

# Putting Data on the Map

Edited by

Stephen Kobourov<sup>1</sup>, Alexander Wolff<sup>2</sup>, and Frank van Ham<sup>3</sup>

1 University of Arizona, Tucson, US, [kobourov@cs.arizona.edu](mailto:kobourov@cs.arizona.edu)

2 Universität Würzburg, DE,

[http://www1.informatik.uni-wuerzburg.de/en/staff/wolff\\_alexander](http://www1.informatik.uni-wuerzburg.de/en/staff/wolff_alexander)

3 IBM Software Group, NL, [frankvanham@nl.ibm.com](mailto:frankvanham@nl.ibm.com)

---

## Abstract

This report documents the program and the outcomes of Dagstuhl Seminar 12261 “Putting Data on the Map”.

**Seminar** June 24–29, 2012 – [www.dagstuhl.de/12261](http://www.dagstuhl.de/12261)

**1998 ACM Subject Classification** F.2.2 Nonnumerical Algorithms and Problems

**Keywords and phrases** Information Visualization, Cartography, GIScience, Computational Geometry, Graph Drawing, Cognition

**Digital Object Identifier** 10.4230/DagRep.2.6.51

**Edited in cooperation with** Philipp Kindermann

## 1 Executive Summary

*Stephen Kobourov*

*Alexander Wolff*

*Frank van Ham*

**License**  Creative Commons BY-NC-ND 3.0 Unported license  
© Stephen Kobourov, Alexander Wolff, and Frank van Ham

Visualization allows us to perceive relationships in large sets of interconnected data. While statistical techniques may determine correlations among the data, visualization helps us frame what questions to ask about the data. The design and implementation of algorithms for modeling, visualizing and interacting with large relational data is an active research area in data mining, information visualization, human-computer interaction, and graph drawing.

*Map representations* provide a way to visualize relational data with the help of conceptual maps as a data representation metaphor. In a narrow sense, a map representation of a graph is a contact graph representation where the adjacency of vertices is expressed by regions that share borders. Such representations are, however, limited to planar graphs by definition. We can extend the notion of a map representation to non-planar graphs by generalizing the idea as follows: clusters of well-connected vertices form countries, and countries share borders when neighboring clusters are tightly interconnected.

*Information spatialization and cartograms* also connect the notions of data with those of maps. Cartograms redraw an existing geographic map such that the country areas are proportional to some metric (e.g., population), an idea that dates back to a paper by Raisz in 1934 and is still popular today. *Spatialization* is the process of assigning two- or three-dimensional coordinates to abstract data points, ideally such that the spatial mapping has much of the characteristics of the original high-dimensional space. Multi-dimensional scaling or principal component analysis are techniques that allow us to spatialize high-dimensional



Except where otherwise noted, content of this report is licensed under a Creative Commons BY-NC-ND 3.0 Unported license

Putting Data on the Map, *Dagstuhl Reports*, Vol. 2, Issue 6, pp. 51–76

Editors: Stephen Kobourov, Alexander Wolff, and Frank van Ham



Dagstuhl Reports

Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

data. Techniques like information landscapes can then be used to convert the resulting two-dimensional coordinates into meaningful three-dimensional landscapes.

Providing efficient and effective data visualization is a difficult challenge in many real-world software systems. One challenge lies in developing algorithmically efficient methods to visualize large and complex data sets. Another challenge is to develop effective visualizations that make the underlying patterns and trends easy to see. And finally, we need to allow users to interactively access, analyze, and filter these patterns in an intuitive manner. All of these tasks are becoming increasingly more difficult due to the growth of the data sets arising in modern applications, as well as due to their highly dynamic nature.

### Topics of the Seminar

Graph representations of side-to-side touching regions tend to be visually appealing and have the added advantage that they suggest the familiar metaphor of a geographical map. Traditional maps offer a natural way to present geographical data (continents, countries, states) and additional properties defined with the help of contours (topography, geology, rainfall).

An important difference between drawings of graphs and maps is the following: graphs are usually *drawn on* the plane (using small placeholder symbols for vertices and curves for edges), whereas maps *fill* the plane (or a sufficiently large area). We want to explore this new paradigm.

In the process of data mining and data analysis, clustering is an extremely important step. It turns out that maps are very helpful in dealing with clustered data. There are several reasons why a map representation of clusters can be helpful. First, by explicitly defining the boundary of the clusters and coloring the regions, we make the clustering information clear. Second, as most dimensionality-reduction techniques lead to a two-dimensional positioning of the data points, a map is a natural generalization. Finally, while it often takes us considerable effort to understand graphs, charts, and tables, a map representation is intuitive, as most people are very familiar with maps and even enjoy carefully examining maps.

When designing algorithms to produce maps for abstract data, we can leverage cartography and GIS expertise in order to answer critical questions such as how regions and geographic networks (such as street or river networks) are represented on traditional geographic maps, how they are labeled (an interesting problem in its own right) and how (boundary) lines are simplified (through a process called *cartographic generalization*), or even schematized, in order to focus on important features. Therefore, participation of people from several diverse areas is essential for the success of our seminar.

### Aims of the Seminar

The main goal of this seminar was to foster co-operation between researchers with interests in data visualization coming from the information visualization, human-computer interaction, data mining, graph drawing, and GIS communities. The specific aims of the Dagstuhl seminar were:

1. To bring together researchers working on visualization from a theoretical point of view (graph theory, computational geometry), from a practical point of view (information visualization, HCI), and from a map point of view (cartography, GIS).

2. To identify specific theoretical and practical problems that need to be solved in order to make it possible to create full-fledged conceptual maps as an interactive and scalable data-representation metaphor and to begin working on these problems at the seminar.
3. To formulate the findings as a first step to the solutions of the problems under consideration and to define future research directions.

In order to promote the communication and cooperation between the diverse set of participants, we used a non-traditional format, which included survey presentations, open problem sessions, demo sessions, open mic sessions, problem solving sessions, as well as an exhibition of map-based visualizations. The exhibition entitled “Beyond the Landscape” was organized by seminar participant Maxwell Roberts and by seminar co-organizer Alexander Wolff. It was opened on June 26 by the scientific director of Schloss Dagstuhl, Prof. Reinhard Wilhelm.

### Achievements of the Seminar

The achievements in the seminar were numerous and varied. Some of the more important ones can be summarized as follows:

1. On Monday and Tuesday, we enjoyed five survey lectures; see Section 3. Jason Dykes discussed geographic data visualization. Sara Fabrikant presented the cartographic and geovisual perspective. Stephen Kobourov talked about visualizing relational data with the help of the map metaphor. Stefan Felsner illustrated connections with geometry and graph theory. Falko Schmid discussed maps and the interaction with geographic data on small mobile devices. Beyond the survey lectures, a highlight of the seminar was the Friday morning lecture by psychology and perception expert Barbara Tversky; see Section 3.6.
2. We also had a number of stimulating presentations and demos of new software. In particular, new approaches to the layout of large and/or dynamic graphs as well as new visualization paradigms were presented.
3. A number of relevant open problems were formulated early in the seminar and working groups formed around related open problems. The groups then worked by themselves; formalizing and solving their specific theoretical and practical challenges. Below is a list of the working group topics.
  - a. Geometric properties of cartograms; convex cartograms
  - b. Evaluation of maps and graphs
  - c. Metro map visualization
  - d. Semantic word cloud visualization
  - e. Edge bundling problems
  - f. Multi-dimensional temporal data on maps
  - g. Map distortion based on (dis)similarity
  - h. Work flow for creating maps out of relational data
    - i. Maps based on space-filling curve ordering
    - j. Multi-scale map generalizations

The last three days of the seminar were dedicated to working group effort. Several of the groups kept their focus on the original problems as stated in the open problem session, while other groups modified and expanded the problems; see Section 4. On the last day of the seminar we heard progress reports from all but two of the groups. We are expecting several research publications to result directly from the Seminar.

Arguably the best, and most-appreciated, feature of the seminar was the opportunity to engage in discussion and interactions with experts in various fields with shared passion about maps. The aforementioned exhibition “Beyond the Landscape” made topics of the seminar visible and raised new questions.

In summary, it is our impression that the (56!) participants enjoyed the great scientific atmosphere offered by Schloss Dagstuhl and profited from the scientific program. We are grateful for having had the opportunity to organize this seminar. We thank Philipp Kindermann for helping us to put this report together.

## 2 Table of Contents

### Executive Summary

*Stephen Kobourov, Alexander Wolff, and Frank van Ham* . . . . . 51

### Overview of Talks

(Geo)Visualization at the giCentre  
*Jason Dykes* . . . . . 57

Connecting the Dots: A Cartographic Perspective  
*Sara Fabrikant* . . . . . 61

From Data to Maps  
*Stephen Kobourov* . . . . . 61

Graph Representations: Rectangles, Squares and Prescribed Area  
*Stefan Felsner* . . . . . 62

Removing Data From The Map – How Information Reduction and Tailored Interaction Makes Maps Usable  
*Falko Schmid* . . . . . 62

Cognitive Tools (brief talk)  
*Barbara Tversky* . . . . . 63

### Working Groups

A Note on Representing Data on Maps  
*Mohammed Jawaherul Alam, Walter Didimo, Stefan Felsner, Ferran Hurtado, Marc van Kreveld, Giuseppe Liotta, Pavel Valtr, and Kevin Verbeek* . . . . . 63

Computation of Wordles with Semantic Constraints  
*Sara Irina Fabrikant, Stephen Kobourov, Anna Lubiw, Martin Nöllenburg, Yoshio Okamoto, Claudio Squarcella, Torsten Ueckerdt, and Alexander Wolff* . . . . . 64

The Role of Animation and Interaction for Exploring Geotemporal Data  
*Pierre Dragicevic, Sara Irina Fabrikant, Christophe Hurter, William Mackaness, and Barbara G. Tversky* . . . . . 65

A Taxonomy of Temporal Visualizations Based on Space-Time Cube Operations  
*Daniel Archambault, Sheelagh Carpendale, Pierre Dragicevic, Christophe Hurter, and Ying Yang* . . . . . 65

A Transdisciplinary Survey of Multiscale Rendering Techniques  
*Sheelagh Carpendale, Pierre Dragicevic, Christoph Hurter, William Mackaness, and Monika Sester* . . . . . 66

Putting Maps on a Curve  
*David Auber, Sergi Cabello, Fabrizio Frati, Herman Haverkort, Martin Gronemann, Michael Kaufmann, and Ignaz Rutter* . . . . . 66

A Generic Work-Flow for Making Maps out of Graphs  
*Robert P. Biuk-Aghai, Joe Fowler, Jan-Hendrik Haunert, Petra Mutzel, Frank van Ham, and Marc van Krefeld* . . . . . 68

Area Labeling <i>Jan-Henrik Haunert, Bernhard Jenny, Philipp Kindermann, Sergey Pupyrev, and Falko Schmid</i> . . . . .	71
A Metro Map Problem <i>Maxwell J. Roberts, Andreas Reimer, Yoshio Okamoto, and Therese Biedl</i> . . . . .	72
Showing Dissimilarity Data on Cartographic Maps <i>Sheelagh Carpendale, Tim Dwyer, Jason Dykes, Nathalie Henry-Riche, Arlind Nocaj, and Bettina Speckmann</i> . . . . .	73
Evaluating Bundling Quality Using Image-Based Techniques <i>Stephan Diehl, Seok-Hee Hong, Quan Nguyen, Monika Sester, and Alexandru Telea</i>	74
<b>Participants</b> . . . . .	76

### 3 Overview of Talks

#### 3.1 (Geo)Visualization at the giCentre

Jason Dykes (City University London, GB)

License     Creative Commons BY-NC-ND 3.0 Unported license  
© Jason Dykes

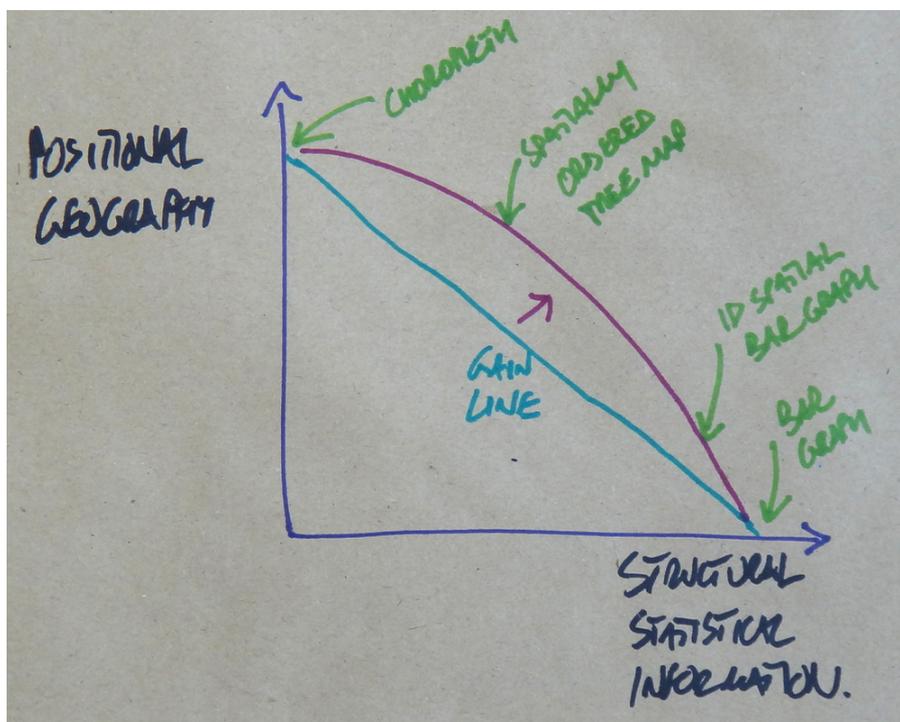
Joint work of Wood, Jo; Slingsby, Aidan

In my survey talk I drew upon a number of examples of geo and information visualization to explore. The objective was to demonstrate and situate the giCentre approach to putting data on various kinds of maps and map like graphics.

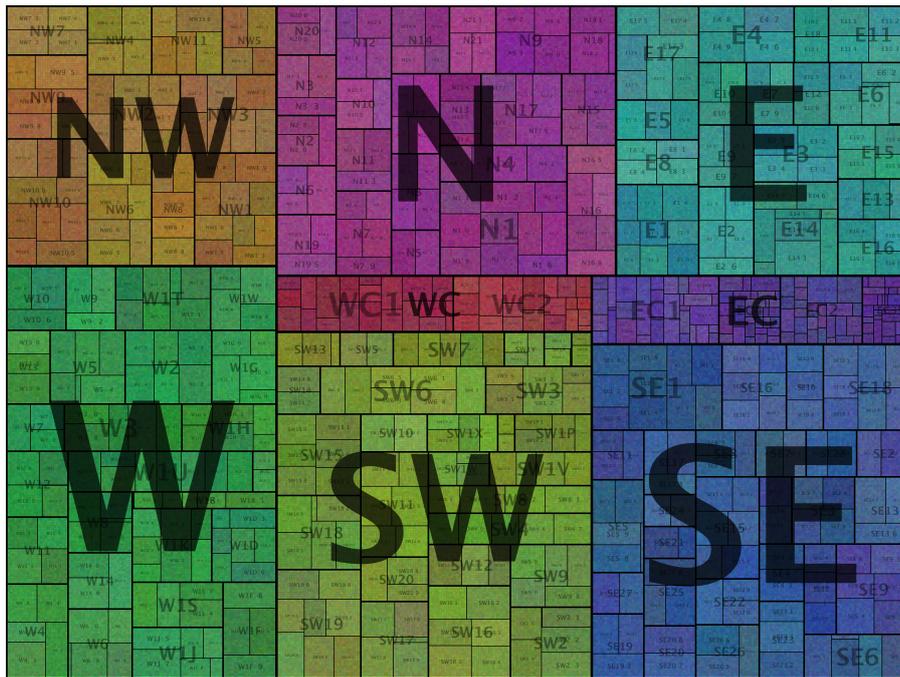
I began by showing how the *cdv* [1] and *panorama* [2] applications used linked graphics to relate spatial, semi-spatial and aspatial views of data.

I then drew attention to the different emphases in 'mapping' as we try to combine aspects of the geographic and aspects of the statistical or structural information in our graphics. Examples included aspatial bar charts, 1-dimensional spatial bar charts, choropleth alternatives, etc.

This can be depicted in a 'back of the envelope' sketch showing the trade-off between positional geography and statistical / structural information; see Figure 1. The objective of much of the design activity involved in 'putting data on the map' is to identify viable positions above the line depicting the trade-off. How can we show the geography that we need to achieve the tasks for which our 'maps' are designed whilst providing adequate statistical and / or structural information to support this activity?



■ **Figure 1** Positional Geography vs. Structural / Statistical Information. Ways in which maps and statistical graphics address this trade-off and ideas for crossing the gain line to show both effectively.



■ **Figure 2** A spatially ordered TreeMap of postcodes: London postal areas, districts, sectors and units are sized by the number of postal addresses.

Spatially ordered TreeMaps [11] were introduced as a means of adding geographic information to hierarchical representations of data and establishing a potentially useful position along or beyond this line; see Figure 2.

Distortions in the mapping between geographic space graphic space were discussed and ways of visualizing and addressing them introduced with a focus on Bernhard Jenny's work with MapAnalyst [5, 6, 7].

Difficulties in associating places with geographic spaces and the personal, emotional and uncertain nature of place were introduced [8] along with some ways of using community contributed information to gain some insights into these [4, 9].

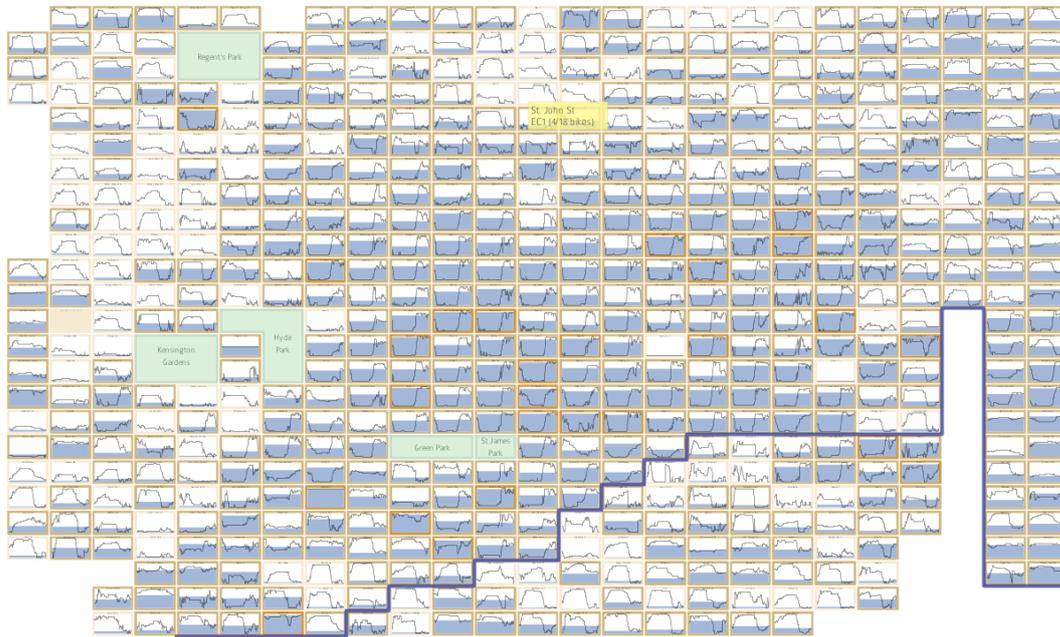
These perspectives lead to questions about what we are mapping. Many of these are open:

- How much geography do we need in our maps?
  - Which maps? Which spaces / places?
- Can we learn new geographies?
  - Who? Which geographies? How personalised are these?
- Can effective symbolism and interactive cartography help us with this learning process?
  - How? Who? When?!

I showed how at the giCentre we have been using hierarchical grids to show information about the London Cycle Hire scheme and bias in the London elections [13].

BikeGrid shows current and historical data relating to the stations comprising the London Cycle Hire Scheme in a semi-spatial grid view with geographic clues and animated transitions [3] to help orientate the map reader; see Figure 3.

BallotMaps show voting patterns in semi-spatial and aspatial views with animated transitions [3] to help orientate the map reader; see Figure 4.



■ **Figure 3** BikeGrid uses an semi-spatial view to aid geographic comparisons of bike station capacity over time.



■ **Figure 4** A BallotMap showing the boroughs of London arranged in a semi-spatial configuration. Each borough is equally sized and split and ordered by party, position of candidate on the ballot paper within party and electoral success and coloured by party and electoral success.

Using these semi-spatial geographies at multiple levels of the graphic hierarchy enables us to create maps within maps. This enables us to see the flows between origins and destinations through OD Maps [12].

Each of these views can be described using the hierarchical visualization expression language HiVE [10].

I concluded by contending that crossing the 'gain line' (see Figure 1) involves:

- Cognitively informed symbolism
- Task, user and data dependent solutions
- Effective (novel?) interaction
- Learning
- Ideation with data and technology
- Evaluation with informed users

To put data more usefully on more effective maps it seems to me that we need to know more about crossing the gain line:

- When and where this is achieved
- Computational and numeric approaches to support this
- Effects of learning
- Means of supporting learning
- With a variety of: phenomena; people; data sets; tasks
- In a range of application domains

## References

- 1 J. A. Dykes. Cartographic visualization: Exploratory spatial data analysis with local indicators of spatial association using Tcl/Tk and cdv. *The Statistician*, 47(3):485–497, 1998.
- 2 J. A. Dykes. An approach to virtual environments for fieldwork using linked geo-referenced panoramic imagery. *Computers, Environment and Urban Systems*, 24(2):127–152, 2000.
- 3 J. Heer and G. Robertson. Animated transitions in statistical data graphics. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1240–1247, 2007.
- 4 L. Hollenstein and R. Purves. Exploring place through user-generated content: Using flickr tags to describe city cores. *Journal of Spatial Information Science*, 1:21–48, 2010.
- 5 B. Jenny. Geometric distortion of schematic network maps. *Bulletin of the Society of Cartographers*, 40:15–18, 2006.
- 6 B. Jenny, T. Patterson, and L. Humi. Graphical design of world map projections. *International Journal of Geographical Information Science*, 24(11):1687–1702, 2010.
- 7 B. Jenny, A. Weber, and L. Hurni. Visualising the planimetric accuracy of historical maps with mapanalyst. *Cartographica*, 42(1):89–94, 2007.
- 8 D. R. Montello, M. F. Goodchild, J. Gottsegen, and P. Fohl. Where's downtown?: Behavioural methods for determining referents of vague spatial queries. *Spatial Cognition and Computation*, 3(2–3):185–204, 2003.
- 9 T. Rattenbury and M. Naaman. Methods for extracting place semantics from flickr tags. *ACM Transactions on the Web*, 3(1):1, 2009.
- 10 A. Slingsby, J. Dykes, and J. Wood. Configuring hierarchical layouts to address research questions. *IEEE Transactions on Visualization and Computer Graphics*, 16(6):977–984, 2009.
- 11 J. Wood and J. A. Dykes. Spatially ordered treemaps. *IEEE Transactions on Visualization and Computer Graphics*, 14(6):1348–1355, 2008.
- 12 J. Woods, J. A. Dykes, and A. D. Slingsby. Visualization of origins, destinations and flows with od maps. *The Cartographic Journal*, 47(2):117–129, 2010.
- 13 J. Woods, A. D. Slingsby, and J. A. Dykes. Visualizing the dynamics of london's bicycle hire scheme. *Cartographica*, 46(4):239–251, 2011.

### 3.2 Connecting the Dots: A Cartographic Perspective

*Sara Fabrikant (Universität Zürich, CH)*

License  Creative Commons BY-NC-ND 3.0 Unported license  
© Sara Fabrikant

For more than 5000 years cartographers have systematically transformed collected (commonly multivariate) spatial data into a two-, three- or four-dimensional visuo-spatial displays. This process is typically performed by applying scientific (i.e., systematic, transparent, and reproducible) cartographic design methods, as well as aesthetic expressivity. In recent years, cartographers and GIScientists have become involved in extending geographic concepts and cartographic design approaches to the depiction of massive, non-geographic data archives. These so-called information spaces also incorporate explicit geographic metaphors with the intention to create a graphic representation that is easier to comprehend for information seekers.

In this presentation I propose an empirically validated design framework for the construction of cartographically sound spatialized network displays based on spatial metaphors. As empirical studies on spatialized networks suggest, basic geographic principles and cartographically informed design guidelines enable information designers to not only construct conceptually robust and usable semantic network spaces, but also allow information seekers to more efficiently extract knowledge buried in large digital data archives.

### 3.3 From Data to Maps

*Stephen Kobourov (University of Arizona, Tucson, US)*

License  Creative Commons BY-NC-ND 3.0 Unported license  
© Stephen Kobourov

**Main reference** E. Gansner, Y. Hu, S. G. Kobourov, “Visualizing Graphs and Clusters as Maps,” IEEE Computer Graphics and Applications, vol. 30, no. 6, p. 54–66, 2010.

**URL** <http://www.cs.arizona.edu/~kobourov/PROJECTS/maps.html>

Information visualization can be invaluable in making sense out of large data sets. However, traditional visualization methods often fail to capture the underlying structural information, clustering, and neighborhoods. Our approach for visualizing relational data as a map provides a way to overcome some of the shortcomings with the help of the geographic map metaphor. While graphs, charts, and tables often require considerable effort to comprehend, a map representation is more intuitive, as most people are very familiar with maps and even enjoy carefully examining maps. The effectiveness of the map representation algorithm is illustrated with applications in recommendation systems for TV shows, movies, books, and music. Several interesting and challenging geometric and graph theoretic problems underlie this approach of creating maps from graphs. Specifically, we review recent progress on contact representations, rectilinear cartograms, and maximum differential coloring.

### 3.4 Graph Representations: Rectangles, Squares and Prescribed Area

*Stefan Felsner (TU Berlin, DE)*

**License**  Creative Commons BY-NC-ND 3.0 Unported license  
© Stefan Felsner

**Main reference** S. Felsner, “Rectangle and Square Representations of Planar Graphs,” to appear in “Thirty Essays in Geometric Graph Theory” edited by J. Pach.

**URL** <http://page.math.tu-berlin.de/~felsner/Paper/geom-rep.pdf>

In the first part of this survey we consider planar graphs that can be represented by a dissections of a rectangle into rectangles. In rectangular drawings the corners of the rectangles represent the vertices. The graph obtained by taking the rectangles as vertices and contacts as edges is the rectangular dual. In visibility graphs and segment contact graphs the vertices correspond to horizontal or to horizontal and vertical segments of the dissection. Special orientations of graphs turn out to be helpful when dealing with characterization and representation questions. Therefore, we look at orientations with prescribed degrees, bipolar orientations, separating decompositions, and transversal structures.

In the second part we ask for representations by a dissection of a rectangle into squares. We review results by Brooks et al. (1940), Kenyon (1998) and Schramm (1993), and discuss a technique of computing squarings via solutions of systems of linear equations.

In the third part we report on contact representations of planar graphs with regions of prescribed area (cartogram representations). In joint work with Alam, Biedl, Kaufmann, Kobourov and Ueckerdt, we have recently shown that planar triangulations admit such a representation with rectilinear polygons with 8 corners. The proof is based on Schnyder woods and the notion of area universal layouts.

### 3.5 Removing Data From The Map – How Information Reduction and Tailored Interaction Makes Maps Usable

*Falko Schmid (Universität Bremen, DE)*

**License**  Creative Commons BY-NC-ND 3.0 Unported license  
© Falko Schmid

**Joint work of** Schmid, Falko; Kuntzsch, Colin; Winter, Stephan; Kazerani, Aisan; Preissig, Benjamin;  
**Main reference** F. Schmid, C. Kuntzsch, S. Winter, A. Kazerani, B. Preissig, “Situating Local and Global Orientation in Mobile You-Are-Here Maps,” in *Proceedings of the Int’l Conf. on Mobile Human-Computer Interaction (MobileHCI’10)*, pp. 83–92. ACM, 2010.

**URL** <http://dx.doi.org/10.1145/1851600.1851617>

In my talk I gave a brief survey of topics related to Small Display Cartography. I covered topics such as transformations, off-screen visualization, map-based interaction, and schematization.

I showed that instead of adding data on the map, one has to consider the TEAR model of map creation: by considering the task  $T$ , the environment  $E$ , and the agent or user  $A$ , it is possible to create a tailored representation  $R$ .

### 3.6 Cognitive Tools (brief talk)

*Barbara Tversky (Stanford University, US)*

License  Creative Commons BY-NC-ND 3.0 Unported license  
© Barbara Tversky

Main reference B. Tversky, “Visualizations of thought,” *Topics in Cognitive Science*, 3 (2011) 499–535.

URL <http://dx.doi.org/10.1111/j.1756-8765.2010.01113.x>

Thought quickly overwhelms the mind and uses the world, as words, gestures, diagrams. Diagrams and sketches map elements and relations of ideas that are spatial or metaphorically spatial to elements and relations in space. Space carries meaning, e.g., vertical is more, good, power, health, strength. Horizontal is more neutral, but reading order provides direction. Elements carry meaning, concrete and abstract. Points are places, ideas; lines connect/relate them in road, social, brain, etc. networks. Gestures use analogous ways of expressing meanings. Both diagrams and gestures (actions) are instrumental to thought. The designed world—on shelves, buildings, streets, etc.—is a diagram created by actions and carries abstract meanings, 1–1, hierarchies, symmetries, embeddings.

## 4 Working Groups

### 4.1 A Note on Representing Data on Maps

*Mohammed Jawaherul Alam, Walter Didimo, Stefan Felsner, Ferran Hurtado, Marc van Kreveld, Giuseppe Liotta, Pavel Valtr, and Kevin Verbeek*

License  Creative Commons BY-NC-ND 3.0 Unported license  
© Mohammed Jawaherul Alam, Walter Didimo, Stefan Felsner, Ferran Hurtado, Marc van Kreveld, Giuseppe Liotta, Pavel Valtr, and Kevin Verbeek

Here we address problems related to cartography, where we are given a planar map and we want to redraw the map so that the face-areas in the map represent some prescribed data. Formally, let  $M$  be a planar map and  $F$  be the set of faces in  $M$ . A weight function  $w: F \rightarrow \mathbb{R}^+$  on the faces assigns a positive weight to the faces of  $M$ . We want to redraw  $M$  so that each face  $f \in F$  has area proportional to  $w(f)$ . We discussed two approaches for these problems.

In the first approach we first fix a “centroid” point inside each face of  $M$  and then we compute a weighted voronoi diagram or power diagram with respect to these centroid points in order to obtain a final map  $M'$ .

In the second approach, we consider the following problem: we are given a planar map  $M$  drawn inside a fixed outer boundary  $B$  and a weight function  $w$  on the faces of  $M$ . We want to redraw  $M$  in such a way the topology and the outer boundary remain fixed (vertices may move around the boundary) and the area of each face  $f$  is proportional to  $w(f)$ . In a restricted version of the problem the given map is convex and we also want to maintain the convexity of the regions.

In both the problems we have some positive as well as negative results. Furthermore, we pointed out open questions and future direction in both these problems. The details can be found at [http://www.cs.arizona.edu/~mjalam/dagstuhl\\_12261/group\\_report.pdf](http://www.cs.arizona.edu/~mjalam/dagstuhl_12261/group_report.pdf).

## 4.2 Computation of Wordles with Semantic Constraints

*Sara Irina Fabrikant, Stephen Kobourov, Anna Lubiw, Martin Nöllenburg, Yoshio Okamoto, Claudio Squarcella, Torsten Ueckerdt, and Alexander Wolff*

License  Creative Commons BY-NC-ND 3.0 Unported license  
 © Sara Irina Fabrikant, Stephen Kobourov, Anna Lubiw, Martin Nöllenburg, Yoshio Okamoto, Claudio Squarcella, Torsten Ueckerdt, and Alexander Wolff

A “wordle” is a visual representation for text data where prominent features of words (e.g., their relative frequency or importance) are shown with font size or color [5]. Wordles are generally regarded as beautiful visualizations for websites and blogs, and are usually employed to show the main contents or topics of a web page.

During the Dagstuhl Seminar “Putting Data on the Map” our working group focused on algorithms and methods to compute wordles with specific semantic and geometric constraints, such as the proximity between words that are related to each other. We strive to create wordles that transfer more information, such as a chronological order of the words or close semantic relations between certain words.

Wordles can be created with a number of online and offline services (see, e.g., [1]). An algorithm for the computation of appealing wordles is described in [4]: it features a heuristic for the computation of wordles and focuses on aesthetic criteria.

We proposed and studied various constraints for the computation of wordles. In all of our settings we apply some simplifications to the underlying problem. First of all, each word (or group of words) is considered as a rectangle. In a wordle no two rectangles overlap. Height and width are given, but sometimes we allow 90-degree rotations.

As a first step we imagined to draw wordles on a time axis, fixing each rectangle by time interval and packing all the rectangles using the minimum height (“time flow wordle”). The formulation is equivalent to a scheduling problem [2, 3], which is NP-hard. We independently found counterexamples for greedy algorithms, sketched a proof for weak NP-hardness in the general case and found algorithms for more constrained cases.

We also tackled the problem of realizing a wordle with semantic proximity between selected pairs of words, i.e., where two words must touch in the wordle if they are given as semantically close to each other. We compared this problem with known problems (e.g., rectangular dual of a planar graph) and focused on specific aspects like the presence of holes between words.

We restricted the problem to proximity relations that are represented by trees and sketched NP-hardness proofs for fixed, as well as, free embedding:

- We sketched an NP-hardness proof reducing from 3-PARTITION for the case when the embedding of the tree is not given.
- We sketched a weak NP-hardness proof reducing from SUBSETSUM for the case when the embedding of the tree is given.

We further restricted ourselves to hierarchical trees (directed trees with unique source) and required that every contact between two words must be horizontal with the hierarchically superordinated word being on top.

- We sketched an NP-hardness proof for the case when the embedding of the hierarchical tree is not given.
- We developed a polynomial time algorithm for the case when the embedding of the hierarchical tree is given.
- We extended the polynomial time algorithm to four hierarchies (corresponding to the four sides top, bottom, left and right) with a common root.

We thank Michael Kaufmann for discussions.

## References

- 1 Wordle – beautiful word clouds. <http://www.wordle.net>.
- 2 A. L. Buchsbaum, H. Karloff, C. Kenyon, N. Reingold, and M. Thorup. *OPT* versus *LOAD* in dynamic storage allocation. *SIAM Journal on Computing*, 33(3):632–646, 2004.
- 3 H. A. Kierstead. A polynomial time approximation algorithm for dynamic storage allocation. *Discrete Mathematics*, 87(2-3):231–237, 1991.
- 4 F. B. Viegas, M. Wattenberg, and J. Feinberg. Participatory visualization with Wordle. *IEEE Transactions on Visualization and Computer Graphics*, 15:1137–1144, 2009.
- 5 Wikipedia. Tag cloud. [http://en.wikipedia.org/wiki/Tag\\_cloud](http://en.wikipedia.org/wiki/Tag_cloud).

## 4.3 The Role of Animation and Interaction for Exploring Geotemporal Data

*Pierre Dragicevic, Sara Irina Fabrikant, Christophe Hurter, William Mackaness, and Barbara G. Tversky*

License  Creative Commons BY-NC-ND 3.0 Unported license  
© Pierre Dragicevic, Sara Irina Fabrikant, Christophe Hurter, William Mackaness, and Barbara G. Tversky

Visualizing geotemporal data is challenging. Known approaches include

1. static views (e.g., small multiples),
2. animated views (e.g., animated maps) and
3. interactive views (e.g., dynamic queries).

The relative effectiveness of these approaches – and of animation in particular – is subject of debate. Numerous studies have been carried out in various fields but are difficult to compare and generalize. In order to better understand which approach works when and why, we propose to place animation and interaction on two ends of a continuum where the locus of control is shared between the system, the end user, and possibly third-party users. We expect that a study comparing various loci of control (i.e., automatic animation vs. manual exploration vs. showing explorations performed by another user) will help us understand if some animations are ineffective due to their poor design or due to a lack of user control.

## 4.4 A Taxonomy of Temporal Visualizations Based on Space-Time Cube Operations

*Daniel Archambault, Sheelagh Carpendale, Pierre Dragicevic, Christophe Hurter, and Ying Yang*

License  Creative Commons BY-NC-ND 3.0 Unported license  
© Daniel Archambault, Sheelagh Carpendale, Pierre Dragicevic, Christophe Hurter, and Ying Yang

Although there are a number of surveys and textbooks on temporal data visualizations, there is still a need to clearly structure the large amount of previous work in the area. Taxonomies and typologies are essential for reflecting on existing techniques, designing new techniques, and teaching information visualization. A possibly useful classification is based on geometrical operations on a space-time cube. Although the space-time cube has been used as a visualization metaphor, it has never been used to classify visualization techniques. This framework can capture most known techniques for representing dynamic 2D data (e.g.,

geotemporal data), 2D and 3D visualization techniques alike. The goal of this transversal project is to write a taxonomy and/or a survey article on temporal visualizations based on space-time cube operations or a similar concept.

## 4.5 A Transdisciplinary Survey of Multiscale Rendering Techniques

*Sheelagh Carpendale, Pierre Dragicevic, Christoph Hurter, William Mackaness, and Monika Sester*

License  Creative Commons BY-NC-ND 3.0 Unported license  
© Sheelagh Carpendale, Pierre Dragicevic, Christoph Hurter, William Mackaness, and Monika Sester

The idea of displaying data differently depending on scale is an important research topic in many disciplines, including in cartography, HCI and infovis. However, each discipline uses a different terminology and (mostly) keeps ignoring the work from other disciplines. The goal of this project is to connect concept and ideas for a better transdisciplinary awareness and for a faster progress on the problem of multi scale rendering in general. The goal will be to

1. identify disciplines that use multi scale rendering (possibly inviting experts from fields not represented in the Dagstuhl seminar),
2. identify the conceptual overlaps between the terms used in each field (e.g., semantic zooming vs. visual aggregation vs. map generalization),
3. identify the techniques used in each field (e.g., reduction, etc.),
4. connect these techniques conceptually and algorithmically and
5. discuss the possibility of re-using techniques and algorithms across disciplines.

## 4.6 Putting Maps on a Curve

*David Auber, Sergi Cabello, Fabrizio Frati, Herman Haverkort, Martin Gronemann, Michael Kaufmann, and Ignaz Rutter*

License  Creative Commons BY-NC-ND 3.0 Unported license  
© David Auber, Sergi Cabello, Fabrizio Frati, Herman Haverkort, Martin Gronemann, Michael Kaufmann, and Ignaz Rutter

We study several problems that arise in the context of drawing clustered graphs with a map metaphor. We prove that it is NP-hard to find representations where all vertices that are adjacent in the input graph touch in the map. Due to this inherent difficulty, we suggest a heuristic approach based on alternating between a 1-dimensional and a 2-dimensional layout problem and study the complexity of the individual steps in this heuristic.

One way to draw a clustered graph  $G = (V, E)$  in a map metaphor, where the clustering is given as a tree  $T$  whose leaves are the vertices of  $G$ , is to choose an embedding of the tree, which determines a linear ordering  $\mathcal{L}$  of the vertices of the graph. The map  $M(G, \mathcal{L})$  is then generated by laying out the landmass for each leaf along a space-filling curve in 2-dimensional space according to the ordering  $\mathcal{L}$ . This has the advantage that the clusters given by the tree are automatically connected. However, it gives little guarantees on the relative positions of countries. In particular, adjacent vertices of the graph may be placed far apart in the map. We study the problem of finding linear orderings  $\mathcal{L}$  represented by the cluster tree  $T$  such that adjacent vertices are close in  $M(G, \mathcal{L})$ .

### 4.6.1 Our Work

We study the complexity of finding a linear ordering  $\mathcal{L}$  represented by  $T$  that optimizes the proximity of vertices that are adjacent in  $G$ . We showed that it is NP-complete to decide whether an ordering exists such that landmasses of adjacent vertices touch each other. Hence it is not even possible to efficiently approximate the distances.

On the theoretical side, we consider the more restricted case, where we just put the ordering  $\mathcal{L}$  on the real line, trying to minimize proximity of adjacent vertices. Two natural objective functions are to either measure the total length of the edges, when putting points at unit distance on the real line, or to measure the number of crossings when adding the edges of  $G$  to the embedding of  $T$  without crossing  $T$ . The first problem is called `LinearArrangement` and is well-studied. It is generally NP-hard, admits a  $O(\log n)$ -approximation and can be solved by dynamic programming for balanced bounded-degree trees [1]. We denote the second problem of minimizing the number of crossings by `MinCrossing`. It is at least as hard as the closely related tanglegram problem [2]. However, the machinery showing that `CrossingNumber` is FPT with respect to the number of crossings [3] can be applied to `MinCrossing` as well.

On the practical side, to circumvent the inherent difficulty to preserve distances on the 2D space-filling curve, we suggest a more heuristic approach. The idea is to alternate between the layout along a 2D space-filling curve (optimizing proximity, possibly violating representability by  $T$ ) and a more tractable 1D ordering problem that enforces representability by  $T$ .

More precisely, we start out with an initial ordering  $\mathcal{L}$  represented by  $T$  and consider the corresponding map  $M(G, \mathcal{L})$  and iterate the following steps.

1. Optimize the 2D layout using a force-directed approach with attracting forces for landmasses that should be close. Call the resulting map  $M'$ .
2. Traverse the space-filling curve and collect the landmasses as they occur along the curve. This yields a linear ordering  $\mathcal{L}'$  of the landmasses and hence of the vertices of  $G$ .
3. Compute an ordering  $\mathcal{L}''$  that is represented by  $T$  and as similar as possible to  $\mathcal{L}'$ .
4. Start over with step (1) and map  $M(G, \mathcal{L}'')$ .

The main issue, aside from choosing suitable forces for step (1), is the implementation of step (3). Inspired by the linear ordering problems mentioned in the introduction, we suggest three ways to measure the similarity of two linear orderings  $\mathcal{L}$  and  $\mathcal{L}'$ . Namely, either the sum of all displacements, the maximum displacement and the number of transpositions. We call the corresponding problems `TotalDisplacement`, `MinMaxDisplacement` and `MinTranspositions`. We showed that using approaches similar to the one for `LinearArrangement` [1], one can obtain efficient algorithms for all three problems on balanced bounded-degree trees.

### References

- 1 R. Bar-Yehuda, G. Even, J. Feldman, and J. Naor. Computing an optimal orientation of a balanced decomposition tree for linear arrangement problems. *Journal of Graph Algorithms and Applications*, 5(4):133–160, 2001.
- 2 K. Buchin, M. Buchin, J. Byrka, M. Nöllenburg, Y. Okamoto, R. I. Silveira, and A. Wolff. Drawing (complete) binary tanglegrams — hardness, approximation, fixed-parameter tractability. *Algorithmica*, 62(1–2):309–332, 2012.
- 3 K. Kawarabayashi and B. Reed. Computing crossing number in linear time. In *Proceedings of the 39th Annual ACM Symposium on Theory of Computing (STOC'07)*, pages 382–390, 2007.

## 4.7 A Generic Work-Flow for Making Maps out of Graphs

*Robert P. Biuk-Aghai, Joe Fowler, Jan-Hendrik Hawnert, Petra Mutzel, Frank van Ham, and Marc van Krefeld*

License  Creative Commons BY-NC-ND 3.0 Unported license  
 © Robert P. Biuk-Aghai, Joe Fowler, Jan-Hendrik Hawnert, Petra Mutzel, Frank van Ham, and Marc van Krefeld

GMap [2] is a new approach to visualizing graphs and clusters as maps. Depending on the application, one may have different goals for an alternative map framework to GMap. No one framework is necessarily ideal, and often there will be trade offs to consider. Our proposed framework primarily tries to address the needs of the Wikipedia graph, since GMap (as currently implemented in GraphViz as `gvmap`) does not meet several aesthetic criteria including

- *fixed areas of each country*—based on the logarithm of the size of the topic,
- *logical adjacencies and proximities*—very similar topics should be adjacent or nearby, while unrelated topics need to be sufficiently separated, and
- *contiguous boundaries*—i.e., non-fragmented countries.

The well-defined real-world graph whose map representation (once fully realized to meet its constraints) should have broad appeal, validating our approach (if successful).

### 4.7.1 Approach and Aims

We focused on producing a recursive rectangular cartogram representation that could potentially meet some specific guarantees, such as area, width, and/or aspect-ratio requirements. We proposed a framework that addresses several of GMap’s potential shortcomings (depending on the application), namely,

- *fragmentation,*
- *highly irregular boundaries,* and
- *no control over relative positioning.*

All three shortcomings are consequences of using force-directed/MDS placement with standard node overlap removal to create Voronoi-based maps.

### 4.7.2 Identified Issues with the GMap Framework

We considered a variety of alternatives including removing edges with high betweenness to unroll a graph once a force-directed algorithm is applied. However, unless we used some type of tree-map or cartogram approach to then take that embedding to produce a map, we were left with four problems inherent to the GMap framework:

1. *How does one modify the node-overlap removal algorithm to respect clustering and relative placement?*  
 We want to eliminate overlapping labels that can produce useless fragmentation.
2. *How does one reposition or eliminate points so that the Voronoi diagram has less irregular boundaries?*  
 Repositioning points may violate country adjacencies. Having too few points can produce blocky maps, while having too many can yield highly jagged boundaries.

3. *How does one achieve guarantees regarding area or aspect ratio?*  
Current method is to specify font size and shore line depth that has weak bounds.
4. *How does one best ensure having the desired adjacencies and non-adjacencies?*  
Resulting Voronoi cells are highly dependent on the outcome of node overlap removal algorithm where adjacencies can be drastically altered.

### 4.7.3 Proposed Generic Graph-to-Map Framework

Rather than directly addressing these questions, we opted to employ a non-force directed method using recursive rectangular cartograms, with SPQR-trees used to efficiently extract a suitable planar subgraph (with no separating triangles) and find an embedding to reduce overall edge lengths. This circumvents the four problems above, giving an alternative to the GMap force-directed/MDS paradigm. More specifically, our framework takes a weighted hierarchical clustered graph as input and has the following steps:

1. Extract a planar subgraph with no separating triangles prioritizing edges with low betweenness and greater weights.
2. Using SPQR-trees, greedily determine an embedding that reduces lengths of edges with higher priority.
3. Fully triangulate graph by inserting dummy nodes and edges without creating separating triangles.
4. Given that the graph is now 4-connected, obtain a rectangular dual.
5. Solve for a proportional representation using iterative linear programming for desired areas, aspect ratios and proximity.
6. Recurse on each rectangular region representing a non-trivial cluster.

### 4.7.4 Open Problems with Possible Approaches

In developing this framework, we were confronted with the following several open problems for which we formulated tentative high-level approaches.

1. *How can SPQR-trees be efficiently used to extract a maximal planar subgraph while avoiding separating triangles?*  
Use an SPQR-tree approach to find a maximal planar subgraph (in terms of  $\frac{\text{betweenness}}{\text{edge weight}}$ ), and then attempt to swap edges and/or expand nodes locally. Given that separating triangles are uncommon, in practice one may always eliminate these. Ideally, one would wish to obtain an efficient method to guarantee their removal.
2. *How do you then use the same SPQR-tree data structure to search overall all possible combinatorial embeddings to reduce edge crossings of heavily weighted edges?*  
This problem is akin to crossing minimization, where one can also use various types of integer linear programming techniques to optimally solve. Also, there may be reasonable approximation algorithms that may be suitable.
3. *What is the best strategy of inserting dummy nodes and weighting their incident edges so as to not introduce separating triangles?*  
Insert chains of length 2 in lieu of edges to fully triangulate the graph, where then the dummy node would also be adjacent to the two neighbors of its endpoints.  
Weight the dummy node according to the dissimilarity between endpoints so as to produce a lake or a river whose width would be dependent on the similarity of opposing adjacent countries.

4. *How can we find a sufficiently proportional rectangular representation (permitting recursive subdivision)?*

We can solve for a desirable proportional representation in terms of area, aspect ratios, and/or adjacencies/proximities with given error bounds as follows:

- a. Fix horizontal segments while allowing vertical segments to shift left or right.
- b. Solve associated linear program for optimal  $x$ -coordinates of vertical segments.
- c. Fix vertical segments while allowing the horizontal segments to shift up or down.
- d. Solve associated linear program for optimal  $y$ -coordinates of horizontal segments.
- e. Repeat steps 1 to 4 until solutions of steps 2 and 4 converge.

If the solution space is convex, it may be possible to restate this problem in terms of quadratic equations with linear constraints permitting a polynomial-time solution (not requiring this iterative back and forth) for a given error bound. Typically, though, the solution space is not convex, or even connected.

Furthermore, a 4-connected graph can have many alternate rectangular representations. For the best result, we may have to perform the iterative shift approach for each of them. Optimizing criteria over rectangular representations was considered by Buchin et al. [1].

5. *What guarantees can we ensure for the final map representation?*

For this to be viable alternative, we should demonstrate that this method has advantages, preferably with guaranteed constraints in terms of area, aspect ratio, proximities, relative positioning, and/or adjacencies within some fixed error, over the current GMap framework. The advantages should be sufficiently compelling to adopt, compensating for having a more complex implementation than GMap.

Our method (as proposed) requires using an involved SPQR-tree data structure in conjunction with linear programming techniques, which may not be efficient for large instances. This is in contrast to using a simple force-directed algorithm with Voronoi overlay in the case of GMap, which in its most efficient implementation runs in time proportional to  $O(n \log n)$  using a force-directed algorithm using a Barnes-Hut approach with quadtrees to approximate distant force.

#### 4.7.5 Conclusion

If we can obtain a reasonable, time-efficient implementation that can create reasonably proportional maps for most real-world weighted graphs, then we can clearly demonstrate the viability of this proposed framework. Moreover, in developing this framework, we have the opportunity to address several open theoretical problems, whose solutions have practical applications. We can further strengthen our results by comparing such an implementation against existing GMap implementations, and evaluate how well our proposed framework fares both in terms of visual aesthetics and desired constraints.

#### References

- 1 K. Buchin, B. Speckmann, and S. Verdonschot. Optimizing regular edge labelings. In U. Brandes and S. Cornelsen, editors, *Proceedings of the 18th International Conference on Graph Drawing (GD'10)*, volume 6502 of *Lecture Notes in Computer Science*, pages 117–128, Berlin, Heidelberg, 2011. Springer-Verlag.
- 2 E. R. Gansner, Y. Hu, and S. G. Kobourov. Gmap: Visualizing graphs and clusters as maps. In *Proceedings of the IEEE Pacific Visualization Symposium (PacificVis'10)*, pages 201–208, 2010.

## 4.8 Area Labeling

*Jan-Henrik Haunert, Bernhard Jenny, Philipp Kindermann, Sergey Pupyrev, and Falko Schmid*

License  Creative Commons BY-NC-ND 3.0 Unported license  
© Jan-Henrik Haunert, Bernhard Jenny, Philipp Kindermann, Sergey Pupyrev, and Falko Schmid

An instance to the area-labeling problem is given with a geographic region, its name, and a set of obstacles.

- A geographic region may be given as a polygon. Additionally, we consider regions with uncertain boundaries (e.g., as proposed by Jones et al. [5]). A region  $R$  with an uncertain boundary is represented with a function  $f_R: \mathbb{R}^2 \rightarrow \mathbb{R}$  that assigns a membership value to every point in the plane;  $f_R(p)$  measures to what degree point  $p$  belongs to  $R$ .
- In our basic problem, each name consists of a single word of multiple letters.
- An obstacle is an object (e.g., a road or a house) that already lies in the map.

A solution should satisfy the following criteria.

- The label should follow a support line  $\ell$  of low curvature that approximates the given region  $R$ . More precisely, if  $R$  is a polygon,  $\ell$  should approximate the medial axis of the region. If  $R$  is given with a membership function  $f_R$ ,  $\ell$  should cover a preferably large set of points of high membership values.
- The label should preferably be centered on the support line  $\ell$ , that is, the distance between the center point  $c$  of  $\ell$  and the center  $\gamma$  of the label should be small.
- Each two consecutive letters must have the same distance. This distance is a variable that we denote by  $\delta$ .
- The distance  $\delta$  should be close to  $d$ , which is the distance between two consecutive letters if they are evenly spread over the whole extent of  $\ell$ .
- A letter should not be placed on an obstacle; this may be a hard constraint, but we also consider a variant where we aim to minimize a cost function that depends on how many (and which) obstacles become occluded by a letter.

### 4.8.1 Related Work

The problem of labeling a geographic region has been discussed by multiple authors [1, 2, 3, 4].

Barrault [1] has developed an automatic method that first selects a discrete set of candidates for the support line of the label. Then, for each candidate line  $\ell$ , an optimal label position is searched by explicitly testing a large number of values for three variables, namely the offset of the first letter from the begin of  $\ell$ , the distance between two consecutive letters, and the distance between two consecutive words. Each solution is assessed based on multiple criteria, also considering overlaps between letters and other map objects. The overall best solution is returned.

While the method of Barrault only tests a finite number of solutions, we propose an approach by continuous optimization that for a given support line optimizes over all possible solutions. Furthermore, Barrault has not considered geographic regions with uncertain boundaries and his method for selecting candidates for the support line of the label, which is based on a skeleton of the input polygon, cannot easily be generalized to regions with uncertain boundaries. We therefore propose a (heuristic) force-directed method that tries to move a randomly sampled initial support line to favorable points while keeping its curvature low.

### 4.8.2 Outline of the Algorithm

1. For a given object and a label  $L$  of  $n$  letters, compute a good set of candidates for the support line, i.e., a set of lines on which  $L$  may be placed.
2. For each candidate line  $\ell$  from step 1, find all  $\mathcal{O}(n^2m^2)$  solutions (with  $m$  being the number of obstacles intersected by  $\ell$ ) in which at least two letters touch an obstacle.
3. For each solution from step 2, optimally adjust letter positions by quadratic programming.
4. Return the best solution found in step 3.

For the special case that the region is a polygon and the text is required to be horizontal, we have an efficient sweep-line algorithm that finds all interesting support lines. This replaces our heuristic force-directed method.

### References

- 1 M. Barrault. A methodology for placement and evaluation of area map labels. *Computers, Environment and Urban Systems*, 25(1):33–52, 2001.
- 2 H. Freeman. Automated cartographic text placement. *Pattern Recognition Letters*, 26(3):287–297, 2005.
- 3 E. Imhof. Positioning names on maps. *Cartography and Geographic Information Science*, pages 128–144, Oct. 1975.
- 4 C. B. Jones, R. S. Purves, P. D. Clough, and H. Joho. Modelling vague places with knowledge from the web. *International Journal of Geographical Information Science*, 22(10):1045–1065, 2008.
- 5 I. Pinto and H. Freeman. The feedback approach to cartographic areal text placement. In *Proceedings of the 6th International Workshop on Advances in Structural and Syntactical Pattern Recognition*, SSPR '96, pages 341–350, London, UK, 1996. Springer-Verlag.

## 4.9 A Metro Map Problem

Maxwell J. Roberts, Andreas Reimer, Yoshio Okamoto, and Therese Biedl

**License**  Creative Commons BY-NC-ND 3.0 Unported license  
© Maxwell J. Roberts, Andreas Reimer, Yoshio Okamoto, and Therese Biedl

This problem concerns optimizing a linear schematic of a rail network by identifying the angles most congruent with the structure of the network. By choosing the most appropriate angles, trajectory corrections to the lines can be minimized, and hence these can be presented more simply. However, the use of multiple angles is potentially a source of complexity in its own right. Hence the point of view of automation of schematic map design, is it possible to identify the minimum number of angles necessary in order to provide well-optimized line trajectories, maintaining network topology and without unacceptable levels of topographical distortion.

The group tackled this problem bottom-up, trying to identify the stages that a human designer would go through in schematizing a network and answering the question, trying to provide the straightest lines with the minimum angles. In doing so, a task analysis was created, identifying which specific stages can potentially be addresses using current, or slightly modified computer algorithms, versus stages for which new computer algorithms will need to be developed.

- Identify priority region of the network (typically the Central Business District).
- Apply suitable transformation (to enlarge the center in relation to suburbs).

- Identify difficult regions (dense stations), lines with complex versus simple trajectories.
- Identify local traditions (e.g., key landmarks) and focal points.
- Apply edge straightening routines to line trajectories.
- Straighten edges further in conjunction with adjusting trajectories in line with coherence criteria (parallel lines etc.).
- Identify a grid, rotate approximately to it.
- Snap remaining angles to grid.

In the process of outlining these stages, the following observations were made:

- Identifying and applying coherence principles almost makes the map self-organizing.
- Parallel lines may be more important than precisely intervalled angles.
- The Importance of parallel lines may proportional to edge length, and the number of edges may also be a factor (e.g., it is more important that four long lines are parallel than two short ones).
- Automation may be easier to implement if human design techniques are followed, perhaps permitting complex networks to be schematized for the first time.
- Vienna is an octilinear city, but not Barcelona.
- A very abstract distorting schematic might be acceptable if clear benefits of simplification.

Overall, the following 'next steps' were identified:

- Perform similar analyses for other cities of similar complexity
- If there is no clear solution, revisit edge simplification and coherence transformations
- Identify hierarchy of design criteria
- Develop algorithms to maximize parallel lines on graphs
- Develop algorithms to snap to grid, not a trivial adjustment even for a few degrees

## 4.10 Showing Dissimilarity Data on Cartographic Maps

*Sheelagh Carpendale, Tim Dwyer, Jason Dykes, Nathalie Henry-Riche, Arlind Nocaj, and Bettina Speckmann*

License  Creative Commons BY-NC-ND 3.0 Unported license  
© Sheelagh Carpendale, Tim Dwyer, Jason Dykes, Nathalie Henry-Riche, Arlind Nocaj, and Bettina Speckmann

Given a dataset containing pairwise dissimilarities for points on a map, we wish to explore various ways to distort or overlay the map in order to best show those dissimilarities in the context of the geography. Our workshop discussions on this topic turned into three separate lines of inquiry that we will pursue further in the coming months:

1. Inspired by the “Wrap/Distort/Cut” project of Van Wijk [1], we take the original geographic placement and map to the surfaces of sphere, torus, etc. Using, for example, SOM, we distort on surface in order to minimize dissimilarity error. We find cut paths through the surface in order to:
  - minimize dissimilarity error in final projection.
  - split geographically uninteresting places (e.g., deserts, oceans)
2. A framework for combining overlays and simple global distortion for showing dissimilarity on maps. We want to experiment with trading-off distortion of the map with overlays (such as “ridges” or “bridges”) between points to show the degree of their similarity in the underlying dataset.

3. There is scope for the development of an entire taxonomy, mapping out the design space around this problem and hopefully leading to a reasoned exploration of possible alternatives for displaying dissimilarity data on the map.

### References

- 1 J. J. van Wijk. Unfolding the Earth: Myriahedral projections. *The Cartographic Journal*, 45(1):32–42, 2008.

## 4.11 Evaluating Bundling Quality Using Image-Based Techniques

*Stephan Diehl, Seok-Hee Hong, Quan Nguyen, Monika Sester, and Alexandru Telea*

License  Creative Commons BY-NC-ND 3.0 Unported license  
© Stephan Diehl, Seok-Hee Hong, Quan Nguyen, Monika Sester, and Alexandru Telea

In the last few years, over a dozen algorithms were proposed to simplify complex graph drawings by *edge bundling*. Edge bundling, at a high level, can be seen as a technique that trades off overdraw for clutter, *i.e.*, generates graph drawings where, on the one hand, several edges share the same screen space (thus, the overdraw), but on the other hand one obtains more empty space between groups of edges, or bundles (thus, the clutter reduction). Such drawings thus can successfully convey the coarse structure of a graph, *e.g.*, the main communication paths in a network.

Although many techniques exist that produce bundled edge drawings with different looks, it is not clear what are the properties of a *good* bundled edge drawing. Moreover, although significant work exists on the aesthetics and desirable properties of straight-line graph drawings, such properties cannot be directly taken over to bundled edge drawings.

One promising new avenue would be to use *image-based techniques* to analyze the quality of a bundled edge drawing. For this, we analyze the final image (in which bundled edges have been drawn) rather than the geometric (polyline) representation of the bundled graph. One of the advantages of this approach is that we can directly use many existing image-processing techniques, such as filtering, segmentation, and edge detection.

Relevant questions identified in this working group are as follows:

1. *What is a bundled edge drawing?*

We need to answer this question before we are able to propose desirable properties thereof as well as quality criteria. We propose an analogy between a bundled graph and an image based on the following elements (in the table below, the left column shows data structures and algorithms for bundled graphs; the right column shows image-processing operations).

Edge bundling drawing	Image properties
an edge / a bundle	a pixel / an image segment
the bundling process	image segmentation
bundle smoothing	image blurring
edge curvatures	segment contour smoothness

2. *How to model the quality of an edge bundling?*

Following the above graph-image analogy, we can now encode various bundled-graph quality aspects into the corresponding image characteristics:

Edge bundling quality factors	Image properties
smoothness and continuity	texture richness
ink/whitespace ratio	image histogram sharpness
bundling strength	image-edge histogram sharpness
disorder/entropy	image histogram flatness
pattern segregation	image histogram inter-peak distance

3. *How to quantify the quality of an edge bundling?*

Besides modeling the quality of a bundled graph by image properties, we need to *quantify* it. For this, we propose a simple but flexible quality metric  $B$ .  $B$  should have two ingredients: a distance-metric term (encoding the domain knowledge and user requirements to what a good bundling is); and a distribution sharpness term (encoding how well the bundles are segregated in the image according to the earlier-mentioned distance metric). A simple instance of  $B$  would be the ratio between the *bundling strength*, *e.g.* the ink/non-ink ratio in the image, to the image-space difference between the original and bundled drawings.

4. *Optimizing an edge bundling*

Now that we have a model for bundling quality, we can propose techniques to exploit this metric to optimize an edge bundling. Two promising directions are: (a) showing the bundling error  $B$  of a given graph, to let users understand how much the drawing deviates from the “ideal” one; and (b) Computing an optimal graph bundling by using (non)parametric optimization techniques for increasing the value of  $B$ .

## Participants

- Muhammad Jawaherul Alam  
Univ. of Arizona – Tucson, US
- Daniel Archambault  
University College Dublin, IE
- David Auber  
Université Bordeaux, FR
- Peter Bak  
IBM – Haifa, IL
- Therese Biedl  
University of Waterloo, CA
- Robert P. Biuk-Aghai  
University of Macau, CN
- Sergio Cabello  
University of Ljubljana, SI
- Sheelagh Carpendale  
University of Calgary, CA
- Walter Didimo  
University of Perugia, IT
- Stephan Diehl  
Universität Trier, DE
- Pierre Dragicevic  
Université Paris Sud, FR
- Tim Dwyer  
Microsoft – Redmond, US
- Jason Dykes  
City University – London, GB
- Sara Irina Fabrikant  
Universität Zürich, CH
- Stefan Felsner  
TU Berlin, DE
- Martin Fink  
Universität Würzburg, DE
- Joe Fowler  
Univ. of Arizona – Tucson, US
- Fabrizio Frati  
Università di Roma III, IT
- Martin Gronemann  
Universität Köln, DE
- Jan-Henrik Haunert  
Universität Würzburg, DE
- Herman J. Haverkort  
TU Eindhoven, NL
- Nathalie Henry Riche  
Microsoft – Redmond, US
- Seok-Hee Hong  
The University of Sydney, AU
- Ferran Hurtado  
UPC – Barcelona, ES
- Christophe Hurter  
ENAC – Toulouse, FR
- Bernhard Jenny  
Oregon State University –  
Corvallis, US
- Michael Kaufmann  
Universität Tübingen, DE
- Philipp Kindermann  
Universität Würzburg, DE
- Stephen Kobourov  
Univ. of Arizona – Tucson, US
- Jan Kratochvil  
Charles University – Prague, CZ
- Giuseppe Liotta  
University of Perugia, IT
- Anna Lubiw  
University of Waterloo, CA
- William A. Mackaness  
University of Edinburgh, GB
- Petra Mutzel  
TU Dortmund, DE
- Quan Nguyen  
The University of Sydney, AU
- Arlind Nocaj  
Universität Konstanz, DE
- Martin Nöllenburg  
KIT – Karlsruhe Institute of  
Technology, DE
- Yoshio Okamoto  
Univ. of Electro-Communications  
– Tokyo, JP
- Sergey Pupyrev  
Ural State Univ. –  
Ekaterinburg, RU
- Andreas Reimer  
GFZ Potsdam, DE
- Maxwell J. Roberts  
University of Essex, GB
- Ignaz Rutter  
KIT – Karlsruhe Institute of  
Technology, DE
- Falko Schmid  
Universität Bremen, DE
- Monika Sester  
Leibniz Univ. Hannover, DE
- Bettina Speckmann  
TU Eindhoven, NL
- Claudio Squarcella  
Università di Roma III, IT
- Alexandru C. Telea  
University of Groningen, NL
- Barbara Tversky  
Stanford University, US
- Torsten Ueckerdt  
Charles University – Prague, CZ
- Pavel Valtr  
Charles University – Prague, CZ
- Frank van Ham  
IBM Software Group, NL
- Marc van Kreveld  
Utrecht University, NL
- Kevin Verbeek  
TU Eindhoven, NL
- Dorothea Wagner  
KIT – Karlsruhe Institute of  
Technology, DE
- Alexander Wolff  
Universität Würzburg, DE
- Jing Yang  
University of North Carolina at  
Charlotte, US

