urban residential environment of a European city; thus, it remains to be seen whether our results

can be applied to other types of environments. On the one hand, a close visual match between the mobile map display and the environment could have directed participants' visual attention away from the display and toward the environment, thus helping participants to learn the environment while mitigating cognitive load. On the other hand, the increased visual information on the realistic 3D landmarks might have increased participants' cognitive load due to the higher amount of visual information to be processed [5]. Therefore, examining participants' cognitive load (i.e., recorded with electroencephalography) and visual attention allocation (i.e., by using mobile eye-tracking) could complement the interpretation of these behavioral results. Even though we did not find differences in the landmarks seen in the environment and route direction knowledge, there might be differences between the conditions regarding survey knowledge acquisition. Specifically, viewing the route from a bird's-eye view and remembering 3D realistic landmarks on the mobile map display better could mean that participants' survey knowledge is better with the realistic 3D compared to the abstract 3D landmarks. Therefore, for the next steps in our research program, we are interested in analyzing the relationships between participants' map interactions, familiarity with the study area, survey knowledge acquisition, visual attention allocation, and cognitive load during this real-world navigation task.

References

1 H. Couclelis, R.G. Golledge, N. Gale, and W. Tobler. Exploring the anchor-point hypothesis of spatial cognition. *Journal of Environmental Psychology*, 7(2):99–122, 1987. doi:10.1016/S0272-4944(87)80020-8.

2 B. Elias and V. Paelke. User-Centered Design of Landmark Visualizations. In *Map-based Mobile Services*. Springer, 2008. doi:10.1007/978-3-540-37110-6_3.

3 A.L. Gardony, T.T. Brunyé, and H.A. Taylor. Navigational Aids and Spatial Memory Impairment: The Role of Divided Attention. *Spatial Cognition & Computation*, 15(4):246–284, 2015. doi:10.1080/13875868.2015.1059432.

4 T. Ishikawa. Satellite Navigation and Geospatial Awareness: Long-Term Effects of Using Navigation Tools on Wayfinding and Spatial Orientation. *The Professional Geographer*, 71(2):197–209, 2019. doi:10.1080/00330124.2018.1479970.

5 A. Kapaj, S. Lanini-Maggi, and S. I. Fabrikant. The impact of landmark visualization style on expert wayfinders' cognitive load during navigation. *Abstracts of the ICA*, 3:138, 2021. doi:10.5194/ica-abs-3-138-2021.

6 A. Kapaj, S. Lanini-Maggi, and S.I. Fabrikant. The influence of landmark visualization style on expert wayfinders' visual attention during a real-world navigation task. *UC Santa Barbara: Center for Spatial Studies*, 2021. doi:10.25436/E2NP44.

7 H. Liao, W. Dong, C. Peng, and H. Liu. Exploring differences of visual attention in pedestrian navigation when using 2D maps and 3D geo-browsers. *Cartography and Geographic Information Science*, 44(6):474–490, 2017. doi:10.1080/15230406.2016.1174886.

8 D.R. Montello. Landmarks are Exaggerated. *KI - Künstliche Intelligenz*, 31(2):193–197, 2017. doi:10.1007/s13218-016-0473-5.

9 C. Nothegger, S. Winter, and M. Raubal. Selection of Salient Features for Route Directions. *Spatial Cognition & Computation*, 4(2):113–136, 2004. doi:10.1207/s15427633scc0402_1.

10 M.A. Plesa and W. Cartwright. Evaluating the Effectiveness of Non-Realistic 3D Maps for Navigation with Mobile Devices. In *Map-based Mobile Services*. Springer, 2008. doi:10.1007/978-3-540-37110-6_5.

11 M. Raubal and S. Winter. Enriching Wayfinding Instructions with Local Landmarks. In *Geographic Information Science*, 2002. doi:10.1007/3-540-45799-2_17.

12 K-F. Richter and S. Winter. *Landmarks*. Springer, 2014. doi:10.1007/978-3-319-05732-3.

13 W.P. Tanner and J.A. Swets. A decision-making theory of visual detection. *Psychological Review*, 61(6):401–409, 1954. doi:10.1037/h0058700.

14 F. Wenczel, L. Hepperle, and R. von Stülpnagel. Gaze behavior during incidental and intentional navigation in an outdoor environment. *Spatial Cognition & Computation*, 17(1-2):121–142, 2017. doi:10.1080/13875868.2016.1226838.

15 J. Wilkening and S.I. Fabrikant. How Do Decision Time and Realism Affect Map-Based Decision Making? In *Spatial Information Theory (COSIT)*, 2011. doi:10.1007/978-3-642-23196-4_1.

Smart Crowd Management: The Data, the Users and the Solution

Laure De Cock[1]
Ghent University, Belgium

Steven Verstockt
imec, Ghent University, Belgium

Christophe Vandeviver
Ghent University, Belgium

Nico Van de Weghe
Ghent University, Belgium

Abstract

This research project is situated in the domain of smart crowd management, a domain that is gaining importance because of the challenges that arise from urbanization, but also the opportunities that come with smart cities. While our cities become more crowded every day, they also become smarter, for example by employing pedestrian tracking sensors. However, the datasets that are generated by these sensors do not allow smart crowd management yet, because they are sparse and not linked to the perception of the crowd. This research will tackle these issues in three steps. First, pedestrian counts will be estimated on streets that have no tracking data by use of deep learning and space syntax data. Next, the perception of crowdedness within the crowd will be linked to the objective pedestrian counts by conducting two user studies, and finally, the resulting subjective pedestrian counts will be used as weights for a routing algorithm. The last step has already been developed as a proof of concept. The routing algorithm, that uses partly simulated data and partly real-time tracking data, has been embedded in a webtool to show stakeholders the potential and goal of this innovative project.

2012 ACM Subject Classification Software and its engineering → Software design engineering; Information systems → Sensor networks; Information systems → Location based services

Keywords and phrases crowd tracking, crowd modeling, space syntax, deep learning, perception, routing

Digital Object Identifier 10.4230/LIPIcs.COSIT.2022.16

Category Short Paper

Supplementary Material *Software (Source Code)*: `https://github.com/laudcock/Smart_crowd_management`

Funding *Laure De Cock*: BOF/24J/2021/289.

1 Introduction

As the gathering of a crowd can lead to hazardous situations, public safety is a major concern for local authorities [21]. In the past 20 years, more than 100 stampedes occurred with over 5000 deaths [10]. These numbers highlight the need for a flexible system that can monitor crowd dynamics in an urban environment. A need that will be even higher in the future, as the population growth in urban areas is projected to increase to 68% of the world's population by 2050 [4, 19]. Urbanization will go hand in hand with the development of smart cities,

[1] corresponding author

which will facilitate urban sensing [4]. This opens the door to a smart solution for crowd management, because crowds can be tracked throughout the urban environment. However, this sensor data has a few limitations. First of all, the sensors that are employed in urban environments are sparse as it is unfeasible to place them in every street [5, 16]. This means that pedestrian tracking datasets almost never fully cover an urban street network. Second, the objective number of pedestrians that is counted by tracking sensors does not necessarily reflect the subjective feeling of crowdedness within the crowd [1]. This research project aims to tackle the issues mentioned above by focusing on three components: the data, the users and the solution. In the remainder of this section, the state of the art of each component will be outlined.

The data

As pedestrian tracking data with full-coverage is rare, models are being developed to simulate the dynamics of a crowd. There exists a wide variety of models, but no model has all criteria to enable crowd modeling in practice, leading to a gap between theory and implementation [24]. There are two reasons for this [9]:

- Human behavior and decision making are influenced by numerous individual factors that are difficult to capture in a general set of model rules and equations.
- The varying environmental context makes it hard to introduce universal models that work in every context.

In this research project a model will be developed that provides an answer to both issues. The first will be resolved by using data-driven techniques, such as deep learning. Unlike theoretical models and simulations (e.g., cellular automata), deep learning algorithms do not need previous assumptions on the data, which has already been shown to be an advantage in several studies. Wang et al. (2019) found, for example, that deep learning methods were better than traditional ones when crowd movements during an evacuation experiment were more complex [23]. Given these promising first results of deep learning for crowd modeling, this project will use this technique, but enhance it with geodata to resolve the second issue mentioned above, i.e., the varying environmental context. Raubal et al. (2020) distinguish two types of geodata: tracking data and context data. While the availability of tracking datasets for research purposes remains limited, there is an increasing number of urban context data sources [18]. One specific type of context data known to correlate particularly well with pedestrian movement flows is space syntax [12]. The space syntax theory was defined in 1984 by Hillier and Hanson to "quantify space in a way we socially experience it" [11]. Its representations (e.g., axial lines, isovists, visibility graphs) and measures (e.g., integration, connectivity, occlusivity, controllability) have been used by designers, spatial planners and researchers to quantify the structure of cities and buildings ever since (e.g., [7, 22]). Space syntax has been proven very useful for crowd modeling purposes as well [20]. Zampieri et al. (2009), for example, combined space syntax and other spatial data with deep learning to estimate pedestrian counts [25]. The resulting model had a correlation coefficient of more than 90% for both the training and testing set, but they did not use crowd tracking sensors as the pedestrians were manually counted. Therefore, in this research project space syntax and deep learning will be combined for the first time to estimate pedestrian counts based on urban sensor data. The resulting model will estimate crowdedness on every street of an urban network.

The users

As opposed to the data and modeling component of crowd management, there is few research on the perception component [1]. However, several researchers agree that the perceived crowdedness can substantially differ from the objective density (i.e., the number of people per unit of space) [6, 14, 17]. Li and Hensher (2013), for example, compared the results of a survey which measures the passenger loads with the results of a survey which asked 2500 train commuters for feedback on the rail services. According to the first survey there was no substantial crowding problem, while in the second survey 55% of the commuters indicated there was [14]. Besides in public transport, this mismatch between objective counts and perception can also be found in urban green spaces. Campagnaro et al. (2020) found, for example, that crowding in a park was experienced as a negative thing for most participants, while other studies show that moderate crowding increases the feeling of safety [3]. This does not only show that there is a difference between objective and subjective crowding, but also that their relationship is complex and not linear [6]. This is because the perception of crowding is influenced by numerous factors. The two most obvious factors are space and people, which is why some authors differentiate between spatial density and social density [2, 17]. This shows that we must go beyond the physical space when discussing crowding and also analyze the behavioral or cognitive space, something that is rarely done within the field of crowd modeling, but more common within the field of space syntax [17]. As the environment has an important influence on the link between crowd counts and crowdedness perception, we feel that quantifying space by use of space syntax might be an essential step to determine this link. Once this link has been made, the perceived crowdedness can be deduced from the objective counts, which are generated by the model described in the previous section.

The solution

Determining the link between objective and perceived crowdedness is important as Li and Hensher (2011) have shown that the willingness to pay for reduced crowding is often as high as for reduced travel time [13]. However, most routing algorithms today minimize travel time or distance by calculating the shortest or the fastest path, even though scholars agree that users of navigation systems do not always prefer these paths [15]. Muller et al. (2017) state that many subjective parameters determine the route choice, but that it is not straightforward to include these parameters as weights in a routing algorithm, because that would require data of the user's perception [15]. The lack of data is one of the reasons that crowdedness has long been overlooked as a weight for pedestrian routing algorithms, although it clearly can be a decisive factor. This research project aims to fill this gap, by generating perceived crowding data on a city-scale and incorporating it in a routing algorithm. This algorithm will generate the least perceived crowded route, which can be used by local authorities and safety officials for crowd management purposes, but also by pedestrians as a route planning service. In the next section is explained how we will reach this result.

2 Materials and methods

The data

In the first step of the methodology a model will be generated that interpolates pedestrian counts in between sensor locations. In this modeling phase two concepts will be combined for the first time in the domain of sensor-based crowd tracking: space syntax and deep

learning. Both local (area, perimeter, compactness, vista length, occlusivity) and global (integration, depth, control, controllability) space syntax measures will be calculated for every street with the isovists.org software. Tracking data will be obtained from Telraam, a Belgian citizen-science project that helps citizens to install a tracking device in their front window (for more information on the technology and sensor locations, see `https://telraam.net`). The data of these devices can be freely imported in custom applications through an API. In a next step, a graph will be created for each timestamp of the street network of the study area, and the space syntax measures and Telraam counts of a certain timestamp will be added as attributes to the edges. The motivation to use a graph is twofold: first, the spatially enhanced graphs will serve as input for a graph convolutional network (GCN) to estimate the counts on the graph edges without Telraam sensor, and next, the resulting graphs with an estimated pedestrian count for every edge will be used as input for a shortest path algorithm.

The users

In the first step, objective tracking data was used as input for the model, but in the second step the bridge will be made to subjective crowding data. Therefore, the link between objective and subjective crowding must first be determined. To this end two user studies will be conducted: one in the field and one in virtual reality (VR). For the first study, we will ask pedestrians how crowded they find the street where they are walking on two days in three different contexts: the main shopping street in Ghent (Langemunt), the main bar street in Ghent (Overpoort) and the main metro station in Antwerp (Groenplaats). For all three locations, objective tracking data will be collected by the radio frequency sensors of Crowdscan during the survey (for more information on the technology, see `https://www.crowdscan.be`) [8]. This way, both objective and subjective crowding data will be obtained for three different activities (shopping, nightlife and public transport) on two different days. The second user study will be conducted in VR, as influencing parameters (e.g., the weather) can be easier controlled with this medium than in a field study. Different scenarios and levels of crowding will be simulated for participants while they will be asked the same question: how crowded do you find this place? Based on the datasets, resulting from the two user experiments, and the space syntax measures of the environment the link between objective and subjective crowding will be determined, so that for each Telraam count we can calculate the corresponding general subjective crowdedness. In summary, the result of this step will be a graph with subjective crowdedness levels for every edge.

The solution

The resulting graph of step 1 and 2 can be used as input for a routing algorithm. As a proof of concept, the routing algorithm has already been developed and embedded in a webtool. The tool uses real-time Telraam pedestrian counts and simulates the interpolated data that will be generated by the model in the future. In the next section the code is explained in more detail.

3 Results

The code and link for the webtool can be found on `https://github.com/laudcock/Smart_crowd_management` and the general workflow is summarized in Figure 1.The webtool allows you to choose two points on the map in the city of Aalst in Belgium (the case study of the webtool) and shows two routes between these points: the shortest one and the least

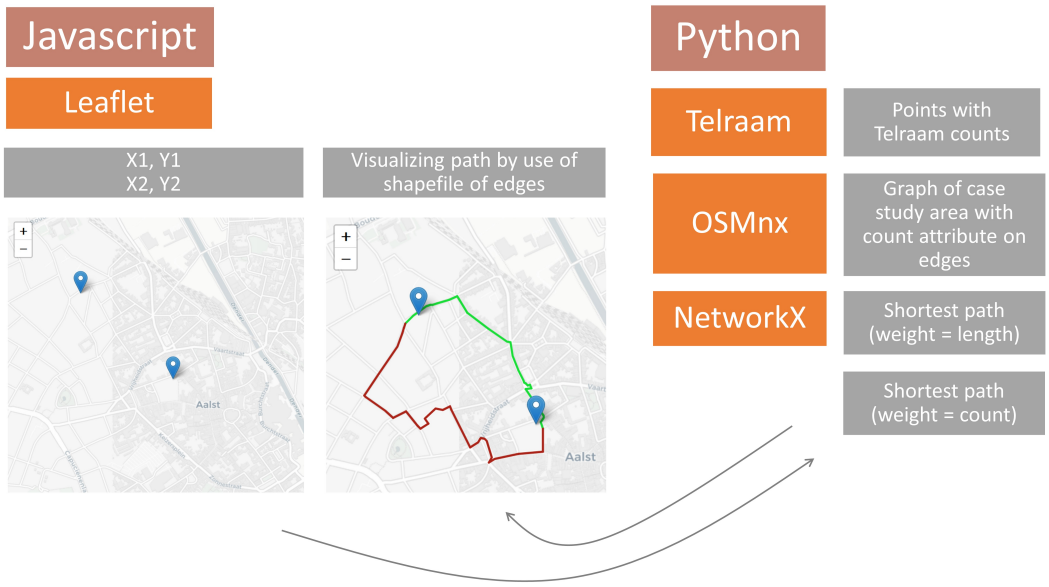


Figure 1 General workflow of the crowd routing webtool.

crowded one. The coordinates of origin and destination and the visualization of the routes is handled by a javascript code, for the calculation of the routes a python script is called. First, the real-time Telraam counts are loaded through an API. Next, a precalculated graph of the study area is loaded and the count attribute is added to the edges. The 25 edges that correspond with a Telraam sensor in Aalst case study get the real-time count, the edges without Telraam sensor get a random number between the mean – SD and mean + SD of the Telraam counts, to simulate the interpolation of crowd data in between sensors. It is important to note that each time a new route is requested, the edges get new (real-time or random simulated) count values. Finally, two shortest paths are calculated, one with the length attribute of the edges as weight and one with the count attribute.

4 Discussion and future research

As this research project is in an early phase, there is a substantial amount of future research to be done. It might seem extraordinary that we started developing the final phase of the project (i.e., "the solution"), but this had two reasons. First of all, it gives quite a good idea to stakeholders of the outcome. Second, by starting from "the end product" the prerequisites for the preceding steps of the workflow become clear. For example, the result of the modeling phase (i.e., "the data") will have a graph-like structure as we decided to use a GCN. Although large parts of the methodology have been clarified by starting from the solution, there are still some questions that need answering. For example, it might be hard to find a good model that explains the trend in the data without overfitting. We aim to maximize the success rate by choosing a type of model that fits the data (GCN), but we might have to look into other deep learning algorithms as well. Additionally, it will be challenging to determine the link between the objective and subjective crowding as besides the environment numerous contextual and personal factors influence this link. Therefore we conduct two studies: a first explorative one in the field to identify the influencing parameters, and a second one in VR the determine the specifics of the relationship in a controlled environment. By anticipating possible fallbacks, we tried to maximize the success rate of our proposed method.

5 Conclusion

Pedestrian routing algorithms almost always use length or travel time as the weight of the path, while crowdedness can also be a decisive factor for the route choice. However, crowd-steered routing is not that straightforward because of a lack of crowd data. Crowd tracking sensors in smart cities might be an answer to this issue, but as it is infeasible to place sensors on every corner of every street, this tracking data remains sparse. Moreover, it is hard to make decisions based on this objective data because it does not necessarily reflect the feeling of crowdedness. This research project aims at resolving these issues, by using deep learning and space syntax data to interpolate pedestrian counts in between sensors and then linking these objective counts to the subjective perception of the crowd itself. The resulting subjective pedestrian counts of every street of a network can be used as weights for a crowd-steered routing algorithm. The coding and implementation of this last step has been presented in this paper, by use of partly simulated and partly real-time tracking data. The result is a webtool that generates both the shortest and least crowded path from a graph of the case study city. This outcome has been used to inform stakeholders and finetune the methodology of the remaining parts of this project.

References

1 Mohammed Alkhadim, Kassim Gidado, and Noel Painting. Risk management: The effect of fist on perceived safety in crowded large space buildings. *Safety Science*, 108(November 2017):29–38, 2018. doi:10.1016/j.ssci.2018.04.021.

2 A. Celil Cakici, Nurhayat Iflazoglu, and Levent Altinay. Impact of crowded restaurant perception on affectivity and behavioral intentions. *Tourism*, 69(3):429–492, 2021. doi:10.37741/T.69.3.8.

3 Thomas Campagnaro, Daniel Vecchiato, Arne Arnberger, Riccardo Celegato, Riccardo Da Re, Riccardo Rizzetto, Paolo Semenzato, Tommaso Sitzia, Tiziano Tempesta, and Dina Cattaneo. General, stress relief and perceived safety preferences for green spaces in the historic city of Padua (Italy). *Urban Forestry and Urban Greening*, 52(March):126695, 2020. doi:10.1016/j.ufug.2020.126695.

4 Clayson Celes, Azzedine Boukerche, and Antonio A.F. Loureiro. Crowd management: A new challenge for urban big data analytics. *IEEE Communications Magazine*, 57(4):20–25, 2019. doi:10.1109/MCOM.2019.1800640.

5 Achituv Cohen, Sagi Dalyot, and Asya Natapov. Machine learning for predicting pedestrian activity levels in cities. In Anahid Basiri, Georg Gartner, and Haosheng Huang, editors, *Proceedings of the 16th International Conference on Location Based Services (LBS 2021)*, 2021.

6 Tom Cox, Jonathan Houdmont, and Amanda Griffiths. Rail passenger crowding, stress, health and safety in britain. *Transportation Research Part A: Policy and Practice*, 40(3):244–258, 2006. doi:10.1016/j.tra.2005.07.001.

7 Laure De Cock, Nico Van de Weghe, Kristien Ooms, Nina Vanhaeren, Matteo Ridolfi, Eli De Poorter, and Philippe De Maeyer. Taking a closer look at indoor route guidance; usability study to compare an adapted and non-adapted mobile prototype. *Spatial Cognition and Computation*, pages 1–23, February 2021. doi:10.1080/13875868.2021.1885411.

8 Stijn Denis, Ben Bellekens, Maarten Weyn, and Rafael Berkvens. Sensing thousands of visitors using radio frequency. *IEEE Systems Journal*, 15(4):5090–5093, 2021. doi:10.1109/JSYST.2020.3019189.

9 Milad Haghani and Majid Sarvi. Crowd behaviour and motion: Empirical methods. *Transportation Research Part B: Methodological*, 107:253–294, 2018. doi:10.1016/j.trb.2017.06.017.

10 Dirk Helbing, Lubos Buzna, Anders Johansson, and Torsten Werner. Self-organized pedestrian crowd dynamics: Experiments, simulations, and design solutions. *Transportation Science*, 39(1):1–24, 2005. doi:10.1287/trsc.1040.0108.

11 Bill Hillier and Julienne Hanson. *The social logic of space*. Cambridge University Press, 1984.

12 Bin Jiang. Ranking spaces for predicting human movement in an urban environment. *International Journal of Geographical Information Science*, 23(7):823–837, 2009. doi:10.1080/13658810802022822.

13 Zheng Li and David A. Hensher. Crowding and public transport: A review of willingness to pay evidence and its relevance in project appraisal. *Transport Policy*, 18(6):880–887, 2011. doi:10.1016/j.tranpol.2011.06.003.

14 Zheng Li and David A. Hensher. Crowding in public transport: A review of objective and subjective measures. *Journal of Public Transportation*, 16(2):107–134, 2013. doi:10.5038/2375-0901.16.2.6.

15 Manuel Müller, Christina Ohm, Florin Schwappach, and Bernd Ludwig. The path of least resistance: Calculating preference adapted routes for pedestrian navigation. *KI - Kunstliche Intelligenz*, 31(2):125–134, 2017. doi:10.1007/s13218-016-0472-6.

16 Itzhak Omer and Nir Kaplan. Using space syntax and agent-based approaches for modeling pedestrian volume at the urban scale. *Computers, Environment and Urban Systems*, 64:57–67, 2017. doi:10.1016/j.compenvurbsys.2017.01.007.

17 Amos Rapoport. Toward a redefinition of density. *Environment and behavior*, pages 133–158, 1975.

18 Martin Raubal, Dominik Bucher, and Henry Martin. Geosmartness for personalized and sustainable future urban mobility. In W. Shi, M. F. Goodchild, M. Batty, and M.-P. Kwan, editors, *Urban Informatics*, pages 59–83. Springer Singapore, 2021. doi:10.1007/978-981-15-8983-6_6.

19 Andreas Schadschneider, Mohcine Chraibi, Armin Seyfried, Antoine Tordeux, and Jun Zhang. Pedestrian dynamics: From empirical results to modeling. In Livio Gibelli and Nicola Bellomo, editors, *Crowd Dynamics, Volume 1*, volume 1, page 292. Birkhauser, 2018.

20 Samia Sharmin and Md Kamruzzaman. Meta-analysis of the relationships between space syntax measures and pedestrian movement. *Transport Reviews*, 38(4):524–550, 2018. doi:10.1080/01441647.2017.1365101.

21 Utkarsh Singh, Jean François Determe, François Horlin, and Philippe De Doncker. Crowd monitoring: State-of-the-art and future directions. *IETE Technical Review*, 2020. doi:10.1080/02564602.2020.1803152.

22 Timmermans H. Teklenburg J., Borgers A. Space syntax as a design support system: evaluating alternative layouts for shopping centres. In A. D. Seidel, editor, *Banking on design: proceedings of the 25th annual conference of the Environmental Design Research Association*, pages 220–228, 1994.

23 Ke Wang, Xiupeng Shi, Algena Pei Xuan Goh, and Shunzhi Qian. A machine learning based study on pedestrian movement dynamics under emergency evacuation. *Fire Safety Journal*, 106(April):163–176, 2019. doi:10.1016/j.firesaf.2019.04.008.

24 Nanda Wijermans, Claudine Conrado, Maarten van Steen, Claudio Martella, and Jie Li. A landscape of crowd-management support: An integrative approach. *Safety Science*, 86:142–164, 2016. doi:10.1016/j.ssci.2016.02.027.

25 Fábio Lúcio Zampieri, Décio Rigatti, and Cláudio Ugalde. Evaluated model of pedestrian movement based on space syntax, performance measures and artificial neural nets. In Daniel Koch, Lars Marcus, and Jesper Steen, editors, *7th International Space Syntax Symposium*, pages 1–8, 2009. doi:ISSN1402-7453ISRNKTH/ARK/FP-09:01-SEISBN978-91-7415-347-7.

A Weather-Aware Framework for Population Mobility Modelling

Vanessa Brum-Bastos
Wroclaw University of Environmental and Life Sciences, Poland
University of Canterbury, Christchurch, New Zealand

Kamil Smolak
Wroclaw University of Environmental and Life Sciences, Poland

Witold Rohm
Wroclaw University of Environmental and Life Sciences, Poland

Katarzyna Sila-Nowicka
The University of Auckland, New Zealand
Wroclaw University of Environmental and Life Sciences, Poland
University of Glasgow, UK

Abstract

The widespread availability of GPS-enabled mobile devices has contributed towards an unprecedented volume of data on human movement. Human mobility data are the key input for developing accurate mobility models that can support decision-making in, for example, urban planning, transportation planning and disease spread. However, the increasing geoprivacy concerns have been limiting the use of and access to such data. For this reason, the WHO-WHERE-WHEN (3W) model, a privacy-protective model for generating synthetic mobility data, has been developed. However, human mobility is affected by multiple factors that must be accounted for to produce synthetic mobility trajectories that accurately simulate the fluctuations of population in a study area. The 3W model already considers four main factors affecting human mobility: size and shape of activity spaces, circadian rhythm, and home and work locations. Yet, meteorological factors are known to affect human mobility patterns but, to our knowledge, there is not a model that accounts for weather conditions. In this paper, we propose a theoretical framework to extend the 3W model to a 4W model: WHO-WHERE-WHEN-WEATHER. We hypothesise that accounting for weather conditions in human mobility predictions will increase the overall accuracy of predicted mobility patterns.

2012 ACM Subject Classification Applied computing; Human-centered computing

Keywords and phrases movement analytics, human movement, mobility models, context-awareness

Digital Object Identifier 10.4230/LIPIcs.COSIT.2022.17

Category Short Paper

Funding *Vanessa Brum-Bastos*: The Polish National Agency for Academic Exchange (NAWA) Ulam Fellowship PPN/ULM/2020/1/00207/DEC/1 and The Royal Society International Exchanges Grant IES/R1/201127.
Kamil Smolak: The research was financed under the National Science Centre, Poland research grant "Explanation and mitigation of the bias in human mobility predictions" no. 2021/41/N/HS4/03084.
Katarzyna Sila-Nowicka: Supported by the Marsden Fund Council from Government funding, managed by Royal Society Te Apārangi – UOA2037. Also supported by The Royal Society International Exchanges Grant IES/R1/201127.

1 Introduction

Mobility models are developed with data on human movements to identify and predict mobility patterns. However, human movements do not take place in a vacuum but are rather embedded and influenced by the surrounding environment, also known as movement

context [6]. Thus, recent mobility studies have been linking movement data to contextual information to gather insights into commuting behaviour [24], tourist behaviour [11] and how weather influences human movement patterns [7]. Similarly, mobility models have also been incorporating movement context by considering work and home regions [16], commuting distance [17] or social media data [10].

Mobility models do not account for weather variables, yet human movement behaviour is impacted by meteorological conditions: [7] and [29] found that wind-speed and direction had an effect on the proportional distribution of time spent in different activities as well as promoted changes in the choice of transportation modes. [9] and [14] described an effect of rain on the proportion of vehicular and walking trips. [23] studied the effect of wind and rain on peoples shopping behaviour. [11] and [30] found a positive effect of increasing temperatures on walking. [5] discovered that increased rainfall led to higher use of public transportation in Bergen, Norway. [22] found that weather conditions influence urban mobility patterns.

Despite the increasing body of evidence supporting the role played by weather in human mobility patterns, up-to-date to our knowledge no mobility model accounts for meteorological variables when modelling nor predicting human movement patterns. Therefore, the goal of this paper is laying the foundations of a framework for weather-aware human mobility modelling at the population level. To consider weather conditions we chose to extend the WHO-WHERE-WHEN (3W) model, a privacy-protective model for generating synthetic mobility data [25], by incorporating ERA5 reanalysis data from the European Centre for Medium-Range Weather Forecasts (ECWMF).

2 The WHO-WHERE-WHEN (3W) model

The 3W is an agent-based model that generates synthetic trajectories by imitating real spatio-temporal characteristics of movement data [15, 19]. Such models can be used to simulate hypothetical scenarios, which are useful in disease spread prediction and urban planning. Additionally, in models such as 3W, the generation of synthetic trajectories from real data safeguards individual's privacy while also providing synthetic data that are still useful for analyses due to their realism [26, 13].

3 The WHO-WHERE-WHEN-WEATHER (4W) model concept

In this section we overview the conceptual framework for the 4W model (Figure 1) and describe the required input data, trajectory sampling process, synthetic trajectory generation process and respective output.

3.1 Inputs

3.1.1 Mobility data

We will use Near® unidentifiable mobility data, which is an aggregate of mobile location data collected from smartphone apps. When a user installs an app, they are asked to share their location data. Users can opt in or out to location sharing, but when they opt in, their phone collects location data and shares it with the app provider and Near®. Latitude-longitude coordinate pairs are collected by smartphones as they move through time and space. These lat-long pairs are associated with the specific device through a unique identifier also available in the dataset. The data are provided as a table containing latitude, longitude, timestamp and device id.

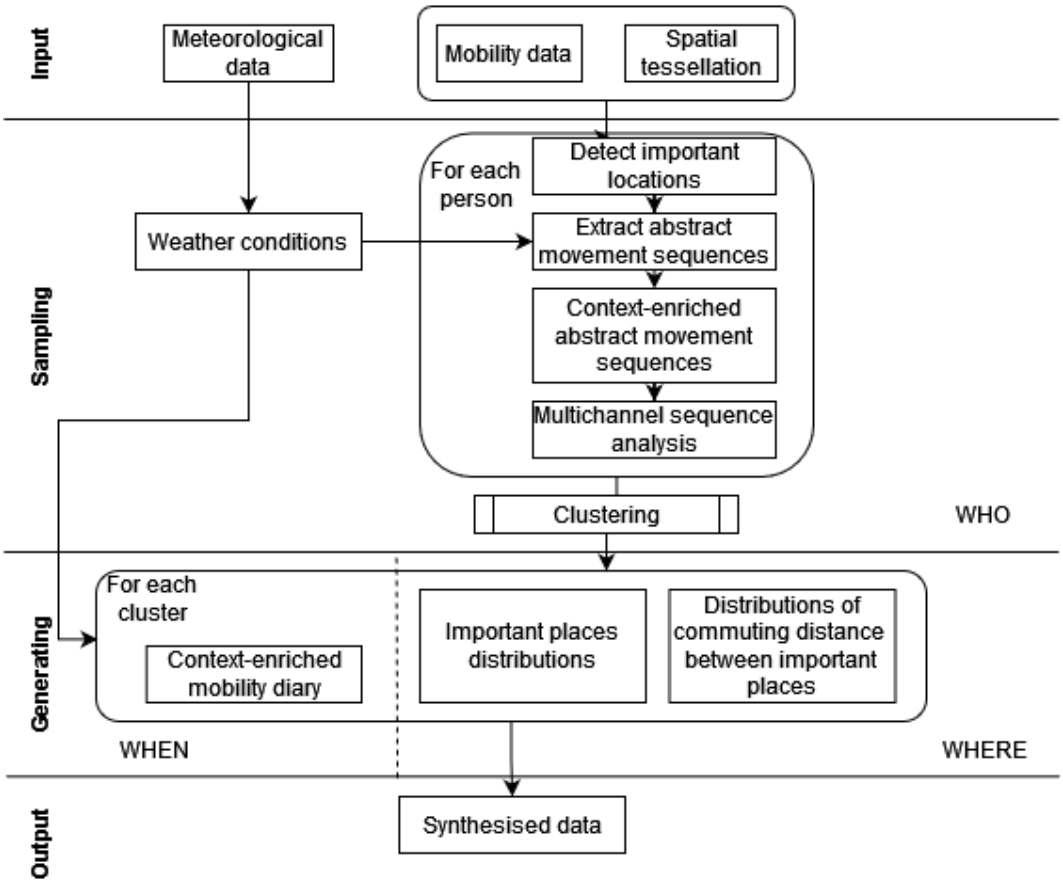


Figure 1 Concept overview for the WHO-WHERE-WHEN-WEATHER (4W) model.

The Near® mobility dataset we use was generated by about 4.7 million devices and it covers Rio de Janeiro, a city of approximately 6.7 million people in Brazil, between May 2019 and January 2021. The sampling rate of location data varies from milliseconds to hours or days depending on the device.

3.1.2 Weather data

The weather data we plan to use comes from the fifth generation ECMWF atmospheric reanalysis model (ERA5) of the global climate covering the period from January 1950 to present. ERA5 provides hourly global estimates of atmospheric, land and oceanic climate variables on a 30 km grid from the surface up to a height of 80km. The outputs from ERA 5 reanalysis have been validated and performed with high accuracy when compared to data from meteorological stations in Brazil [4]. Even though ERA5 reanalysis model provides a large number of atmospheric variables, we are only interested in using the ones that have potential direct effect on human mobility patterns. More specifically, we will look at wind speed components at ten meters height from the surface, temperature at two meters height from the surface, total cloud cover, total precipitation, rain rate and accumulated snowfall. We will also take into account the daylight conditions by also annotating the trajectories with sunrise and sunset information from the daylight Python package [1].

3.1.3 Spatial tessellation

The spatial tessellation is an aggregation layer that divides the simulation area, into a set of non-overlapping polygons, each with a unique identifier and centroid coordinates [20]. In our case, we will use a common approach, which is constructing a grid over the whole area of simulation [17, 8]. The resolution of the grid should not be higher than the resolution of the input data. We plan to test our model using various grid resolutions.

3.1.4 Data pre-processing

The raw data were not sorted, therefore we first split them into separate files for each unique device identifier. Furthermore, to facilitate sample selection for further analyses each device data file was also annotated with metadata on data incompleteness. The incompleteness q is expressed by the total number of missing observations in one-hour time intervals [27]. In many cases, we observed large gaps between consecutive records in movement trajectories, reaching up to a few months. Therefore, we calculated the highest possible q that could be achieved from a one-month-long sample of the movement trajectory for a given device ID.

At this stage, mobility data have a form of a set of raw movement trajectories P_i of i devices. Each P_i is a set of x, y coordinates and timestamps t ordered by time. In the next step, we will select a sample of at least 5000 movement trajectories with $q < 0.2$. In order to reduce noise in the data, we will extract stay-points from the movement trajectories by grouping data points from each P_i into clusters of spatio-temporally neighbouring points [25]. Stay-points are defined as locations where a device spends more than tau time within a range of $delta$. In our case, we plan to apply commonly used values of $tau = 10$ minutes and $delta = 300$ metres [18, 25]. and Then, we will assign all detected stay-points for each mobility trajectory P_i to the respective cells in the spatial tessellation, creating a movement trajectory where the coordinates are the centroids of the spatial tessellation cell. Finally, we will create movement sequences by aggregating data to one-hour time-bins. To achieve that, movement trajectories will be transformed into a sequence of cells visited every hour, where at each time interval the cell visited for the longest period will be selected as the location of a device for the one-hour time-bin .

3.2 Sampling

In the sampling stage, input data are used to estimate mobility-related probability distributions. This stage of the 4W model is called the WHO module. First, we use the mobility dataset to detect important places for each individual. Important places are defined as regularly visited locations that have particular importance or function to an individual [27]. In the 3W model, important places refer to work and home only, i.e. the first and second most significant locations. In the 4W model we will expand this to the n^{th} most important locations. Introducing a 3$^{\text{rd}}$ or even 4$^{\text{th}}$ most important location can help accounting for other regular trips, such as attending an evening course or going to a gym, which could improve the accuracy of the model [10].

We will assign a rank to detected important locations, where the highest rank indicates the most significant place. Based on that, we will replace locations recorded in a movement sequence of each individual by the important rank for this place. Locations that are not given a rank will be denoted as “other”. This process creates an abstract movement sequence consisting of “abstract locations” [20], i.e., a temporal sequence of places without geographic coordinates and in which the spatial component is represented by places, such as “Home”, “Work”, “n^{th} most significant place” and “Other”.

In the next step, we will enrich abstract movement sequences with the chosen meteorological data. Each record of the abstract movement sequence will be annotated with information on the weather conditions at the time and location where it was registered. Similarly to [7], we will translate weather conditions into sequences of qualitative states. The combination of the sequences describing weather conditions and the abstract mobility trajectory will create a contextualised abstract mobility trajectory that can be represented by multiple sequences of strings aligned in time (Fig 2). These are the so-called multi-channel sequences (MCSA), which is a bioinformatics technique for sequencing and analysing human genome [3]. However, MCSA has been extensively used for longitudinal studies in social sciences [2] and more recently in animal and human mobility studies [12, 7].

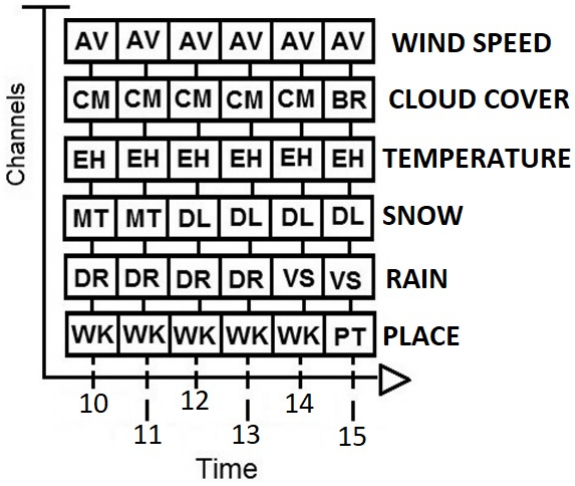


Figure 2 Schematic representation of one context-enriched abstract movement trajectory in the form of a multi-channel sequence. Each channel represents a property of movement or movement's surroundings. Here we have one channel for each meteorological variable and one channel representing places, which is our geographical dimension here. The strings in each sequence represent a specific state for that channel. For example, "AV" indicates average wind speed, "EH" indicates extremely high temperature, "DR" indicates dry weather.

All context-enriched abstract movement sequences will be used to find groups of individuals with similar mobility patterns while also taking into account the weather conditions affecting movement. We will use multi-channel sequence analysis (MSCA), optimal matching and hierarchical clustering to group contextualised abstract movement trajectories according to mobility behaviour and weather condition. The number of clusters of mobility behaviour will be estimated using Silhouette Coefficient criterion. The Silhouette Coefficient criterion is a partitioning metric that takes into account the mean intra-cluster distance and the mean distance to the nearest cluster for each sample [21]. We will use MCSA to calculate pairwise dissimilarity between all context-enriched abstract movement sequences. Dissimilarity is computed based on the minimal cost of replacing, deleting or inserting strings in the channels representing "place" and "weather conditions" so that one sequence is identical to another. This will generate a dissimilarity matrix that can then be used as a distance metric for clustering [7].

Clustered abstract trajectories will be used to calculate conditional probabilities expressing how likely it is for a person from one of the behavioural groups to visit a particular abstract location (identified by its rank) at a given time of the day and under certain weather

conditions. The set of these probabilities is called a mobility diary. We will use these conditional probabilities in the generation stage of the model to create synthetic movement trajectories. Clustered abstract trajectories will also be used to determine the proportion of agents in each behavioural group for the generation process. The proportion will be established as equal to the one found in the original data.

Finally, we will calculate the distribution of significant locations and the distributions of commuting distances between important locations for each behavioural group. The distribution of significant places indicates how likely it is to find the important location of a particular rank in each cell of spatial tessellation. Such distribution is referred to as a weighed spatial tessellation [20]. Distributions of commuting distances indicate the median distance between locations of a particular rank, in all possible combinations without repetitions.

3.3 Generating

The generation process will start initializing agents. Agents will be assigned a mobility diary from clusters in a ratio corresponding to the ratio of individuals clustered in each group. After that, each agent will be assigned important locations based on sampled spatial distributions, following the procedure shown in Fig. 3. Important places will be generated for each agent, using the weighted spatial tessellation and distributions of commuting distance from the agent's cluster. The first significant location will be selected using the spatial distribution of the most significant locations. The next places, if considered within the model, will be selected using weighed spatial tessellation for a currently generated rank of places and all combinations of commuting distance distributions. For example, when generating the second place, the commuting distance between the first and the second place will be considered (see Fig. 3). The second location will be selected from an underlying weighed spatial tessellation within the commuting distance from the first place. When the third place is generated, its location will be selected from its weighted spatial tessellation in the area which lies within commuting distances from the first and the second place. The process will follow the same logic for the fourth, fifth and n^{th} most significant places. The location of the fourth place, for example, will be selected in the area which lies within commuting distances from the first, second and third places.

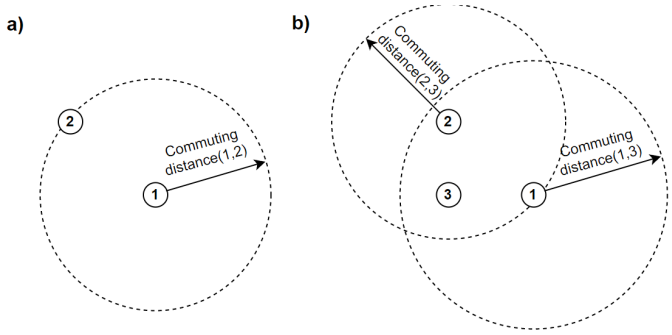


Figure 3 Important places generation process on the example of the second place (a) and the third place (b) selection. The second place is selected within commuting distance from the first place using underlying weighted spatial tessellation for the second place. The third place is selected within a commuting distance from the first and the second place using underlying spatial tessellation for the third places.

After assigning important places to every agent, the data generation process will start. That process will be controlled by the WHERE and WHEN modules, which separately simulate temporal and spatial aspects of mobility. This will be an iterative process, where each location for every agent will be generated using its mobility diary and considering current weather conditions. The WHEN module will select the next abstract location for an agent, which can be either one of the important places or another location. The WHERE module, depending on the selection, will generate the next data point in the location of a selected important place or will use the Exploration and Preferential Return (EPR) mechanism [28]to select other location. The EPR mechanism selects the next location to be an exploration or return to a previously visited place. This decision is based on the number of locations already visited by the agent and accounts for exploratory behaviour in human mobility. Locations for exploration will be selected using their attractiveness, which is defined as the population distribution in a given location divided by a distance friction function. The distance friction function determines that locations further away are less attractive than closer locations. The location of return will be selected using a probability distribution based on the frequency of previous visits.

The above process will be completed when the desired number of agents and data records are generated. The output from the model will have the form of a raw movement trajectory file, where each row consists of an identifier, timestamp and geographic coordinates. Generation accuracy will be assessed by comparing the results from the 4W model to real data.

4 Expected results and final considerations

We introduced a theoretical framework for taking weather into account when performing human mobility modelling. More specifically, we lay the foundations for extending the WHO-WHERE-WHEN (3W) model [26] into a WHO-WHERE-WHEN-WEATHER (4W) model. The 4W model will not only introduce weather information, but also experiment with varying the number of most important places. We hope that, by introducing weather information and expanding the number of most significant locations, we will be able to improve the accuracy of predictions, which we will test by comparing the results from the 4W model to the results from the original 3W model and real data.

References

1 Daylight python library. URL: https://pypi.org/project/daylight/.

2 Andrew Abbott. Sequence Analysis: New Methods for Old Ideas. *Source: Annual Review of Sociology Annu. Rev. Sociol*, 21(21), 1995. URL: http://www.jstor.org/stable/2083405.

3 Andrew Abbott and Angela Tsay. Sequence Analysis and Optimal Matching Methods in Sociology. *Sociological Methods & Research*, 29(1):3–33, August 2000. doi:10.1177/0049124100029001001.

4 Lucas Eduardo de Oliveira Aparecido, Glauco de Souza Rolim, Jose Reinaldo da Silva Cabral de Moraes, Guilherme Botega Torsoni, Kamila Cunha de Meneses, and Cicero Teixeira Silva Costa. Accuracy of ECMWF ERA-interim reanalysis and its application in the estimation of the water deficieny in paraná, Brazil. *Revista Brasileira de Meteorologia*, 34(4):515–528, 2019. doi:10.1590/0102-7786344066.

5 Kenneth Apeland. *Analysis Using Machine Learning*. PhD thesis, Western University of Applied Sciences, Bergen, June 2020. URL: https://bora.uib.no/bora-xmlui/bitstream/handle/1956/23882/A-Weather-Mobility-Analysis-using-Machine-Learning.pdf?sequence=1.

6 Vanessa Brum-Bastos, Marcelina Łoś, Jed A. Long, Trisalyn Nelson, and Urška Demšar. Context-aware movement analysis in ecology: a systematic review. *International Journal of Geographical Information Science*, 36(2):405–427, 2022. doi:10.1080/13658816.2021.1962528.

7 Vanessa S. Brum-Bastos, Jed A. Long, and Urška Demšar. Weather effects on human mobility: A study using multi-channel sequence analysis. *Computers, Environment and Urban Systems*, 70:1–17, September 2018. doi:10.3233/AIC-2008-0431.

8 F. Calabrese, G. Di Lorenzo, and C. Ratti. Human mobility prediction based on individual and collective geographical preferences. In *13th International IEEE Conference on Intelligent Transportation Systems*, pages 312–317, 2010. doi:10.1109/ITSC.2010.5625119.

9 Dandan Chen, Yong Zhang, Liangpeng Gao, Nana Geng, and Xuefeng Li. The impact of rainfall on the temporal and spatial distribution of taxi passengers. *PLOS ONE*, 12(9), September 2017. doi:10.1371/journal.pone.0183574.

10 Eunjoon Cho, Seth A Myers, and Jure Leskovec. Friendship and mobility: user movement in location-based social networks. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1082–1090, 2011.

11 M. Cools, E. Moons, L. Creemers, and G. Wets. Changes in Travel Behavior in Response to Weather Conditions: Whether Type of Weather and Trip Purpose Matter? *Journal of the Transportation Research Board*, 2157:22–28, 2010. doi:10.3141/2157-03.

12 Johannes De Groeve, Nico Van de Weghe, Nathan Ranc, Tijs Neutens, Lino Ometto, Omar Rota-Stabelli, and Francesca Cagnacci. Extracting spatio-temporal patterns in animal trajectories: An ecological application of sequence analysis methods. *Methods in Ecology and Evolution*, 7(3):369–379, 2016. doi:10.1111/2041-210X.12453.

13 Marco Fiore, Panagiota Katsikouli, Elli Zavou, Mathieu Cunche, Françoise Fessant, Dominique Le Hello, Ulrich Matchi Aivodji, Baptiste Olivier, Tony Quertier, and Razvan Stanica. Privacy in trajectory micro-data publishing: A survey. *Transactions on Data Privacy*, 13(2):91–149, 2020. arXiv:1903.12211.

14 Zhan Guo, Nigel Wilson, and Adam Rahbee. Impact of Weather on Transit Ridership in Chicago, Illinois. *Transportation Research Record: Journal of the Transportation Research Board*, 2034:3–10, December 2007. doi:10.3141/2034-01.

15 Andrea Hess. A Data-driven Human Mobility Modeling: A Survey and Engineering Guidance for Mobile Networking. *ACM Computing Surveys (CSUR)*, 48(8), 2015. doi:10.1145/0000000.0000000.

16 Sibren Isaacman, Richard Becker, Ramón Cáceres, Stephen Kobourov, Margaret Martonosi, James Rowland, and Alexander Varshavsky. Identifying important places in people's lives from cellular network data. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 6696 LNCS, pages 133–151. Springer, Berlin, Heidelberg, 2011. doi:10.1007/978-3-642-21726-5_9.

17 Sibren Isaacman, Richard Becker, Ramón Cáceres, Margaret Martonosi, and I Computing Methodologies Simulation. Human Mobility Modeling at Metropolitan Scales: Spatial and Temporal Parameters for Mobility Modeling. In *MobiSys '12 Proceedings of the 10th international conference on Mobile systems, applications, and services*, pages 239–252, 2012.

18 Shan Jiang, Joseph Ferreira, Marta C González, Fei Wang, Hanghang Tong, Phillip Yu, Charu S Aggarwal Jiang, J Ferreira, and M C González. Clustering daily patterns of human activities in the city. *Data Min Knowl Disc*, 25:478–510, 2012. doi:10.1007/s10618-012-0264-z.

19 Massimiliano Luca, Gianni Barlacchi, Bruno Lepri, and Luca Pappalardo. Deep Learning for Human Mobility: a Survey on Data and Models. *CoRR*, 2020. arXiv:2012.02825.

20 Luca Pappalardo and Filippo Simini. Data-driven generation of spatio-temporal routines in human mobility. *Data Mining and Knowledge Discovery*, 32(3):787–829, 2018.

21 Peter J. Rousseeuw. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20:53–65, 1987. doi:10.1016/0377-0427(87)90125-7.

22 G. Sánchez-Barroso, J. González-Domínguez, J. García-Sanz-Calcedo, and M. Sokol. Impact of weather-influenced urban mobility on carbon footprint of Spanish healthcare centres. *Journal of Transport and Health*, 20, March 2021. doi:10.1016/j.jth.2021.101017.

23 Katarzyna Siła-Nowicka and A. Stewart Fotheringham. Calibrating spatial interaction models from GPS tracking data: An example of retail behaviour. *Computers, Environment and Urban Systems*, 74:136–150, March 2019. doi:10.1016/j.compenvurbsys.2018.10.005.

24 Katarzyna Sila-Nowicka, Jan Vandrol, Taylor Oshan, Jed A. Long, Urška Demšar, and A. Stewart Fotheringham. Analysis of human mobility patterns from GPS trajectories and contextual information. *International Journal of Geographical Information Science*, 30(5):881–906, May 2016. doi:10.1080/13658816.2015.1100731.

25 Kamil Smolak, Witold Rohm, Krzysztof Knop, and Katarzyna Siła-Nowicka. Population mobility modelling for mobility data simulation. *Computers, Environment and Urban Systems*, 84(January), 2020. doi:10.1016/j.compenvurbsys.2020.101526.

26 Kamil Smolak, Katarzyna Sila-Nowicka, and Witold Rohm. Towards anonymous mobility data through the modelling of spatiotemporal circadian rhythms. In *LBS 2019; Adjunct Proceedings of the 15th International Conference on Location-Based Services/Gartner, Georg; Huang, Haosheng*. Wien, 2019.

27 Chaoming Song, Tal Koren, Pu Wang, and A-L Barabasi. Modelling the scaling properties of human mobility. *Nature Physics*, 6(10):1–6, 2010. doi:10.1038/NPHYS1760.

28 Chaoming Song, Tal Koren, Pu Wang, and Albert-László Barabási. Modelling the scaling properties of human mobility. *Nature physics*, 6(10):818–823, 2010.

29 Victor W Stover, Nygaard Consulting Associates, and Seattle D Edward McCormack. The Impact of Weather on Bus Ridership in Pierce County The Impact of Weather on Bus Ridership in Pierce County, Washington. *Journal of Public Transportation*, 15(1):95–110, 2012. URL: https://www.nctr.usf.edu/wp-content/uploads/2012/04/JPT15.1Stover.pdf.

30 P. Tucker and J. Gilliland. The effect of season and weather on physical activity: A systematic review. *Public Health*, 121(12):909–922, 2007. doi:10.1016/j.puhe.2007.04.009.

Qualitative Spatial Reasoning over Questions

Mohammad Kazemi Beydokhti[1]
Department of Geospatial Science, RMIT University, Melbourne, Australia

Matt Duckham
Department of Geospatial Science, RMIT University, Melbourne, Australia

Yaguang Tao
Department of Geospatial Science, RMIT University, Melbourne, Australia

Maria Vasardani
Department of Geospatial Science, RMIT University, Melbourne, Australia

Amy Griffin
Department of Geospatial Science, RMIT University, Melbourne, Australia

Abstract

Although geospatial question answering systems have received increasing attention in recent years, existing prototype systems struggle to properly answer qualitative spatial questions. In this work, we propose a unique framework for answering qualitative spatial questions, which comprises three main components: a geoparser that takes the input questions and extracts place semantic information from text, a reasoning system which is embedded with a crisp reasoner, and finally, answer extraction, which refines the solution space and generates final answers. We present an experimental design to evaluate our framework for point-based cardinal direction calculus (CDC) relations by developing an automated approach for generating three types of synthetic qualitative spatial questions. The initial evaluations of generated answers in our system are promising because a high proportion of answers were labelled correct.

2012 ACM Subject Classification Theory of computation → Semantics and reasoning

Keywords and phrases Qualitative spatial reasoning, geospatial question answering, Qualitative spatial questions

Digital Object Identifier 10.4230/LIPIcs.COSIT.2022.18

Category Short Paper

Supplementary Material *Software (Source Code)*: `https://github.com/MohammadUT/QSR-QA` archived at `swh:1:dir:fbe89a2479977a64c2fb15a1a10a7592fe3bd1ab`

1 Introduction

Qualitative spatial reasoning (QSR) in knowledge representation and reasoning (KR) deals with knowledge about the discrete, imprecise, and non-numerical properties of space and time. Humans' common-sense understanding of space is more connected with the concept of qualitative reasoning, as opposed to quantitative reasoning. Qualities are conceptually simpler than quantities (e.g., "tall" versus 1.93m), they can be obtained from quantities (e.g., bearing of 274 ◦ is "left"), and they generally correspond to discontinuities that are salient to humans (e.g., "left" versus "right," "in front" versus "behind") [4, 10].

Recently, a thorough classification of GeoQA systems based on the type of questions they can answer has been proposed [7]. Answering natural language qualitative spatial questions, which is the focus of the current study, has been studied using approaches that are mainly based on retrieving answers over linked geospatial data sources, such as DBpedia, GADM,

[1] Corresponding author

and OSM [8, 11]. In this paper, we go beyond these approaches by proposing a qualitative spatial reasoning framework for answering such questions, considering certain evidence in qualitative spatial scenes. For example, in the question 'Is A left of C?', if the supporting evidence is certain, then the qualitative spatial inferences may be certain (e.g., if A is left of B and B is left of C, then A is certainly left of C).

In this work, we propose a framework for answering qualitative spatial questions in which a reasoning system is integrated into a GeoQA system. This framework has three main components, including: 1) A semantic geoparser which gets the natural language questions and returns the place entities and spatial relationships in a triple format. 2) A reasoning system which reasons based on an evidence database. 3) Answer extraction which post-processes the generated answers from the reasoning system to return the best possible solutions. In addition, we provide an automated approach to generate three types of synthetic qualitative spatial questions, including Finding Relations (QType-1), Finding Places (QType-2), and Yes/No (QType-3). To make useful inferences over the queries, this study assumes that enough evidence is fed to the reasoning engine.

2 Related works

Expressions of spatial relations serve as the basis of reasoning on spatial data. Researchers have proposed extensive qualitative spatial calculi on the expression of spatial relations. A thorough survey of qualitative spatial and temporal calculi along with their computational properties has been developed by [3]. The inherently imprecise nature of many types of spatial information can be represented and modelled with qualitative spatial reasoning. Qualitative spatial calculi provide the basis for automated qualitative spatial reasoning with complex spatial scenes. For example, the RCC represents the topological relation between two regions without needing their precise quantitative locations or geometries.

Different approaches have been proposed for answering qualitative spatial questions. Early work answered three classes of qualitative and quantitative questions, including proximity, containment, and crossing, by reasoning over DBpedia [11]. This work was further developed by [8], who developed a GeoQA architecture with three main components: identifying the required instances from each question in their own proposed benchmark dataset, translating each formulated question into a SPARQL/GeoSPARQL query, and finally executing each correctly generated query over the linked geospatial data sources. Their proposed system did not perform well as it could only correctly answer 22% of all questions.

Extracting geospatial information that is widely used in queries (e.g., place names, spatial relationships, and place types [9, 6] from unstructured natural language text is one of the main reasons for analyzing questions. [5] proposed a semantic encoding approach in which various semantics such as place names and spatial relationships, among others, are extracted from natural text. This approach, derived from part-of-speech tagging and pretrained named entity recognition (NER) models in the AllenNLP library, has performed well for extracting all their considered entities except events, which are completely missing.

In this study, we integrated the above components into one framework that takes qualitative spatial questions as inputs, extracts the spatial information from the question in the form of triples, performs qualitative spatial reasoning under certain situations, and finally, generates the best possible answer(s) to each question.

3 Methodology: Experimental design

Synthetic question dataset

A question corpus is important to this research because it serves two main purposes: 1) understanding what types of questions are asked and what are their typical answers; and 2) serving as a gold standard for QA system evaluation. The ideal question corpus to use in this study should satisfy two conditions: having a sufficient number of qualitative spatial questions and having an accompanying supporting evidence database including spatial relations between spatial entities in the questions.

In order to prepare a question corpus that met our criteria, we developed an automated approach to generate three types of synthetic qualitative spatial questions: asking for the spatial relation between two features (Q-Type1), asking for feature(s) of a given class that have a particular spatial relation with another feature (Q-Type2), and asking whether a feature has a spatial relation with another feature (Q-Type3) (Table 1). These are three basic and widely used categories that have been also discussed in previous research by [8]. In our question-generation approach, 500 place names in and around the Melbourne CBD area are selected from the Gazetteer of Australia database[2]. Next, the recursive algorithm randomly selects a small number of key places and checks whether the points are well distributed. It is important to have a well distributed set of points as it is more in accordance with the distribution of the rest of the points that we are going to find in our answer.

To generate our corpus, the algorithm uses four randomly selected places, test whether they are well distributed, and if so, generates a simulated query for each question pattern following its corresponding general question template presented in Table 1. For the sake of testing our system, 1000 synthetic queries based on the CDC qualitative spatial logic [1] were generated for each question type.

Table 1 General question template for each question type, illustrated with an example.

Question type	Question template	Example
Q-Type1	What is the spatial relation between *<feature>* and *<feature>?*	What is the spatial relation between London and Manchester?
Q-Type2	Which features of type <X> are *<spatial relation>* of *<feature>*	Which county is east of county Dorset?
Q-Type3	Is *<feature>* *<spatial relation>* of *<feature>?*	Is Hampshire north of Berkshire?

Supporting evidence database

The qualitative spatial reasoning process described in this paper considers a crisp reasoning scenario, where we have certain questions and certain evidence sets, but the answers could be uncertain. In this case where the relations between features are exactly known, we store all the CDC relations between all places in each configuration using the qualification process proposed in [10]. The certain evidence database is finalized by applying the following post-processing steps: 1) Removing mutually inferable relations (composition relations). For example, if we have "A is north of B" and "B is north of C" relations, then the relation between A and C is not stored. 2) Removing converse relations. For instance, if we have the relation between A and B, then the relation of B and A is not stored. 3) Removing any relations that we want to infer in the questions. For example, in the question "Is A southeast of B", the relation between A and B is not included in the evidence database.

[2] `https://placenames.fsdf.org.au/`

Extracting spatial semantics from text

Here, we extract the required spatial information from questions and then presented them in a triple format, including geographical objects and the spatial relation between them. Place names and CDC relations are the required semantics that need to be extracted for each question. There are several available pre-trained NER models developed by open-source NLP libraries. We used the BERT-based model from the DeepPavlov library, which can recognize up to 19 different entities [2].

Qualitative spatial reasoning

To reason over the synthetic queries and make inferences for each of them, we used the open-source SparQ toolbox. for the crisp reasoning scenario when both questions and evidence sets are certain. As the SparQ shell commands are directly callable and executable in Python, this allowed us to integrate it with our GeoQA system. The SparQ reasoning system infers any possible relation among each pair of entities in each configuration when their relations are not known in the evidence database.

Answer extraction

The inference results in the previous step need to be refined in order to identify the best possible answer(s) for each question pattern. In addition, this stage enables us to better evaluate the results of our system for each question type with the benchmark question corpus containing actual answers. For QType-1 where the two place names that we want to infer are mentioned in the questions, we extract the generated relations between the place names in SparQ. For QType-2 where one place name and the cardinal relation are known in the question, all possible relations between the known place name and each of the key place names in each configuration are generated by SparQ. Then, any place names contained in the known relation in the question are retrieved. For QType-3, where two place names and their cardinal relation are known in the question, we extract the generated relations between the place names in SparQ and then check whether the known relation in the question is in agreement with them, and if yes, it returns 'YES', and vice versa.

4 Results and discussions

Because we have stored all the relations between place names for each configuration in the evidence database, answering any of the three types of qualitative spatial questions for the place names in the database is an information retrieval (IR) task, which is not in line with the goal of this work, making inferences over qualitative spatial queries. To address this, we add a new place name (Place P) in the study area which does not exist in our place name database. In addition, its cardinal relation is only known to one or two existing place names in each configuration and no information is available about the relation of other place names with Place P. The aim then is to infer these unknown relations based on the known evidence database. By including the new Place P, all of the simulated questions require a qualitative spatial reasoning process in order to answer them, and the answers are not directly retrievable from the evidence database. These simulated questions are freely available in our GitHub repository[3].

[3] https://github.com/MohammadUT/QSR-QA

Next, we extract place names and spatial relations from text and then structure them in a triple format as *(place name 1, place name 2, relation)* and do the following analysis. We refer to this triple as a 'query sentence' in equation 1. The evaluation of the semantic encoding tool's accuracy is conducted at the sentence level and is defined as the ratio of the number of correctly extracted query sentences from each question to the total number of questions.

Here, we only address the accuracy for QType-1 and QType-3 questions, because we do not have any place names in QType-2 because that question type is about finding places that are in a known spatial relation with the unknown Place P. We used the NER model to extract a query sentence for each question, then matched them with true labels to measure the extraction accuracy. Our model performs well for correctly detection of query sentences, where the accuracies obtained for QType-1 and QType-3 are 79.10% and 82.30%, respectively.

SparQ requires the certain evidence sets for each configuration from the evidence database as well as the extracted query sentence from the NER models to generate possible answers to each qualitative spatial question in each question pattern. We would expect certain and uncertain answers from SparQ. The excerpt outputs of SparQ for the first five configurations are provided in Tables 3 to 5 for QType-1, QType-2, and QType-3, respectively. For the sake of simplicity, place IDs have been utilized instead of their corresponding place names.

Table 2 shows that for QType-1, SparQ predicted all possible relations for the requested query sentence with equal level of importance. For example, the generated relations between places 499 and 501 in C4 could be south, southeast, or southwest and no priority ordering has been made here. By comparing these results with the corresponding actual answer, we get southeast, meaning the correct answer is a subset of the answers identified by SparQ. Blank rows indicate that the NER model was unable to successfully extract the query sentence, which results in null values for the reasoning step of the methodology as well.

Table 2 Inference results for the first five configurations of QType-1 for the SparQ system along with their actual answers and extracted query sentences.

Configurations	Extracted query sentence	SparQ inferences	Actual answers
C1	(18, 501,?relation)	(18, 501, E EQ N NE NW S SE SW W)	(18, 501, NW)
C2	–	–	(415, 501, SE)
C3	–	–	(343, 501, SE)
C4	(499, 501,?relation)	(499, 501, S SE SW)	(499, 501, SE)
C5	(484, 501,?relation)	(484, 501, E EQ N NE NW S SE SW W)	(484, 501, NW)

Considering the results of QType-2 in Table 3, SparQ generates all possible places that could be in a specific relation with the unknown place 501. Taking the C2 configuration as an example, 415 and 360 are possible places inferred from SparQ that could be southeast of 501.

Based on the QType-3 results presented in Table 4, in some cases, SparQ infers all possible relations for each extracted query sentence. For example, to check whether 365 is northwest of 501 in C4, SparQ infers with north, northeast, and northwest relations. Cross checking these outputs with 'YES' as the actual answer, the final answer from the reasoner could be YES, as northwest is one of the generated relations.

In the final stage, we evaluated the accuracy of the generated answers obtained from the SparQ system for the three question patterns in terms of their closeness to the actual answers. To accomplish this, we have defined three categories to characterize the correctness

Table 3 Inferred places from SparQ for the first five configurations of QType-2, along with the actual answers and the extracted query sentence.

Configurations	Extracted query sentence	SparQ inferences	Actual answers
C1	(?Places, 501,NW)	(18, 501, NW) (379, 501, NW)	(18, 501, NW)
C2	(?Places, 501,SE)	(415, 501, SE) (360, 501, SE)	(415, 501, SE)
C3	(?Places, 501, NW)	(343, 501, NW) (326, 501, NW)	(326, 501, NW)
C4	(?Places, 501,SE)	(499, 501, SE)	(499, 501, SE)
C5	(?Places, 501,SE)	(484, 501, SE) (343, 501, SE)	(343, 501, SE)

Table 4 I Inference results from the first five configurations of QType-3 for the SparQ system along with the actual answers and extracted query sentences.

Configurations	Extracted query sentence	SparQ inferences	Actual answers
C1	(379, 501, NE)	(379, 501, E EQ N NE NW S SE SW W)	(379, 501, NO)
C2	–	–	(360, 501, NO)
C3	(326, 501, SW)	(326, 501, E EQ N NE NW S SE SW W)	(326, 501, NO))
C4	(365, 501, NW)	(365, 501, N NE NW)	(365, 501, YES)
C5	–	–	(343, 501, NO)

of generated answers, including Correct, Incorrect, and Uninformative answers. A Correct answer is defined when the answer generated for each question is either a complete or partial match with the relevant actual answer. Taking C4 in Table 3 as an example, the SparQ inference results are labeled as correct as the SparQ-generated relations partially match with the actual relation. An Incorrect answer is tallied when the actual answer is neither a complete match with generated answer nor is found among the generated answers, for instance, when the SparQ inference result for two places is (Place1, Place 2, 'n ne nw') but its corresponding actual answer is se. Finally, an uninformative answer is those cases in which the generated answers do not provide any useful information about whether they are correct or incorrect. Considering C1 in Table 3 as an example, SparQ infers that the relation between the queried places could be any of the cardinal directions.

We represent the number of questions in each type that fall under the three answer correctness categories for the DeepPavlov-based NER model[4]. By considering Figure 3 in general, a high proportion of generated answers in all question types is labelled correct, meaning that the overall performance of each was acceptable. By focusing on each question pattern in particular, in QType-1, where the answers are relations between an unknown place and known places, SparQ generated answers that are either correct or uninformative, with no incorrect answers generated. For QType-2 where the places retrieved from the reasoners are checked with the corresponding actual places, SparQ answered all questions correctly. Finally, in QType-3 where the inferred YES/NO answer is checked with the actual answer, three forms of answers are found here, but the incorrect category contained the fewest answers. The overall performance of the system depends highly on the place semantics extraction step, which means the more questions from which we could successfully extract place semantics, the greater the chance that the reasoner will return a correct answer.

[4] https://github.com/MohammadUT/QSR-QA/blob/main/Figure%203.jpg

5 Conclusion and Future Work

This paper has addressed the problem of answering qualitative spatial questions by presenting a GeoQA system based on deductive spatial reasoning. To achieve this goal, this system begins by taking three types of qualitive spatial questions, extracting toponyms and spatial relations from the question text, applies crisp qualitative spatial reasoner to each question, and finally, generating final answers for each question type. To evaluate our system, we have compared the results obtained from all simulated question types with a benchmark including correct answers. The results have shown that our system performed well with in all three types of questions when the SparQ reasoner has been fed by a sufficient number of evidence sets. Initial results showed that there is the possibility of adapting our approach to addressing other spatial relation logics, a result of which is that more diverse types qualitative spatial questions can be answered.

References

1 UF Andrew, D Mark, and D White. Qualitative spatial reasoning about cardinal directions. In *Proc. of the 7th Austrian Conf. on Artificial Intelligence. Baltimore: Morgan Kaufmann*, pages 157–167, 1991.

2 Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint*, 2018. `arXiv:1810.04805`.

3 Frank Dylla, Jae Hee Lee, Till Mossakowski, Thomas Schneider, André Van Delden, Jasper Van De Ven, and Diedrich Wolter. A survey of qualitative spatial and temporal calculi: algebraic and computational properties. *ACM Computing Surveys (CSUR)*, 50(1):1–39, 2017.

4 Antony Galton et al. *Qualitative spatial change*. Oxford University Press on Demand, 2000.

5 Ehsan Hamzei, Haonan Li, Maria Vasardani, Timothy Baldwin, Stephan Winter, and Martin Tomko. Place questions and human-generated answers: A data analysis approach. In *International Conference on Geographic Information Science*, pages 3–19. Springer, 2019.

6 Mohammad Kazemi Beydokhti, Matt Duckham, Amy Griffin, and Vedran Kasalica. Geo-event question answering systems: A preliminary research study. In *Proceedings of the 11th International Conference on Geographic Information Science (GIScience 2021)*, page 6, 2021.

7 Gengchen Mai, Krzysztof Janowicz, Rui Zhu, Ling Cai, and Ni Lao. Geographic question answering: Challenges, uniqueness, classification, and future directions. *AGILE: GIScience Series*, 2:1–21, 2021.

8 Dharmen Punjani, Kuldeep Singh, Andreas Both, Manolis Koubarakis, Iosif Angelidis, Konstantina Bereta, Themis Beris, Dimitris Bilidas, Theofilos Ioannidis, Nikolaos Karalis, et al. Template-based question answering over linked geospatial data. In *Proceedings of the 12th Workshop on Geographic Information Retrieval*, pages 1–10, 2018.

9 Mark Sanderson and Janet Kohler. Analyzing geographic queries. In *SIGIR workshop on geographic information retrieval*, volume 2, pages 8–10, 2004.

10 Diedrich Wolter and Jan Oliver Wallgrün. Qualitative spatial reasoning for applications: New challenges and the sparq toolbox. In *Geographic Information Systems: Concepts, Methodologies, Tools, and Applications*, pages 1639–1664. IGI Global, 2013.

11 Eman MG Younis, Christopher B Jones, Vlad Tanasescu, and Alia I Abdelmoty. Hybrid geo-spatial query methods on the semantic web with a spatially-enhanced index of dbpedia. In *International Conference on Geographic Information Science*, pages 340–353. Springer, 2012.

Transcepts: Connecting Entity Representations Across Conceptual Views on Spatial Information

Eric J. Top[1]
Department of Human Geography and Spatial Planning, Utrecht University, The Netherlands

Simon Scheider
Department of Human Geography and Planning, Utrecht University, The Netherlands

Abstract

Analysts interpret geographic and other spatial data to check the validity of methods in reaching an analytical goal. However, the meaning of data is elusive. The same data may constitute one concept in one view and another concept in another. For example, the same set of air pollution points may be regarded as field values if they are considered pollution measurements and objects if they are considered locations of measurement devices. In this work we adopt a framework of conceptual spaces and viewpoints and show how entity representations in one semantic interpretation may be related to entity representations in others in terms of what we call transcepts. A transcept captures which things represent the same entity. We define and use transcepts in the framework to explain how different views of geographic data may relate to one another.

2012 ACM Subject Classification Computing methodologies → Knowledge representation and reasoning; Computing methodologies → Spatial and physical reasoning

Keywords and phrases Transcept, Spatial Information, Knowledge Representation, Conceptual Space, View, Point Of View, Viewpoint, Object, Event, Network, Field, Relation

Digital Object Identifier 10.4230/LIPIcs.COSIT.2022.19

Category Short Paper

Funding This work was supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement no. 803498 (QuAnGIS)).

1 Introduction

Geodata analysts usually have a choice between multiple valid conceptualizations of their data. As a result, different analysts may have different interpretations, which could lead to disagreement about the underlying concepts. Considering conceptual discussions are usually at high levels of complexity and abstraction, finding common ground is challenging. Also, for the automation of analytical tasks, e.g., with artificial intelligence, knowledge representations need to align with the conceptual view of the analyst. Understanding the interpretations of analysts and how they align is important for both these problems. Next to knowledge representation, there is a need for entity representation. In other words, we need *transcepts*.

Before we explain what transcepts are, briefly consider the word *concept*. It can be traced back to the Latin verb *concipere*. This verb can be dissected into the prefix *con-*, which means approximately *with* or *together*, and the verb *cipere* (or *capere*), which roughly translates to *take*, *take on* or *take in*. According to this, a concept can thus be understood as something that is, e.g., *taken with*, *taken together* or *taken on with*. Similar constructions of a prefix and the suffix *-cept* are found in the words *deception*, *perceptual*, *receptor* and *acceptance*, and in each case the prefix seems to add additional meaning to the process of *taking*.

The word *transcept* also has this structure. The prefix *trans-* is best translated to *across* or *over*. For example, the term *transdisciplinary* means *across* or *over disciplines* and the term *transgender* means *across* or *over genders*. In similar fashion, a transcept can be

[1] Corresponding author

understood as something that is taking [things] across [something or someplace]. The notion of transcepts is useful as a connection between different representations of the same entities or phenomena in different interpretations. For example, a well-known comic circulating the internet (See Figure 1) depicts how two valid conceptualizations of a shape are incompatible. In one conceptualization the shape seems to form the symbol *6*, while in another it forms the symbol *9*. In a single interpretation, these conceptualizations are not compatible. An interpretation of the shape as *6* contradicts any interpretation of the shape as *9*. However, it is still useful to note that the shape may be both *6* or *9*, e.g., when considering possibilities or hypothetical scenarios. Things that form a transcept do not necessarily contradict one another. It could for example also be useful to know that two people standing side-by-side interpret the shape as being *6*.

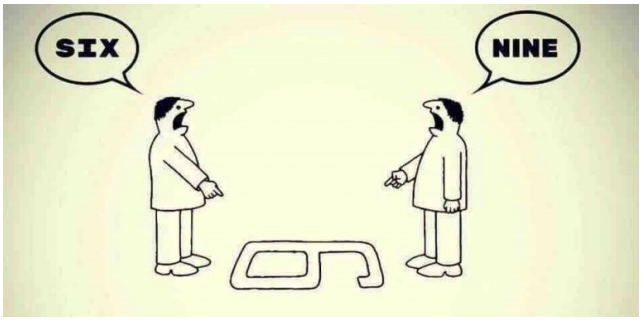


Figure 1 Two incompatible concepts in two interpretations represent one shape on the ground. The concepts *6* and *9* form a transcept across the two interpretations. (Author of image unknown.)

In the practice of geographic information it also occurs that a single entity or phenomenon can be interpreted in multiple valid ways. We give three (hypothetical) example cases and we elaborate on them after we establish the conceptual space framework. The example cases are:

- **Volcano eruption**: A two-dimensional cartographic view may hold *where* a volcano eruption took place and a temporal view (e.g., with a calendar) may hold *when* the eruption took place. In terms of conventional geographic information concepts, in the first view the volcano eruption is an object and in the second view it is an event. However, the object on the cartographic map and the event on the calendar represent the same eruption.
- **Trees in the Amazon rain forest**: If we assume that the Amazon rain forest is identifiable by the set of its trees, then it can be identified by collecting all trees that are part of it [7]. However, the set of all trees is not equal to the rain forest because the latter is atomic (e.g., half of the Amazon rain forest is not the Amazon rain forest). Nonetheless, they do represent the same phenomenon.
- **Road network**: A road network can be considered a relation over a set of street junctions. It can also be considered a set of objects, because each of these roads are tangible and have qualities, e.g., they may be paved with concrete. The relation and the set of objects both represent the roads.

Hautamäki [9] proposes a knowledge representation framework with a conceptual space that may be partitioned into various views. A conceptual space is a geometric structure that may be used for knowledge representation and views are in this respect partial or incomplete substructures of the conceptual space. In this paper we introduce transcepts as a notion in geo-analytical cognition in context of conceptual spaces. We first give a quick background of conceptual spaces and shortly reflect on the importance of concepts to geographic information.

Following, we provide examples of transcepts in geography. We then define transcepts in the philosophical framework proposed by [9]. In this framework a transcept serves as a connection between different things that represent the same thing across these various partitions. We then show how the conceptual space framework can be used to model the transcept examples.

2 Background: Conceptual spaces and geographic concepts

Gärdenfors [5] introduces conceptual spaces as an alternative to the symbolic and associative approaches to knowledge representation. The idea of a conceptual space is that concepts can be represented as regions of objects in a space consisting of one or more quality dimensions. For example, a quality dimension "taste" could host the qualities "salty", "sour", "sweet", "bitter" and "umami" and a quality dimension "physical state" could have the qualities "solid", "liquid", "gaseous" and "plasma". Then the concept of sweet liquids would be the set of all objects with the qualities "sweet" and "liquid". A conceptual space is a metric space, meaning there is a notion of distance between qualities and concepts, and thereby implicitly a topological space, meaning there is a notion of neighborhoods of concepts.

Conceptual space-based learning models are shown to outperform models based on multidimensional feature spaces [12] and conceptual spaces have been used for a variety of applications, including spatial cognition [1], AI-learning, and (vague) classification and categorization [3]. Hautamäki [9] provides an alternative framework that makes it possible to partition a conceptual space into multiple "points of view". We use Hautamäki's framework to model how analysts may hold different views with regard to the same data.

A search for concise and correct conceptualizations characterizes many theoretical contributions to GI-science. This search particularly took off after the publication of Couclelis' work [2], which redirected a discourse on syntactic data types to one on objects and fields, two important semantic concepts of geographic information. Goodchild, Yuan and Nova [6] argue that all concepts in geographic information science are generalizable to so-called geo-atoms. Galton [4] extents the discourse beyond spatial concepts and suggests a temporal framework of concepts with a "*process-priority view*" (p.1). Kuhn [10] and Kuhn and Ballatore [11] introduce a set of core concepts of geographic information.

3 Conceptual spaces and views

A conceptual space is an abstract notion that encompasses all concepts given some determination base, i.e., a base structure that establishes the building blocks of concepts and relations between them. A view is a structure that limits the elements in the determination base to those "within view". Those "outside of view" either merge into a single element or simply stay out of consideration. If two different views are based on the same determination base, they can be compared by means of the elements of the base. In this section conceptual spaces and views are defined in more detail using Hautamäki's work [9].

A conceptual space can be defined with respect to a *determination base*, which is a structure $\langle I, D, E, S\rangle$, where I is a set of *determinables*, D is a set of *determinates*, E is a set of *entities* and S is the so-called *state function* $S : E \to D^I$. The codomain D^I is the *conceptual space* for the entities of E. The notation of D^I denotes a set of functions from the determinables I to the determinates D, i.e., $D^I := \{f | f : I \to D\}$. An example of this is a function $Belgium : I_B \to D_B$ where $I_B \subseteq I$ and $D_B \subseteq D$ such that for example $Size \in I_B$ and 30689 $km^2 \in D_B$. Then an instance of this function could be $Belgium(Size) = 30689\ km^2$. An element of D^I is called a *state* and any set of states, i.e., any subset of D^I, is called a *concept*. Note that concepts can be subsets of other concepts. As such, they form a concept lattice (c.f., [13]).

According to Hautamäki [9], a conceptual space can be approached from different points of view or viewpoints. Viewpoints can be defined as structures of subsets of determinables and theories. We choose to refer to viewpoints simply as *views*. More specifically, a *view* relative to some determination base is a structure $V = \langle K, T \rangle$ where $K \subseteq I$ and $T \subseteq D^K$. The set of functions T is called a *theory* and the set D^K is a *subspace* of the conceptual space D^I. For example, a temporal view is one where some time determinable (e.g., *date*) is an element of K and the state function and the theory relates certain entities (e.g., birthdays) to certain determinates, e.g., some dates. A view-specific state function S_K for the subspace D^K is defined as follows: $S_K := \{(x, y) \in S | x \in K\}$.

Each view has a *scope*, which is the set of entities that have some distinguishable state within the view. The notion of scope is necessary because entities may be indistinguishable in some views. For example, if a conceptual space has no temporal states, it is impossible to distinguish over time, meaning the observation of a tree at 10 o'clock is indistinguishable from an observation of that same tree at 11 o'clock. A subtle consequence of this is that in a view a single state may represent more than one entity. For example, time could be aggregated to years, meaning that each (non-leap year) year state represents 365 day entities. Two states from different views may correlate with one another, which means they have the same entities in their scope. For example, Peking and Beijing both refer to the capital of China and may be considered synonymous, although the former is of a more historical view while the latter is of a more contemporary view.

Objects, fields, events and networks are common concepts in geographic information [2, 10, 6, 4]. In the context of conceptual spaces these concepts can be defined as mathematical objects from or structures over the elements in the determination base along with a semantic interpretation. With respect to a conceptual space, an object can simply be defined as a state with some geospatial definition. Let x be any state and g the concept of all geospatial things. If $x \in g$, then x is an object. Similarly, x is an event if $x \in t$ where t is the concept of all temporal things. Note that in both cases x is a function. For example, if x represents the 2022 winter olympics, then $x(City) = Beijing$, where $City$ is a determinable and $Beijing$ is a determinate.[2]

Fields and networks can both be defined as relations between sets of states, i.e., between concepts. Any relation between any concept on the one hand and a spatial concept on the other is a field, whereas a relation of a concept to itself is a network. For example, a relation between the concept of temperatures and the concept of locations in Spain could be a temperature measurement function (Conventionally considered a field) and a relation on locations in Spain to itself could indicate connections between those locations (A characteristic of, e.g., any road network).

With respect to the conceptual space framework, we define a transcept as a set of any multitudes of states and concepts and relation tuples between them. We denote a transcept with θ. A transcept suggests that any of its elements represents the same entity in E as all other transcept elements. For example, the state of a particular $Crowd$ and the concept $\{Person_1, Person_2, ..., Person_{500}\}$ may represent the same crowd entity, so a transcept of them could be $\theta\{Crowd, \{Person_1, Person_2, ..., Person_{500}\}\}$. Transcepts thus link representatives of the same entity across different views. If one view includes the state and not the concept and another view includes the concept and not the state, then the transcept serves as a "bridge" between the views.

[2] This second example with Beijing shows how something can be a state in one context and a determinate in another. The instances $x(City) = Beijing$, $Beijing(Country) = China$ and $Country(Beijing) = China$, where $Beijing$ takes each of the three possible roles, could occur within the same view.

4 Modeling the examples with the framework

We can now define a transcept for the volcano eruption case across the cartographic and temporal views. Let *Space* be the cartographic view $\langle A, X\rangle$ where $lat, long \in A$ are latitude and longitude determinables and where $x \in X$ is the state representing the volcano eruption in space. Also, let *Time* be the temporal view $\langle B, Y\rangle$ where B has the determinable *Date* as element and where $y \in Y$ is the state representing the volcano eruption in time. Across these two views $\theta\{x, y\}$ is the transcept of the volcano eruptions. Figure 2 shows a schematization of the example. The determinates that relate to the determinables form quality dimensions and are indicated by the corresponding determinables. Because respectively x is in a spatial view and y is in a temporal view, x is an object and y is an event.

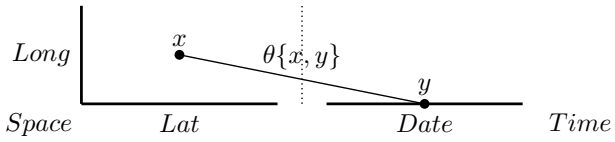


Figure 2 States in two views connected by a transcept.

The example shows how a transcept can bundle multiple states that represent the same entity. That is the case if x and y are actually the same state in the conceptual space. If x and y represent the exact same entity (i.e., $\forall e \in E\ S(e) = x \iff S(e) = y$), then $x = y$ and $\theta\{x\} = \theta\{y\} = \theta\{x, y\}$. However, x and y can be different states because an entity that is represented by one state in one view may need to be represented by multiple states in another. This becomes apparent in the next example.

The Amazon rain forest example can be modeled either using one view or two views. We start by modeling the example with two views. Let $V = \langle K, T\rangle$ be a view where $lat, long \in K$. For the sake of the example, assume that $t_1, t_2, t_3 \in T$ are all the trees in the Amazon rain forest. The concept of the three trees $\{t_1, t_2, t_3\}$ may be denoted as *Trees*. Let $V' = \langle K, T'\rangle$ be another view where $Amazon \in T'$. Then $\theta\{Trees, Amazon\}$ is a transcept across V and V'. The example is visualized in Figure 3. To see that they can also be in one view, let $V* = \langle K, T^*\rangle$ be a third view where $Amazon, t_1, t_2, t_3 \in T^*$. Then $\theta\{Trees, Amazon\}$ is also a transcept across the single view V^*.

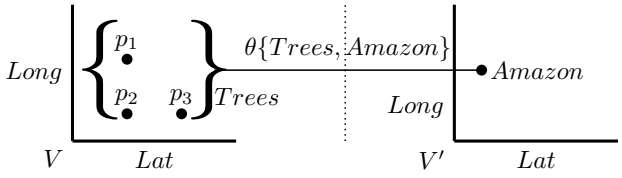


Figure 3 Between state and concept transcept.

This example shows how a concept and a state can represent the same entity. This is useful for understanding how qualities are assigned to collections. It would be confusing to assign qualities *"endangered"* to a set of trees to indicate the Amazon rain forest is endangered, because it is then unclear whether the trees or the forest is endangered. With transcepts the distinction between the set of parts and the whole can be made, thereby handling mereological problems (c.f., [8]) without losing the information that they represent the same thing.

There are at least three different views with which the road network can be modeled. In the first the road network is viewed as a relation over vertices, in the second it is viewed in terms of road objects and in the third it is viewed as a single roads object. Roads can also be modeled as sets of location states (e.g., to model lines), but we choose not to do so here. Let $lat, long \in K$ and let the views be respectively $V^J = \langle K, J \rangle$, $V^O = \langle K, O \rangle$ and $V^N = \langle K, N \rangle$ where $j_1, j_2, j_3 \in J$ are road vertices, $Road_1, Road_2 \in O$ are roads and where $Roads \in N$ is the entire road network. Multiple transcepts can be defined across these views. Across V^J, V^O and V^N we find the transcepts $\theta_1 = \theta\{(j_1, j_2), Road_1, Roads\}$ and $\theta_2 = \theta\{(j_2, j_3), Road_2, Roads\}$. These two transcepts are visualized in Figure 4. Another transcept is found by creating concepts. That is, the concepts $\{(j_1, j_2), (j_2, j_3)\}$ and $\{Road_1, Road_2\}$ and the state $Roads$ all represent the same entity, so we can also define a transcept $\theta_3 = \theta\{\{(j_1, j_2), (j_2, j_3)\}, \{Road_1, Road_2\}, Roads\}$. This third transcept is visualized in Figure 5.

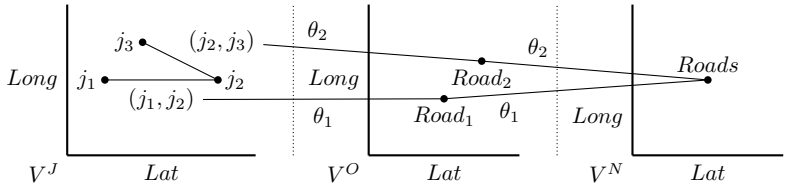


Figure 4 Transcepts of road entities.

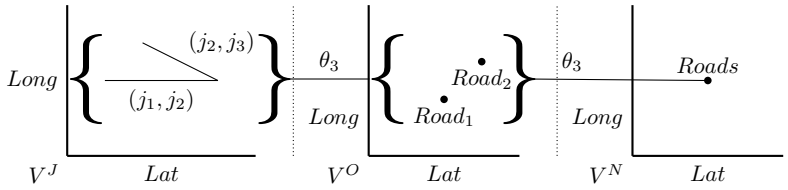


Figure 5 Transcept of network entity.

This example stresses how transcepts are different from concepts. Where a concept is a set of states representing different entities that are understood as one theoretical thing, a transcept is a set of states representing exactly one entity across many understandings of their theory. In the example, $Road_1$ and $Road_2$ are bundled in a set while they represent two different road entities in a single view. The same is impossible for a transcept. On the other hand, in θ_1, $Road_1$ and $Roads$ are part of the same transcept. No concept within the views can have all these elements because no view includes all these elements.

Furthermore, the set of tuples in V^J can be considered an example of the network concept in geographic information, which is characterized by connections between nodes [10].

5 Discussion and conclusion

Transcepts seem to be a new notion in the context of conceptual spaces. Where concepts are instrumental to knowledge representation, transcepts seem to be central notions of entity representation. They are useful for talking about different representations of the same entity without either confusing or forfeiting the semantics of those representations, which are captured in concepts. Interestingly, some of the well-known concepts of geographic information seem to be implicitly represented in Hautamäki's framework. Any objects and events can both be modeled as states in respectively spatial and temporal views, networks

arise from relation tuples between states in the same view, and determinables seem to have applicability comparable to fields. It may prove worthwhile to further investigate these resemblances in future work, as well as to further develop the framework. For instance, it may be useful to have a theory of transcept functions and to extend the notion of concept lattices to transcepts. Also, Hautamäki proposes a logic of points of view which has mostly been ignored in this work, but which could also increase the applicability of transcepts in knowledge and entity representation tasks.

References

1. Benjamin Adams and Martin Raubal. A metric conceptual space algebra. In *COSIT 2009*, volume 5756, pages 51–68. Springer, 2009. doi:10.1007/978-3-642-03832-7_4.
2. Helen Couclelis. People manipulate objects (but cultivate fields): Beyond the raster-vector debate in GIS. In *Theories and methods of spatio-temporal reasoning in geographic space*, volume 639 of *Lecture Notes in Computer Science*, pages 65–77. Springer, 1992. doi:10.1007/3-540-55966-3_3.
3. Igor Douven. Vagueness, graded membership, and conceptual spaces. *Cognition*, 151:80–95, 2016. doi:10.1016/j.cognition.2016.03.007.
4. Antony Galton and Riichiro Mizoguchi. The water falls but the waterfall does not fall: New perspectives on objects, processes and events. *Appl. Ontology*, 4(2):71–107, 2009. doi:10.3233/AO-2009-0067.
5. Peter Gardenfors. Conceptual spaces as a framework for knowledge representation. *Mind and matter*, 2(2):9–27, 2004. doi:10.1017/S0140525X04280098.
6. Michael F. Goodchild, May Yuan, and Thomas J. Cova. Towards a general theory of geographic representation in GIS. *Int. J. Geogr. Inf. Sci.*, 21(3):239–260, 2007. doi:10.1080/13658810600965271.
7. Nicola Guarino and Christopher A. Welty. Identity, unity, and individuality: Towards a formal toolkit for ontological analysis. In *ECAI 2000*, pages 219–223. IOS Press, 2000.
8. Giancarlo Guizzardi. On the representation of quantities and their parts in conceptual modeling. In *Formal Ontology in Information Systems, 2010*, volume 209, pages 103–116. IOS Press, 2010. doi:10.3233/978-1-60750-535-8-103.
9. Antti Hautamäki. Points of view: a conceptual space approach. *Foundations of Science*, 21(3):493–510, 2016. doi:10.1007/s10699-015-9422-2.
10. Werner Kuhn. Core concepts of spatial information for transdisciplinary research. *Int. J. Geogr. Inf. Sci.*, 26(12):2267–2276, 2012. doi:10.1080/13658816.2012.722637.
11. Werner Kuhn and Andrea Ballatore. Designing a language for spatial computing. In *AGILE 2015*, pages 309–326. Springer, 2015. doi:10.1007/978-3-319-16787-9_18.
12. Ickjai Lee and Bayani Portier. An empirical study of knowledge representation and learning within conceptual spaces for intelligent agents. In *ICIS 2007*, pages 463–468. IEEE Computer Society, 2007. doi:10.1109/ICIS.2007.57.
13. Rudolf Wille. Concept lattices and conceptual knowledge systems. *Computers & mathematics with applications*, 23(6-9):493–515, 1992. doi:10.1016/0898-1221(92)90120-7.

A Computational Method for the Classification of Mental Representations of Objects in 3D Space

Samuel S. Sohn
Rutgers University, Piscataway, NJ, USA

Panagiotis Mavros[1]
Singapore-ETH Centre, Future Cities Laboratory, CREATE campus, 1 CREATE Way, #06-01 CREATE Tower, 138602, Singapore

Mubbasir Kapadia
Rutgers University, Piscataway, NJ, USA

Christoph Hölscher
Chair of Cognitive Science, ETH Zürich, Switzerland
Future Cities Laboratory, Singapore-ETH Centre, Singapore

Abstract

The structure mapping task is a simple method to test people's mental representations of spatial relationships, and has recently been particularly useful in the study of volumetric spatial cognition such as the spatial memory for locations in multilevel buildings. However, there does not exist a standardised method to analyse such data and structure mapping tasks are typically analysed by human raters, based on criteria defined by the researchers. In this article, we introduce a computational method to assess spatial relationships of objects in the vertical and horizontal domains, which are realized through the structure mapping task. Here, we reanalyse participants' digitised structure maps from an earlier study (N=41) using the proposed computational methodology. Our results show that the new method successfully distinguishes between different types of structure map representations, and is sensitive to learning order effects. This method can be useful to advance the study of volumetric spatial cognition.

2012 ACM Subject Classification General and reference → Metrics; General and reference → Experimentation; General and reference → Evaluation

Keywords and phrases mental representations of space, spatial cognition, structure mapping task, 3D space, volumetric space

Digital Object Identifier 10.4230/LIPIcs.COSIT.2022.20

Category Short Paper

Funding The research was conducted at the Future Cities Lab, Singapore-ETH Centre and supported by the National Research Foundation, Singapore under the CREATE programme, as well as NSF Awards: IIS-1703883, IIS-1955404, IIS-1955365, RETTL-2119265, and EAGER-2122119.

1 Introduction

For many species, navigation entails movement not only in the horizontal plane, but also, to some extent, in the vertical domain. Aerial species need to coordinate flight to find their nest or food on trees. Underwater species coordinate movement in different depths to seek shelter and resources. Terrestrial species may traverse undulating terrain or climb atop objects and surfaces. As humans, we also climb surfaces, live in multi-storey structures, and routinely organise objects and other information in vertical space.

[1] Corresponding Author

Previous research in spatial cognition has shown that people organise spatial information hierarchically [13], grouping objects based on proximity, visual, or semantic salience [3]. This process is also called "regionalisation", and people use the resulting hierarchies, or regions, to plan spatial behaviour [13]. While extensive research has focused on the cognition of 2D surfaces (layout rooms, buildings, or neighbourhoods), recently research has turned to how people perceive, form mental representations of, and reason about 3D and volumetric spaces; here we are particularly interested in navigating through buildings.

Various methods have been developed to access and assess mental representations of space. These include onsite and offsite pointing, identifying novel shortcuts, sketch-mapping, and others [7]. To assess how people perceive, understand, and utilise 3D spatial relationships, previous research in volumetric spatial cognition has relied mostly on 3D pointing tasks [16, 9] or navigational tasks [6, 8, 4]. One limitation of pointing tasks is that they examine pairwise (or triplewise in the case of judgements of relative direction) spatial relationships; the overall spatial organisation of multiple locations is only indirectly assessed through the accuracy in multiple trials between separate locations. Similarly a limitation of multilevel navigation tasks is that the mental representation is confounded with the immediate spatial information available to the individual, e.g., the visibility of spaces, decision points, landmarks, and signs.

An alternative approach that captures in an abstract manner the spatial organisation of multiple objects is the structure mapping task (SMT). In the SMT, as it has been applied to study volumetric spatial cognition [1, 10], individuals are provided with a set of representations for physical locations (e.g. a small card or object) and are asked to place them on a two-dimensional (2D) surface in a manner that represents how these locations are organised in space.

In this study, we seek to develop a new, computational approach to analyse 2D representations of 3D spatial information. This approach is motivated by and applied to the analysis of structure mapping tasks. In the following sections we describe the analytical framework developed to distinguish between horizontally-biased and vertically-biased representations, and we apply this to behavioural data obtained from a previous study. Below, we summarise key aspects of the previous study, articulate in detail the mathematical description of the new metrics, and use them to assess the representations produced by human participants.

2 Methods

2.1 Data collection

Data collection was conducted as part of a larger experiment which is presented in more detail in [10]; here we include relevant information to assist readers understand the structure mapping task analysis. Participants (N = 41) learned the layout and the location of twelve goal locations (shops) spread across four floors of a large, complex, multilevel building (Figure 1).

Participants were randomly assigned in two spatial learning groups. The horizontal training group learned the locations by walking to all locations in a floor before moving to the next floor. By contrast, the vertical training group walked to the locations by first visiting all locations of a vertical cluster (using the escalators) before moving to the locations of the next cluster (Figure 1). This allowed us to test the effect of spatial knowledge acquisition mode (learning) on mental representation structure. After the training phase, they completed a structure mapping task which is the focus of this paper, and then proceeded with the rest of the experiment [10].

For the structure mapping task, 12 cards representing each goal were shuffled and given to them, with theinstruction *"how do you think they are arranged in space"* without at time limit. Later, they self-reported their sense of direction using SBSOD [5], prior familiarity with the building, as well as demographic information.

3 Analysis

Each goal location (landmark) from the structure mapping task corresponds to a recalled 2D position in each participant's representation of the building. The same landmark is also associated with a real 3D position computed from the digital floorplan of the building. However, the relationships between the recalled positions and real positions learned by participants remain unknown, e.g., whether their representations were tied to a particular perspective of the building. In order to uncover these relationships and assess their accuracy, we propose two analysis techniques catered to the structure mapping task.

Morphological analysis uses *recalled* positions to determine how well a participant's representation conforms to canonical organisations of *real* landmark positions, e.g., by the floor of the building. This requires that an experimenter make assumptions about which landmarks should be canonically grouped. On the other hand, functional analysis uses clustering algorithms to group a participant's recalled positions and determines what type of information the clusters are useful for distinguishing about real positions. This technique requires the experimenter to make assumptions about the utility of clusters instead of the clusters themselves, making it complimentary to morphological analysis.

3.1 Morphological Analysis

A canonical organisation of landmarks is defined as a set of clusters. For multilevel buildings, we consider two types of canonical clusters based on the floors of the building (labeled 1-4) and corridors (labeled A-D), which in our case are consistent between floors. Henceforth, we refer to floor-based clusters as horizontal clusters C^H and corridor-based clusters as vertical clusters C^V. For a given participant, we encoder their clusters $C \in \{C^V, C^H\}$ as a sequence of m clusters and each cluster C_i as a sequence of recalled 2D positions $C_{i,k}$. The morphological analysis quantifies two characteristics of a participant's representation for each type of canonical clustering C: the distinctiveness $\mathcal{D}(C)$ and the alignment $\mathcal{A}(C)$.

3.1.1 Distinctiveness

The distinctiveness metric $\mathcal{D}(C)$ measures the separateness between clusters by considering them as polygons (Eq. 1). In particular, we represent each cluster as a convex hull, i.e., the smallest convex polygon that encompasses all of its recalled positions. The separateness between clusters can then be computed using function $\Delta(C_i, C_j)$, which outputs 0 when the convex hulls of clusters C_i and C_j are intersecting and 1 otherwise. This means that two clusters are only considered separate when their recalled positions are visually separable by a line. The distinctiveness metric $\mathcal{D}(C) \in [0, 1]$ encodes the probability that a cluster is visually separable from all other clusters.

$$\mathcal{D}(C) = \frac{1}{m} \sum_{i=1}^{m} \prod_{j=1}^{m} \Delta(C_i, C_j) \;\Big|\; i \neq j \tag{1}$$

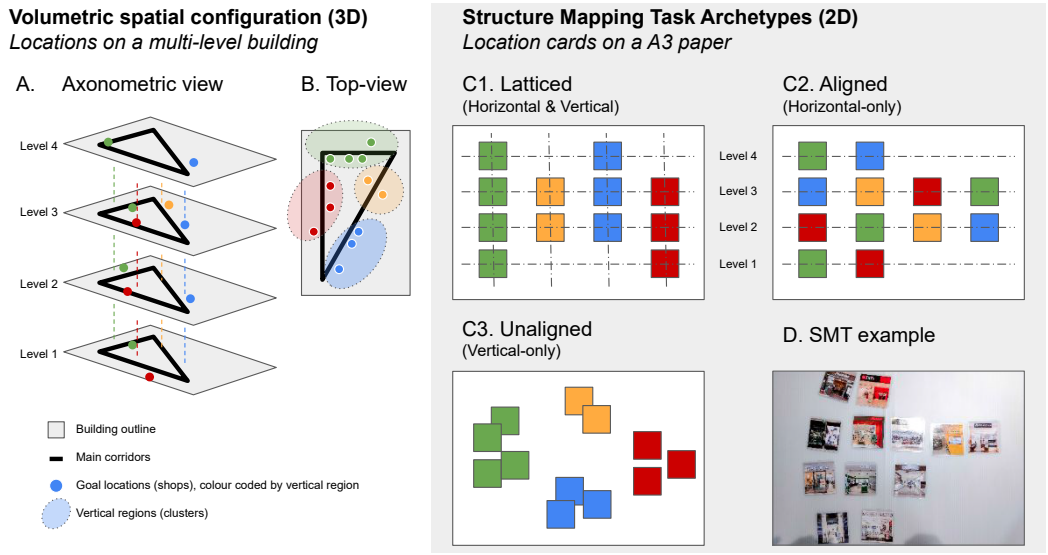


Figure 1 Schematic diagrams of the physical volumetric space, a large multilevel building, in axonometric-view (A) and top-view (B). There were 2 goal locations on level 1 (ground-floor), 4 on level 2, 4 on level 3, and 2 on level 4. (C) Three major archetypes of structure mapping task products (C1–3) that capture horizontal and vertical rationalisations of the same 3D spatial configuration.

3.1.2 Alignment

The alignment metric $\mathcal{A}(C)$ averages two terms: the eccentricity of clusters and the similarity between their orientations (Eq. 2). Both terms rely on Principal Component Analysis (PCA) [15] to compute the principal components $\Psi(C_i)$ and eigenvalues $\Lambda(C_i)$ of each cluster C_i using the distribution of its recalled positions. The first principal component $\Psi(C_i)_1$ for cluster C_i is the axis or unit vector along which the variance in recalled position $\Lambda(C_i)_1$ is maximized. The second principal component $\Psi(C_i)_2$ is orthogonal to the first, and it accounts for the next highest variance $\Lambda(C_i)_2$. In effect, the principal components act as the major and minor axes of an ellipse fitted to the cluster [14].

$$\mathcal{A}(C) = \frac{1}{2m}\sum_{i=1}^{m}\sqrt{1-\frac{\Lambda(C_i)_2^2}{\Lambda(C_i)_1^2}} + \frac{1}{m(m-1)}\sum_{i=1}^{m}\sum_{j=i+1}^{m}\left|\Psi(C_i)_1 \cdot \Psi(C_j)_1\right| \tag{2}$$

The eccentricity term measures the average elongatedness of a cluster using the formula for the eccentricity of an ellipse [12]. For cluster C_i, this is computed as the square root of 1 minus the ratio between $\Lambda(C_i)_2^2$ and $\Lambda(C_i)_1^2$. This can be thought of as the ratio between the major and minor axes of an ellipse. The similarity term computes the average angular similarity between all unique pairs of clusters. The angular similarity between clusters C_i and C_j is formulated as the absolute value of the dot product between their first principal components, $\Psi(C_i)_1$ and $\Psi(C_j)_1$. The absolute value ensures that vectors facing in opposite directions are considered equal. Since alignment $\mathcal{A}(C)$ averages the eccentricity and similarity terms, it is maximized when clusters are both linear and parallel to each other.

3.1.3 Structure Mapping Task Archetypes

Distinctiveness is first measured for both horizontal and vertical clusters to assess whether participants are cognizant of them. For participants with distinctive representations, alignment is then measured to classify the appearance of the representation. When only one

canonical type is distinctive, an *aligned* representation configures the clusters into either rows or columns, and an *unaligned* representation separates clusters into regions independent of any axis. It is possible for a representation to be aligned with respect to multiple canonical types. When these types have very little in common between their clusters, e.g. horizontal and vertical clusters, their axes of alignment can become orthogonal to each other. This causes horizontal clusters to become rows and vertical clusters to become columns or vice versa. We consider this special case as a *latticed* representation. We consider a *non-canonical* representation to be nondistinctive ($\mathcal{D}(C) < 1$) with respect to all canonical types. These archetypes align with the manual classification of structure maps done by Mavros et al. [10].

3.2 Functional Analysis

For participants with non-canonical representations according to the morphological analysis, functional analysis can offer an explanation. First, clusters are extracted from a participant's representation using a set of methods chosen by the experimenter. In our case, the methods are all instances of HDBSCAN [2] using different distance functions based on archetypes from the morphological analysis: Euclidean distance for clustering unaligned representations, Manhattan distance for latticed representations, and two variants of Manhattan distance for aligned representations, which weight the distances along the x- and y-axes with $\langle 0.8, 0.2\rangle$ and $\langle 0.2, 0.8\rangle$ before summing them. Manually annotated clusters by either the experimenter or participants can also be used as a method.

$$\mathcal{I}(C, \Phi) = \frac{1}{m} \sum_{i=1}^{m} \prod_{j=1}^{m} \Phi(C_i, C_j) \;\Big|\; i \neq j \tag{3}$$

The clusters C of each method are then judged by an informativeness metric $\mathcal{I}(C, \Phi)$ (Eq. 3) with respect to a function Φ, and only the best method is considered. $\Phi(C_i, C_j)$ is defined as any function that outputs 1 when two clusters C_i and C_j are separable according to a metric over *real* landmark positions (not *recalled* positions) and outputs 0 otherwise. This means that $\mathcal{I}(C, \Phi)$ evaluates the probability that a cluster can be distinguished from all other clusters according to function Φ, which is comparable to the formulation of distinctiveness $\mathcal{D}(C)$. We consider two metrics for Φ: the horizontal Euclidean distance along the xy-plane (used by Φ^H) and the vertical distance along the z-axis (used by Φ^V). Both functions Φ^H and Φ^V first compute the real centroid (i.e., the average real position) of landmarks in C_i and then compute whether all landmarks in C_i are closer to the centroid than those in C_j using the respective metric. With more data about the landmarks and the building, more metrics can be explored, e.g., the metabolic energy used when travelling between landmarks [11]. However, some complex metrics may be better suited for morphological analysis, because they are difficult to formulate computationally, e.g., the visual or semantic saliency of landmarks.

4 Results

Morphological analysis revealed that among the 41 participants, 7 people (17.1%) had aligned representations ($\mathcal{A}(C) \geq 0.9$), 5 people (12.2%) had unaligned representations ($\mathcal{A}(C) < 0.9$), 3 people (7.3%) had latticed representations, and 26 people (63.4%) had non-canonical representations ($\mathcal{D}(C^H) < 1$ and $\mathcal{D}(C^V) < 1$). For non-canonical representations, we apply functional analysis and compare $\mathcal{D}(C^H)$ with $\mathcal{I}(C, \Phi^H)$ and $\mathcal{D}(C^V)$ with $\mathcal{I}(C, \Phi^V)$. On average, the probability increased from distinctiveness to informativeness by 9.0% for horizontal and vertical comparisons, meaning that functional analysis offers a better

explanation for non-canonical representations than morphological analysis. Furthermore, 9 people (22.0%) had a gain of at least 50% for one of the two comparisons, and 4 people (9.8%) had perfectly informative representations ($\mathcal{I}(C, \Phi) = 1$), which morphological analysis could not account for. Between aligned and informative representations, 7 people (17.1%) respected horizontal regionalisation and 6 people (14.6%) respected vertical regionalisation.

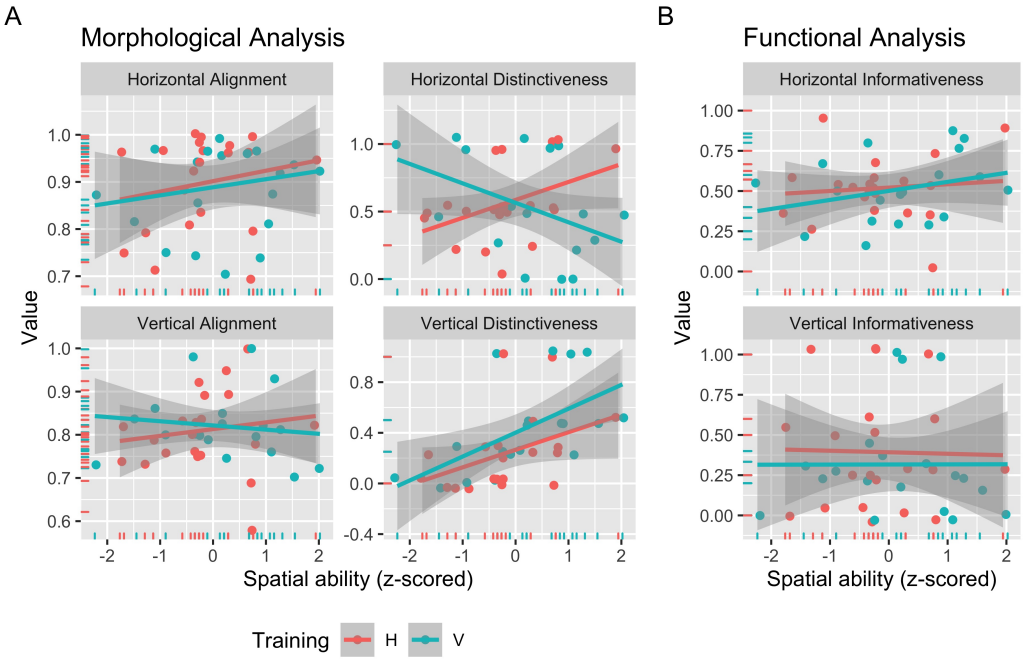


Figure 2 Scatter-plots of the morphological (A) and functional (B; i.e. bottom-up) analysis of the structure mapping task data, as a function of spatial ability (normalised SBSOD score). We observe that morphological analysis is sensitive to individual differences with regards to perception/externalisation of vertical relationships.

Figure 2 shows scatter plots of the 6 metrics (on the y-axis) from both analyses compared against the standardised spatial ability (SBSOD, on the x-axis) of participants. We observe an interaction between training and spatial ability, with respect to horizontal distinctiveness. Participants high in spatial ability trained horizontally (floor-by-floor) produced more horizontally distinctive structure maps; whereas high-spatial participants trained vertically had low horizontal distinctiveness. However, these participants produced higher vertical distinctiveness than high-spatial participants with horizontal training.

5 Discussion and Conclusion

In this article, we sought to establish a mathematical framework to capture spatial relationships from 2D structure maps. More specifically to measure spatial regionalisation of multilevel (i.e. 3D) environments. Our proposed method captures vertical versus horizontal alignment and distinctiveness of groups of objects —here landmarks in a building. We first demonstrated that this method can capture key aspects of structure mapping tasks, such as horizontal and vertical clustering (regionalisation). We then applied this methodology on human structure mapping task data, which helped identify a relationship between spatial abilities and the comprehension of horizontal and vertical spatial relationships.

Various objective methods have been developed to analyse externalisations (products) of mental representations of 2D space, e.g., measuring the angular error in a pointing task, or performing bi-dimensional regression to analyse configurational tasks like sketch maps [7]. Structure mapping tasks can also be analysed in similar terms when they represent two-dimensional configurations, such as objects on a table or locations on a city. To the best of our knowledge, the analysis of structure mapping data from 3D spatial configurations has not been sufficiently studied. As our example illustrates, 3D spatial configurations can be organised in both horizontal and vertical regions (or hierarchies) based on building floors, staircases, and other vertical spatial features. When people produce structure maps of such configurations, they have to perform a series of mental transformations, e.g., adopting an elevation-, top-, or other viewpoint and decoding spatial relationships in 3D space.

To conclude, in the present work we have proposed a novel mathematical framework to capture 3D spatial relationships from 2D structure maps. We applied this methodology on a set of human structure mapping task data, which helped identify a relationship between spatial abilities and the comprehension of vertical spatial relationships. Our results suggest that we can mathematically analyse the spatial configuration of 3D locations, by decomposing the two-dimensional structure maps into top-down as well as bottom-up spatial hierarchies that capture 3D spatial relationships (such as vertical columns and planes). This method provides a complementary approach to previously used classification by human raters [1, 10]. Further work will seek to compare how this method performs in comparison to human ratings. This method can be useful to advance the study of volumetric spatial cognition.

References

1 Simon J. Büchner et al. Path choice heuristics for navigation related to mental representations of a building. In *Proceedings of the Euro CogSci*, 2007.

2 R. Campello et al. Density-based clustering based on hierarchical density estimates. In *Pacific-Asia conference on knowledge discovery and data mining*. Springer, 2013.

3 Helen Couclelis et al. Exploring the anchor-point hypothesis of spatial cognition. *Journal of Environmental Psychology*, 7, 1987.

4 Yan Feng et al. Wayfinding behaviour in a multi-level building: A comparative study of HMD VR and Desktop VR. *Advanced Engineering Informatics*, 51, 2022.

5 Mary Hegarty et al. Development of a self-report measure of environmental spatial ability. *Intelligence*, 30:425–447, 2002.

6 Christoph Hölscher et al. Up the down staircase: Wayfinding strategies in multi-level buildings. *Journal of Environmental Psychology*, 26(4):284–299, 2006.

7 Rob Kitchin and Marc Blades. *The Cognition of Geographic Space*. I.B. Tauris, 2002.

8 Saskia F. Kuliga et al. Exploring Individual Differences and Building Complexity in Wayfinding: The Case of the Seattle Central Library. *Environment and Behavior*, 2019.

9 Yi Lu and Yu Ye. Can people memorize multilevel building as volumetric map? A study of multilevel atrium building. *Environment and Planning B: Urban Analytics and City Science*, 46(2):225–242, 2019. doi:10.1177/2399808317705659.

10 Panagiotis Mavros et al. Human Navigation in a Multilevel Travelling Salesperson Problem. *PsyArXiv Pre-prints*, pages 1–38, 2022.

11 M. Schwartz. Human centric accessibility graph for environment analysis. *Automation in Construction*, 127:103557, 2021.

12 G.B. Thomas. *Calculus and analytic geometry*. Addison-Wesley Publishing, 1968.

13 J.M. Wiener and Hanspeter A. Mallot. "Fine-to-Coarse" Route Planning and Navigation in Regionalized Environments. *Spatial Cognition and Computation*, 3(4):331–358, 2003.

14 Sudanthi Wijewickrema and Andrew Papliński. Principal component analysis for the approximation of an image as an ellipse. In *WSCG'2005, Plzen, Czech Republi*, 2005.

15 Svante Wold et al. Principal component analysis. *Chemometrics and intelligent laboratory systems*, 2(1-3):37–52, 1987.

16 Andreas Zwergal et al. Anisotropy of Human Horizontal and Vertical Navigation in Real Space: Behavioral and PET Correlates. *Cerebral Cortex*, 26(11):4392–4404, 2016.

A Comparison of Geographically Weighted Principal Components Analysis Methodologies

Narumasa Tsutsumida[1]
Saitama University, Japan

Daisuke Murakami
Institute of Statistical Mathematics, Tokyo, Japan

Takahiro Yoshida
The University of Tokyo, Japan

Tomoki Nakaya
Tohoku University, Sendai, Japan

Binbin Lu
Wuhan University, China

Paul Harris
Rothamsted Research, Harpenden, UK

Alexis Comber
University of Leeds, UK

Abstract

Principal components analysis (PCA) is a useful analytical tool to represent key characteristics of multivariate data, but does not account for spatial effects when applied in geographical situations. A geographically weighted PCA (GWPCA) caters to this issue, specifically in terms of capturing spatial heterogeneity. However, in certain situations, a GWPCA provides outputs that vary discontinuously spatially, which are difficult to interpret and are not associated with the output from a conventional (global) PCA any more. This study underlines a GW non-negative PCA, a geographically weighted version of non-negative PCA, to overcome this issue by constraining loading values non-negatively. Case study results with a complex multivariate spatial dataset demonstrate such benefits, where GW non-negative PCA allows improved interpretations than that found with conventional GWPCA.

2012 ACM Subject Classification Mathematics of computing → Multivariate statistics

Keywords and phrases Spatial heterogeneity, Geographically weighted, Sparsity, PCA

Digital Object Identifier 10.4230/LIPIcs.COSIT.2022.21

Category Short Paper

Funding This research was funded by the Joint Support Center for Data Science Research at Research Organization of Information and Systems (ROIS-DS-JOINT) under Grant 006RP2018, 004RP2019, 003RP2020, and 005RP2021.

1 Introduction

A principal components analysis (PCA) summarizes multi-dimensional data [5], by reducing the number of dimensions of the dataset. It provides a purely mathematical means of highlighting key sources of variation. Due to its form, spatial effects of spatial autocorrelation and spatial heterogeneity are not considered in transforming the multi-dimensional spatial data. For spatial data, some PCA methods have been developed to consider these spatial

[1] corresponding author

autocorrelation and heterogeneity [2]. Spatial PCA, referred to as sPCA, takes spatial autocorrelation into account to reveal spatial patterns [6]. sPCA yields global principal components (PCs) similar to the conventional PCA, but its scores underline the spatial autocorrelation; thus, for example, the spatial distribution of its first PC score is positively high autocorrelated. Similarly, geographically weighted (GW) PCA takes spatial heterogeneity into account. Analogous to GW regression (GWR) [1], GWPCA assembles local PCAs by applying a moving window weighted kernel and yields spatially varying eigenvalues, loadings, and percentage of total variance captured by each PC [4, 7]. The outputs of GWPCA are extensive but unique in terms of their spatial variations, allowing an investigation into the spatial data structure [2]. However, in some cases, GWPCA gives spatially discontinuous loadings (from positive to negative, for example, see [10]), and this presents problems of interpretation, and espcially in terms of relating to its global, conventional PCA counterpart. This issue arises because GWPCA assembles local PCAs that are independent of each other. To deal with this issue, we consider a GW non-negative PCA (GWnnegPCA) to constrain all loadings non-negatively [11]. This results in that at any local PCA calibration point, the input variables are linearly summed to build new PCs, but where loadings are only zero or positive, providing a more intuitive interpretation for the GWnnegPCA mapped outputs as a whole.

2 Methods

GWPCA assembles a series of local PCAs where each PCA is constructed using nearby, spatially-weighted data according to a moving window kernel. Given a $n \times m$ matrix $\overline{\mathbf{X}}$ which consists of m variables at n observation sites and each variable is re-scaled to zero-mean and unit-variance, GWPCA decomposes the GW variance-covariance matrix of $\overline{\mathbf{X}}$ at the p-th location with coordinates (u_p, v_p), which is defined by $\mathbf{\Sigma}_p = \overline{\mathbf{X}}^T \mathbf{W}_p \overline{\mathbf{X}}$, by

$$\mathbf{L}_p \mathbf{V}_p \mathbf{L}_p^T = \mathbf{\Sigma}_p, \tag{1}$$

where $\mathbf{L}_p$ is a GW loading matrix and $\mathbf{V}_p$ is a diagonal matrix of GW eigenvalues at the p-th location. $\mathbf{W}_p$ is a diagonal matrix of geographic weights that can be generated using a given kernel function. In this case study, we used the bisquare kernel function for the q-th location:

$$\omega_{pq} = \begin{cases} \left(1 - \left(\frac{d_{pq}}{b}\right)^2\right)^2 & \text{if } |d_{pq}| < b, \\ 0 & \text{otherwise,} \end{cases} \tag{2}$$

where the bandwidth size b is any distance for a given number of nearest observations (i.e., 100 or 10% of the total observation), and d_{pq} is the euclidean distance between spatial locations of the p-th and q-th data locations in this work.

The first loading $\mathbf{l}_p^{(1)}$ for the GWPCA at the p-th location is applied so that:

$$\underset{\mathbf{l}_p^{(1)}}{argmax}\ \mathbf{l}_p^{(1)T} \mathbf{\Sigma}_p \mathbf{l}_p^{(1)}, \quad subject\ to\ \|\mathbf{l}_p^{(1)}\|_2 = 1 \tag{3}$$

The subsequent loading maximizes the variance under the constraint that it is orthogonal to the previous component(s) [9]: $\mathbf{l}_p^{(i)T} \mathbf{\Sigma} \mathbf{l}_p^{(i)}$ for all $i \in \{2, ..., m\}$, $subject\ to\ \|\mathbf{l}_p^{(i)}\|_2 = 1$ and $\mathbf{l}_p^{(i-1)} \mathbf{l}_p^{(i)} = 0$.

A robust GWPCA has also been proposed to reduce the effect of anomalous observations on the output [3]. This uses a local covariance matrix by using the robust minimum covariance determinant (MCD) estimator [8].

In this context, we develop the GWnnegPCA which applies the following additional restriction to the equation (3):

$$subject\ to\ \ \|\mathbf{l}_p^{(1)}\|_0\ \leq k,\ \ \mathbf{l}_p^{(1)} \geq \mathbf{0}, \tag{4}$$

where k ($\leq m$) is the number of non-zero variables and $\|\mathbf{l}_p^{(1)}\|_0$ is the cardinality of $\mathbf{l}_p^{(1)}$. This makes all the loadings at any location non-negative. However, as the order of local PC axes determined by the eigenvalues, the order at the p-th location from GW-based PCA may not be the same as that from the conventional PCA due to high spatial heterogeneity. Thus, by using the result of the conventional and non-negative PCA loadings, we rearranged the order of local loadings by GW-based PCAs. The Pearson's correlation matrix between the absolute values of the global and the local loadings at every location was calculated, then the local loadings were reordered according to the correlation coefficient. This modification is expected to make global and local loadings comparative.

The local PC scores at the p-th location, $\mathbf{S}_p$, are represented by:

$$\mathbf{S}_p = \overline{\mathbf{X}}_p\mathbf{L}_p \tag{5}$$

It is noted that to introduce the non-negativity, the cardinality in equation (4) requires a minimum angle between components and thus the orthogonality constraint of the PCs is relaxed [9]. This means that components are quasi-orthogonal amongst the PCs. In this sense, loadings from non-negative PCA and GW non-negative PCA are regarded as quasi-eigenvectors.

Bandwidth is a critical parameter in the GW framework as its size determines the localness of the analysis and whether the given process is indeed non-stationary. The bandwidth of GWPCA is optimized by a leave-one-out cross-validation [4] and in this study, the optimized bandwidth by GWPCA is also used for robust GWPCA and GWnnegPCA to be comparative. The bandwidth size used in this study was 45.3% of the total.

3 Case study

For our case study, we build 21 variables using census statistics of Japan in 2005 (Table 1). These variables describe the urban social structure of Tokyo within the 3,134 *chocho-aza* units (the smallest administrative unit in Japan) of the 23 special wards, Tokyo. All 21 variables were standardized (zero mean with a unit variance) before being input into our non-spatial and GW models.

4 Results

Spatial distribution maps of a loading ($HIGHEDU$) for PC1-3 by GWPCA, robust GWPCA, and GWnegPCA were shown in Figure 1 as an example. The loading values of this variable for conventional PCA were 0.30, 0.19, and -0.37, respectively, and those for non-negative PCA were 0.42, 0.02, and 0.11, respectively. All results from the GWPCA, robust GWPCA, and GWnnegPCA are associated with the conventional and non-negative PCAs, while GW-based PCAs show spatial distributions of loading values. The loading map for GWPC1 represents the higher value surrounding the center of Tokyo compared to other regions. The map for GWPC2 represents slightly lower negative values in the east and the north of the regions, while positive values are found in the central part. The negative values are also found with a strip shape within the central part. Furthermore, the map for GWPC3 shows a clear discontinuous spatial pattern with positive/negative patches. Such loading maps make us confuse how to interpret them, resulting in the difficulty of using GWPCA.

Table 1 Variable descriptions used in this study.

Abbreviation	Variable Descriptions
AGE0014	Num. of persons in the age of 0 – 14 / The total pop.
AGE1529	Num. of persons in the age of 15 – 29 / The total pop.
AGE3044	Num. of persons in the age of 30 – 44 / The total pop.
AGE4564	Num. of persons in the age of 45 – 64 / The total pop.
AGE65	Num. of persons in the age over 65 / The total pop.
UNIVPOP	Num. of university students / The total pop.
HIGHEDU	Num. of university graduates / The total pop.
NUCFAMR	Num. of nuclear families / The total households
MULTIFAM	Num. of extended families / The total households
SINGHR	Num. of single households / The total households
OWNHR	Num. of owned housing households / The total households
SELFEMP	Num. of self employments / The total worker pop.
WHITER	Num. of white-coloured employees / The total worker pop.
BLUER	Num. of blue-coloured employees / The total worker pop.
PRIMR	Num. of the prime sector employees / The total worker pop.
SECR	Num. of the second sector employees / The total worker pop.
TERR	Num. of the third sector employees / The total worker pop.
SHORTCOM	Num. of commuters ($\leq$ 30 min) / The total num. of commuters
LONGCOM	Num. of commuters ($\geq$ 1 hour) / The total num. of commuters
NEWCOMER	Num. of in-migrant pop. ($\geq$ 5 yrs.) in 5 yrs. / Total pop. ($\geq$ 5 yrs.)
WPRATIO	Num. of workers / The total pop.

The loading maps for robust GWPCA show a more continuous spatial pattern than those for GWPCA, but at the middle west part of the region, a positive patch is found in the loading map for robust GWPC2. Such a pattern can be found in the loading map for robust GWPC3 such that a negative patch in regions with positive values.

GWnnegPCA also provides loading maps, and those spatial patterns are simpler. The loading map for GWnnegPC1 represents a positive value in large parts of the area. The map for GWnnegPC2 shows zero value in almost all areas, corresponding with the result of non-negative PCA. The map for GWnnegPC3 represents positive values in the middle and western regions while zero in others. These loading values for GWnnegPCA tell in which areas this $HIGHEDU$ variable contributes to each PC.

5 Discussion

This study demonstrates that GWPCA can give quite discontinuous spatial patterns for its localized loadings. This characteristic can be challenging to interpret spatially compared to the corresponding conventional global PCA outputs. This is especially pertinent as GW methods are often employed to (hopefully) reveal smoothly changing spatial drifts of model parameters, not discontinuous ones. The discontinuous issue would come from the flexibly rotated axes of local PCs at each observation point. It is applicable to fix the first axis direction to be all positive in GWPCA by applying a similar modification as the *fix_sign* option in the *princomp* R function. However, such an approach may not work straightforwardly for the subsequent components for GW-based PCA because the reminders of data variance after applying the equation (3) are not constant over space: that is, the

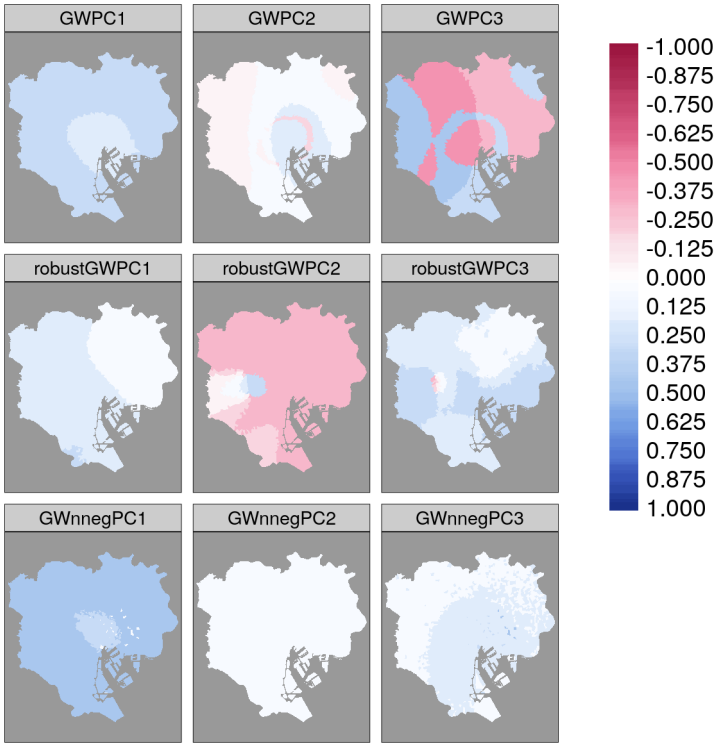


Figure 1 Loading maps of the HIGHEDU variable for the first, the second, and the third principal component by geographically weighted principal components analysis, robust geographically weighted principal components analysis, and geographically weighted non-negative principal components analysis as an example.

explained variation in the data for the GWPC1 varies spatially. Thus, the second component axis (and subsequent component axes as well) at each observation point cannot be determined in a particular manner, and such rotations at the p-th location cannot be coordinated with each other. GWnnegPCA has overcome this issue significantly. Non-negativity of local PCAs fixes the flexible rotation of PC axes of GWPCA, and thus the loading maps represent spatially varying patterns. Figure 1 showed a clear difference of spatially varying patterns of loading maps of an example variable. GWPCA has been used in many case studies to obtain the largest absolute value of loading at every location as the winning variable [4], and is valuable to see the spatial heterogeneity in the input data structure. It is however challenging to investigate the spatial pattern of each loading due to the discontinuous problem as seen in this study, even applying the robust way. In this context, GWnnegPCA has the potential to show the continuous spatial variation smoothly that would contribute to further interpretations of the result.

Similar to the discussion on the validity of installing the non-negativity for PCA, GWnnegPCA also inherits the technical issue of heavy computation and relaxing orthogonality [9]. In addition, we found noisy patterns in loading maps of GWnnegPCA, which may come from the sparsity and outliers in input data. Further investigations will be expected to be the stability of the result.

6 Conclusions

We demonstrated in the case study that geographically weighted principal components analysis (GWPCA) and robust GWPCA yield discontinuous spatial patterns for its localized loadings and introduced geographically weighted non-negative principal components analysis (GWnnegPCA) to overcome this issue. GWnnegPCA would be a reasonable choice to show spatially varying loading values in multivariable spatial data so that the degree of variable(s) contribution to a principle component varies spatially. This allows us to interpret data locally to understand regional characteristics in the spatial data as we expect GW approach. We will work on interpreting regional changes in loading patterns and handling noises on loading maps for further developments.

References

1 Chris Brunsdon, A Stewart Fotheringham, and Martin E Charlton. Geographically Weighted Regression: A Method for Exploring Spatial Nonstationarity. *Geogr. Anal.*, 28(4):281–298, September 1996. doi:10.1111/j.1538-4632.1996.tb00936.x.

2 Urška Demšar, Paul Harris, Chris Brunsdon, A. Stewart Fotheringham, and Sean McLoone. Principal Component Analysis on Spatial Data: An Overview. *Ann. Assoc. Am. Geogr.*, 103(1):106–128, January 2013. doi:10.1080/00045608.2012.689236.

3 Isabella Gollini, Binbin Lu, Martin Charlton, Christopher Brunsdon, and Paul Harris. GWmodel : An R Package for Exploring Spatial Heterogeneity Using Geographically Weighted Models. *J. Stat. Softw.*, 63(17):85–101, April 2015. doi:10.18637/jss.v063.i17.

4 Paul Harris, Chris Brunsdon, and Martin Charlton. Geographically weighted principal components analysis. *Int. J. Geogr. Inf. Sci.*, 25(10):1717–1736, October 2011. doi:10.1080/13658816.2011.554838.

5 Ian Jolliffe. Principal Component Analysis. In *Wiley StatsRef Stat. Ref. Online*. John Wiley & Sons, Ltd, Chichester, UK, September 2014. doi:10.1002/9781118445112.stat06472.

6 T. Jombart, S. Devillard, A. B. Dufour, and D. Pontier. Revealing cryptic spatial patterns in genetic variability by a new multivariate method. *Heredity (Edinb).*, 101(1):92–103, 2008. doi:10.1038/hdy.2008.34.

7 Christopher D. Lloyd. Analysing population characteristics using geographically weighted principal components analysis: A case study of Northern Ireland in 2001. *Comput. Environ. Urban Syst.*, 34(5):389–399, 2010. doi:10.1016/j.compenvurbsys.2010.02.005.

8 Peter Rousseeuw. Multivariate Estimation with High Breakdown Point, 1985.

9 Christian D. Sigg and Joachim M. Buhmann. Expectation-maximization for sparse and non-negative PCA. In *Proc. 25th Int. Conf. Mach. Learn. - ICML '08*, pages 960–967, New York, New York, USA, 2008. ACM Press. doi:10.1145/1390156.1390277.

10 Narumasa Tsutsumida, Paul Harris, and Alexis Comber. The Application of a Geographically Weighted Principal Component Analysis for Exploring Twenty-three Years of Goat Population Change across Mongolia. *Ann. Am. Assoc. Geogr.*, 107(5):1060–1074, 2017. doi:10.1080/24694452.2017.1309968.

11 Narumasa Tsutsumida, Daisuke Murakami, Takahiro Yoshida, Tomoki Nakaya, Binbin Lu, and Paul Harris. Geographically Weighted Non-negative Principal Components Analysis for Exploring Spatial Variation in Multidimensional Composite Index. *GeoComputation 2019*, September 2019. doi:10.17608/k6.auckland.9850826.v1.

Abnormal Situation Simulation and Dynamic Causality Discovery in Urban Traffic Networks

Yadi Wang
School of Software and Microelectronics, Peking University, Beijing, China

Yicheng Pan
School of Electronics Engineering and Computer Science, Peking University, Beijing, China

Meng Ma[1]
National Engineering Research Center for Software Engineering, Peking University, Beijing, China

Ping Wang[2]
School of Software and Microelectronics, Peking University, China
National Engineering Research Center for Software Engineering, Peking University, Beijing, China

Abstract

Various participants in urban traffic systems intertwine a highly complicated coupling network. An interpretable analysis of underlying correlations is one of the keys to understanding traffic anomalies. Unfortunately, abnormal situation analysis in real scenarios faces severe limitations in negative sample deficiency, data integrity, and verifiability. In view of this, we developed a simulation tool – the Traffic Anomaly Situation Simulator (TASS). Through configurable scripts, TASS simulates real traffic networks by road editing, data collection, and fault injection. Given the generated cases, we designed a dynamic causal discovery algorithm, Dycause-Traffic, to demonstrate the features of causality in traffic anomalies.

2012 ACM Subject Classification Applied computing → Transportation

Keywords and phrases SUMO simulation, dynamic causality discovery, congestion propagation

Digital Object Identifier 10.4230/LIPIcs.COSIT.2022.22

Category Short Paper

Funding The National Natural Science Foundation of China (92167104, 62072006) and Science and Technology on Communication Networks Laboratory (6142104200103) support this work.

1 Introduction

Abnormal traffic situation such as congestion keeps increasing in metropolitan areas, contributing to great economic losses and causing delay, compromised levels of service, and discomfort during traveling [10]. Since the 1980s, enormous progress has been made in predicting traffic, but current studies lack the analysis of important flow relationships and evolution trends of important urban structures [3, 8], and the dynamic causal association information hidden in the dataset has not been fully explored. Besides, most algorithms only consider the existence of causalities while ignoring their strength.

The traditional traffic analysis primarily relies on traffic time series data. For example, by redesigning the distance measurement and clustering methods in the clustering algorithm, the urban traffic flow patterns and urban structures are revealed from the spatiotemporal data [9, 12, 17]. The association rule and clustering method are often combined into a

[1] Corresponding author. Meng Ma is also a visiting researcher with the Science and Technology on Communication Networks Laboratory.
[2] Corresponding author. Ping Wang is also with the Key Laboratory of High Confidence Software Technologies (PKU), Ministry of Education, China.

multi-step model [15]. But less consideration is given to the topological relationship between roads and the interaction between road segments. Graph-based algorithms that can be used to discover such associations [18] mainly include CGC [5] and PSTE [14] based on Granger causality theory [4], SVAR-FCI [11] and PCMCI [16] based on conditional constraints, and other models [13, 19] built on GNN. However, the upper limit of the number of conditional independence tests of the algorithm increases with the maximum degree of any node.

High-quality urban traffic flow data is the basis for analyzing abnormal urban traffic situations. Some papers provide relevant research with good actual data [2, 6, 7]. But the data generated in real scenarios are insufficient. The timely feedback of traffic intervention cannot be achieved quickly and it's challenging to excavate the influencing factors under the dynamically changing traffic conditions. Analyzing urban traffic anomalies requires controllable and reproducible data, and simulation tools provide this possibility.

Our contributions are manifolds: *1.Urban traffic simulation.* We build an urban traffic simulation system based on the relevant technical framework of SUMO [1], which provides a basis for the design and implementation of simulation experiments. *2.Traffic abnormal situation injection.* During the simulation, controllable anomalies, such as signal lights, lanes, and flow, are injected into the traffic road network to reproduce controllable traffic congestion. *3.Abnormal situation propagation analysis.* We use the dynamic causal correlation algorithm to mine the abnormal situation propagation path and intensity of traffic and analyze road congestion's impact on the overall road network and other roads.

2 Traffic abnormal situation simulation

Given the requirement of adjusting the traffic lights, intersections, road facilities, and lanes in the abnormal urban traffic situation, we design the Traffic Anomaly Situation Simulator (TASS) based on SUMO [1], and additionally various options for traffic anomaly injection in the case of simulating a given traffic demand consisting of a single-vehicle.

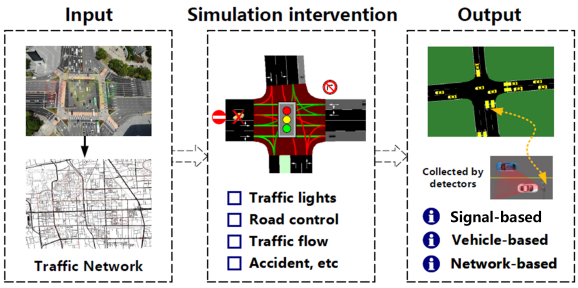


Figure 1 Overall pipeline of TASS.

The overall simulation flow of TASS is shown in Fig. 1. Please find the source code, sample cases, and dataset of TASS in this anonymous `github` repository.

2.1 TASS input

The input of TASS includes the following elements. In this study, we leverage the OSMWebwizard program to generate the network from the satellite map information of OpenStreetMap (OSM), including the traffic network structure of selected area as well as the subway, light rail, and other public transport systems.

(1) *Road attributes.* The basic network structure of the road, the allowed and limited direction, the types of vehicles that are allowed to pass, etc.
(2) *Vehicle attributes.* Vehicle type, route planning, driving preference and proficiency of the driver, etc., are mainly described by the dynamic model of the vehicle.
(3) *Interaction between vehicles.* The driving behavior of vehicles during operation is not only affected by road conditions and route planning but also by other vehicles, resulting in various traffic phenomena, such as traffic anomalies and traffic congestion.

2.2 Simulation intervention

The core task of TASS is the reproduction and intervention of traffic demand, that is, the data and related information of traffic flow. By intervening in the simulation process through the Traci script, TASS can realize the fixed-point timing injection of the abnormal situation. Urban traffic anomalies can be generally divided into three categories:

(1) *Abnormal state of lanes or edges.* Lanes or edges cannot meet the normal traffic needs of vehicles and pedestrians. For example, traffic accidents, vehicle lane changes, traffic control, and other abnormal driving behaviors cause lane blockages, resulting in lane prohibition. The maximum travel speed for fixed lanes can be set to 0 via Traci in TASS.
(2) *Abnormal traffic facilities.* For example, the traffic light is damaged, or the alternating cycle of the traffic light is abnormal. The Traffic Lights States (TLS) can be modified after specifying the number of simulation steps during the simulation. After recovery, the TLS can be changed to the normal state.
(3) *Abnormal driving state of vehicles.* It is generally reflected in the abnormal increase of pedestrians and vehicles caused by extreme weather, peak travel period, road control, and low-speed vehicles. The maximum driving speed of the lanes in the area can be reduced as required by Traci.

2.3 TASS output

In practice, the traffic simulation focuses on three output values to solve the traffic problem: vehicle-based information, network-based element (such as stations, traffic-lights, cross-linking, and other factors, which are related to the geometric layout of roads), and hidden cost (such as waiting time, time loss and depart delay). In this paper, we collect the three outputs but consider the vehicle-based data primarily. The vehicle-based information includes positions, speeds, acceleration of all vehicles for every simulation step, emission values, battery usage, collisions, lane-changing events, and trajectory data. Data collected from detectors provide lane/edge-based performance measurements such as speed, road occupancy, and traffic flow.

3 Dynamic Causality Discovery – Dycause-Traffic

3.1 Problem statement

We formally denote a traffic network as a directed graph $G = (V, E)$, where V is the set of $N = |V|$ vertices representing nodes on the road network (i.e., detectors in the following experiments), and E is the set of edges representing the connectivity among vertices. At each time step t, there is a graph signal $\mathbf{X}^{(t)} \in \mathbb{R}^{N \times C}$ on graph G, where C is the number of features of the input signal (such as traffic flow, speed, occupancy). The superscript (t) is the sample time index in a short period. To reveal the dynamic causalities between locations in traffic anomalies, we design the Dycause-traffic algorithm to examine Granger causal intervals with sliding windows, generating the dynamic causality curves $C_{i,j}(t)$ for each pair

of detectors. The dynamic causality curve $C_{i,j}(t)$ depicts the time-varying strengths of the correlation from detector v_i to v_j. Thus, we transform the anomaly propagation analysis task into the computing of the dynamic causality dependencies between vertices denoted as $\hat{Y} = (\hat{\mathbf{X}}^{t_{p+1}}, \hat{\mathbf{X}}^{t_{p+2}}, ..., \hat{\mathbf{X}}^{t_{p+q}})$ from the historical simulation output of time series of p time steps. Note that the calculation is limited by the overall simulation steps q.

3.2 Temporal dynamic causality discovery

We build Dycause-traffic based on the Granger causality test. Particularly, we denote the time series of the node i and node j as x_i and x_j. Our test constructs two linear regression models (the partial $\mathcal{M}_p$ and the full $\mathcal{M}_f$) to predict $x_j(t), t = 1, 2, ..., T$ with the past observations. Please note that one of them has more independent variables $x_i(t-l), 1 \leq l \leq L_{lag}$, representing the additional information from the node i.

$$\mathcal{M}_p \colon \hat{x}_j(t) = \sum_{l=1}^{L_{lag}} \alpha_l^p x_j(t-l) + b^p, \mathcal{M}_f \colon \hat{x}_j(t) = \sum_{l=1}^{L_{lag}} \left(\alpha_l^f x_j(t-l) + \beta_l^f x_i(t-l) \right) + b^f \quad (1)$$

Here, $\hat{x}_j(t)$ denotes the prediction and α_l^p, b^p, α_l^f, β_l^f and b^f are the parameters. By computing the sums of squared errors (SSE_p, SSE_f) from the two models, which is calculated by $\sum_t (\hat{x}_j(t) - x_j(t))^2$, we test the null hypothesis using the F-distribution. When the null hypothesis cannot be rejected, we accept that the extra information can not improve the prediction and x_i has no causal impact on x_j. To do that, we calculate the value $F = \frac{(SSE_p - SSE_f)/(d_f - d_p)}{SSE_f/(T - d_f - 1)}$ to estimate the probability of the null hypothesis based on the F-distribution of $\mathcal{F}_{d_f - d_p, T - d_f - 1}$ and reject it lower than a significant threshold α, which in turn indicates the Granger causality from x_i to x_j. Here, d_p and d_f are the degrees of freedom of the two models, which are $L_{lag}, 2L_{lag}$, respectively. T is the total number of samples. Given the fact that the causalities change dynamically in anomalies, we conduct the test on enumerated sliding windows $\{[s_q : e_q] \mid 0 \leq s_q < e_q \leq T\}$ instead of a single-window $[0, T]$. Finally, Dycause-traffic generates the causality curve $C_{i,j}$ according to Algorithm 1.

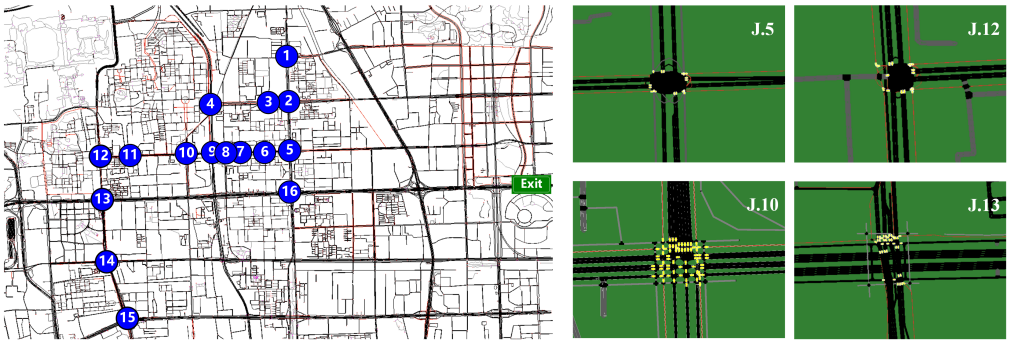
Algorithm 1 Dycause-Traffic.

Input: Metric data during time intervel T: $\mathbf{X} \in \mathbb{R}^{N \times T}$, Granger causality significance value α, basic interval length L_b
Output: Dynamic causality curves $C \in \mathbb{R}^{N \times N \times T}$.

```
begin
    for each node pair i → j do
        C_{i,j} ⟵ {0}^T
        for s = 0; s ≤ T − L_b; s += L_b do
            for e = s + L_b; e ≤ T; e += L_b do
                Estimate the ⌈F⌉ and ⌊F⌋ using previous regression models
                if ⌈F⌉ and ⌊F⌋ indicate pruning then
                    Continue
                else if ⌈F⌉ and ⌊F⌋ indicate causality then
                    Increase C_{i,j}(s: e) by 1
```

4 Experiments

We experimentally select a 50 km^2 dense traffic network in Beijing, China. As shown in Fig. 2(a), we number the main junctions from J.1 to J.16. The horizontal and vertical lanes at J.13 and J.16 are urban arterial roads. The horizontal line is the Fourth Ring Road in Beijing, and the surrounding infrastructure density is low. Schools, residential areas, and commercial areas are scattered around other junctions. A certain number of detectors are distributed at the traffic inflow position at each junction, which can provide traffic flow, vehicle speed, and lane occupancy data during the simulation period.



(a) Map and intersection number (b) Distribution of traffic detectors

Figure 2 Map and detectors in our experiment.

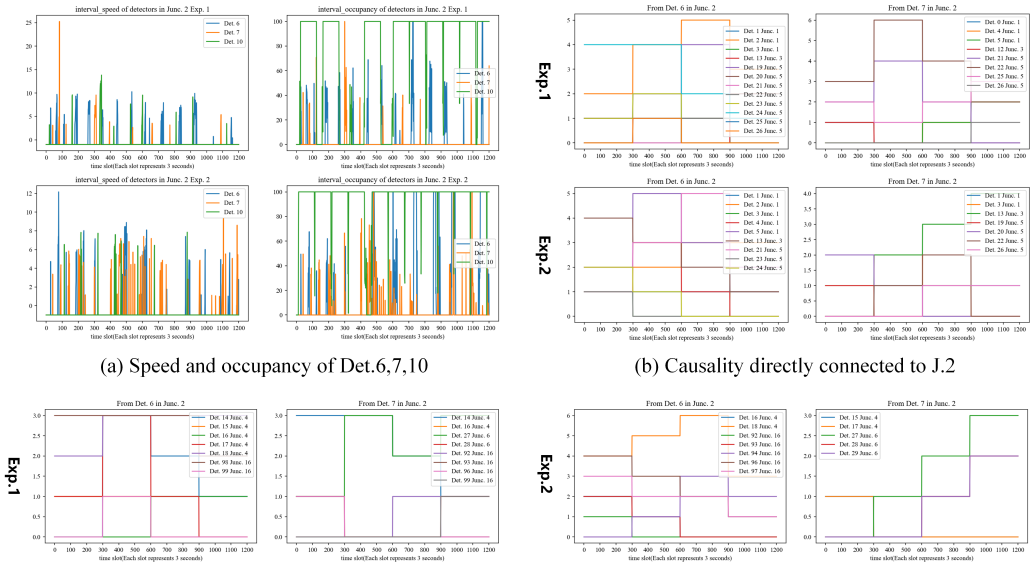
The distribution of detectors at junctions is shown in Fig. 2(b). The detectors are placed in the direction of the traffic entering the junction. J.5 and J.12 are on the urban arterial road. There are fewer connections to the secondary roads around the junction. In other words, the traffic network structure is simpler where the detector density is also lower. Correspondingly, the road network structure of J.10 and J.13 is more complicated, which is related to the surrounding infrastructure to a certain extent.

We conduct TASS to the following experimental simulations, considering traffic light failure, road failures, and abnormal traffic flow. Details of each experiment are summarized in Table 1. The simulation time is 3600 seconds, i.e., one hour in total. The fault injection period is 600 to 1800 seconds, lasting for 20 minutes. To improve the efficiency of model training, we select the raw data of 100 detectors, from Det.0 to Det.99.

Table 1 Simulation details of each experiment.

Experiment ID	Simulation Details	Failure Type
Exp. 1	Original traffic flow, the detector sampling frequency is set to 3 seconds	No failure.
Exp. 2	The traffic light at intersection Junc. 2 is faulty, only green	Traffic light failures
Exp. 3	The traffic light at intersection Junc. 2 is faulty, only red	
Exp. 4	The traffic light at intersection Junc. 12 is faulty, only green	
Exp. 5	The traffic light at intersection Junc. 12 is faulty, only red	
Exp. 6	The inbound route of the north-south road at junction Junc. 2 is faulty	Lane failures. Only the north-south edges are modified.
Exp. 7	The outbound route of the north-south road at junction Junc. 2 is faulty	
Exp. 8	The inbound route of the north-south road at junction Junc. 5 is faulty	
Exp. 9	The outbound route of the north-south road at junction Junc. 5 is faulty	
Exp. 10	Double the traffic flow of some lanes in Experiment 1	Abnormal traffic flow. The lanes selected are the same.
Exp. 11	Reduce the traffic flow of some lanes in Experiment 1	

Fig.3 shows part of the experimental results where we observe several features of anomaly propagation in traffic networks.



(a) Speed and occupancy of Det.6,7,10

(b) Causality directly connected to J.2

(c) Causality secondary connected to J.2

Figure 3 Example of experimental results.

No traffic lights or a single permit(red or green only)can cause traffic anomalies. Fig. 3(a) shows the speed and occupancy rate of Det.6,7,10 in Exp.1,2. When there is only a green light, the vehicle's average speed decreases. During the time slot 200 to 600 (i.e., 600s to 1800s), the speed has a short-term increase compared to Exp.1. Still, the green light increases the traffic volume while also increasing the road load, and the final result is a negative effect. When there are red lights at J.2, the vehicle runs normally only when there is no abnormal increase in the early period at J.2. Even if the accident is cleared, it does not quickly return to the initial condition.

Anomaly propagation is restricted by road capacity and road structure. J.3 is a junction directly linked with J.2, and the maximum road load is smaller than that of J.2. Therefore, the failure propagation from J.2 to J.3 is relatively stable, as shown in Fig. 3(b), and the results of all experiments have relatively minor differences. Since the capacity from the main arterial road to the secondary arterial road is limited by the secondary arterial road, the J.2 continues to transfer the flow to the J.3. Still, at the same time, once the traffic flow exceeds the threshold, the increase will not cause additional congestion for J.3.

Upstream and downstream traffic can affect each other. The upstream traffic flow parameters have a transmission effect on the downstream traffic flow parameters through the road carrier, and vice versa. As a result, they both will experience a time delay. Dotted failure can be achieved by injecting specified road anomalies or adding specific traffic flows. It will be propagated with the overflow or outflow mode, thereby creating linear or area failure. But it should be noted that the reproduction of overflow or outflow mode needs to be better reproduced in the case of meeting the road conditions.

Secondary road connection structure affects anomaly's propagation. Fig. 3(c) shows how the secondary junction connected with J.2 is affected. Traffic anomalies in the same place caused by different factors have inconsistent characteristics. Considering the diffusion of traffic flow, the problem of secondary anomalies caused by local anomalies cannot be ignored. When severe traffic anomalies occur locally, the traffic conditions of upstream and downstream sections and junctions will be affected along with secondary road links. A worse situation could lead to the collapse of large areas of the road network.

5 Conclusion

This study presents a novel tool, TASS, to simulate anomalies in urban traffic networks, including traffic light failures, lane failures, and abnormal flow. We design the Dycause-Traffic algorithm to reveal the anomaly's propagation mechanism. Experiments verified that the length of lanes, the connection of roads, and the location of faults are the main factors affecting the propagation of anomalies.

References

1 Michael Behrisch et al. Sumo–simulation of urban mobility: an overview. In *Proceedings of SIMUL 2011, The Third International Conference on Advances in System Simulation*. ThinkMind, 2011.

2 Gleb Beliakov et al. Measuring traffic congestion: an approach based on learning weighted inequality, spread and aggregation indices from comparison data. *Applied Soft Computing*, 67:910–919, 2018.

3 Etienne Come et al. Spatio-temporal analysis of dynamic origin-destination data using latent dirichlet allocation: Application to vélib'bike sharing system of paris. In *TRB 93rd Annual meeting*, page 19p. Transportation Research Board, 2014.

4 Clive WJ Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: journal of the Econometric Society*, pages 424–438, 1969.

5 Meng Hu et al. A copula approach to assessing granger causality. *NeuroImage*, 100:125–134, 2014.

6 Bin Jiang et al. Topological analysis of urban street networks. *Environment and Planning B: Planning and design*, 31(1):151–162, 2004.

7 Zihan Kan et al. Traffic congestion analysis at the turn level using taxis' gps trajectory data. *Computers, Environment and Urban Systems*, 74:229–243, 2019.

8 Felix Kling et al. When a city tells a story: urban topic analysis. In *Proceedings of the 20th international conference on advances in geographic information systems*, pages 482–485, 2012.

9 Xi Liu et al. Revealing travel patterns and city structure with taxi trip data. *Journal of transport Geography*, 43:78–90, 2015.

10 Qiong Lu et al. The impact of autonomous vehicles on urban traffic network capacity: an experimental analysis by microscopic traffic simulation. *Transp. Lett.*, 12(8):540–549, 2020.

11 Daniel Malinsky et al. Causal structure learning from multivariate time series in settings with unmeasured confounding. In *Proceedings of 2018 ACM SIGKDD workshop on causal discovery*, pages 23–47. PMLR, 2018.

12 Feng Mao et al. Mining spatiotemporal patterns of urban dwellers from taxi trajectory data. *Frontiers of Earth Science*, 10(2):205–221, 2016.

13 Meike Nauta et al. Causal discovery with attention-based convolutional neural networks. *Machine Learning and Knowledge Extraction*, 1(1):312–340, 2019.

14 Angeliki Papana et al. Detecting causality in non-stationary time series using partial symbolic transfer entropy: Evidence in financial data. *Computational economics*, 47(3):341–365, 2016.

15 Felix Rempe et al. Spatio-temporal congestion patterns in urban traffic networks. *Transportation Research Procedia*, 15:513–524, 2016.

16 Jakob Runge et al. Detecting and quantifying causal associations in large nonlinear time series datasets. *Science advances*, 5(11):4996, 2019.

17 Xiaoying Shi et al. Exploring the evolutionary patterns of urban activity areas based on origin-destination data. *IEEE Access*, 7:20416–20431, 2019.

18 Huijun Sun et al. Urban traffic congestion spreading in small world networks. *International Journal of Modern Physics B*, 19(28):4239–4246, 2005.

19 Zhang Zhang et al. A general deep learning framework for network reconstruction and dynamics learning. *Applied Network Science*, 4(1):1–17, 2019.

Spatial and Spatiotemporal Matching Framework for Causal Inference

Kamal Akbari[1]
Faculty of Engineering and Information Technology, The University of Melbourne, Australia

Martin Tomko
Faculty of Engineering and Information Technology, The University of Melbourne, Australia

Abstract

Matching is a procedure aimed at reducing the impact of observational data bias in causal analysis. Designing matching methods for spatial data reflecting static spatial or dynamic spatio-temporal processes is complex because of the effects of spatial dependence and spatial heterogeneity. Both may be compounded with temporal lag in the dependency effects on the study units. Current matching techniques based on similarity indexes and pairing strategies need to be extended with optimal spatial matching procedures. Here, we propose a decision framework to support analysts through the choice of existing matching methods and anticipate the development of specialized matching methods for spatial data. This framework thus enables to identify knowledge gaps.

2012 ACM Subject Classification Mathematics of computing → Causal networks; Information systems → Spatial-temporal systems; Information systems → Data analytics

Keywords and phrases Framework, Spatial, Spatiotemporal, Matching, Causal Inference

Digital Object Identifier 10.4230/LIPIcs.COSIT.2022.23

Category Short Paper

1 Introduction

When collecting data to analyze causal effects of an intervention, Randomised Controlled Trials (RCTs) are the theoretical best practice. Yet RCTs are costsly and complex [15, 18]. Quasi-experimental methods on observational data have been proposed, i.e. on data collected for different purposes and only re-analyzed to identify causal effects of interventions. Causal studies on observational data lack control over the design and data collection process, making it impossible to manage the selection and confounding bias. Matching is the analytical step that aims to reduce such bias by controlling for the imbalance between the characteristics of the units in the treated and control groups based on the distribution of covariates [16].

Matching on spatial data is an emerging topic in spatial causal inference [1], where theoretical assumptions of independent random processes do not hold due to first-order effects (spatial heterogeneity) and second-order effects (spatial lags) [10]. Matching methods find pairs of units from the treated and control groups based on the similarity measured from baseline covariates. For non-spatial data, established methods are available, e.g., Propensity Score (PS) [12] and Mahalanobis Distance (MD) matching [13]. These methods do not consider the effects of spatial heterogeneity and spatial autocorrelation and may lead to biased estimates of causal effects in spatial data.

Here we investigate the requirements for considering spatial data characteristics in *spatial matching*, addressing the question: *How to measure the similarity and match spatial units of control and treated groups in spatial causal inference?* We describe the challenges along the analytical process of (1) spatiotemporal dependence estimation, (2) covariate selection and

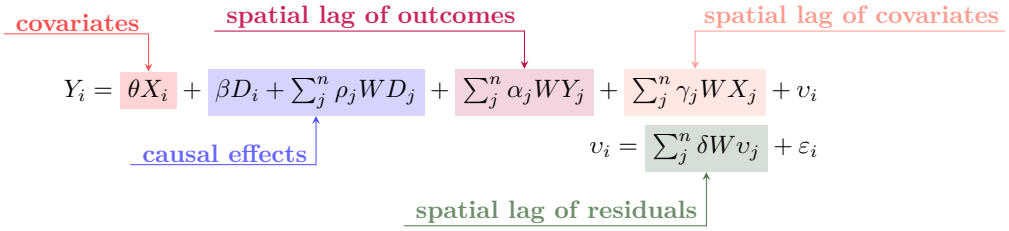
[1] Corresponding Author

prioritization, (3) similarity measurement, and finally, (4) the optimal matching and pairing of spatial units. We present an outline of a decision-making workflow, enabling to reduce bias in spatial causal analysis.

2 Spatial/Spatiotemporal Matching Framework

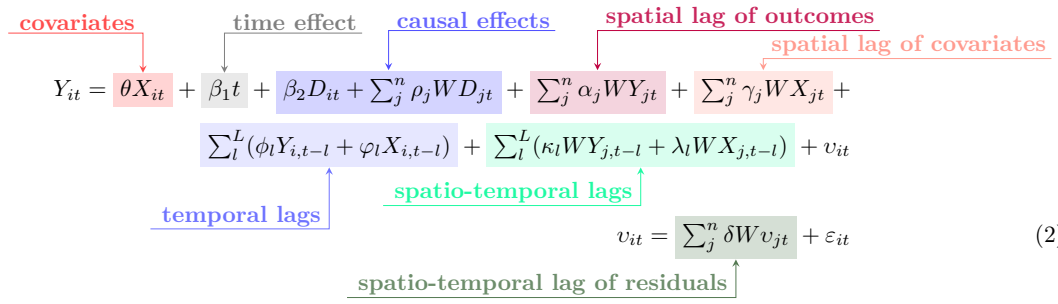
2.1 Data Generation Processes

Causal inference on spatial data should lead to better insights into the spatial data generation process, whether static (i.e., single snapshot) or dynamic (spatio-temporal change). Static spatial data are cross-sectional, i.e., reflecting spatial dependence in the system but not capturing change over time. A general static spatial causal model can be expressed by Equation 1, where β and ρ_j are the coefficients that quantify the direct and indirect causal effects, respectively. Y, D, W, X, and ε refer to the outcome variable, treatment, weight matrix, observed covariates, and the error term sequentially. Correspondingly, $\beta, \theta, \rho, \alpha, \gamma, \delta$ are coefficients.



$$Y_i = \theta X_i + \beta D_i + \sum_j^n \rho_j W D_j + \sum_j^n \alpha_j W Y_j + \sum_j^n \gamma_j W X_j + \upsilon_i$$
$$\upsilon_i = \sum_j^n \delta W \upsilon_j + \varepsilon_i \quad (1)$$

In dynamic spatial systems, temporal dependence should be considered in addition to spatial dependence. Data capturing dynamical systems are *spatial panel data*, and these require spatiotemporal causal models. Equation 2 shows the structure of the general spatiotemporal data generation process resulting in spatial panel data. There, β_2 and ρ_j are the coefficients that need to be determined to quantify the direct and indirect causal effects sequentially. t is the time variable, where Y_{jt} and X_{jt} are the outcome variable and covariates of adjacent neighbors for spatial unit i at time t. $Y_{i,t-l}$ and $X_{i,t-l}$ are the outcome and covariates of time lag l of spatial unit i where L is the number of effective temporal lags chosen by the researcher. Similarly, $Y_{j,t-l}$ and $X_{j,t-l}$ are the outcome and covariates of different time lags of neighbours of the spatial unit i. Spatial and spatiotemporal observational data generation processes thus hold different characteristics which should be reflected in matching strategies.



$$Y_{it} = \theta X_{it} + \beta_1 t + \beta_2 D_{it} + \sum_j^n \rho_j W D_{jt} + \sum_j^n \alpha_j W Y_{jt} + \sum_j^n \gamma_j W X_{jt} +$$
$$\sum_l^L (\phi_l Y_{i,t-l} + \varphi_l X_{i,t-l}) + \sum_l^L (\kappa_l W Y_{j,t-l} + \lambda_l W X_{j,t-l}) + \upsilon_{it}$$
$$\upsilon_{it} = \sum_j^n \delta W \upsilon_{jt} + \varepsilon_{it} \quad (2)$$

2.2 The Spatial and Spatiotemporal Matching Process

Imagine the need to measure the causal effect of an intervention, e.g., the effect of opening new train stations on the average suburb property prices. There are two groups of suburbs: suburbs with new stations (treated group) and those without (control group). Because we lack control over the assignment of suburbs to treated and control groups, a matching process is needed to manage the selection and confounding biases on the measured causal effects, i.e., price increases may be due to other effects. Figure 1 shows a simple matching process for five suburbs. A similarity index is computed based on the values of different covariates, and similar suburbs from the treated and control groups are then matched (suburb 1 $\longrightarrow$ suburb 4, and suburb 2 $\longrightarrow$ suburb 5). During the matching process, some units may be pruned (e.g., suburb 3).

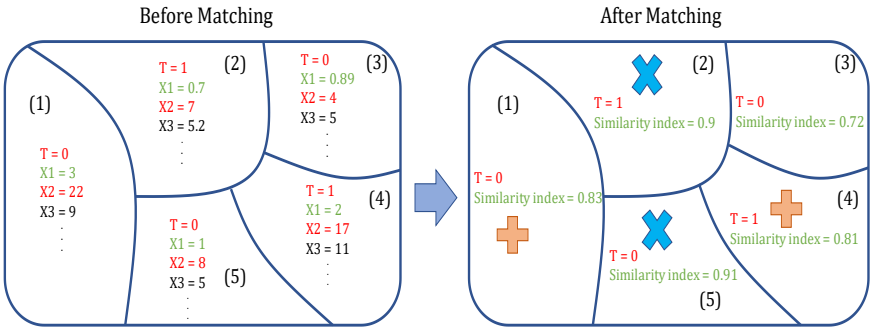


Figure 1 Matching process for suburbs (spatial units); T - treatment, X_i - covariates.

In spatial and spatiotemporal matching (S/STM), we are analyzing data that are sets of variables $\{X_i\}_{i\in S}$ over a study area (S) partitioned into a set of spatial units (locations) in the d-dimensional Euclidean space R_d (where is typically 2). X_i refers to a measured variable at unit i [9]. For each study area S, we may observe a distinct spatial process. The outcome variable $\{Y_i\}_{i\in S}$ then captures a multivariate spatial process. Moreover, the similarity index measured based on the spatial processes $\{X_i\}_{i\in S}$ will be, in turn, a type of a multivariate spatial process, too. We define the similarity index of spatial units $\{Sim_i\}_{i\in S}$ as a multivariate spatial process in S, where $\{Sim_i\}_{i\in S} = g(\{X_i\}_{i\in S})$. g is a similarity measurement function based on the baseline spatial processes for each unit i. For spatial matching we define M with a function of pairing and matching based on the measured Sim_i for the treated and control units, $M = f(k, j \in S|\{Sim_k\}_{k\in S_1}, \{Sim_j\}_{j\in S_2}, W)$. f is a function that matches, i.e., pairs of units – a unit k from the treated group (S_1) and a matching unit j from the control group (S_2), conditional on the similarity based on the values of covariates and the values of a spatial function W, representative of the variation in spatial dependence and heterogeneity.

Optimal matching seeks to find pairs close to a hypothetical exact matching process. The difference between the output of a defined matching procedure and exact matching is known as *imbalance* (I). If I_S is a degree of imbalance for the whole set of spatial units before matching and I_z the imbalance after matching for a subset of the entire dataset z, we aim to reduce imbalance, $I_z < I_S$. Imbalance can thus be defined as a mean of absolute differences of similarity index values between spatial treated and control units, $I(z_m) = \frac{\sum_{k,j\in z_m}(|Sim_k - Sim_j|)}{n}$. Where z_m refers to the set of spatial observations after

matching by method m. Sim_k and Sim_j are similarity index values of units in the treated and control groups, respectively, and n is the number of matched pairs identified using method m.

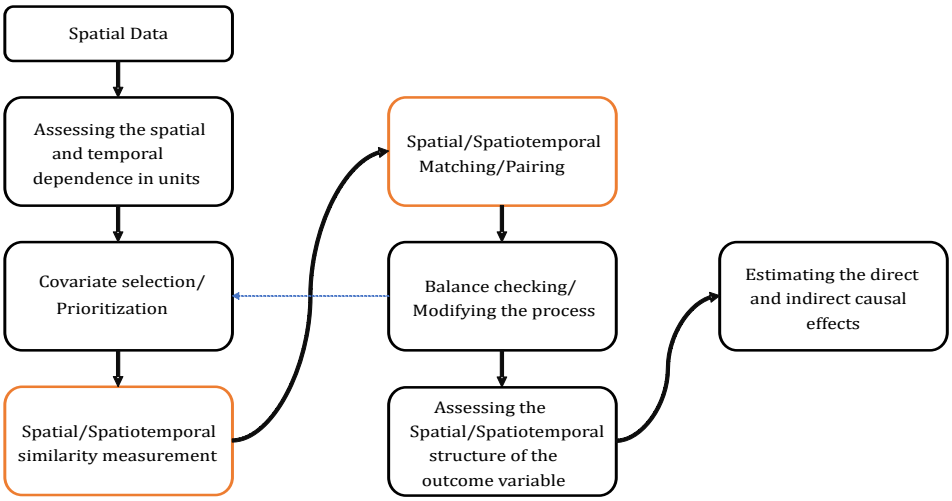


Figure 2 Framework of Spatial Matching Process.

2.3 Components of the Framework

Figure 2 shows our proposed framework of an S/STM procedure to reduce bias in the estimation of spatial causal effects. We now discuss the challenges of each step.

Spatial and Temporal Dependence. Spatial or temporal autocorrelation in covariates is an important factor that must be considered in the S/STM. A similarity index based on spatial covariates of the spatial units may also result in a multivariate spatial process. The spatial dependence in the covariates can be effective in measuring the value of the similarity index. Failing to consider spatial or temporal dependence in the similarity measurement and pairing steps may result in imbalanced and biased matches.

Covariate Prioritisation. Covariate prioritization (CP) is a procedure that enables bringing the qualitative domain expertise of experts to the process of matching and thus minimizing the imbalance. An exact matching method based on categorical covariates (one of the simplest methods for CP) and the caliper matching methods (predefined threshold on a covariate) are both established methods of CP [7]. CP based on expert input can effectively contribute to the matching process and causal effect inference, yet caution is required to avoid bringing human bias into the matching process.

Similarity Measurement of Spatial Units. The effect of the spatial structure of covariates on the value of similarity indices is a challenge for S/STM, as is the effect of time dependence when matching spatial panel data. We defined the spatial similarity with Equation 2.2 as a multivariate spatial process that may be affected by spatial dependence. PS has been suggested as a similarity index for spatial units in a matching procedure [2, 11]. However, King and Nielsen [8] recently showed that PS may lead to matching imbalance, known as the Propensity Score Paradox. Mahalanobis distance has also been applied to quantify the

similarity between spatial units, yet it is problematic when applied on units with a large number of covariates that are not normally distributed [5]. Hence, specialized similarity indices should be developed for spatial and spatiotemporal similarity quantification.

Spatial/Spatiotemporal Pairing. Following similarity measurement, the process of finding a suitable match (i.e., pairing) remains one of the main hurdles of the overall process for S/STM. Current pairing methods in S/STM include Nearest Neighbour search, caliper methods, and threshold distance-based search. Matching pairs only based on distance or search radius threshold between the spatial units is not optimal, as it neglects spatial/temporal autocorrelation and heterogeneity. Such simplistic pairing processes may lead to incorrect matches and, consequently, incorrect estimates of causal effects. An integrated pair search method is required for S/STM, including the multidimensional consideration of similarity across time and geographical distance and neighbourhood evaluation. As shown in Equation 2.2, after measuring similarity, the definition of a suitable spatial function W is essential for a match with low imbalance.

Balance Checking. Ultimately, the most influential factor in the outcomes of causal analysis is having a balanced set of treated and control group members after matching. Therefore, balance checking is a critical validation step that enables to have unbiased causal inference. Equation 2.2 shows the quantification of imbalance for a given matching model. The goal is to achieve an imbalance close to the imbalance of a theoretical exact matching process. Typically, a threshold value of imbalance is set for the imbalance index. Metrics for assessing imbalance in the matching process include standardized mean differences [4], and variance ratios [14]. For spatial data, new metrics may be needed enabling to better assess the balance in the matching process on S/STM data, with special consideration for the characteristics of static and dynamic spatial data.

Spatial/Spatiotemporal Structure of Outcome Variables. Common spatial data generation processes include the General Nesting Spatial Model, Spatial Autoregressive Combined Model, Spatial Durbin Model, Spatial Durbin Error Model, Spatial Autoregressive Model, Spatial Lag of X Model, Spatial Error Model, and Ordinary Least Squares Model [6]. In the case of dynamic data, the spatiotemporal data generation processes we may consider the General Nesting Spatiotemporal Model, Vector Autoregressive Model [17], Spatiotemporal Autoregressive Combined Model, Spatiotemporal Durbin Model, Spatiotemporal Durbin Error Model, Spatiotemporal Autoregressive Model, Spatiotemporal Lag of X Model, and Spatiotemporal Error Model [3]. The nuanced selection of the right model applicable to the data generation process will allow for better quantification of the causal effect, which is the final step in our framework.

Estimation of Direct and Indirect causal Effects. The direct and indirect treatment effects resulting from spatial dependence between units must be considered to quantify the true effect of treatment accurately (e.g., a policy intervention applied to different spatial units, for instance, administrative regions). The indirect effects occur when treated spatial units are adjacent to untreated spatial units. In Equations 1 and 2, the first and second terms of the causal effect component relate to the direct and indirect effects, respectively. After matching, an assessment of the structure of the underlying data generation process is needed before the quantification of causal effects.

3 Conclusion

We discussed the challenges of the spatial and spatiotemporal data matching process, a critical analytical step in causal analysis. We proposed an outline of a framework for spatial matching. We discussed why spatial dependence and spatial heterogeneity challenge the matching process on spatial data. The effects of temporal autocorrelation in panel spatial data further complicate matching. We discussed, in particular, the issue of imbalance in matching results, including when applying similarity measurement methods that do not explicitly consider the spatio-temporal structure in the matching process (e.g., PS or MD matching), failing to capture the effects of spatial dependence and heterogeneity. We reflected upon the need to explore nuanced unit similarity measurement in space. We next will address the process of supporting analysts through this matching framework computationally, aiming to investigate whether the interpretation of the structures in the data may be automated to the extent that analysts can be guided through the process.

References

1 Kamal Akbari, Stephan Winter, and Martin Tomko. Spatial causality: A systematic review on spatial causal inference. *Geographical Analysis*, 2021.

2 André Luis Squarize Chagas, Rudinei Toneto, and Carlos Roberto Azzoni. A spatial propensity score matching evaluation of the social impacts of sugarcane growing on municipalities in brazil. *International Regional Science Review*, 35(1):48–69, 2012.

3 J. Paul Elhorst. Dynamic spatial panels: models, methods, and inferences. *Journal of Geographical Systems*, 14(1):5–28, 2012.

4 Bernhard K Flury and Hans Riedwyl. Standard distance in univariate and multivariate analysis. *The American Statistician*, 40(3):249–251, 1986.

5 Xing Sam Gu and Paul R Rosenbaum. Comparison of multivariate matching methods: Structures, distances, and algorithms. *Journal of Computational and Graphical Statistics*, 2(4):405–420, 1993.

6 Solmaria Halleck Vega and J Paul Elhorst. The slx model. *Journal of Regional Science*, 55(3):339–363, 2015.

7 Luke Keele and Dylan S. Small. Comparing covariate prioritization via matching to machine learning methods for causal inference using five empirical applications. *The American Statistician*, 75(4):355–363, 2021.

8 Gary King and Richard Nielsen. Why propensity scores should not be used for matching. *Political Analysis*, 27(4):435–454, 2019.

9 Dirk P Kroese and Zdravko I Botev. Spatial process simulation. In *Stochastic geometry, spatial statistics and random fields*, pages 369–404. Springer, 2015.

10 David O'Sullivan and David Unwin. *Geographic information analysis*. John Wiley & Sons, 2014.

11 Georgia Papadogeorgou, Christine Choirat, and Corwin M Zigler. Adjusting for unmeasured spatial confounding with distance adjusted propensity score matching. *Biostatistics*, 20(2):256–272, 2019.

12 Paul R Rosenbaum and Donald B Rubin. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55, 1983.

13 Donald B Rubin. Bias reduction using mahalanobis-metric matching. *Biometrics*, pages 293–298, 1980.

14 Donald B Rubin. Using propensity scores to help design observational studies: Application to the tobacco litigation. *Health Services and Outcomes Research Methodology*, 2(3):169–188, 2001.

15 Henrik Toft Sorensen, Timothy L. Lash, and Kenneth J. Rothman. Beyond randomized controlled trials: A critical comparison of trials with nonrandomized studies. *Hepatology*, 44(5):1075–1082, 2006.

16 Elizabeth A Stuart. Matching methods for causal inference: A review and a look forward. *Statistical science: A review journal of the Institute of Mathematical Statistics*, 25(1), 2010.

17 Mark W Watson. Vector autoregressions and cointegration. *Handbook of econometrics*, 4:2843–2915, 1994.

18 Liuyi Yao, Zhixuan Chu, Sheng Li, Yaliang Li, Jing Gao, and Aidong Zhang. A survey on causal inference. *ACM Trans. Knowl. Discov. Data*, 15(5):74:1–74:46, 2021. doi:10.1145/3444944.

An Entropy-Based Model for Indoor Self-Localization Through Dialogue

Kimia Amoozandeh[1]
The University of Melbourne, Parkville, Australia

Ehsan Hamzei
The University of Melbourne, Parkville, Australia

Martin Tomko
The University of Melbourne, Parkville, Australia

Abstract

People can be localized at a particular location in an indoor environment using verbal descriptions referring to distinct visible objects (e.g., landmarks). When a user provides an incomplete initial location description their location may remain ambiguous. Here, we consider a dialogue initiated to update the initial description, which continues until the updated description can be related to a location in the environment. In each interaction, the wayfinder is incrementally asked about the visibility of a particular object to update the initial description. This paper presents an entropy-based model to minimize the number of interactions. We show how this entropy-based model leads to a significant reduction of interactions (i.e., reduction of conversation length, measured by the number of additional referents) compared to baseline models. Moreover, the effect of the initial description, i.e., the first set of visible objects with different combinations, is investigated.

2012 ACM Subject Classification Human-centered computing → Interaction paradigms; Human-centered computing → Interactive systems and tools; Information systems → Location based services

Keywords and phrases Indoor self-localization, Dialogue, Entropy

Digital Object Identifier 10.4230/LIPIcs.COSIT.2022.24

Category Short Paper

Supplementary Material *Software (Source Code and Data)*: `https://github.com/hamzeiehsan/dialogue-based-self-localization`

Funding The support by the Australian Research Council grant DP210101156 is acknowledged.

1 Introduction

Self-localization is a fundamental prerequisite of navigation systems for indoor environments [8, 7]. Existing indoor localization systems often rely on infrastructures (e.g., WiFi, beacons) [6]. Yet, such systems are not universally available and are associated with significant costs of installation, affected by disruptions, and require maintenance by operators [11]. Infrastructure-independent approaches for localization based on verbal communication may help overcome these challenges.

People can be localized within a space through verbal descriptions containing references to a set of visible objects (e.g., landmarks) [1]. In this approach, the space is decomposed into a set of discrete regions that are each characterized by a set of visible objects in the environment known as *visibility signature*.

Yet, when self-localizing with verbal descriptions, people may provide incomplete descriptions, resulting in ambiguous specifications of their position in the environment. As a solution, we envisage the initiation of dialogue to resolve description ambiguity (see Figure 1).

[1] Corresponding author

This localization dialogue thus involves two agents (i.e., computer and human): a *guide* and a *wayfinder*, who interact to achieve a common goal – i.e., identifying the location of the wayfinder [2]. For this purpose, Gintner et al. [3] introduced a method to improve the localization process for visually impaired people by interacting with the user to derive the direction of the user and the particular sidewalk of the street on which they are. From another point of view, Sernani et al.[9] introduced a chatbot system based on voice and written interaction in museums combining indoor localization and chatbots to offer customized visits. In their approach, wearable sensors and a mobile map are required for localization with Ultra-WideBand (UWB) radio technology. This paper proposes a dialogue-based localization approach independent of any particular infrastructure.

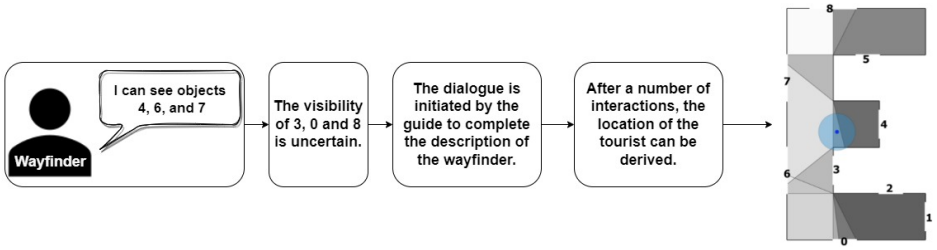


Figure 1 The human description of their location, through reference to an incomplete set of surrounding objects, i.e., facades. A dialogue is initiated to update the description and identify the location of the wayfinder in the environment.

Hua et al. [5] introduced a qualitative place map to enable landmark-based interactive localization. In their approach, the agent describes the nearest landmarks based on different qualitative relations, such as direction and distance. The agent's location is then derived based on the description matched to an existing qualitative graph map. The challenge of their [5] approach is that since the vertices of the place graph are landmarks and turning points, the location of the agent can only be assigned to these vertices. Hence, in some complex environments (e.g., airports), the resulting position determination may not be accurate enough.

In order to enable indoor localization through dialogue in our approach, we assume that the *guide* has access to a map of decomposed regions based on location signatures, as recently proposed by Amoozandeh et al. [1]. We further assume that the *wayfinder* has a 360-degree view of the environment and the ability to identify visible objects that are captured in the signatures. The agents then interact and collaborate to locate the wayfinder through dialogue [2]. The wayfinder initiates the dialogue with reference to the first set of visible objects. If incomplete, this initial description needs to be updated through dialogue to add references until an unambiguous match to a signature of one of the decomposed regions is eventually found. This dialogue may consist of multiple iterative rounds of interaction.

In this paper, we hypothesize that the number of interactions in the dialogue (i.e., dialogue length) can be effectively minimized using a measure of *entropy* evaluating references selected for the update of the initial description. This paper addresses the following research question:

- How can the number of interactions for localizing a person in an environment be minimized?
- How does the amount of references in the initial description affect the number of interactions in the entropy-based reference search compared to the baselines?

Since *entropy* measures the quantity of information held by a quantum of data [10], in order to minimize the number of interactions, objects with a high value of entropy are hypothesized to lead to less ambiguous descriptions faster (Section 2). In order to answer the

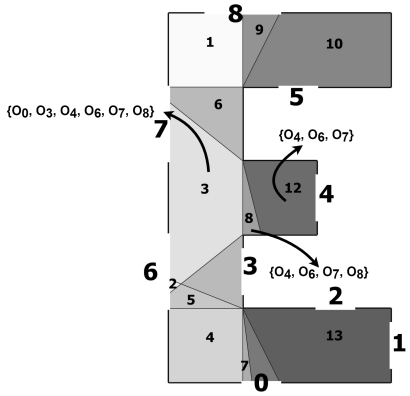
research questions, we compare the *entropy*-based approach to two baseline models, which update the initial description querying about the visibility of objects that are not included in the initial description either randomly or in a supervised (guided) manner.

The three approaches are tested with different combinations of initial descriptions (Section 3). The results show that our method performs similarly to the baselines with complete initial information. In contrast, with incomplete initial information, the method outperforms the baselines and significantly reduces the dialogue length (Section 3).

2 Approach

To initiate a dialogue for indoor self-localization, an initial description of the location of the *wayfinder* is required. The initial description is a set of references to visible objects in the space (e.g., I can see objects O_4, O_6, and O_7 according to the conversation in Figure 2b). This set of references is supposed to identify the wayfinder's location in the space uniquely. Here, a location is synonymous with one of the regions identified by decomposition of space by visibility signatures of a set of visible objects (landmarks), located along the periphery of the space through which the wayfinder navigates [1]. We note that these are not point-like objects but objects with a linear extent.

In Figure 2a we show a hypothetical environment with arbitrary objects (the extent of which is shown as along the boundaries). The grey-shaded polygons are the decomposed regions from each of which a distinct set of objects is visible. The set of objects O_4, O_6, O_7 is visible from region number R_{12}. If the initial description is incomplete, a dialogue is initiated to update and complete the description (see Figure 3). In each interaction between the *wayfinder* and the *guide*, the guide asks whether a particular object is visible to the *wayfinder*. Based on the answer (i.e. visible or not visible), the set of references in the initial information is updated (e.g., can you see object O_3?) and, again, matched to the signatures (see Figure 2b).



(a)

Question	Answer	Potential regions
What can you see?	Ini_Desc = O_4, O_6, O_7	R_2, R_3, R_8, R_{12}
Can you see 3?	No (Update_1 = O_4, O_6, O_7)	R_8, R_{12}
Can you see 8?	Yes (Update_2 = O_4, O_6, O_7, O_8)	R_8

(b)

Figure 2 (a) An example of *visibility signatures* for decomposed regions in an E_shaped environment with ten arbitrary objects. For instance, from R_12, the set of objects O_4, O_6, O_7 is visible, and (b) A sample dialogue between the *guide* and the *wayfinder* to localize the wayfinder in the environment depicted in Figure 2a. O: object, R: Region, Ini_Desc is the initial description.

In each interaction, the *guide* needs to choose an object and ask about its visibility. This object can be chosen in three ways: randomly, supervised between remaining objects, or based on the *entropy* measure among the remaining objects, i.e., the objects that are not

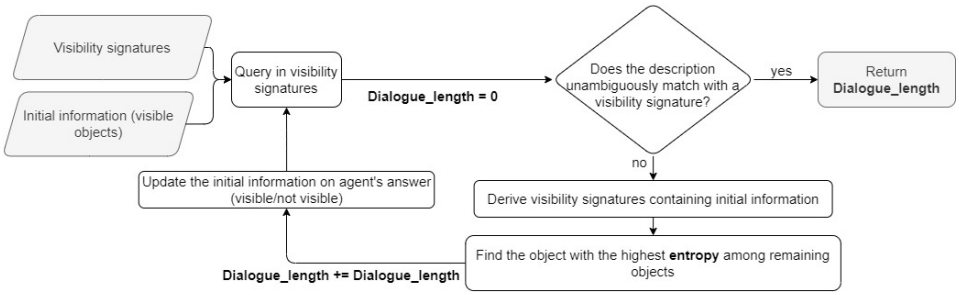


Figure 3 Overview of the entropy-based search to complement and match an initial location description to the environment.

included in the initial set of visible objects and visible from the potential regions according to the initial information. A sample dialogue based on the *entropy* measure is shown in Figure 2b. Assuming that the *wayfinder* is in region R_8, the initial description from the *wayfinder* is the set O_4, O_6, O_7. According to Figure 2a, the *wayfinder* can be in regions R_2, R_3, R_8, or R_{12}. The visibility of objects O_0, O_1, O_2, O_3, O_5, and O_8 can be asked in the next interaction.

The entropy will help to identify candidate object that invalidates the most candidate regions with visibility signatures partially matching the initial description (see Equation1). We treat these regions as candidate regions. For instance, based on the initial description in Table 2b (objects O_4, O_6, and O_7), we can identify candidate regions based on the partially matching visibility signatures (Table 1). The entropy of remaining objects O_0, O_1, O_2, O_3, O_5, and 8 is calculated based on Equation 1 and shown in Figure 4.

$$Entropy_{O_k} = -(P(v)\log_2 P(v) + P(nv)\log_2 P(nv)) \tag{1}$$

In Equation 1, $P(v)$ is the probability that the object O_k is visible in the decomposed region matching the initial description, and $P(nv)$ is the probability that O_k is not visible in the decomposed region matching the initial description. The object with the highest entropy will be enquired about next, to update the initial description. Amongst the remaining objects in the example in Table 1, objects O_3 and O_0 have the highest entropy. Choosing one of these objects means that only one other question is required to derive the wayfinder's location, regardless of the answer. Different possible combinations of answers and questions are shown in Figure 4.

Table 1 *Visibility signatures* of the potential regions based on the initial description of the conversation in Table 2b.

$\text{Region}_{\text{id}}$	O_0	O_1	O_2	O_3	O_4	O_5	O_6	O_7	O_8
R_2	1	1	0	1	1	0	1	1	1
R_3	1	0	0	1	1	0	1	1	1
R_8	0	0	0	0	1	0	1	1	1
R_{12}	0	0	0	0	1	0	1	1	0

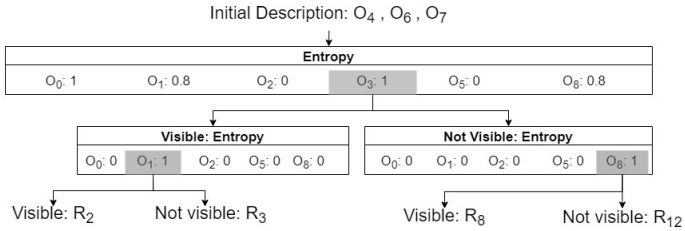


Figure 4 Choosing the questions based on the entropy of remaining objects in the conversation shown in Figure 2b.

To show the effect of entropy on the reduction of the dialogue length, the entropy-based model is compared with two baseline models enabling to update the initial description: references chosen randomly and a guided supervised search among the set of objects that are not in the initial description, i.e., remaining objects that includes *only* the objects that are visible from each candidate region but not all of them. For instance, the next answer object in our scenario can be selected randomly among all the remaining objects in the environment, i.e., objects O_0, O_1, O_2, O_3, O_5, and O_8. Alternatively, since objects O_2 and O_5 cannot be seen from the regions identified as candidates by the initial description (Table 1), the next object may be chosen in a supervised (guided) manner among objects O_0, O_1, O_3, and O_8. All three models are tested with different numbers of references in the initial description and combinations of initial descriptions to illustrate the effect of the initial description on the number of interactions.

3 Results

The three methods have been applied to six different hypothetical environments, generated based on the font outlines of san-serif characters from the capitalised Roman alphabet (Figure 5). For each decomposed region set as the *wayfinder*'s location, different combinations of visible objects have been introduced as the initial description (a subset of the visibility signature). Next, the dialogue was initiated, and an additional object's visibility has been enquired about among the remaining objects.

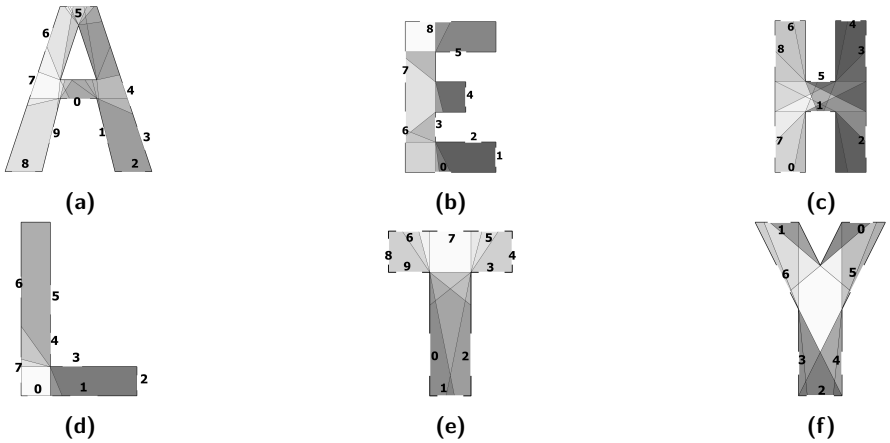


Figure 5 Different environments with distinct visible objects along the boundary. The discrete gray shaded polygons in the environments are the decomposed regions. A set of objects is visible from each decomposed region [1].

In the random model, the object was chosen randomly amongst the remaining objects in the environment. In the supervised model, the object has been chosen among the objects in the candidate regions (i.e., partially matching the original signature). Finally, in the entropy-based model, the object was chosen from the same set matching candidate regions but in order of decreasing entropy.

Figure 6 illustrates the effect of the amount of initial information and the model used on the length of the dialogue for the test environments (Figure 5). Using the entropy-based model leads to a shorter dialogue than in the random or supervised models. Further, the more objects are referred to in the initial description; the lesser is the effect of the model (incl. entropy model) on the dialogue length. In other words, the entropy model outperforms in minimal information situations, but the improvement is less notable for comprehensive initial descriptions.

Moreover, in some cases, such as in Figure 5b, d and e, the number of interactions (dialogue length) remains constant with the increasing amount of initial information when using the entropy-based model. It means that the spatial relation of objects and the configuration of the space can affect the number of interactions in the dialogue, which will be investigated in future works. Thus, the information provided by references to additional objects may be excessive.

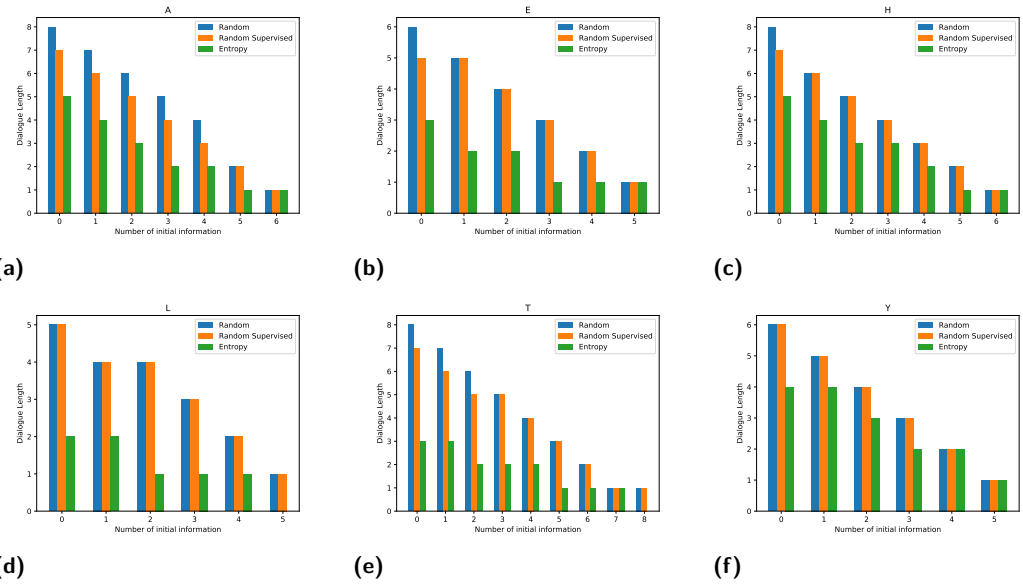


Figure 6 Comparison of the dialogue length in different environments based on the amount of initial information and using baseline approach and entropy approach.

4 Conclusion and Future Works

In this paper, we introduced a dialogue-based approach to localize wayfinders in indoor environments based on the visibility of objects (e.g., landmarks). We show that to minimize the number of interactions in a localization dialogue (dialogue length) between the wayfinder and a guide, the proposed entropy-guided method can be used effectively. Our approach significantly reduces the required dialogue length compared to a random or supervised baseline, even with minimal initial information.

In future work, we will test the entropy-based approach in real indoor environments and evaluate its performance with human wayfinders in realistic scenarios. We aim to enhance the approach to derive additional information from each interaction. The current approach relies only on prompted verification questions (i.e., can you see object O_4?), while other types of questions (e.g., disjunctive questions: can you see object O_3 or O_4?) may be used for more informative interaction. Using diverse types of questions [4], we expect to gain improvements and thus further reduce dialogue length. Moreover, we did not consider the intrinsic differences among the landmarks (e.g., some are more salient than others). However, in a real-world scenario, such differences are essential, and their differences should be considered for a dialogue-based localization. Finally, the complexity of the environment and its impact on dialogue length is not comprehensively investigated in this preliminary study and provides promising avenues for future work.

References

1 Kimia Amoozandeh, Stephan Winter, and Martin Tomko. Space decomposition based on visible objects in an indoor environment. *Environment and Planning B: Urban Analytics and City Science*, 49(3):883–897, 2022. doi:10.1177/23998083211037347.

2 Harm de Vries, Kurt Shuster, Dhruv Batra, Devi Parikh, Jason Weston, and Douwe Kiela. Talk the walk: Navigating new york city through grounded dialogue. *arXiv preprint*, 2018. arXiv:1807.03367.

3 Vojtech Gintner, Jan Balata, Jakub Boksansky, and Zdenek Mikovec. Improving reverse geocoding: Localization of blind pedestrians using conversational ui. In *2017 8th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)*, pages 000145–000150. IEEE, 2017.

4 Arthur C. Graesser and Natalie K. Person. Question asking during tutoring. *American Educational Research Journal*, 31(1):104–137, 1994. doi:10.3102/00028312031001104.

5 Hua Hua, Peng Zhang, and Jochen Renz. Qualitative place maps for landmark-based localization and navigation in gps-denied environments. In *Proceedings of the 27th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pages 23–32, 2019.

6 Germán Martín Mendoza-Silva, Joaquín Torres-Sospedra, and Joaquín Huerta. A Meta-Review of Indoor Positioning Systems. *Sensors*, 19(20), 2019. doi:10.3390/s19204507.

7 Daniel R Montello. Navigation. In *The Cambridge Handbook of Visuospatial Thinking.*, pages 257–294. Cambridge University Press, 2005. doi:10.1017/CBO9780511610448.008.

8 Urs-Jakob Rüetschi and Sabine Timpf. Modelling Wayfinding in Public Transport: Network Space and Scene Space. In *Spatial Cognition IV. Reasoning, Action, Interaction*, pages 24–41, 2005.

9 Paolo Sernani, Sergio Vagni, Nicola Falcionelli, Dagmawi Neway Mekuria, Selene Tomassini, and Aldo Franco Dragoni. Voice interaction with artworks via indoor localization: A vocal museum. In *International Conference on Augmented Reality, Virtual Reality and Computer Graphics*, pages 66–78. Springer, 2020.

10 C. E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27(3):379–423, 1948. doi:10.1002/j.1538-7305.1948.tb01338.x.

11 Stephan Winter, Martin Tomko, Maria Vasardani, Kai-Florian Richter, Kourosh Khoshelham, and Mohsen Kalantari. Infrastructure-Independent Indoor Localization and Navigation. *ACM Comput. Surv.*, 52(3), June 2019. doi:10.1145/3321516.

Collaborative Wayfinding Under Distributed Spatial Knowledge

Panagiotis Mavros[1]
Singapore-ETH Centre, Future Cities Laboratory, CREATE campus, 1 CREATE Way, #06-01 CREATE Tower, 138602, Singapore

Saskia Kuliga
Singapore-ETH Centre, Future Cities Laboratory, CREATE campus, 1 CREATE Way, #06-01 CREATE Tower, 138602, Singapore
German Center for Neurodegenerative Diseases (DZNE) Witten site and Faculty of Health, University of Witten/Herdecke, 58453 Witten, Germany

Ed Manley
School of Geography, University of Leeds, Leeds, U.K.

Hilal Rohaidi Fitri
Singapore-ETH Centre, Future Cities Laboratory, CREATE campus, 1 CREATE Way, #06-01 CREATE Tower, 138602, Singapore

Michael Joos
Singapore-ETH Centre, Future Cities Laboratory, CREATE campus, 1 CREATE Way, #06-01 CREATE Tower, 138602, Singapore

Christoph Hölscher
Chair of Cognitive Science, D-GESS, ETH Zürich, Switzerland
Singapore-ETH Centre, Future Cities Laboratory, CREATE campus, 1 CREATE Way, #06-01 CREATE Tower, 138602, Singapore

Abstract

In many everyday situations, two or more people navigate collaboratively but their spatial knowledge does not necessarily overlap. However, most research to date, has investigated social wayfinding under either 1-sided or fully shared spatial information. Here, we present the pilot experiment of a novel, computerised, non-verbal experimental paradigm to study collaborative wayfinding under the face of spatial information uncertainty. Participants (N=32) learned two different neighbourhoods individually, and then navigated together as dyads (D=16), from one neighbourhood to the other. Our pilot results reveal that overall participants share navigational control, but are in control more when the task leads them to a familiar destination. We discuss the effects of spatial ability and motivation to lead, as well as the outlook of the paradigm.

2012 ACM Subject Classification Applied computing → Psychology; General and reference → Experimentation; General and reference → Empirical studies

Keywords and phrases navigation, wayfinding, collaboration, dyad, online

Digital Object Identifier 10.4230/LIPIcs.COSIT.2022.25

Category Short Paper

Funding The research was partly conducted at the Future Cities Laboratory at the Singapore-ETH Centre, which was established collaboratively between ETH Zurich and the National Research Foundation Singapore (NRF) under its Campus for Research Excellence and Technological Enterprise (CREATE) programme. The research was also supported by the Chair of Cognitive Science, ETH Zürich. S. Kuliga at the time of preparing the study held a DAAD Postdoc Fellowship at Singapore-ETH Centre and currently is affiliated at DZNE Witten / University of Witten/Herdecke.

[1] Corresponding author

Acknowledgements We are grateful for the advice of Dr Iva Barisic on social wayfinding, and the support of the ETH Zürich Decision Science Lab team: Stefan Wehrli, Lea Imhof, and Salome Egli.

1 Introduction

In everyday life, we often navigate together with other people, a process called "social wayfinding". In some cases, we may share similar environmental knowledge, for instance while walking together with a friend around an area we both live in. In other cases, we may have different environmental knowledge than the other person. We define this as *spatial information uncertainty*, where two (or more) people have non-overlapping, complementary spatial knowledge – in other words, each of them only knows parts of the area to be navigated. This is the case, for example when guiding a taxi driver how to find our home, which we need to communicate to them. This phenomenon is called "social wayfinding" [5], and here we focus on the *synchronous/strong* type of social interaction. This is the case of pilot and co-pilot in air-crafts, driver and passenger cars, and of course pedestrians. In such cases, the interpersonal dynamics between then two or more individuals influence both the nature of their interactions as well as their wayfinding performance and outcome.

Previous studies have shown a clear influence of personality traits (spatial ability, leadership tendencies) of each individual, the spatial information available, and the wayfinding agency (who is in command of the steering wheel). Passengers in a car who consider themselves as having a better wayfinding ability than the driver, they are more likely to support the driver (collaborate) providing input for tasks such as pre-planning, en-route wayfinding directions or visual search [3]. Groups of people take more time than individuals to complete the same navigational task [2]. Imbalance of spatial ability in navigating dyads may lead to additional communication and dissatisfaction [6]. Males tend to assume more control in mixed-sex dyads [8]. Groups varied in leadership, some having a navigational leader and others not; this was explained by an overall level of *conscientiousness* of the dyad; however within the dyad *conscientiousness* was not related with leading [1]. Navigational control has been related with individuals' leadership tendencies [2], a construct called *motivation to lead* (MTL) [4]. However, unlike individual-navigation, currently there is not an experiment paradigm to study collaborative navigation in a controlled or networked manner. In addition, in previous studies the navigators (dyads or teams) either all or only one had access to the same amount of information (e.g. a map). Thus, it is not known how people would collaborate if they have known different parts of the environment.

In this pilot study, we investigate these two questions: a) how to study collaborative navigation using a computer-based paradigm, and b) how dyads collaborate under spatial information uncertainty. Specifically, we implement a computerised, non-verbal collaborative navigation paradigm to explore: which player of each dyad assumes more navigation control during a wayfinding trial (RQ1); how taking navigation control is moderated by individual characteristics, i.e., sense of direction and motivation to lead (RQ2); and which strategies individuals use in collaborative navigation (RQ3).

2 Methods

2.1 Participants

Thirty-two participants (N = 32, 17 female; mean age = 23.7 years old) took part in this pilot study, and two (2) more dropped-out due to motion-sickness. They were recruited from the Decision Science Laboratory of ETH Zürich (DeSciL) participant pool, and testing took

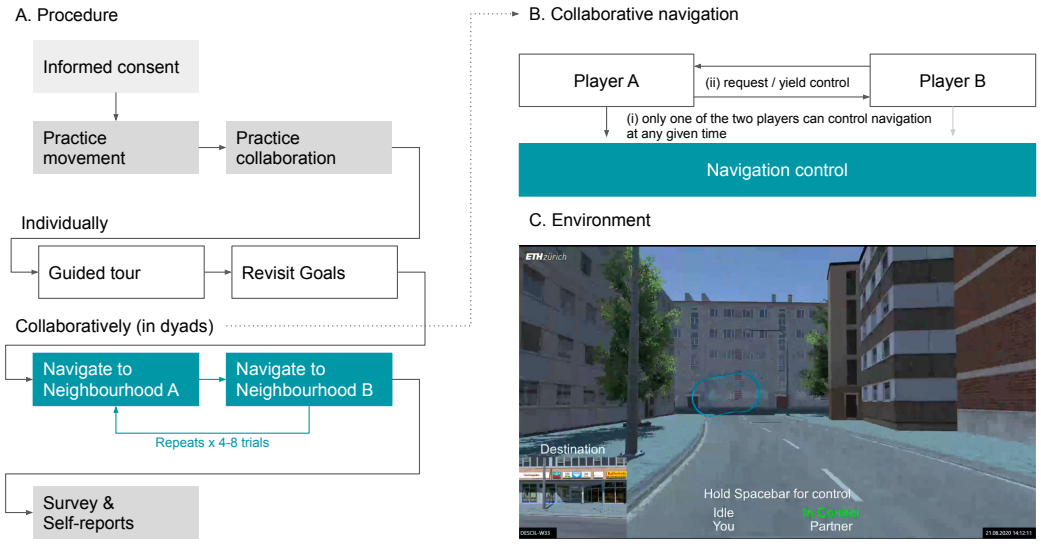


Figure 1 (A) Experimental procedure. (B) Only one of the two players in the dyad is in control of the navigation at any point in time (i), and the other player can request control. (C) Example screenshot of the environment during one navigation task. Goal locations remain visible during the entire task. Participants can always see who is in control. Also note that communication is non-verbal and the only message participants exchange is to request navigation control.

place in DeScil which allows simultaneous data collection during a single session. Participants were compensated 25 Swiss Francs (approximately 26 US$). The study was approved by the ETH Ethics Committee (Project ID: B EK 2019-N-179). All procedures were performed according to the declaration of Helsinki.

2.2 Experimental design

For this pilot study, we adopted single factorial repeated measures design, with a single factor: spatial knowledge of the starting area or the goal area. Specifically, participants learned individually to navigate to 6 locations in one out of two neighbourhoods of a small virtual city (cf. 2.4 and appendix). Subsequently, they were assigned to dyads of distributed spatial knowledge – in other words participants were paired up by the researcher, so that if one participants learned neighbourhood A, the other learned neighbourhood B. We explored whether spatial ability, personality (motivation to lead) or spatial knowledge determined which participant was in charge of navigation during the wayfinding tasks, and also explored their strategies.

2.3 Procedure

Participants were received at DeScil, briefed by the experimenter, and provided consent. They received instruction and practise time on how to look around, navigate, obtain and yield control of the navigation interface.

Training Phase. During this phase, participants were randomly assigned to a neighbourhood (A or B) and they learned individually the location of six (6) goal locations in their assigned neighbourhood (A or B) through a guided tour (passive learning). Afterwards, to assess the quality of their spatial learning, they were immediately teleported to the start of the tour,

and asked to revisit all six goal locations again, navigating independently. After revisiting all goals, they could continue exploring their assigned neighbourhood; note that they could not cross or see their partner's neighbourhood or any large distal landmarks.

Testing phase. Participants were paired in dyads (one from A and one from B). Navigating dyads were asked to complete a series of wayfinding tasks (4 trial for pilot session 1 and 2; 8 tasks for pilot session 3). They were explicitly told they were paired with another partner, and shown how to get and give control. In each trial, the dyad was teleported to a location in one neighbourhood and asked to "find the shortest path" to a location in the other. Each player was sitting on separate monitors and could not see who their partner is. Only one of pair could be in control of the navigation at any given time. The player *not* in control could request control with a simple message (initiated by pressing a button); no other form of communication was allowed.

2.4 Environment and Data logging

Figure 1 C shows an image from the testing environment which was also created in Unity 3D (Unity Technologies, USA). The virtual city was split in two neighbourhoods of approximately similar area and number of streets (see also Figure 4 in the Appendix). Because the two neighbourhoods have similar visual appearance, we enhanced their differentiation with two additional visual cues: one neighbourhood had red and the other blue paving, while one had trees and the other not. Six locations distributed across the entire area of each of the two neighbourhoods were used as goals / destination (12 in total). These included, for example, a fountain, a gas-station, and various shops. The two virtual neighbourhoods were located next to each other, with connecting streets. Testing collaborative navigation in a network virtual setup allows the precise control of environment conditions and confounds, as well as provides a scalable paradigm for spatial cognition research.

Both the experiment procedure and data collection were implemented in Unity 3D, relying on the *UNet* framework (now deprecated) to enable multiplayer sessions. During the (pilot) experiment, behavioural data were recorded at a sampling rate 5Hz, including: the coordinates (x,y,z) of the player (during training) or the dyad (during testing), which player out of the two was in control, as well as the stage of the experiment (training or testing), task order, and other auxiliary data (computer id).

3 Behavioural and self-reported measures

Training performance. Defined as the (virtual) distance travelled to revisit all six landmarks a participant learned during the guided tour. This produced one value per participant.

Testing performance. Defined as the (virtual) distance travelled while performing the dyadic navigation tasks. This produced one value per dyad (for each trial).

Total proportion in control of navigation. Defined as the proportion (measured as number of logs/rows) that each player was in control of the navigation during each task. This produced one value per participant (for each trial).

Proportion in control of navigation per area. Defined as the proportion (measured as number of logs/rows) that each player was in control of the navigation during each task, *and* within each of the two regions. This captured how much a player was in control in the area they had learned versus the unknown area. This produced two values per participant (for each trial).

This data were joined with two standardised questionnaires measuring the self-reported sense of direction (spatial ability) and motivation to lead. We used the Santa Barbara Sense of Direction (SBSOD) to assess participants' self-reported spatial ability [7], which consists of

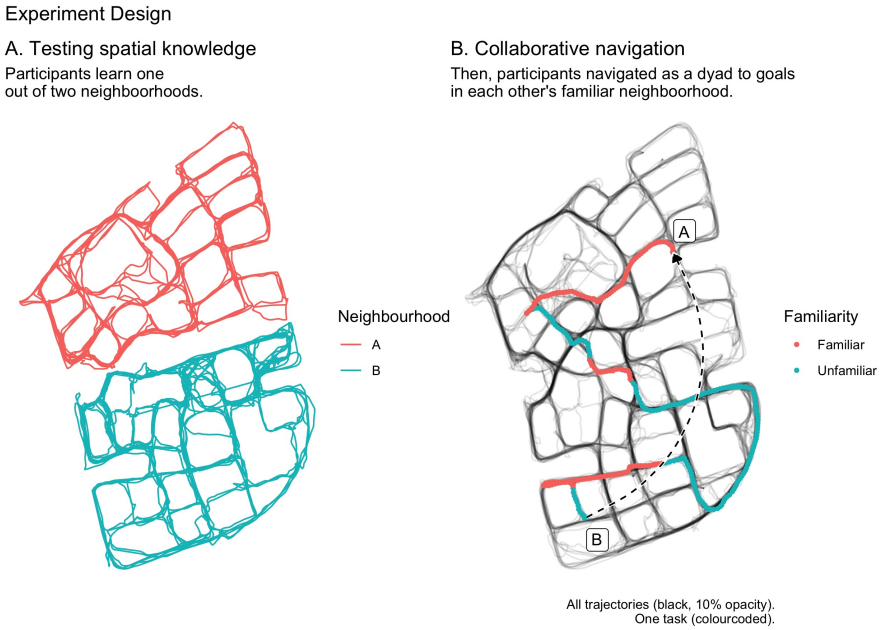


Figure 2 Maps of key behavioural data from wayfinding during the individual training phase (landmark revisit task; left) and the testing phase (dyadic navigation; right). Notice that each participant learns one out of the two neighbourhoods (A or B). They thus have different spatial knowledge of a region. Subsequently, each collaborative wayfinding trial starts in one neighbourhood and ends in the other; thus during the navigation phase, players alternate between being in a familiar environment. We can notice that Player A (unfamiliar with neighbourhood B) started navigating first, before handing over control to Player B (familiar); the pair kept exchanging navigating; eventually Player A (familiar with with neighbourhood A) led them to the destination.

15 items such as "I am very good at giving directions", rated in a 7-point scale. We also used the *motivation to lead* (MTL) scale [4], which includes 27 statements, such as "I usually want to be the leader in the groups I work in", that belong in 3 separate factors (see Introduction).

3.1 Data pre-processing and Analysis

Overall, during 3 pilot sessions with 32 participants we obtained a total of 247,549 rows of data, consisting of training and testing data from both participants of each data (i.e. data logs are recorded on both sides of a dyad for redundancy). During the pre-processing step, the data were split into training phase (i) guided tour and (ii) goal revisit, and testing phase (iii) the dyadic navigation tasks. Based on these data, we computed the behavioural measures defined above. All data processing and analyses were performed in R (see Appendix for list of packages used).

4 Results

4.1 Manipulation check

As an initial test of our experimental paradigm, participants self-reported that they understood the task and interface (see Figure 6 in the Appendix). The majority (75%) also reported that they realised if they were teleported to their "own" (i.e. familiar) or the "other" neighbourhood at the beginning of each trial, although almost half did not notice the visual cues (i.e. trees and pavement colour).

4.2 Spatial learning

As an initial test of how well players learned the environments, we examined their training performance. Note that this was only recorded for pilot session 3 (12/32 participants). Overall, participants varied in their goal-revisit performance, and a bayesian correlation test revealed anecdotal evidence (BF = 1.14) that it was inversely correlated with Sense of Direction ability (Rho: -0.28; 95%Cr.I. = [-0.6, 0.08], probability of direction = 0.87).

4.3 Collaborative navigation

Examining the research questions, we first explored which player of each dyad assumed more navigation control during a wayfinding trial (RQ1). A bayesian ANOVA revealed evidence for a substantial effect of spatial knowledge on navigation control (BF = 2288.885), showing that the player familiar with the destination spent more time in control (see also Figure 5). Note that we did not explicitly tell participants the other player knows the area they do not; so this effect suggests an implicit understanding that if they do not know the location, the other person might do.

We then examined whether the proportion in control measure was moderated by individual participants' characteristics (RQ2): *spatial ability* (sense of direction), and *motivation to lead* (3 subscales: *social-normative*, *non-calculative*, and *affective identity*). Figure 3 shows a trend line overlayed on a scatterplot for each parameter; note this figure includes repeated measurements per participant. The results of linear mixed-effects regression (see formula in Appendix) reveal that in line with the previous analysis, *familiarity* with the Goal area had a significant, negative effect on proportion in control ($b = -0.240, SE = 0.064, p < 0.05$). While a positive trend can be observed for *affective identity* and *non-calculative* (motivation to lead sub-scales) the effect was not significant; however we observed a trend for an interaction of *social-normative* and familiarity ($b = -0.074, SE = 0.039, p = 0.066$).

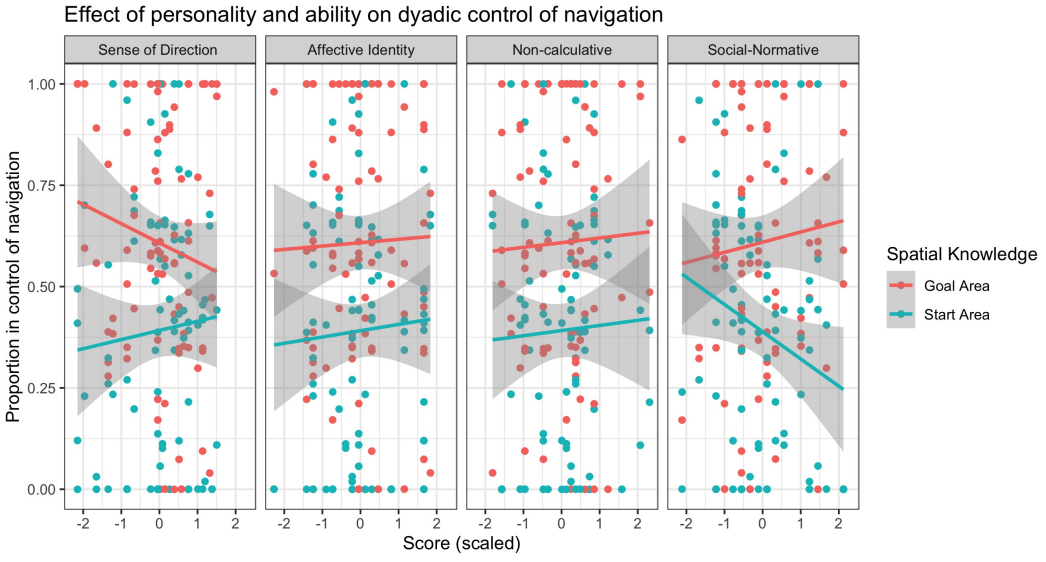


Figure 3 Scatter-plot showing the relationships between each players' characteristics and proportion in control of the navigation in each trial.

Finally, we also explored participants' responses to the post-experiment survey with regards to strategies for collaborative navigation (RQ3) (appendix: Figure 6 bottom). While responses vary, these are inline with the behavioural observation that participants took over

control when navigating to a destination they were familiar with (64.5%), and that they yielded/exchanged control when they felt disoriented (78.8%). We can also note that 37.5% of participants report *not* strictly navigating in the area they learned.

5 Discussion and conclusions

In this pilot study, we demonstrated the feasibility and operation of a novel, non-verbal, online, collaborative navigation paradigm. We explored how people navigate collaboratively in cases where spatial knowledge is distributed, rather than shared equally, among a team (here dyads). Results indicated a positive effect of spatial knowledge of the destination area, as well as interactions between spatial knowledge with spatial ability and *social normative* motivation to lead. These interactions will be further explored in future studies. Two limitations are that non-verbal and remote collaborative navigation may differ from everyday situations; also a larger sample will enable more elaborate statistical analyses. Future work could consider additional analyses of navigation trajectories, beyond shortest paths.

To summarise, here we report on the pilot of a novel, computer-based study paradigm to test non-verbal collaborative navigation under spatial information uncertainty. Our results demonstrate the effectiveness of this paradigm, and open up new opportunities for the study of collaborative navigation, such as large online experiments. Our future work includes a comprehensive experiment to understand navigation under spatial information uncertainty.

References

1 Crystal Bae and Advisor Daniel R Montello. Route Planning and Situated Navigation in Collaborative Wayfinding. In *14th International Conference on Spatial Information Theory (COSIT 2019)*, 2019.

2 Iva Barisic. *Social and spatial factors of wayfinding usability* . PhD thesis, ETH Zürich, 2019.

3 Kelly Jane Bryden, Judith Charlton, Jennifer Oxley, and Georgia Lowndes. Older driver and passenger collaboration for wayfinding in unfamiliar areas. *International Journal of Behavioral Development*, 38(4):378–385, 2014. doi:10.1177/0165025414531466.

4 Kim Yin Chan and Fritz Drasgow. Toward a theory of individual differences and leadership: Understanding the motivation to lead. *Journal of Applied Psychology*, 86(3):481–498, 2001. doi:10.1037/0021-9010.86.3.481.

5 Ruth Conroy Dalton, Christoph Hölscher, and Daniel R. Montello. Wayfinding as a social activity. *Frontiers in Psychology*, 10(FEB):1–14, 2019. doi:10.3389/fpsyg.2019.00142.

6 Gengen He, Toru Ishikawa, and Makoto Takemiya. Collaborative Navigation in an Unfamiliar Environment with People Having Different Spatial Aptitudes. *Spatial Cognition and Computation*, 15(4):285–307, 2015. doi:10.1080/13875868.2015.1072537.

7 Mary Hegarty, AE Richardson, and DR Montello. Development of a self-report measure of environmental spatial ability. *Intelligence*, 30:425–447, 2002. URL: http://www.sciencedirect.com/science/article/pii/S0160289602001162.

8 Pekka Kallioniemi, Tomi Heimonen, Markku Turunen, Jaakko Hakulinen, Tuuli Keskinen, Laura Pihkala-Posti, Jussi Okkonen, and Roope Raisamo. Collaborative navigation in virtual worlds: How gender and game experience influence user behavior. In *Proceedings of the ACM Symposium on Virtual Reality Software and Technology, VRST*, volume 13-15-Nove, pages 173–182, 2015. doi:10.1145/2821592.2821610.

A Appendix

A.1 Virtual environment & navigation tasks

Figure 4 The virtual town, goal locations, and navigation tasks.

The virtual town, consisting of two neighbourhoods (red or blue paving), of approximately equal number of streets and surface area. Green circles indicate goal locations (numbered). Dashed arrows indicate the starting point and the destination of each navigation task, each starting in one neighbourhood and ending on the other. Start locations were chosen to allow multiple alternatives to the goal (i.e. not along the same street). Goal locations (destinations) were chosen so that they that in each neighbourhood 3 goals were closer to the boundary with the opposite neighbourhood and 3 goals towards the edge of virtual town.

A.2 Proportion in control

Proportion in control was influenced by spatial knowledge.

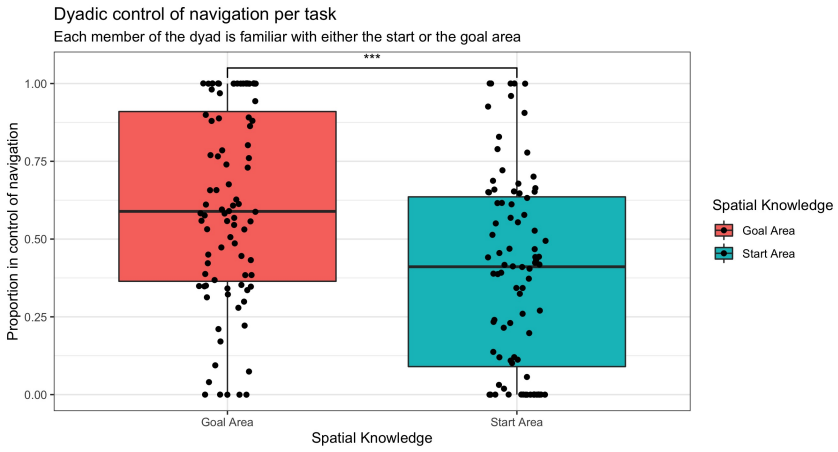


Figure 5 Boxplot showing which participant was more in control during a wayfinding trial, as a proportion of the entire trial. We can observe that overall, the player of the dyad that was familiar with the neighbourhood of the destination (goal area) assumed more navigation control.

A.3 Analysis tools

All analyses were performed in *R* (version 4.1.1), using the following R-packages: *dplyr* (version 1.0.7), *lubridate* (version 1.7.10), *sjPlot* (version 2.8.10), *purrr* (version 0.3.4), *tidyR* (version 1.1.3). Graphics were made using *ggplot2* (version 3.3.5), *sjPlot* (version 2.8.10) and *patchwork* (version 1.1.1). Statistical analyses were performed using the R-package *BayesFactor* (version 0.9.12-4.2) and *correlation* (version 0.6.1).

A.4 Linear mixed-effects model

We fitted a linear mixed effects regression model to understand the interaction between spatial knowledge and inviduals' characteristics. The with formula: $proportion \sim Familiarity + sbsod.score.s : Familiarity + Affective.Identity.s + Noncalculative.s + Social.Normative.s : Familiarity + (1 + Familiarity|pair : ParticipantID) + (1|Order)$

A.5 Self-reports

After completing the experiment, participants were asked to self-report if they found the task and instructions clear, whether they perceived the visual cues distinguishing the two neighbourhoods, and their navigation control strategies (Figure 6).

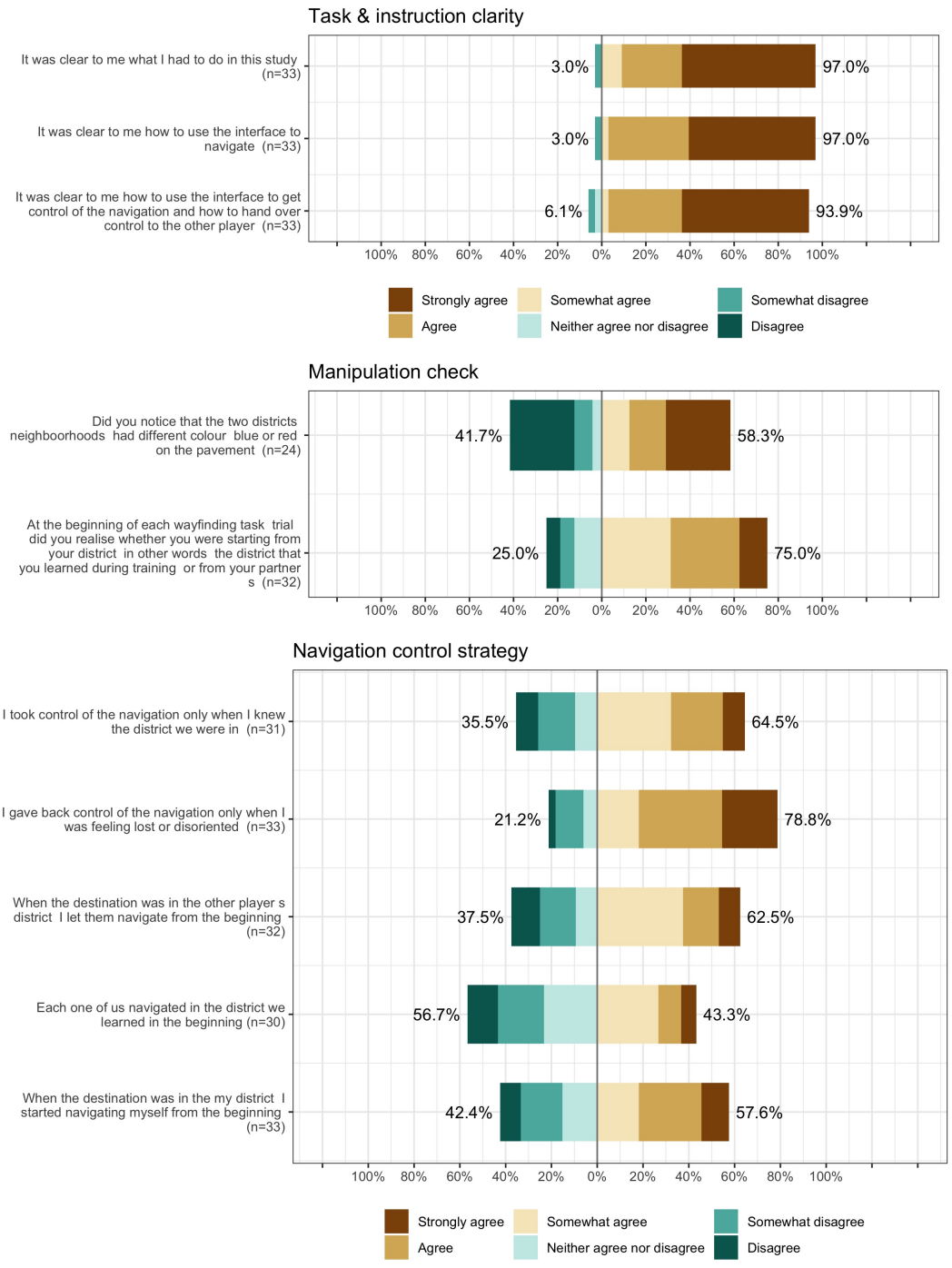


Figure 6 Top: Manipulation check; after completing the experiment, participants were asked if they understood the instructions. Bottom: Participants post-experiment self-reports reveal various strategies of collaborative navigation.

Abnormal Trajectory-Gap Detection: A Summary

Arun Sharma[1]
Department of Computer Science & Engineering, University of Minnesota, Minneapolis, MN, USA

Jayant Gupta
Department of Computer Science & Engineering, University of Minnesota, Minneapolis, MN, USA

Shashi Shekhar
Department of Computer Science & Engineering, University of Minnesota, Minneapolis, MN, USA

Abstract

Given trajectories with gaps (i.e., missing data), we investigate algorithms to identify abnormal gaps for testing possible hypotheses of anomalous regions. Here, an abnormal gap within a trajectory is defined as an area where a given moving object did not report its location, but other moving objects did periodically. The problem is important due to its societal applications, such as improving maritime safety and regulatory enforcement for global security concerns such as illegal fishing, illegal oil transfer, and trans-shipments. The problem is challenging due to the difficulty of interpreting missing data within a trajectory gap, and the high computational cost of detecting gaps in such a large volume of location data proves computationally very expensive. The current literature assumes linear interpolation within gaps, which may not be able to detect abnormal gaps since objects within a given region may have traveled away from their shortest path. To overcome this limitation, we propose an abnormal gap detection (AGD) algorithm that leverages the concepts of a space-time prism model where we assume space-time interpolation. We then propose a refined memoized abnormal gap detection (Memo-AGD) algorithm that reduces comparison operations. We validated both algorithms using synthetic and real-world data. The results show that abnormal gaps detected by our algorithms give better estimates of abnormality than linear interpolation and can be used for further investigation from the human analysts.

2012 ACM Subject Classification Information systems → Data mining; Computing methodologies → Spatial and physical reasoning

Keywords and phrases Spatial Data Mining, Trajectory Mining, Time Geography

Digital Object Identifier 10.4230/LIPIcs.COSIT.2022.26

Category Short Paper

Funding This research is funded by an academic grant from the National Geospatial-Intelligence Agency (Award No. HM0476-20-1-0009, Project Title: Abnormal Trajectory Gap Detection). Approved for public release, 22-379.

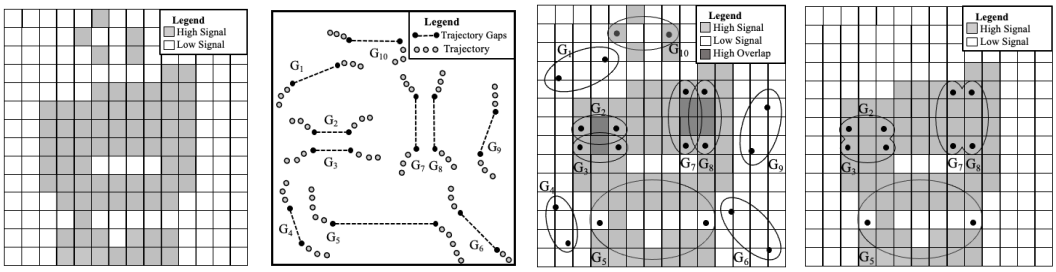
Acknowledgements We also want to thank Kim Koffolt and the spatial computing re- search group for their helpful comments and refinements.

1 Introduction

Given multiple trajectory gaps and a signal coverage map based on historic object activity, we find possible abnormal gaps in activity where moving objects (e.g., ships) may have behaved abnormally such as not reporting their locations in an area where other ships historically did report their location. Figure 1 shows the problem's *input* which includes a map of the signal coverage area (grey cells) for a set of derived historic trajectories (shown in Figure 1a) and trajectory gaps G_1, ... , G_{10} (shown in Figure 1b). Figure 1 also shows the output where

[1] Corresponding author

gaps G_1, G_4 G_6 and G_9 are entirely outside the signal coverage map indicating weak signal coverage. In contrast, the rest of the gaps are overlapping the signal coverage map. The absence of location reporting in an area known to have signal coverage may be interpreted as intentional behavior by a ship that temporarily switched off its location broadcasting device. Figure 1c shows *intermediate output* stage where trajectory gaps (G_1, ... , G_{10}) are modelled in the form of geo-ellipses [12, 10] along with their intersections with the signal coverage map. Figure 1c also shows two-gap pairs, G_2, G_3 and G_7, G_8, that have intersecting regions (i.e., dark grey cells), which suggests two ships in a rendezvous potentially engaging in illegal activity. The gaps entirely outside the signal coverage area have been filtered out. Figure 1d shows the *final output* with gap pairs G_2, G_3 and G_7, G_8 are merged to their overlapping regions and gap G_{10} is filtered out since it did not meet the user defined *priority threshold*.



(a) Input 1: Signal Coverage Map (SCM).

(b) Input 2: Trajectories with Trajectory Gaps (G_i).

(c) Intermediate Output: Abnormal Gap Computation.

(d) Final Output: Summarized Abnormal Gaps.

Figure 1 An illustration of the Abnormal Gap Region Problem (Best in color).

Analyzing trajectory gaps has many societal applications in maritime safety, homeland security, epidemiology, and public safety. Other use cases include tracing comets, tracking marine animals and contact tracing. In this paper, we focus on understanding the potential benefits of signal coverage mapping to improve maritime safety and regulatory enforcement. A signal coverage map helps to reduce false positives by providing historical activity traces for a region, against which abnormal behavior can be detected within a trajectory gap. In addition, current methods assume shortest path in a trajectory gap which leads to many missed patterns since moving objects do not always travel in a straight path. In contrast, our approach is based on a space-time prism, which can accommodate greater movement possibilities around a signal gap where an object of interest could have potentially deviated from the predefined (or known) linear path. We also provide a way to reduce computational cost over a large geographical space.

Related Work. Surveys [2, 18] provide a broad classification of anomaly detection methods in trajectory mining. Riveiro et al. [13] provide a holistic view of maritime anomaly detection but do not cover trajectory gaps. Works that do consider trajectory gaps [11, 3, 8] assume shortest path methods. For instance, the proposed framework in [11] extracts maritime movement patterns assuming shortest path within trajectory gaps.

There are some realistic frameworks [4, 15, 16] that employ reconstruction techniques for modeling uncertainty in trajectory gaps based on space-time prism models [10, 12]. However, they are limited to theoretical simulations and little attention is given to real-world applications [5]. For instance, Winter [16] provide a probabilistic interpretation of a space-time prism. However, the interpretation lacked a real-world validation. A recent work [17]

considers trajectory gaps using a space-time prism but does not consider abnormal behavior within gaps. This paper uses a space-time prism model to capture abnormal gaps based on historical data which allows us to distinguish abnormal gaps for possible anomaly hypothesis.

Contributions. We define an Abnormal Gap Measure (AGM) for modeling abnormal gaps and propose an abnormal gap detection (AGD) algorithm to handle multiple gaps. In addition, we propose a memoized abnormal gap detection (Memo-AGD) algorithm to further improve computational efficiency. We show experimentally that our methods are efficient and have good solution quality. We also present a real-world case study to validate our approach.

Scope and Organization. This work is limited to space-time prisms for computing abnormal gaps and methods such as kinetic prisms [7] are not studied. This paper did not consider acceleration in space-time prism due to data limitations. The use of signal coverage maps based on aerial imagery datasets (e.g., satellite imagery) and it's it's falls outside the scope of this paper. The rest of the paper is organized as follows: Section 2 introduces key concepts and formally defines the abnormal gap detection problem. Section 3 describes the proposed algorithms AGD and refined Memo-AGD. Experiments and Results are presented in Section 4. Finally, Section 5 concludes this work and briefly lists the future work.

2 Problem Definition

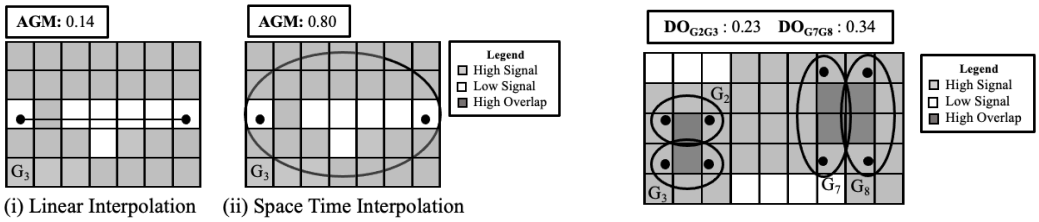
2.1 Basic Concepts

▶ **Definition 1.** *A **signal coverage map (SCM)** is a discretized grid space where cells are color-coded to represent regions with regular historically reported location signals.*

The maps are generated by first computing the total reported ship movement in the area using a set of historic location-traces P_i (i.e., trajectories) within some time interval and then checking if the total is above a given threshold (say θ). Figure 1a shows a synthetic example of signal coverage map based on historical ship movement with binary color-coding. The grey cells in the figure have movement above the threshold (i.e., High Signal), whereas the white cells have low or no reported historical movement (i.e., Low Signal).

▶ **Definition 2.** *An **abnormal gap measure (AGM)** for a gap G_i is the probability that a possible location of the object during the gap (unreported data time interval) has signal coverage. A higher value of AGM indicates anomalous behavior since it means an object is not reporting its location despite having the location signal coverage in the past.*

The probability is computed using interpolated grid cells (GC_{int}) and regions with high historical movement (GC_m). We first compute the overlap between GC_{int} and GC_m, and then normalize the overlap with GC_{int}. Figure 2a provides two examples of computing the abnormal gap measure between two points using cells color-coded based on SCM. The first shows the AGM for linear interpolation, where GC_{int} is the set of cells crossed or touched by a line between the points (i.e., 7) which overlaps with only 1 cell with a high historically reported movement. Thus, the AGM is 0.14 ($\frac{1}{7}$). The second shows the AGM based on (GC_{int}) for space-time interpolation to accommodate additional (movement) possibilities. As shown, GC_{int} is the number of cells crossed or touched by the ellipse (i.e., 35) which overlaps with 28 cells with a high historically reported movement. Thus, the AGM is 0.80 ($\frac{28}{35}$).



(a) Abnormal Gap Measure. **(b)** Degree of Overlap.

Figure 2 Examples of Abnormal Gap Measure (Left) and Degree of Overlap (Right).

▶ **Definition 3.** *Given a pair of gaps (G_A, G_B), the* ***degree of overlap ($DO_{G_AG_B}$)*** *for two pairs is the minimum ratio between the common interpolated cells to the interpolated cells.*

Figure 2b shows two gap pairs G_2, G_3 and G_7, G_8. The degree of overlap for the first pair is $\mathrm{DO}_{G2,G3} = 0.23(min[\frac{2}{9}, \frac{2}{9}])$ and for the second pair $\mathrm{DO}_{G7,G8} = 0.33(min[\frac{4}{12}, \frac{4}{12}])$.

2.2 Problem Formulation

Input.

(1) Trajectory Gaps: Missing location signal(s) between two consecutive points. Fig. 1b represent trajectory gaps G_1 to G_{10} which later modelled as geo-ellipses (Fig. 1c-1d).

(2) Signal Coverage Map (SCM): The primary motivation behind a signal coverage map (Figure 1a) is to improve the accuracy of detecting specific trajectory gaps where a moving object (e.g., ships) may have behaved abnormally. If we consider the entire study area, we may increase the rate of false positives since not reporting locations from moving objects may be due to weak signal coverage in certain regions. Hence, these maps provide a way to narrow down the search space to specific gaps which need further investigation.

(3) Priority Threshold: A threshold value used for extracting abnormal gap based on user's preference. For instance, gaps with AGM scores *above* a priority threshold are extracted and prioritized for further investigation (e.g., G_2, G_3, G_5, G_7, G_8 in Figure 1d).

Output. Summarized abnormal trajectory gaps (as shown in Figure 1d). Here, we first filter out trajectory gaps that are within areas without signal coverage and then use the priority threshold to prioritize gaps with a high AGM score. In addition, we also coalesce the gaps for two or more geo-ellipse intersections, which usually occur in dense regions, and reduce additional scanning of the overlapped area (e.g., Figure 1c) while considering one gap at a time. This results in higher post-processing costs by the human analysts, which can be greatly reduced by merging the common intersection region.

Objective. Our objective is solution quality and computational efficiency. Solution quality can be achieved by reducing the false positive rate and using optimal AGM values to detect abnormal gaps. To enhance computational efficiency, we focus on optimizing gap enumeration during the process of forming clusters of gaps that interact spatially and temporally.

Constraints. The space-time prism do not consider acceleration and deceleration [7] and only consider speed derived from start and end point of a gap. Signal coverage map's stability and effect of weather, environmental conditions, and radio outages are not considered.

3 Proposed Approach

Framework. Our aim is to identify possible abnormal gaps on a given set of trajectory gaps and signal coverage area through a three-phase *Filter* and *Refine* approach. The trajectory signals are first preprocessed to filter out all the trajectory gaps. Then we model gaps as geo-ellipses and apply the proposed algorithms to effectively optimize the spatial interactions by coalescing pair of gaps and reducing redundant linear scans on the overlap as shown in Figure 1c and 2b. Then, the output is a summary of significant abnormal gaps (Figure 1d) which helps a human analyst for ground truth verification via satellite imagery.

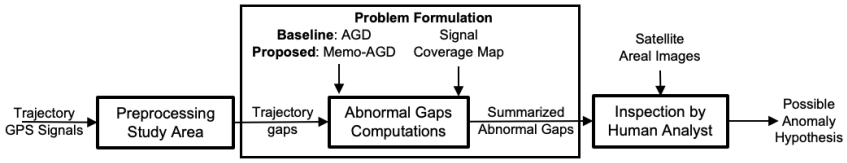


Figure 3 Framework for detecting possible abnormal gaps to reduce manual inspection by analyst.

(1) Abnormal Gap Detection (AGD) Algorithm. Here we describe the abnormal gap detection (AGD) algorithm used for enumerating gaps such that each gap is modeled via a geo-ellipse and an AGM score is computed with a coalescing operation (Figure 1d). First, we sort each gap G_i by time and then check which gap pairs intersect in space and time. We further calculate their Degree of Overlap (DO) and check if DO $\geq \lambda$. Second, we save each qualified gap either as a single ($<G_i>$) or as a subset of gap pairs (e.g., $<G_i,G_j>$). Finally, we merge all subsets of trajectory gaps and compute their AGM scores, which are later extracted via a priority threshold. An execution trace is given in Appendix A.

(2) Memoized Abnormal Gap Detection (Memo-AGD) Algorithm. Since the AGD approach enumerates an exponential number of candidates, we use additional variables such as G_i^{Obs} and G_i^{LU}, where G_i^{Obs} keeps track of the total current elements in an Observed List and G_i^{LU} provides a *lookup table* which allow us to store information which was already involved in a prior intersection with G_i. Such *memoization* avoids unnecessary gap enumeration.

Algorithm 1 Memoized Abnormal Gap Detection (Memo-AGD) Algorithm.

Input : Trajectory Gaps (G_i), Signal Coverage Map (SCM) and DO Threshold λ
Output: Summarized Abnormal Trajectory Gaps

1: **procedure** :
2: **Step 1:** Initialize LookUp Table to G_i^{LU} and copy Observed List to G_i^{Obs}
3: **Step 2:** Check spatiotemporal overlap and avoid gap enumeration
4: **while** $G_i^{Obs} \neq \emptyset \; \forall \; G_j \in G_i^{Obs}$ **do**
5: **while** $G_i \cap G_j \neq \emptyset$ and DO $\geq \lambda$ **do**
6: Save or Update the derived shape from $G_i \cap G_j$ to G_i^{LU} and G_j^{LU}
7: Add $<G_i,G_j>$ to G_i^{LU} and G_j^{LU} and update $G_i^{Obs} \leftarrow G_i^{Obs} - G_j^{LU}$
8: If $G_i \cap G_j = \emptyset$ then update $G_i^{Obs} \leftarrow G_i^{Obs} - G_j^{LU}$
9: **Step 3:** Compute AGM Score $\forall$ elements in LookUp Table G_i^{LU}

First we initialize G_i^{LU} and copy the current Observed List in G_i^{Obs}. After checking for spatiotemporal overlap and the DO $\geq \lambda$ condition, we update and save the resultant shape derived from $G_i \cap G_j$ and subset $<G_i,G_j>$ to G_i^{LU} and G_j^{LU}. For instance, a new gap G_k

only needs to perform one comparison with $G_i \cap G_j$ with subset $<G_i, G_j>$ saved in G_i^{LU} and G_j^{LU}. In addition, G_k will skip a comparison with G_j via G_i^{Obs} - G_j^{LU} providing further computational speedup. An execution of the Algorithm 1 trace is given in Appendix B.

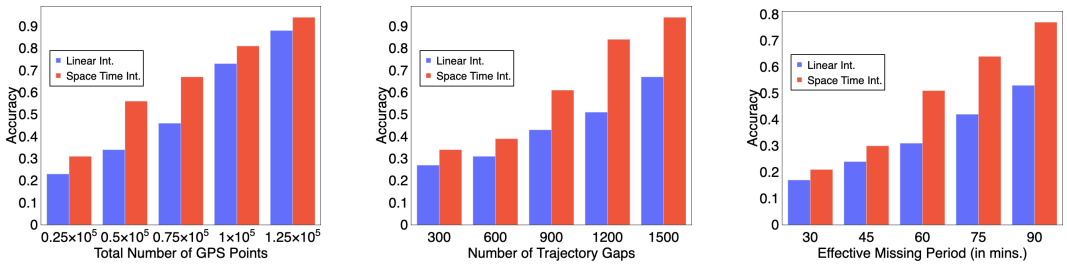
4 Experimental Validation

Synthetic Data Generation. For *solution quality* experiment, we lack ground truth data (i.e., absence of information on whether each gap is abnormal or not). Therefore, we evaluated the proposed algorithms on synthetic data derived from a real-world dataset. First, we gather trajectories gathered on a fixed study area ranging from $179.9W$ to $171W$ degrees in longitude and from $50N$ to $58N$ degrees in latitude in the Bering Sea with 1.25 x 10^5 with 1500 trajectory gaps spanning from 2014 to 2016. Then, for each object, we preprocessed trajectory points with a time-gap duration greater than 30 mins to qualify it as a trajectory gap. Finally, we calculated an AGM score using linear interpolation and proposed methods and classified each gap as abnormal and non-abnormal based on a specific priority threshold (i.e., **0.6**). For instance, gaps with AGM scores greater than **0.6** are considered abnormal.

Real World Data. We used MarineCadastre [1], a real world dataset containing many attributes (e.g., Longitude, Latitude, Speed Over Ground etc.) for $150,000$ objects from 2009 to 2017. In addition, we also used MarineTraffic [9] for a case-study near the Galapagos Islands to verify the effectiveness of the proposed algorithms.

Computing Resources. We performed our experiments on a system with a 2.6 GHz 6-Core Intel Core i7 processor and 16 GB 2667 MHz DDR4 RAM.

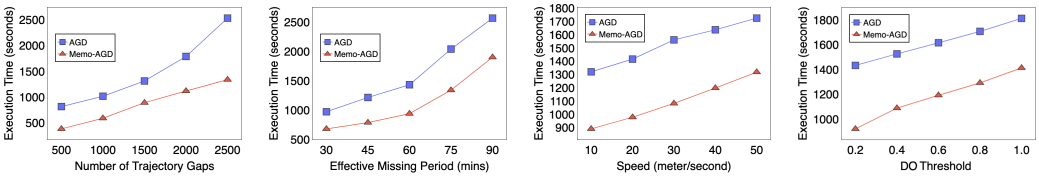
Experiments for Solution Quality. We compared the accuracy of a linear interpolation based method [3, 14] and our space-time interpolation based AGD and Memo-AGD algorithms. The solution quality was based on fixed number of abnormal trajectory gaps resulted from both linear and space-time interpolation methods and a fixed priority threshold. The output gaps are then compared with a predefined normal-abnormal ground truth from synthetic data for computing accuracy. We varied three parameters: number of GPS points, number of trajectory gaps, and effective missing period (EMP), i.e., the total time when a given object was missing. Figure (4a - 4c) shows that space-time interpolation outperforms linear interpolation on all three parameters. The reason is that the AGM scores captured are more accurate in space-time interpolation as compared to linear interpolation based methods.



(a) GPS Points. **(b)** Trajectory Gaps. **(c)** EMP.

Figure 4 Spatio-temporal interpolation is better than linear interpolation for different parameters.

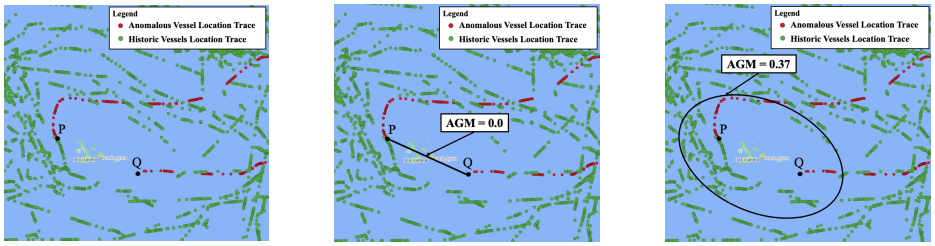
Experiments for Computational Efficiency. Next, we compared Memo-AGD against AGD based on computation efficiency using the MarineCadastre [1] dataset. Figure 5a shows that Memo-AGD consistently outperforms AGD. The reason is that the potential interactions of two or more geo-ellipses increase as we increase the number of trajectory gaps. Figure 5b shows Memo-AGD outperforms AGD as we increase EMP since larger geo-ellipses result in a higher number of potential interactions. Figure 5c also shows MemoAGD is faster than AGD. The reason is that high S_{max} (i.e., maximum possible speed an object can attain during the EMP of its trajectory gap) produces larger geo-ellipses, resulting in more potential interactions. Figure 5d shows that higher DO threshold means gaps are less likely to intersect with each other. This helps in avoiding large coalesced gap pairs which result in higher ground-truth verification cost. Hence, MemoAGD is more efficient than the AGD algorithm.



(a) Trajectory gaps. **(b)** Effect of EMP. **(c)** Effect of S_{max}. **(d)** Effect of λ.

Figure 5 Comparison of execution times of AGD and Memo-AGD under different parameters

Case Study. We conducted a case study of an actual illegal shipping event [6] by applying the proposed Memo-AGD algorithm on a set of real-world data from MarineTraffic [9] data consisting of around 10^5 location trace broadcasted from over 121 fishing vessels for October 2014 (i.e., 31 days) covering the Galapagos marine reserve. Figure 6 shows where a fishing vessel (in red) inside a protected habitat area reported no signals for 15 days because it had switched off its transponder. We applied our Memo-AGD algorithm and a linear interpolation method on the historic location traces (in green). In linear interpolation, the vessels interacted small number of island and did not intersected with any historic location trace (Figure 6b) resulted in an AGM score of **0.0**, indicating no abnormal activity detected. By contrast, the Memo-AGD algorithm returned a geo-elliptical area (Figure 6c) within which some historic trajectories resulted in an AGM score of **0.37**, which accords with the known abnormal activity. The domain experts can set a priority threshold of 0.3 in the flagged area for further investigation by the human analysts.



(a) No Interpolation. **(b)** Linear Interpolation. **(c)** Memo-AGD.

Figure 6 Comparison of Linear Interpolation and Memo-AGD around Galapagos Marine Reserve

Discussion. In this paper, we used a signal coverage map (SCM) as input to significantly reduce the false-positive rate by filtering out gaps in zones with historically weak signal coverage. However, estimating signal coverage is challenging if the vessel enters certain regions (e.g., the arctic) where limited or no historical trace data exists. This result in cold start problem [17]. In addition, the proposed methods results in similar execution times in case of no potential interactions among geo-ellipses (more details described in Appendix C).

In addition, none of the local gaps exceed the time range of the signal coverage map. For instance, Figure 6 shows the time range of the largest signal gap (i.e., 15 days) does not exceed the signal coverage spanned for entire month. Furthermore, given the object disappeared for 15 days in Figure 6, considering speed only at the P and Q Figure 6 while modeling trajectory gaps that may produce relatively smaller ellipse region.

5 Conclusion and Future Work

We investigated Abnormal Gap Measure and proposed AGD and Memo-AGD algorithms. We performed experimental evaluation under varying parameters where the results show that compared to linear interpolation methods, space-time interpolation based methods [12] are better at detecting abnormal gaps in trajectory data. Also, Memo-AGD is computationally more efficient than AGD. In the future, we plan to develop spatiotemporal statistics and conduct a statistical significance test to eliminate chance patterns. We will add other use cases such as a cold-start problem [17] and include acceleration using kinetic prisms [7]. Finally, we plan to explore datasets in network space for other potential societal use-cases.

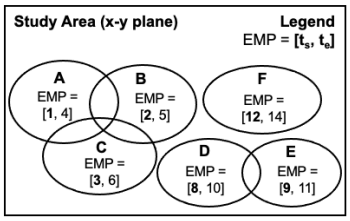
References

1 BOEM and NOAA. Marinecadastre. `https://marinecadastre.gov/ais/`, 2020.
2 V. Chandola et al. Anomaly detection: A survey. *CSUR*, 41(3):1–58, 2009.
3 C. Chen et al. Isolation-based online anomalous trajectory detection. *IEEE TIST*, 2013.
4 R. Cheng et al. Probabilistic verifiers: Evaluating constrained nearest-neighbor queries over uncertain data. In *2008 IEEE 24th ICDE*, pages 973–982. IEEE, 2008.
5 S Dodge et al. Towards a taxonomy of movement patterns. *IV*, 7(3-4):240–252, 2008.
6 S Gibbens. How illegal fishing is being tracked from space. *National Geographic*, 2018.
7 B. Kuijpers et al. Kinetic prisms: incorporating acceleration limits into space time prisms. *IJGIS*, 31(11):2164–2194, 2017.
8 Po-Ruey Lei. A framework for anomaly detection in maritime trajectory behavior. *Knowledge and Information Systems*, 47(1):189–214, 2016.
9 MarineTraffic. Galapagos island case study. `https://www.marinetraffic.com/en/ais/`, 2020.
10 H. J. Miller. Modelling accessibility using space-time prism concepts within geographical information systems. *IJGIS*, 5(3):287–301, 1991.
11 G. Pallotta, M. Vespe, and K. Bryan. Vessel pattern knowledge discovery from ais data: A framework for anomaly detection and route prediction. *Entropy*, 15(6):2218–2245, 2013.
12 D. Pfoser and C. S. Jensen. Capturing the uncertainty of moving-object representations. In *International Symposium on Spatial Databases*, pages 111–131. Springer, 1999.
13 M. Riveiro, G. Pallotta, and M. Vespe. Maritime anomaly detection: A review. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 8(5):e1266, 2018.
14 Robert et. al Skulstad. Dead reckoning of dynamically positioned ships: Using an efficient recurrent neural network. *IEEE Robotics & Automation Magazine*, 26(3):39–51, 2019.
15 Goce Trajcevski et al. Uncertain range queries for necklaces. In *2010 Eleventh International Conference on Mobile Data Management*, pages 199–208. IEEE, 2010.
16 Stephan Winter and Zhang-Cai Yin. Directed movements in probabilistic time geography. *International Journal of Geographical Information Science*, 24(9):1349–1365, 2010.

17 Pengxiang Zhao, David Jonietz, and Martin Raubal. Applying frequent-pattern mining and time geography to impute gaps in smartphone-based human-movement data. *International Journal of Geographical Information Science*, 35(11):2187–2215, 2021.
18 Yu Zheng. Trajectory data mining: an overview. *ACM TIST*, 6(3):1–41, 2015.

A Execution Trace of AGD Algorithm

Figure 7 shows the execution trace of the AGD algorithm with gap (<A>,1)„.,(<F>,1) where each gap set is associated with a single gap (e.g., <A>) and its count (i.e., 1). We then sort each gap G_i based on its start time t_s and at t_s=1 and we add (<A>,1) to the Observed List since the list is empty. At t_s=2, we add (<B>,1) since it satisfies spatiotemporal intersection criteria resulting in (<A>,2),(<B>,2),(<A,B>,2) and incrementing count for each subset by 1. A similar operation is performed at t_s=3, resulting in (<A>,3),(<B>,3),.., (<A,B,C>,3). However at t_s=8, we only add (<D>,1) since it does not interact with any of the elements in the Observed List. At t_s=9, (<E>,1) interacts with (<D>,1) performing similar operations as executed at $t_s = 2$ resulting in (<D>,2),(<E>,2) and (<D,E>,2). At $t_s = 12$, set (<F>,1) is added to the Observed List similar to (<D,1>) since it does not interact with any of the elements in Observed List. Finally we filter out (<A,B,C>,3),(<D,E>,2) and (<F>,1) since the number of elements is equal to their respective counts and we compute AGM via the signal coverage map. We then output the summarized abnormal gaps based on their AGM scores and the priority threshold given by the human analysts.

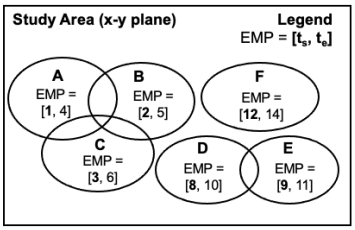


Time	Non-Observed List (G_i)	Observed List (G_j)
t_s = 1	{<B>,1}, {<C>,1}, {<D>,1}, {<E>,1}, {<F>,1}	{<A>,1}
t_s = 2	{<C>,1}, {<D>,1}, {<E>,1}, {<F>,1}	{<A>,2}, {<B>,2}, {<A,B>,2}
t_s = 3	{<D>,1}, {<E>,1}, {<F>,1}	{<A>,3}, {<B>,3}, {<C>,3}, {<A,B>,3} , {<B,C>,3}, {<A,C>,3}, {<A,B,C>,3}
t_s = 8	{<E>,1}, {<F>,1}	{<A>,3}, {<B>,3}, {<C>,3}, {<A,B>,3} , {<B,C>,3}, {<A,C>,3}, {<A,B,C>,3}, {<D>,1}
t_s = 9	{<F>,1}	{<A>,3}, {<B>,3}, {<C>,3}, {<A,B>,3} , {<B,C>,3}, {<A,C>,3}, {<A,B,C>,3}, {<D>,2}, {<E>,2}, {<D,E>,2}
t_s = 12	Empty	{<A>,3}, {<B>,3}, {<C>,3}, {<A,B>,3} , {<B,C>,3}, {<A,C>,3}, **{<A,B,C>,3}**, {<D>,2}, {<E>,2}, **{<D,E>,2}, {<F>,1}**

Figure 7 Execution trace of the Baseline AGD algorithm.

B Execution Trace of Memo-AGD Algorithm

Figure 8 shows the execution trace of the Memo-AGD algorithm. At $t_s = 1$, Step 1 initializes G_i with variables G_i^{LU} and G_i^{Obs} which are later added to the empty Observed List. At $t_s = 2$, $<A^{Obs}>$ copies the *current* Observed List (i.e., only <A>) to $<B^{Obs}>$ after satisfying spatial and temporal overlap conditions. If true, we perform Step 2 where the resultant shape of $<A> \cup <B>$ is saved in both <A> and <B>. The variables LU gets updated with <A,B> for both $<A^{LU}>$ and $<B^{LU}>$. Finally, removing <A> from $<B^{Obs}>$ results in an empty list and the loop terminates. At $t_s = 3$, <C> interacts with the resultant shape of both <A> and <B> (i.e., $<A> \cup <B>$ and result $<A^{LU}>$,$<B^{LU}>$ and $<C^{LU}>$ as <A,B,C> (i.e., maximal sets). However at $t_s = 8$, <D> does not intersect spatially with <A> and will skip intersection operations with other elements of A^{LU} (i.e., <B> and <C>) via G^{Obs} -G^{LU} operation and gets added to the Observed List and result in performance speedup as compared to the AGD algorithm. A similar operation is done at $t_s = 9$, where <E> does not interact with <A> but does interact with <D>, resulting in D^{LU} = <D,E>. Finally, <F> does not interact with any of the elements, causing to be directly saved as maximal set <F>. The rest of the steps remain the same as the baseline except we gather all maximal sets from the lookup table of each G^{LU} in the Observed List.



Time	Non-Observed List (G_i)	Observed List (G_j) with Look Up Table (G_j^{LU})
$t_s = 1$	B^{LU} = {<B>}, C^{LU} = {<B>}, D^{LU} = {<D>}, E^{LU} = {<E>}, F^{LU} = {<F>}	A^{LU} = {<A>}
$t_s = 2$	C^{LU} = {<B>}, D^{LU} = {<D>}, E^{LU} = {<E>}, F^{LU} = {<F>}	A^{LU} = {<A,B>}, B^{LU} = {<A,B>}
$t_s = 3$	D^{LU} = {<D>}, E^{LU} = {<E>}, F^{LU} = {<F>}	A^{LU} = {<A,B,C>}, B^{LU} = {<A,B,C>}, C^{LU} = {<A,B,C>}
$t_s = 8$	E^{LU} = {<E>}, F^{LU} = {<F>}	A^{LU} = {<A,B,C>}, B^{LU} = {<A,B,C>}, C^{LU} = {<A,B,C>}, D^{LU} = {<D>}
$t_s = 9$	F^{LU} = {<F>}	A^{LU} = {<A,B,C>}, B^{LU} = {<A,B,C>}, C^{LU} = {<A,B,C>}, D^{LU} = {<D,E>}, E^{LU} = {<D,E>}
$t_s = 12$	**Empty**	A^{LU} = **{<A,B,C>}**, B^{LU} = **{<A,B,C>}**, C^{LU} = **{<A,B,C>}**, D^{LU} = **{<D,E>}**, F^{LU} = **{<F>}**

Figure 8 Execution Trace of Proposed Memo-AGD Algorithm.

C Asymptotic Time Complexity Analysis

Given N gaps, both algorithms first perform sorting operations in $O(Nlog(N))$ time and perform k operations for k subsets of coalesced gaps. The time complexity of comparison operations for each algorithms are as follows:

AGD Algorithm. Given k gaps that intersect within all the gaps in the Observed List, it is necessary to check the total number of subsets (i.e., $\binom{k}{1} + \binom{k}{2} + \binom{k}{3} + \ldots + \binom{k}{k} = \sum_{i=1}^{k} \binom{k}{i}$). When a new $(k+1)$-th gap is added, it is necessary to check whether it intersects with all the existing subsets. Since checking with the subsets of size-i gap costs i, the cost of each intersection is $\sum_{i=1}^{k} i \times \binom{k}{i} = k \times 2^{k-1}$. Therefore, $|N|$ gaps in the dataset, the total cost for adding is $\sum_{k=1}^{|N|} k \times 2^{k-1} = O(N \times 2^{|N|})$. The best case is when all the gaps are disjoint i.e., $O(N^2)$. The worst case is rare because of the sparsity in real datasets.

Memo-AGD Algorithm. The worst case of Memo-AGD will be similar to *AGD* except we only check maximal sets instead of $\binom{k}{i}$ gaps. Hence, for each N and k, we update their resultant shape and elements in the lookup table (i.e., $O(kN)$) and the cost of updating the lookup table. In contrast, the cost will be similar to best case if no gaps are intersecting.

Improving Pedestrians Traffic Priority via Grouping and Virtual Lanes in Shared Spaces

Yao Li[1]
Institute of Cartography and Geoinformatics, Leibniz Universität Hannover, Germany

Vinu Kamalasanan
Institute of Cartography and Geoinformatics, Leibniz Universität Hannover, Germany

Mariana Batista
Institute of Transportation and Urban Engineering, Technische Universität Braunschweig, Germany

Monika Sester
Institute of Cartography and Geoinformatics, Leibniz Universität Hannover, Germany

Abstract

The shared space design is applied in urban streets to support barrier-free movement and integrate traffic participants (such as pedestrians, cyclists and vehicles) into a common road space. Regardless of the low-speed environment, sharing space with motor vehicles can make vulnerable road users feel uneasy. Yet, walking in groups increases their confidence as well as influence the yielding behavior of drivers. Therefore, we propose an innovative approach to support the crossing of pedestrians via grouping and project the virtual lanes in shared spaces. This paper presents the important components of the crowd steering system, discusses the enablers and gaps in the current approach, and illustrates the proposed idea with concept diagrams.

2012 ACM Subject Classification Human-centered computing → Mixed / augmented reality

Keywords and phrases shared space, urban traffic system, augmented reality, pedestrian grouping

Digital Object Identifier 10.4230/LIPIcs.COSIT.2022.27

Category Short Paper

Related Version *Full Version*: https://arxiv.org/abs/2205.08783

Funding *Yao Li*: Research supported by the Deutsche Forschungsgemeinschaft (DFG; German Research Foundation) – 227198829/GRK1931.
Vinu Kamalasanan: Research supported by DAAD Graduate School Scholarship Programme (GSSP).
Mariana Batista: Research supported by DAAD Graduate School Scholarship Programme (GSSP).

1 Introduction

Shared spaces are mixed traffic environments that aim to minimize traffic control and rely on negotiation-based movement to integrate road users into a common road space [18]. In the absence of traffic control measures, pedestrians are obliged to assess the situation, especially by establishing eye contact with others, before deciding to cross [8, 18]. Other implicit and explicit communication, hand gestures being an example of it, are also used to negotiate the right of way among road users. However, even when such design settings and negotiations are expected to increase safety [18], on-street interviews and surveys indicate that vulnerable road users are not necessarily confident in sharing space with motor vehicles [22, 11].

[1] corresponding author

From an urban planning perspective, designing shared spaces focuses on enhancing pedestrian movement and their perception of domain compared to conventional layouts. The placement of street furniture and design elements, such as a continuous shared level surface, increase the user perception of pedestrian domain and their consequent sense of priority over cars [24]. Although [18] argued that pedestrians feel more assertive and navigate with more control in shared spaces, this unusual setting can be stressful for road users who do not feel confident sharing space with motor vehicles, leading to confusion regarding priority and directly influencing their crossing behavior. It is then crucial to address traffic conflicts arising from these particular situations.

As mentioned in [8], crossing in groups can positively impact crossing assertiveness and vehicle yielding behavior. Therefore, we propose an innovative approach to help support pedestrian movement using virtual lanes and group formation aiming to improve the confidence and safety of vulnerable road users when crossing a shared space.

2 Background

2.1 Crowd steering in public spaces

Crowd steering focuses on the problem of steering the movement of people in public and urban spaces [25], i.e. suggesting people where to move to eventually achieve a desirable global configuration with regard to crowd distribution. It plays an important role in addressing a few real-life problems, e.g. indoor/outdoor evacuation, avoiding peculiar situations (moving across unsafe neighborhoods or extremely polluted streets), visiting a range of attractions with an optimized order, etc.

Most crowd steering applications can be divided into the following steps: firstly, the basic dynamic data like position and orientation are collected by sensors; based on the observed data, applications are able to track multiple persons, predict future trajectories, plan paths or identify groups to accommodate various requests; the desired results are then presented to steered agents via suitable infrastructures. In the following, we listed some recently emerged technologies that enable crowd steering:

Multiple object tracking and trajectory prediction. The task of Multiple Object Tracking (MOT) is largely partitioned into locating multiple objects, maintaining their identities, and yielding their individual trajectories given an input video [21]. Accurate trajectory prediction is a crucial task in different communities, it enables an intelligent system to forecast the behavior of road users based on the behavior to date and make a reasonable and safe decision for the next operation [5].

Grouping and group identification. Groups are often found during pedestrian movements. Here, group is not only restricted to social groups (friends or families) but also contains regulation-based groups (e.g. pedestrians following the same phase of traffic lights). Grouping increases the safety during movements by holding a buddy system [8]. There has been an increasing amount of literature on grouping in recent years: [14] controls the merging and splitting of a single-line group by penalty; [4] used a time-sequence DBSCAN that based on coexisting time and Euclidean distance between pedestrians to detect groups; [20] groups the road user based on similar origin and destination (OD) when they enter a shared space.

Group identification in an intelligent crowd steering environment could potentially help add individuals to existing groups while new or larger groups are formed in the process. Identifying people connected to each other through forms of common interactions (e.g

engaging in similar activities or goals) could help identify potential groups. This would require anticipating every individual's action and intention way ahead of time. Recently large-scale spatio-temporal individual action and social group annotation datasets [7] provide precise data for such identification. If the spatio-temporal information of the tracked pedestrian groups is available, group surfing approaches [6] could be used to increase group size over different time windows. In [6], a robot-based navigation approach with sub-goals was used to promote groups, which could easily be extended to pedestrians.

Infrastructures for displaying result. Recently, researchers have shown an increasing number of physical attempts to lead pedestrians to desirable global configurations, such as dynamic signages[19] and gaze-based approaches [10].

2.2 Visual augmentation to influence pedestrian walking behavior

Enabling potential grouping behavior could be achieved using visual augmentation. Such systems could either be via projecting virtual lanes or using Augmented reality (AR) wearable devices. While projection-based approaches would require the installation of infrastructure to enable this, AR glasses like the Hololens could also be used to visualize virtual content.

An AR-based interface presenting virtual lanes was prototyped [12] for pedestrians which displayed the path and traffic directly into the visual field. A dynamic projection system was used in [3] where pedestrians were detected using the LIDAR sensor mounted on the vehicle, which was then used to show virtual crossing lanes. To prototype, a virtual traffic infrastructure for shared spaces [15] demonstrated how virtual signals can control behavior using wearable AR.

Considering all of the above, we propose a concept that can be applied in crowded situations, or an individual alone situation would have difficulties getting the right of way. In the Chapter 3, we explain our idea in a context of shared space.

3 Gaining priority via collaboration using virtual lanes

Our concept idea of enabling pedestrian priority via augmentation and grouping is more focused on an intelligent traffic environment [13], where the traffic scene is controlled in real-time with sensors and projector displays. While sensors like camera and LIDAR enable live traffic monitoring, projection systems will dynamically show the traffic control signals, such as zebra crossing and virtual pedestrian lanes on the ground to guide/support pedestrians to cross and vehicles to give way once conflicts might happen.

As different traffic agents, for instance, autonomous taxis (ride-sharing or ride-hailing), are potentially interconnected in such an environment, they would share their OD data to a central system which could further enhance traffic signaling and real-time collision management. Our concept focuses on including pedestrian safety and priority to real-time traffic management by incorporating collaborative pedestrian motions. Collaborative pedestrian motions in this context mean pedestrians that are willing to walk together as they are moving to a similar destination [2]. This could also include shared walking approaches if a connected systems are in place. We also made a strong assumption that few/all pedestrians would wear AR glasses for safety and be included in the connected traffic management ecosystems, because current AR devices already dispose of the functionalities of capturing environmental information and projecting recommended routes. Since sensors can collect dynamic data from road users and then send it to the central system, tracking and prediction algorithms could be applied in

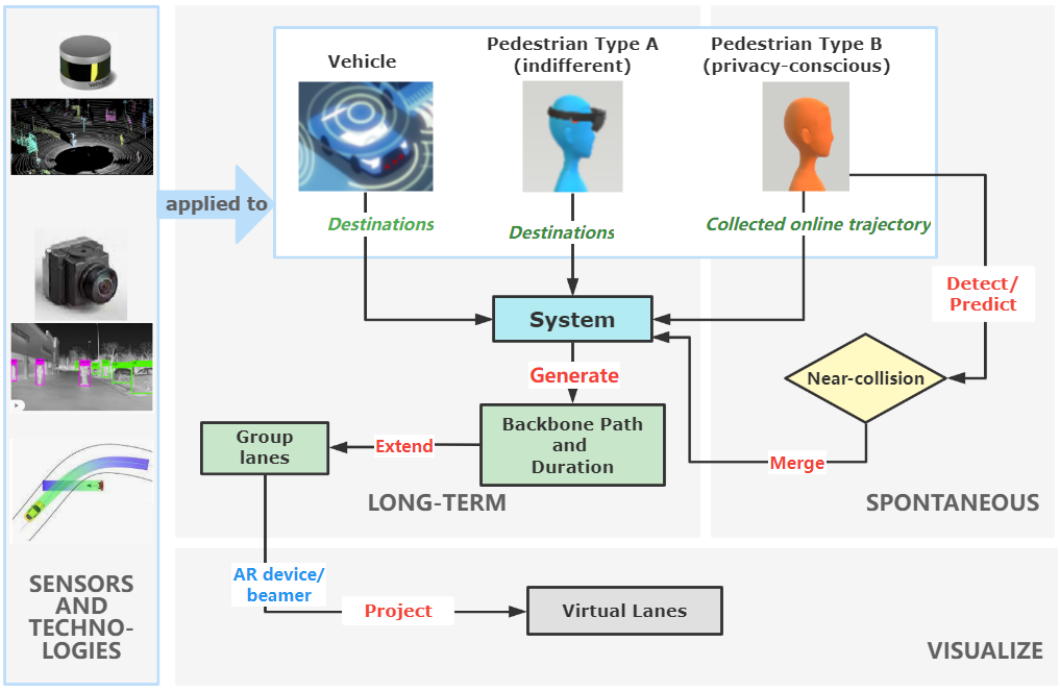


Figure 1 Concept of shared space pedestrian grouping supported by virtual lanes.

the collected data. Meanwhile, the subset of pedestrians who wear connected AR glasses could also share their walking paths and potential crossing points with the central system. We further distinguish the road users according to their privacy settings:

Spontaneous collaborative approach. Spontaneous collaborative approach particularly considers the privacy-conscious road users (Type B) who do not share their destination information with the ecosystem. Once pedestrians are detected by installed sensors, the future motions are predicted to identify overlapping target pedestrian destinations. If the traffic system identifies multiple people walking towards the same goal, people would be motivated to walk/cross together by projecting a virtual lane/zebra crossing based on the predicted context. However, if it identifies a conflict, projected control to avoid collisions would enable centralized management of conflicts.

Long-term collaborative approach. Long-term collaborative approach is suitable for the road users who are willing to share information with the system (so-called "indifferent road users", Type A). These kinds of pedestrians is considered wearing AR glasses or other means also disclose their walking path. Hence the destination is first shared with the central system. As the system is aware of the shared OD data, it collects the road users who have similar ODs at a specific duration (waiting time), forms groups, and gives the crossing priority to larger groups to avoid conflicts and improve efficiency. Group formation can be calculated using an adapted facility location algorithm [20], or edge bundling [1].

To realize virtual walking/crossing paths, the system will calculate the location and duration of each of the routes and will create a path geometry. This process will take the width of a single road user unit (called backbone path) and group size into account [17]. Figure 1 provides an overview of what we envision for this ecosystem.

With our concept idea, we intend to empower pedestrians in shared traffic spaces with 1) groups creating (visual) impact and thus enforcing priority in conflicting scenarios and 2) realizing shared walking and grouping to make it more attractive for other road users to join in and thus take advantage of the joint forces.

4 Discussion

Some of the potential prerequisites and challenges for the realization of this approach are discussed below.

Important parameters for grouping. Important parameters for grouping pedestrians would have to gather at several suitable spots (*group center*) to jointly cross the shared space. They might wait till a sufficiently large group (*group size*) is formed, or till a critical *waiting time* is reached. Waiting time is critical because it protects small groups from unreasonable delays, e.g., isolated pedestrian who do not belong to any groups is considered as a single-member group. Since the system tries to give larger groups priority in crossing, single-member-group may need more time to cross if there is no waiting time limit. The group center would be identified based on the *distance difference* between OD of group members.

Acceptance of virtual infrastructure for crossing. Acceptance of virtual infrastructure for crossing from a traffic planning perspective, setting gap acceptance, crossing speed and duration for virtual lanes to match motor vehicles' speed and yielding behavior can objectively indicate the potential direction and duration of the proposed visualization. Therefore, the acceptance would likely increase with virtual infrastructure that ensures a sense of safety. For that, it should integrate a critical gap acceptance that would project the virtual lane in a suitable distance in space and time from motor vehicles.

Impediment of remote controlling pedestrians. [27] claims that people had stronger intentions to illegally cross as groups when they are not given the right of the way (e.g. wait for over 90s). Thus, similar to the obedience to existing traffic signs, road users would follow the recommended virtual lanes once they experience the benefits of gaining the right of the way. Another concern comes from losing freedom while crossing, as pedestrian locomotion provides more degrees of freedom in terms of how individuals can move. Our approach is aiming to offer a safer and more efficient choice in busy traffic contexts and not limit them.

Impediments of privacy. Low-resolution sensors are sufficient for tracking and prediction algorithms (e.g. in [23]), so individuals are not necessarily identified while crossing. Moreover, the collected data would be dropped once the crossing is completed.

Virtual Infrastructures. Early studies on applying projection-based augmentation by inducing vection [9] have been influential in controlling the behavior of pedestrians. But when such approaches are applied to a larger scale, for example in an outdoor shared space, their effectiveness would depend on the maturity of both wearable AR and projection systems. Such virtual lanes should be visually appealing both during the day and night to serve as effective pedestrian infrastructure. Furthermore, we also believe that AR glasses might become common in near future as smartphones today and perhaps become a mobility aid [16], which can further support the maturity of our concept towards safety.

Trust. The acceptance of our idea depends on the trust of pedestrians in the proposed system, especially vulnerable road users, such as elderly pedestrians who are likely to be less tech-savvy, thus refuse to wear advanced devices. However, pedestrians tend to gather while crossing, thus those pedestrians can also be attracted by the formed groups to follow the virtual lanes. According to [26], pedestrians belonging to the same group tend to automatically adjust the walking speed, making it easier for the vulnerable road users to continue crossing with the group.

Robustness. Misprediction and mis-grouping could happen. However, the user is self-sufficient to judge and make decisions in these situations. Ultimately, the proposed approach is to support vulnerable road users and not bind them. If it leads in the wrong direction, users can simply ignore it. If they still accept it, they will eventually prolong their crossing movement in the shared space. As an alternative to physical infrastructure in shared spaces, pedestrians are informally getting priority from motor vehicles when using virtual lanes, which complement social protocols such as eye contact or a brief gesture.

User study and assessment. A user study to assess the proposed idea's acceptability and investigate objective and subjective parameters in terms of efficiency, safety, and comfort are fundamental in this case. It is paramount to understand users' perceptions of this approach and how they feel crossing such shared environments in a group and through virtual infrastructure to evaluate acceptability. Moreover, analyzing critical waiting time and gap acceptance can further improve the virtual infrastructure design to assure compliance.

5 Conclusion and Outlook

Our approach can help to dynamically and temporally sort and separate the mixed traffic into several virtual lanes, in order to allow for a swift movement of all participants. We also discussed the current research gaps in accomplishing such an approach.

We believe that the challenges around our approach can be addressed more broadly, especially within the spatial information theory research, which can be of interest to the COSIT community for the related field. In future work, we plan to develop a concrete methodology to define and apply pedestrian group formation in shared spaces and evaluate the acceptability of virtual infrastructure. With that, we expect to get insights into and explore solutions for vulnerable road users in urban mixed traffic environments.

References

1 David Auber, Patrick Mary, Morgan Mathiaut, Jonathan Dubois, Antoine Lambert, Daniel Archambault, Romain Bourqui, Bruno Pinaud, Maylis Delest, and Guy Melançon. Tulip: a scalable graph visualization framework. In *Extraction et Gestion des Connaissances (EGC) 2010*, pages 623–624. RNTI, 2010.

2 Debjit Bhowmick, Stephan Winter, Mark Stevenson, and Peter Vortisch. Exploring the viability of walk-sharing in outdoor urban spaces. *Computers, environment and urban systems*, 88:101635, 2021.

3 Steffen Busch, Alexander Schlichting, and Claus Brenner. Generation and communication of dynamic maps using light projection. In *Proceedings of the ICA*, volume 1, page 16, 2018.

4 Hao Cheng, Yao Li, and Monika Sester. Pedestrian group detection in shared space. In *2019 IEEE Intelligent Vehicles Symposium (IV)*, pages 1707–1714, 2019. doi:10.1109/IVS.2019.8813849.

5 Hao Cheng, Wentong Liao, Michael Ying Yang, Bodo Rosenhahn, and Monika Sester. Amenet: Attentive maps encoder network for trajectory prediction. *ISPRS Journal of Photogrammetry and Remote Sensing*, 172:253–266, February 2021. doi:10.1016/j.isprsjprs.2020.12.004.

6 Yuqing Du, Nicholas J Hetherington, Chu Lip Oon, Wesley P Chan, Camilo Perez Quintero, Elizabeth Croft, and HF Machiel Van der Loos. Group surfing: A pedestrian-based approach to sidewalk robot navigation. In *2019 international conference on robotics and automation (ICRA)*, pages 6518–6524. IEEE, 2019.

7 Mahsa Ehsanpour, Fatemeh Saleh, Silvio Savarese, Ian Reid, and Hamid Rezatofighi. Jrdb-act: A large-scale multi-modal dataset for spatio-temporal action, social group and activity detection. *arXiv preprint*, 2021. arXiv:2106.08827.

8 Roja Ezzati Amini, Christos Katrakazas, and Constantinos Antoniou. Negotiation and decision-making for a pedestrian roadway crossing: A literature review. *Sustainability*, 11(23), 2019. doi:10.3390/su11236713.

9 Masahiro Furukawa, Hiromi Yoshikawa, Taku Hachisu, Shogo Fukushima, and Hiroyuki Kajimoto. "Vection field" for pedestrian traffic control. In *Proceedings of the 2nd Augmented Human International Conference*, pages 1–8, 2011.

10 Ioannis Giannopoulos, Peter Kiefer, and Martin Raubal. Gazenav: gaze-based pedestrian navigation. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services*, pages 337–346, 2015.

11 Victoria Hammond and Charles Musselwhite. The attitudes, perceptions and concerns of pedestrians and vulnerable road users to shared space: A case study from the uk. *Journal of Urban Design*, 18:78–97, 2013.

12 Marc Hesenius, Ingo Börsting, Ole Meyer, and Volker Gruhn. Don't panic! Guiding pedestrians in autonomous traffic with augmented reality. In *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct*, pages 261–268, 2018.

13 Lin Hu, Jian Ou, Jing Huang, Yimin Chen, and Dongpu Cao. A review of research on traffic conflicts based on intelligent vehicles. *Ieee Access*, 8:24471–24483, 2020.

14 Tianyu Huang, Mubbasir Kapadia, Norman I Badler, and Marcelo Kallmann. Path planning for coherent and persistent groups. In *2014 IEEE International Conference on Robotics and Automation (ICRA)*, pages 1652–1659. IEEE, 2014.

15 Vinu Kamalasanan and Monika Sester. Behaviour control with augmented reality systems for shared spaces. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 43:591–598, 2020.

16 Kamalasanan, Vinu and Feng , Yu and Sester , Monika. Improving 3D pedestrian detection for wearble sensor data with 2D human pose. https://www.ikg.uni-hannover.de/fileadmin/ikg/Forschung/publications/ppe_fpointnet_submission.pdf.

17 Arno Kamphuis and Mark H Overmars. Finding paths for coherent groups using clearance. In *Proceedings of the 2004 ACM SIGGRAPH/Eurographics symposium on Computer animation*, pages 19–28, 2004.

18 Auttapone Karndacharuk, Douglas J. Wilson, and Roger Dunn. A review of the evolution of shared (street) space concepts in urban environments. *Transport Reviews*, 34(2):190–220, 2014. doi:10.1080/01441647.2014.893038.

19 Norman Langner and Christian Kray. Assessing the impact of dynamic public signage on mass evacuation. In *Proceedings of The International Symposium on Pervasive Displays*, PerDis '14, pages 136–141, New York, NY, USA, 2014. Association for Computing Machinery. doi:10.1145/2611009.2611033.

20 Yao Li and Monika Sester. Group formation in shared spaces. *AGILE: GIScience Series*, 2:1–8, 2021.

21 Wenhan Luo, Junliang Xing, Anton Milan, Xiaoqin Zhang, Wei Liu, and Tae-Kyun Kim. Multiple object tracking: A literature review. *Artificial Intelligence*, 293:103448, 2021.

22 Simon Moody and Steve Melia. Shared space – research, policy and problems. *Proceedings of the Institution of Civil Engineers - Transport*, 167(6):384–392, 2014.

23 Alexandre Robicquet, Amir Sadeghian, Alexandre Alahi, and Silvio Savarese. Learning social etiquette: Human trajectory understanding in crowded scenes. In *European conference on computer vision*, pages 549–565. Springer, 2016.

24 Borja Ruiz-Apilánez, Kayvan Karimi, Irene García-Camacha, and Raúl Martín. Shared space streets: design, user perception and performance. *Urban Design International*, 22(3):267–284, 2017.

25 Andrea Sassi, Claudio Borean, Roberta Giannantonio, Marco Mamei, Dario Mana, and Franco Zambonelli. Crowd steering in public spaces: Approaches and strategies. In *2015 IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing*, pages 2098–2105, 2015. doi:10.1109/CIT/IUCC/DASC/PICOM.2015.312.

26 Kota Yamaguchi, Alexander C Berg, Luis E Ortiz, and Tamara L Berg. Who are you with and where are you going? In *CVPR 2011*, pages 1345–1352. IEEE, 2011.

27 Ronggang Zhou, William J. Horrey, and Ruifeng Yu. The effect of conformity tendency on pedestrians' road-crossing intentions in china: An application of the theory of planned behavior. *Accident Analysis & Prevention*, 41(3):491–497, 2009. doi:10.1016/j.aap.2009.01.007.

Eye Blink-Related Brain Potentials During Landmark-Based Navigation in Virtual Reality

Bingjie Cheng
Department of Geography and Digital Society Initiative, University of Zurich, Switzerland

Enru Lin
Department of Geography and Digital Society Initiative, University of Zurich, Switzerland

Klaus Gramann
Department of Biological Psychology and Neuroergonomics, Technische Universität Berlin, Germany

Anna Wunderlich
Department of Biological Psychology and Neuroergonomics, Technische Universität Berlin, Germany

Abstract

Landmarks support navigation and spatial learning of environments by serving as cognitive anchors. However, little research has been done to investigate how the design of landmarks on mobile maps affects cognitive processing. To address this gap, the present study utilized a within-subjects design to experimentally examine how three different landmark densities (3 vs. 5 vs. 7 landmarks) on mobile maps influence users' spatial learning and cognitive load during navigation. Cognitive load was measured using electroencephalography (EEG). We applied an event-related analysis approach by utilizing eye blinks as naturalistic event markers to segment the EEG data. Results demonstrate that showing five landmarks along a given route to follow on a mobile map, compared to three and seven landmarks, improved spatial learning performance without taxing more cognitive resources. Our study shows that users' cognitive load and spatial learning outcomes should be considered when designing landmark-based navigation assistance systems.

2012 ACM Subject Classification General and reference → Empirical studies; Human-centered computing → Laboratory experiments

Keywords and phrases spatial navigation, landmark, blink-related potentials, spatial learning, cognitive load, mobile map

Digital Object Identifier 10.4230/LIPIcs.COSIT.2022.28

Category Short Paper

Funding *Bingjie Cheng*: H2020 European Research Council (ERC) Advanced Grant GeoViSense [740426]
Enru Lin: H2020 European Research Council (ERC) Advanced Grant GeoViSense [740426]

Acknowledgements We thank Armand Kapaj for his help with data collection, Dr. Ian Ruginski for his help with statistical analysis, and Prof. Sara Irina Fabrikant for her support on this research.

1 Background

1.1 Landmark-based navigation assistance

GPS guidance is increasingly used to facilitate navigation and wayfinding, especially in an unfamiliar environment. Navigators follow turn-by-turn directions given in real time. However, the increased use of mobile maps has been shown to negatively affect landmark and route learning of an environment [4]. Including landmarks in navigation assistance systems has been proposed to facilitate users' learning of their surroundings by serving as cognitive anchors. For example, navigators could use landmarks to determine their current location and remember key decision points along routes. However, using landmarks as mnemonic

devices entails additional cognitive processing, which could additionally affect individuals' cognitive load during navigation. Indeed, previous studies have found that learners have limited cognitive capacity – typically four items (or chunks) and that their cognitive load increased as the number of items to be remembered increased [5]. We thus investigated how the number of landmarks on mobile maps affected navigators' cognitive load during navigation. Based on cognitive capacity theory, we defined low, medium, and high landmark density visualized on mobile maps as three, five, and seven landmarks, respectively.

1.2 Assessing cognitive load through brain activity

Previous research has used performance on dual task or/and pupil dilation to measure cognitive load. However, these measures are an indirect approach to assess cognitive processing. We thus turned to electroencephalography (EEG), an established method that directly measures real-time cognitive load unobtrusively. EEG recordings of brain activity typically require event markers that indicate when notable events such as stimulus presentation or participant responses occur. These markers allow the segmentation of EEG data according to these events for event-related analysis. However, the presentation of additional stimuli may interrupt participants' task performance in naturalistic settings. A different set of event markers is therefore needed when examining brain activity during wayfinding in naturalistic settings.

1.2.1 Eye blinks as event markers in naturalistic settings

Previous research has found that spontaneous eye blinks are suppressed during periods of high cognitive load, and especially during the processing of complex visual scenes [6]. This makes eye blinks particularly useful as indicators of cognitive load in wayfinding, where individuals perform a continuous task without interruption from artificially introduced stimuli [6]. Among the studies that investigated the relationship between eye blinks and cognitive load, consistent evidence has emerged. It was shown that the rate of eye blinks decreases during cognitively demanding tasks [1]. Most research linking eye blinks to cognitive load had focused on characteristics of eye blinks such as blink rate and deflection. Less research studied cognitive load by analyzing brain activity related to eye blinks [6]. Therefore, more research is needed that investigates brain activity during eye blinks when individuals perform cognitive tasks.

1.2.2 Blink event-related potentials (bERPs)

A previous study examined bERPs when participants were performing a cognitive task versus a physical task or during rest [6]. The authors found a significantly more pronounced P1, a positive component 100 ms after blink maximum, in the occipital region (Oz) and N2, a negative component around 200 ms after blink maximum, in the fronto-central region (Fz and FCz) during the cognitive task. An increase in stimulus-evoked P1 amplitude in occipital regions indicates a higher allocation of attentional resources during early visual processing. An increase in stimulus-evoked N2 amplitude is associated with the involvement of cognitive control [3]. Another stimulus-evoked ERP component that has been associated with cognitive load is the P3, a slow wave that appears with a maximum amplitude above the parieto-occipital region (Pz, POz and Oz). Previous studies have shown that the parietal P3 component is a reliable indicator for resource allocation during cognitive processing and a valid index of cognitive load. Increased cognitive load requires more resources for cognitive processing, leading to an increased P3 amplitude (Fig. 1).

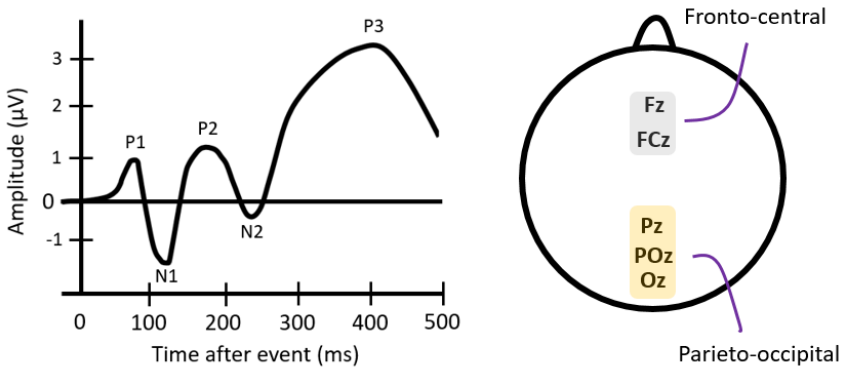


Figure 1 Left panel: A waveform showing ERP components including the P1, N2, and P3. Adapted from `https://en.wikipedia.org/wiki/Event-related_potential`. Right panel: Head map showing the positions of the electrodes of interest in the fronto-central (highlighted in gray) and parieto-occipital (highlighted in orange) regions.

1.3 The present research and hypothesis

The present study investigated how the number of landmarks displayed on a mobile map affects navigators' cognitive load during landmark-based navigation. We hypothesized that a higher number of landmarks displayed on a mobile map would increase cognitive load during a landmark-based navigation task due to increased cognitive resources used to process excess visual and spatial information. Increased cognitive load would be indicated by more pronounced amplitudes in the following blink-related components: the P3 amplitude at the parieto-occipital region, the N2 amplitude at the fronto-central region, and the P1 amplitude at the occipital region. We also hypothesized that spatial learning performance would initially increase from the 3- to 5-landmark conditions and decrease from the 5- to 7-landmark conditions due to increased cognitive load [2].

2 Method

2.1 Participants and experimental design

Forty-eight participants (29 females) with age ranging from 18 to 35 years (M = 25.6 yrs, SD = 4.09) took part in the study. Three participants were excluded because of noisy or missing data due to technical issues, resulting in an analyzed sample of 45 participants. We adopted a within-participant design with three conditions, showing either 3, 5, or 7 landmarks on the mobile map while participants navigated a predefined navigation route (Fig. 2a) in three virtual cities. The three navigation routes consisted each of five intersections and were similar in length (approximately 900 m each). Each route contained seven salient buildings as landmarks: the starting building (home), five landmarks at the five intersections, and the destination building (goal). The landmarks were visualized as either 3D realistic or green rectangles according to landmark-density condition. The three conditions were evenly distributed across the three cities.

2.2 Procedure

Participants were asked to navigate as quickly as possible to a predefined destination and to learn the landmarks displayed on the map. Three virtual cities were designed in ArcGIS City Engine 2018.0 and displayed on a three-sided, stereo cave automatic virtual environment (CAVE) using Unity 2018.4 LTS (Fig. 2b). Participants moved by using a foot-operated controller (Fig. 2c) through the virtual environment displayed in the CAVE. Each city contained a pre-defined route to be followed. The route, including start and destination locations, was shown on a mobile map projected in the center screen of the CAVE during navigation. This map indicated navigators' current location and provided turn-by-turn instructions. The map appeared before and after each intersection, and along straight segments of the followed route. The map rotated along with the navigators' heading direction. After navigating in each city, participants' spatial knowledge was tested using a landmark recognition task, a route direction task, and a Judgements of Relative Direction (JRD) task. While participants were performing the navigation task, their brain activity was measured using a 64-channel EEG device with active electrodes (LiveAmp, Brain Products GmbH, Gilching, Germany). EEG was recorded at a 500 Hz sampling rate with a 131 Hz low-pass filter with input impedance set at below 10 kOhm.

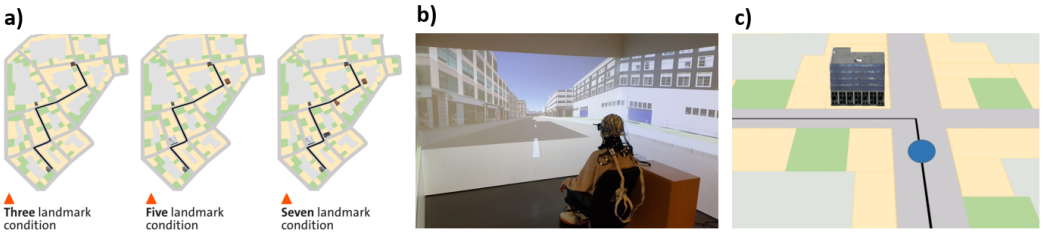


Figure 2 a) Three landmark density conditions in one city. The left, middle, and right figures represent the map condition with three, five, and seven landmarks visualized on the map respectively. b) A participant sat on a chair 30 cm away from the center of the VR system (CAVE), placed her feet on a foot-operated controller, and had her brain activity recorded with EEG during the navigation experiment. c) A track-up map providing a navigator's current location (blue dot), the route direction to follow (black line), and, depending on the landmark density condition, a 3D landmark at the intersection.

2.3 Data processing and analysis

For more details of EEG data preprocessing, please see the appendix A: EEG data preprocessing.

To detect and extract brain activity related to eye blinks, we followed the protocol established by Wunderlich and Gramann [7]. Eye blink events were created by peak detection in the time series of the IC representing vertical eye movements. Next, we removed all independent components from the data that were classified as unlikely to represent brain activity (probability below 30%) and then back-projected the remaining data to the sensor level. To extract bERPs, we used the Unfold toolbox. Information on the different landmark density conditions (3, 5, and 7 landmarks) was entered into the regression formula $y = 1 + cat(landmark)$, which was then solved to obtain the intercepts and beta values (baseline-corrected at -500 to -200 ms preceding the blink event) for each condition. Of the beta values computed, we extracted those corresponding to the bERP components of interest (P1, N2, and P3) at the electrodes of interest (Fz, FCz, Pz, POz, and Oz) for statistical

analysis. The P1 at Oz was extracted from within 110–150 ms after blink maximum. The N2 amplitude was extracted from 250–390 ms after blink maximum and averaged between Fz and FCz. The P3 was extracted from 250–340 ms after blink maximum and averaged between Pz, POz, and Oz. We performed one-way repeated measures ANOVAs with landmark condition as the within-subjects predictor (3 vs. 5 vs. 7 landmarks) on each of the bERP components of interest.

3 Results

3.1 Behavioral results

Multilevel regression modeling was conducted to compare spatial learning performance between the three landmark density conditions in R 4.1.0. The spatial learning result shows that landmark recognition and route direction memory improves when the number of presented landmarks increases from three to five ($beta = 0.51$, 95%CI [0.30, 0.72], $p < 0.001$), while learning performance does not increase further when seven landmarks are depicted on the map ($beta = -0.11$, 95%CI [−0.32, 0.10], $p = 0.31$).There is no significant effect of the number of landmarks on performance on the JRD. More details on results related to behavioral performance are reported in Cheng et al. [2].

3.2 EEG results

The analysis of P1 at Oz shows no significant differences between the conditions, $p = 0.568$. Analysis of the N2 also reveals no significant differences between the conditions $p = 0.660$. Lastly, analysis of the P3 in the parieto-occipital region reveals significant differences between the conditions, $F(2, 44) = 3.72$, $p = 0.028$. Post-hoc contrasts reveal that P3 amplitude in the 7-landmark condition ($M = 3.24$, $SD = 1.99$) is significantly higher compared to the 5-landmark condition ($M = 2.75$, $SD = 1.79$) ($ps = 0.009$), and marginally higher compared to the 3-landmark condition ($M = 2.78$, $SD = 1.82$, $ps = 0.058$; see Fig. 3).

4 Discussion

The present study investigated whether increasing the number of landmarks shown on a mobile map leads to corresponding increases in navigators' cognitive load while they followed a given route in an urban virtual environment. Eye blink-related brain activity was analyzed to reveal cognitive load-dependent changes during map-assisted navigation. We hypothesized that the amplitudes of the P1 at the occipital region, N2 at the fronto-central region, and P3 at the parieto-occipital region would be more pronounced with increased number of landmarks displayed on the mobile map. Our hypothesis on the P3 amplitude was largely supported. It was significantly higher in the 7-landmark condition compared to the 5-landmark condition and marginally higher compared to the 3-landmark condition. However, there were no significant differences between the landmark conditions in the occipital P1 and front-central N2 amplitudes. The behavioral results show that landmark and route knowledge were significantly better in the 5- and 7-landmark condition compared to the 3-landmark condition [2].

The larger P3 amplitude in the 7-landmark condition suggests that participants were allocating more attentional resources to the task, indicating that presenting more landmarks on a mobile map adds to users' cognitive load. Future analysis could use statistical methods such as linear mixed models instead of ANOVAs to reduce inter-subject variance. When

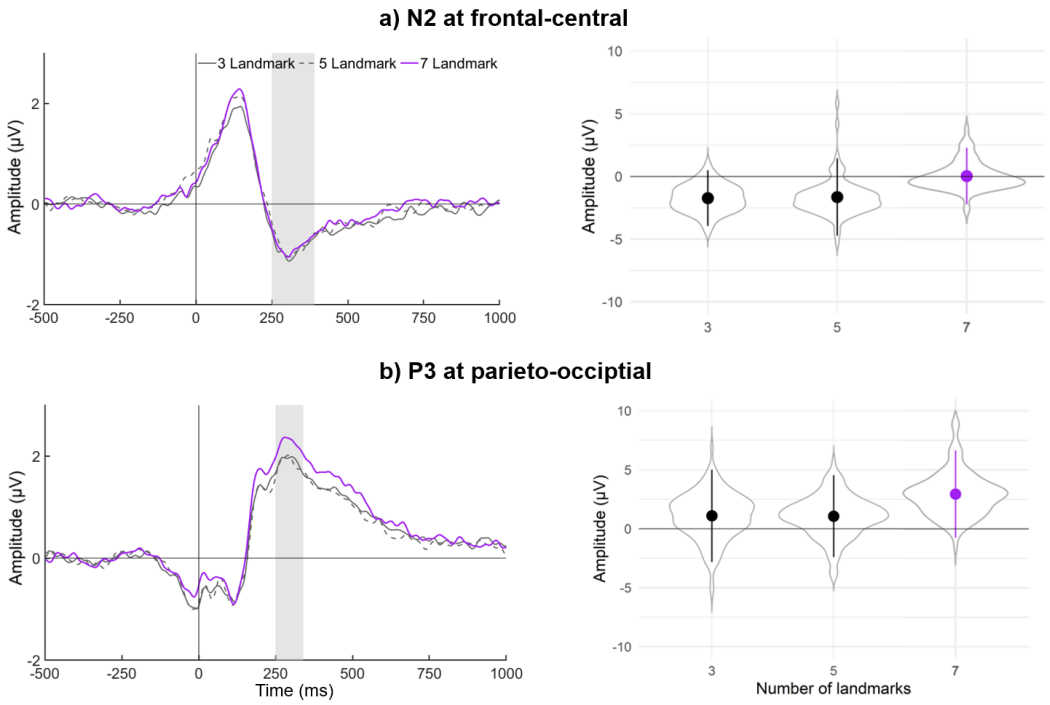


Figure 3 Left panel: Grand averaged amplitudes of bERPs for each landmark condition at the parieto-occipital region (Pz, POz, and Oz). The bERP waves served as visual inspection for individual peak detection – area shaded in gray indicates the time window where the P3 was extracted for each participant. Blink maximum occurred at 0 ms. Right panel: Violin plot displaying the means, standard deviations and distributions of the detected amplitude peaks in each landmark condition for the P3. Vertical lines denote 2× the standard deviation of the mean. Violin widths indicate the probability density of the data at different amplitudes.

considering the behavioral results, displaying five landmarks on the mobile map seems to have the best behavioral outcome without increasing cognitive load. On the other hand, displaying seven landmarks on the mobile map increased cognitive load without improving spatial learning performance [2].

Previous studies on blink-related N2 in the fronto-central region compared N2 amplitude during 1) rest, 2) physical activity, and 3) when performing a cognitive task. This comparison differs from that of the present study, which compared N2 amplitude at different levels of a cognitive task. It is possible that blink-related fronto-central N2 amplitude changes in load vs. no load conditions but is not sensitive to differing levels of cognitive load. This explanation needs further investigation.

5 Conclusion

Our current study provides initial evidence that presenting a greater number of landmarks on mobile maps increases users' cognitive load. Our preliminary results have several implications for the design of map-based navigation assistance systems and the literature on wayfinding. Our study shows that eye blink related potentials are sensitive to cognitive load changes in naturalistic settings. As most of the literature on ERPs use stimulus-evoked or response-related event markers, more research that investigates bERPs is needed. Moreover, designers of mobile maps should consider how the display could influence users' cognitive load during

navigation. Specifically, the amount of information presented on mobile map displays should elicit an optimum level of cognitive load in users without taxing them beyond that used to perform an already cognitively demanding navigation task. The results of our study suggest that showing five landmarks on mobile maps could improve users' spatial learning performance without taxing extra cognitive resources.

References

1 Deborah A Boehm-Davis, Wayne D Gray, and Michael J. Schoelles. The eye blink as a physiological indicator of cognitive workload. In *Proceedings of the XIVth Triennial Congress of the International Ergonomics Association and 44th Annual Meeting of the Human Factors and Ergonomics Association, 'Ergonomics for the New Millennium'*, pages 116–119, 2000. doi:10.1177/154193120004403309.

2 Bingjie Cheng, Ian T Ruginski, and Sara I Fabrikant. The effects of landmark density in mobile maps on spatial learning during pedestrian navigation. In *Spatial Cognition 2020/1*, pages 1–2. University of Lative, Lativa, August 2–4, 2021.

3 Jonathan R Folstein and Cyma Van Petten. Influence of cognitive control and mismatch on the N2 component of the ERP: A review. *Psychophysiology*, 45(1):152–170, 2008. doi:10.1111/j.1469-8986.2007.00602.x.

4 Toru Ishikawa. Satellite Navigation and Geospatial Awareness: Long-Term Effects of Using Navigation Tools on Wayfinding and Spatial Orientation. *Professional Geographer*, 71(2):197–209, 2019. doi:10.1080/00330124.2018.1479970.

5 Steven J. Luck and Edward K. Vogel. The capacity of visual working memory for features and conjunctions. *Nature*, 390(6657):279–284, November 1997. doi:10.1038/36846.

6 Edmund Wascher, Holger Heppner, Sven O. Kobald, Stefan Arnau, Stephan Getzmann, and Tina Möckel. Age-sensitive effects of enduring work with alternating cognitive and physical load. A study applying mobile EEG in a real life working scenario. *Frontiers in Human Neuroscience*, 9(JAN2016), January 2016. doi:10.3389/fnhum.2015.00711.

7 Anna Wunderlich and Klaus Gramann. Eye movement-related brain potentials during assisted navigation in real-world environments. *European Journal of Neuroscience*, 2020. doi:10.1111/ejn.15095.

A EEG data preprocessing

The BeMoBIL pipeline 1.0 was used to preprocess and clean the EEG data using the MATLAB toolbox EEGLAB. We first downsampled the raw EEG data to 250 Hz. Then, we applied a 0.5 Hz high-pass filter to suppress slow drifts in EEG data and removed spectral peaks at 50 Hz, corresponding to power line frequency, using the *ZapLine* plus function. We identified noisy channels using the automated rejection function *cleanartifacts* from EEGLAB with ten iterations. We removed channels that were detected as bad channels more than four times and interpolated them by spherical interpolation of neighboring channels and applied re-referencing to the common average. On the cleaned dataset, we performed an independent component analysis (ICA) using an adaptive mixture independent component analysis (AMICA) algorithm. For each independent component (IC), we computed an equivalent current dipole (ECD) model with the DIPFIT plugin from EEGLAB.

B Analysis on number of eye blinks

The number of eye blinks did not differ by landmark density condition, $F(1, 44) = 1.49$, $p = .229$. Table 1 presents the average numbers of blinks per landmark density condition.

Table 1 Means and standard deviation (SD) of number of blinks in the three landmark density conditions.

	3-Landmark	5-Landmark	7-Landmark
Mean	143.98	130.04	138.87
SD	97	83.13	72.78

Representing Computational Relations in Knowledge Graphs Using Functional Languages

Yanmin Qi
School of Computer Science, University of Nottingham Ningbo, China
State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, China

Heshan Du
School of Computer Science, University of Nottingham Ningbo, China

Amin Farjudian
School of Computer Science, University of Nottingham Ningbo, China

Yunqiang Zhu
State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, China
University of Chinese Academy of Sciences, Beijing, China

Abstract

Knowledge representation is the cornerstone of constructing a geoscience knowledge graph (GKG). The existing representations of spatial and computational relations in GKGs, however, are inadequate. In this paper, we use Dimensionally Extended Nine-Intersection Model (DE-9IM) to represent spatial topological relations. To represent computational relations, we use typed lambda calculus via its implementation in the functional language Haskell, in which functions are first-class primitives. We exemplify our ideas through some basic examples in Haskell.

2012 ACM Subject Classification Theory of computation → Semantics and reasoning

Keywords and phrases spatial relation, computational relation, functional programming, Haskell, geo-knowledge graph

Digital Object Identifier 10.4230/LIPIcs.COSIT.2022.29

Category Short Paper

Funding *Yunqiang Zhu*: National Natural Science Foundation of China [grant number 42050101]

1 Introduction

Knowledge graphs provide a paradigm for representing interconnected information derived from heterogeneous sources [4]. A geoscience knowledge graph (GKG) organizes geoscience knowledge with a structured graph, which is suitable for algorithmic processing [21]. The fundamental idea of establishing a GKG is inspired by the spatial and temporal features contained in geoscience phenomena and processes, and by knowledge representation models illustrated in various disciplines of geoscience [21]. Overall, multi-scale spatial and temporal features are the most significant characteristics of GKGs when compared with knowledge graphs in other disciplines [21].

The cornerstone of constructing a GKG is knowledge representation. Knowledge representation is closely connected with formal ontology, which deals with entities from the given world, together with their properties and relations between them [5]. Normally, two types of representation models are implemented: one being {entity, relation, entity} (e. g., {Ningbo City, within, Zhejiang Province}), and another being {entity, attribute, attribute value} (e. g., {Ningbo City, has population, 8.202 million}) [19]. Reasoning – the

process of inferring conclusions from existing knowledge [3] – provides a salient purpose for knowledge representation. Geoscience knowledge reasoning is carried out over GKGs [21], and is used to shed light on the evolutionary features of geoscience knowledge systems.

It is, however, challenging to represent complex interdisciplinary geoscience knowledge by the existing representation models [21]. Another challenge is posed by the uniform integration of spatial, temporal, and computational relations in GKGs [21]. Three kinds of spatial relations are primarily used in geoscience, which are direction, distance, and topology [20]. Temporal relations signify how events relate to one another in time. Computational relations describe the process of using mathematical or logical methods to generate entities from other entities, and they normally include mathematical formulas and modelling procedures [21]. In the existing models, the representation of a computational relation only expresses the existence of the relation. The formula or logical rule underlying the computational relation, however, is not represented, because the models are not expressive enough.

In this paper, we use Dimensionally Extended Nine-Intersection Model (DE-9IM) to represent spatial topological relations. We use the functional programming language Haskell [8] to represent computational relations and perform reasoning about spatial topological relations. The paper is organized as follows: Section 2 describes the representation model of spatial relations. The Haskell implementations of spatial reasoning and representation of computational relations are exemplified in Section 3. Conclusions and directions for future work are presented in Section 4.

2 Representation of Spatial Topological Relations

According to the Open Geospatial Consortium (OGC), the DE-9IM [17] string code can be a standardized format for the data interchange of the typical spatial predicates [14]. DE-9IM represents the relationship between two spatial objects A and B based on the intersections of their interiors, boundaries, and exteriors. The spatial objects discussed here include points, lines, and 2D regions. For our discussion, the concepts of interior, boundary, and exterior are not the same as those defined in classical topology. For instance, for a given line, we consider the end-points as the boundary, and the (open) segment between the end-points as the interior. All the remaining points in the plane are regarded as the exterior. For a point, the interior is the point itself, the boundary is empty, and the exterior is the entire plane minus the point itself [17]. Dimension values are assigned to spatial objects as follows: 0 for points, 1 for lines, and 2 for 2D regions. All the above values (0, 1 and 2) are "TRUE" values in the spatial predicates as they signify non-empty sets. An empty set is represented by -1, which is the "FALSE" value in the spatial predicates. The eight spatial relationships involved in DE-9IM are "Crosses", "Within", "Contains", "Equals", "Disjoint", "Intersects", "Touches", and "Overlaps", as detailed in [17].

The spatial relations defined in DE-9IM have been implemented in function `ST_Relate`() of PostGIS [6]. For instance, the triple {Ningbo City, Touches, Shaoxing City} semantically describes that the relation between "Ningbo City" and "Shaoxing City" is "Touches". If we call `ST_Relate`("`NingboCity`", "`ShaoxingCity`") in PostGIS, the spatial relation "Touches" will be represented as $[FF2F11212]$, in which `F` represents "FALSE" (-1).

For any particular spatial relation, the existing knowledge graphs use different semantic descriptions, which causes ambiguity. This problem is exacerbated by the ambiguity that is inherent in natural languages. For instance, describing entity A as "nearTo" entity B does not specify how far (say) in meters they are from each other. Therefore, the purpose of using relation matrixes represented by DE-9IM is to quantitatively represent spatial relations, which is more suitable for verified learning and inference of downstream models and reasoning [11].

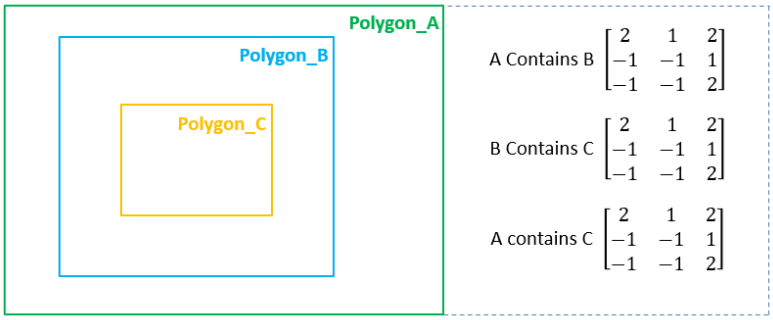


Figure 1 Spatial topological relations among three objects.

3 Representation and Reasoning in Haskell

Haskell is a functional programming language, based on typed lambda calculus. As opposed to imperative languages, the syntax of functional languages resembles mathematical expressions more closely. As such, in functional programming, the focus is mainly on what is being computed, rather than low-level details of how it is computed [7]. Functional syntax has indeed been considered for representation and reasoning in the literature as well. For instance, Shahzad et al. [16] integrated Haskell with relational algebra to create a user model called Universe of Discourse (UoD). Leinberger et al. [9] defined and presented a functional language for handling semantic data. Haskell was also used for integrating conventional sensor information and volunteered geographic information [15].

For our purposes, Haskell provides many advantages over other languages, especially, imperative languages such as Java or Python. The most immediate advantage is that in Haskell, functions are first-class objects, a property shared with other pure functional languages. For the current work, this is the most important feature which enables us to integrate computational relations into GKGs seamlessly. There are also other strong features of Haskell which are vital wherever correctness must be guaranteed [2]. Haskell is a safe and strongly typed language. This enables faithful retention of ontological relations through transformations. In Haskell, variables are immutable, which enables verification of the code through algebraic manipulations [1]. Most importantly, imperative (impure) operations are syntactically separated from the pure ones through the use of monads [12, 18].

Here, we present a simple example to illustrate the process of reasoning about spatial topological relations among three polygons. Then, a computational relation of a case study will be presented.

3.1 Reasoning about Spatial Topological Relation

Figure 1 depicts the spatial topological relations among three polygons. Polygon A contains Polygon B, Polygon B contains Polygon C. The DE-9IM matrices for the corresponding relations are also displayed on the right side of Figure 1. Based on the relation between Polygon A and Polygon B, and the relation between Polygon B and Polygon C, the relation between Polygon A and Polygon C can be inferred.

Although this is a simple deduction, the basic principles can be demonstrated via the corresponding Haskell implementation. Two functions are defined: one is the `compu_sr`() functions to compute the matrix of the spatial topological relation between two objects, and another one is the `reason_sr`(), which infers the matrix that represents the spatial

topological relation based on two input matrices. For function `compu_sr`(), the two input parameters are two spatial objects, the output is the matrix that describes the topological relation between the two input objects. Keyword `data` is applied to define the data types of parameters, while `class` defines the types that share same behaviors, and can be computed by the function declared in the `class`.

```
data Matrix= Matrix [[Int]]

class COMPU_SPATIAL_RELATION object_1 object_2
  where compu_sr:: object_1-> object_2-> Matrix

class REASON_SPATIAL_RELATION matrix_1 matrix_2
  where reason_sr:: matrix_1-> matrix_2-> Matrix
```

Before implementing the defined functions, the data types of entities used as real instance data are defined. The matrices of relations between two pairs of entities ((Polygon_A, Polygon_B) and (Polygon_B, Polygon_C)) are generated.

```
data Entity_A = Polygon_A
data Entity_B = Polygon_B
data Entity_C = Polygon_C

instance COMPU_SPATIAL_RELATION Entity_A Entity_B where
compu_sr (Polygon_A)(Polygon_B) = Matrix [[2,1,2],[-1,-1,1],[-1,-1,2]]

instance COMPU_SPATIAL_RELATION Entity_B Entity_C where
compu_sr (Polygon_B)(Polygon_C) = Matrix [[2,1,2],[-1,-1,1],[-1,-1,2]]
```

Using the matrices computed above, the data types of input matrices are defined, and the matrix of the "contains" relation between Polygon A and Polygon C is generated.

```
data Matrix_M= Matrix_M [[Int]]

instance REASON_SPATIAL_RELATION Matrix_M Matrix_M
  where reason_sr (Matrix_M [[2,1,2],[-1,-1,1],[-1,-1,2]])
                  (Matrix_M [[2,1,2],[-1,-1,1],[-1,-1,2]])
                = Matrix [[2,1,2],[-1,-1,1],[-1,-1,2]]
```

3.2 Computational Relations

Soil erosion is a widespread and major environmental threat to terrestrial ecosystems. To study soil erosion in the Chinese Loess Plateau, we use the revised universal soil loss equation (RUSLE) [10]. RUSLE is a widely accepted method that can be used as the best-fitted model for monitoring soil erosion. The RUSLE equation is as follows:

$$A = R \times K \times LS \times C \times P,$$

in which A is the average soil loss ($t \cdot hm^{-1} \cdot a^{-1}$). Here we take the calculation of factor R as an example of a computational relation in models of GKGs, where R is the rainfall erosivity factor ($MJ \cdot mm \cdot hm^{-1} \cdot h^{-1} \cdot a^{-1}$). The rainfall erosivity factor of a particular month can be obtained using:

$$R_i = 1.735 \times 10^{(1.5 \times lg \frac{p_i^2}{p_a} - 0.8188)} \tag{1}$$

where R_i is the rainfall erosivity factor of a particular month, p_i is the monthly rainfall in a particular year, and p_a is the yearly rainfall. The first step is to construct formal ontology of rainfall. We integrate top-level ontologies in Semantic Web for Earth and Environmental Terminology (SWEET) Ontology [13] with type classes, which are sets of types sharing the same behavior.

```
class THING thing
```

The upper-case `THING` is the type class name, which is the top ontology in SWEET. The parameter `thing` is the type that belongs to the class. To represent subclass relations, the symbol => is used to illustrate the hierarchy of ontology levels. `PHENOMENA` is the subclass of `THING` class in SWEET, we then define the class `RAINFALL`, `AVERAGE_ANNUAL_RAINFALL` and `AVERAGE_MONTHLY_RAINFALL`.

```
class THING phenomena => PHENOMENA phenomena
class PHENOMENA rainfall => RAINFALL rainfall

class RAINFALL average_annual_rainfall =>
  AVERAGE_ANNUAL_RAINFALL average_annual_rainfall

class RAINFALL average_monthly_rainfall =>
  AVERAGE_MONTHLY_RAINFALL average_monthly_rainfall
```

To represent the calculation of factor R, the keyword `data` is applied to define the data type of parameters in (1). The name of the data type is `Value`. There are four value constructors in data type `Value`, for example, constructor `Measure` has the data type `Float`.

```
data Value=Measure Float | Count Int | Boolean Bool | Category String
```

To represent the computational relation described by (1), we first define the type class `RAINFALL_EROSIVITY_FACTOR` (REF), which corresponds to the dependent variable R_i in (1). Two parameters must be included in type class REF: `average_monthly_rainfall` (representing monthly average rainfall p_i), and `average_annual_rainfall` (denoting yearly average rainfall p_a). REF is declared as a multi-parameter type class for the computational relation, defining which parameters are required to calculate the rainfall erosivity factor. The input parameters of the function `rFactor` are average monthly rainfall and average annual rainfall, the data type of the result is one of the data type constructors in `Value`.

```
class RAINFALL_EROSIVITY_FACTOR average_monthly_rainfall
average_annual_rainfall
  where rFactor ::  average_monthly_rainfall->
                    average_annual_rainfall -> Value
```

Next, we use the keyword `instance` to implement the computational relation defined above. Firstly, we define the data type of average monthly rainfall and average annual rainfall as float type. `Avg_monthly_rainfall` and `Avg_annual_rainfall` are the real instance data types that correspond to `average_monthly_rainfall` and `average_annual_rainfall`. The parameters `monthlyrain_avg` and `annualrain_avg` are the parameters with data type `Float`, which are the input parameters for function `rFactor`. Therefore, the constructor `Measure` in `Value` is used to illustrate the data type of rainfall erosivity factor. The computational relation exists among rainfall erosivity factor, average monthly rainfall and average annual rainfall.

```
data Avg_monthly_rainfall = Avg_monthly_rainfall Float
data Avg_annual_rainfall = Avg_annual_rainfall Float

instance RAINFALL_EROSIVITY_FACTOR Avg_monthly_rainfall
Avg_annual_rainfall
  where rFactor (Avg_monthly_rainfall monthlyrain_avg)
                (Avg_annual_rainfall annualrain_avg) =
                Measure (1.735*(10**(1.5* (logBase 10
                ((monthlyrain_avg))**2/(annualrain_avg)))-0.8188))
```

4 Conclusion

Functional languages provide many advantages over their imperative counterparts for representation and reasoning tasks in geoscience knowledge graphs (GKGs). The first step in this direction is to represent computational relations as first-class objects, which enables further reasoning and algorithmic processing in a uniform framework. As such, languages such as Haskell which treat functions as first-class primitives are ideal for augmenting GKGs with computational relations as first-class entities. Higher-order functions can then be used for uniform processing of ground entities and computational relations alike. Furthermore, Haskell is ideal for computational processes that must retain ontological relations faithfully. We have demonstrated through some simple examples how these tasks can be handled in Haskell.

In our future research, further computational relations will be studied. Besides mathematical expressions, logic rules and modelling procedures will also be considered. Moreover, the representation of spatial direction relations and temporal relations will be included. Based on the case study of representing computational relations in this paper, we will investigate efficient reasoning with spatial, temporal, and computational relations in GKGs. This includes seamless integration of functional programming into qualitative spatial and temporal reasoning.

References

1 John Backus. Can programming be liberated from the von Neumann style? A functional style and its algebra of programs. *Commun. ACM*, 21(8):613–641, 1978. doi:10.1145/359576.359579.

2 Thomas Bittner, Jonathan P. Bona, and Werner Ceusters. Ontologies of dynamical systems and verifiable ontology-based computation: Towards a Haskell-based implementation of Referent Tracking. In Roberta Ferrario and Werner Kuhn, editors, *Formal Ontology in Information Systems – Proceedings of the 9th International Conference, FOIS 2016, Annecy, France, July 6-9, 2016*, volume 283 of *Frontiers in Artificial Intelligence and Applications*, pages 313–327. IOS Press, 2016. doi:10.3233/978-1-61499-660-6-313.

3 Xiaojun Chen, Shengbin Jia, and Yang Xiang. A review: Knowledge reasoning over knowledge graph. *Expert Syst. Appl.*, 141, 2020. doi:10.1016/j.eswa.2019.112948.

4 Jiaxin Du, Shaohua Wang, Xinyue Ye, Diana S Sinton, and Karen Kemp. GIS-KG: Building a large-scale hierarchical knowledge graph for geographic information science. *International Journal of Geographical Information Science*, pages 1–25, 2021.

5 Antony Galton. Spatial and temporal knowledge representation. *Earth Sci. Informatics*, 2(3):169–187, 2009. doi:10.1007/s12145-009-0027-6.

6 Leo S Hsu and Regina Obe. *PostGIS in action*. Simon and Schuster, 2021.

7 Zhenjiang Hu, John Hughes, and Meng Wang. How functional programming mattered. *National Science Review*, 2(3):349–370, 2015.

8 Graham Hutton. *Programming in Haskell*. Cambridge University Press, 2nd edition, 2016.

9 Martin Leinberger, Ralf Lämmel, and Steffen Staab. The essence of functional programming on semantic data. In *Programming Languages and Systems – 26th European Symposium on Programming, ESOP 2017*, volume 10201 of *Lecture Notes in Computer Science*, pages 750–776. Springer, 2017.

10 Xun-Gui Li and Xia Wei. Soil erosion analysis of human influence on the controlled basin system of check dams in small watersheds of the Loess Plateau, China. *Expert Syst. Appl.*, 38(4):4228–4233, 2011. doi:10.1016/j.eswa.2010.09.088.

11 Gengchen Mai, Krzysztof Janowicz, Yingjie Hu, Song Gao, Bo Yan, Rui Zhu, Ling Cai, and Ni Lao. A review of location encoding for GeoAI: methods and applications. *International Journal of Geographical Information Science*, pages 1–35, 2022.

12 Eugenio Moggi. Notions of computation and monads. *Information and Computation*, 93(1):55–92, 1991. doi:10.1016/0890-5401(91)90052-4.

13 Robert G Raskin and Michael J Pan. Knowledge representation in the Semantic Web for Earth and Environmental Terminology (SWEET). *Computers & Geosciences*, 31(9):1119–1125, 2005.

14 Ahmet Sayar, Marlon Pierce, and Geoffrey Fox. OGC compatible geographical information systems web services. *Indiana Computer Science Report TR610*, 2005.

15 Sven Schade, Frank O. Ostermann, and Laura Spinsanti. Functional integration for the observation web. In *KEOD 2011 – Proceedings of the International Conference on Knowledge Engineering and Ontology Development, Paris, France, 26-29 October, 2011*, pages 498–504. SciTePress, 2011.

16 Syed K Shahzad, Michael Granitzer, and Klause Tochterman. Designing user interfaces through ontological user model: functional programming approach. In *2009 Fourth International Conference on Computer Sciences and Convergence Information Technology*, pages 99–104, 2009.

17 Christian Strobl. Dimensionally Extended Nine-Intersection Model (DE-9IM). In *Encyclopedia of GIS*, pages 470–476. Springer, 2017. doi:10.1007/978-3-319-17885-1_298.

18 Philip Wadler. How to declare an imperative. *ACM Comput. Surv.*, 29(3):240–263, 1997. doi:10.1145/262009.262011.

19 Shu Wang, Xueying Zhang, Peng Ye, Mi Du, Yanxu Lu, and Haonan Xue. Geographic Knowledge Graph (GeoKG): A formalized geographic knowledge representation. *ISPRS Int. J. Geo Inf.*, 8(4):184, 2019. doi:10.3390/ijgi8040184.

20 Yi Zhang, Yong Gao, LuLu Xue, Si Shen, and KaiChen Chen. A common sense geographic knowledge base for GIR. *Science in China Series E: Technological Sciences*, 51(1):26–37, 2008.

21 Chenghu Zhou, Hua Wang, Chengshan Wang, Zengqian Hou, Zhiming Zheng, Shuzhong Shen, Qiuming Cheng, Zhiqiang Feng, Xinbing Wang, Hairong Lv, et al. Geoscience knowledge graph in the big data era. *Science China Earth Sciences*, 64(7):1105–1114, 2021.