Betweenness Centrality in Spatial Networks: A Spatially Normalised Approach

Christian Werner
Department of Geoinformatics, University of Salzburg, Austria

Martin Loidl
Department of Geoinformatics, University of Salzburg, Austria

Abstract

Centrality metrics are essential to network analysis. They reveal important morphological properties of networks, indicating e.g. node or edge importance. Applications are manifold, ranging from biology to transport planning. However, while being commonly applied in spatial contexts such as urban analytics, the implications of the spatial configuration of network elements on these metrics are widely neglected. As a consequence, a systematic bias is introduced into spatial network analyses. When applied to real-world problems, unintended side effects and wrong conclusions might be the result. In this paper, we assess the impact of node density on betweenness centrality. Furthermore, we propose a method for computing spatially normalised betweenness centrality. We apply it to a theoretical case as well as real-world transport networks. Results show that spatial normalisation mitigates the prevalent bias of node density.

2012 ACM Subject Classification Theory of computation → Theory and algorithms for application domains; Mathematics of computing → Graph algorithms; Mathematics of computing → Network flows; Information systems → Geographic information systems

Keywords and phrases spatial network analysis, edge betweenness centrality, flow estimation, SIBC, spatial interaction, spatial centrality, urban analytics

Digital Object Identifier 10.4230/LIPIcs.GIScience.2023.83

Category Short Paper

Supplementary Material *Other (Code and examples)*: `https://doi.org/10.5281/zenodo.8125632`

1 Introduction

Betweenness centrality is a key metric for assessing node and edge importance in networks. It is based on computing the share of shortest paths that pass each edge or node in relation to the total number of paths in a network. Thereby it reveals the relative importance of edges or nodes for enabling interaction within the network. While centrality metrics can generally be applied to spatial networks, many complex effects occur that are still to be fully understood [2]. As direct consequence of the definition of betweenness centrality, the number of nodes, their location, and their morphological embedment within the network determine centrality. The key aspect to be questioned within spatial applications is the assumption that each pair of nodes has equal influence on centrality. This characteristic implies that the spatial density of nodes strongly influences betweenness centrality.

State of the Art

Due to the generic network science origin of centrality metrics, their focus lies on topological rather than spatial properties of networks. However, important steps for integrating spatial aspects into centrality concepts have been accomplished e.g. by considering the spatial length of edges and paths in betweenness centrality. Further research assessed how different forms of spatial networks influence centrality and how such networks can be characterised through the use of centrality metrics [2]. Various studies applied betweenness centrality for assessing spatial transport networks - for Indian railways[6], Paris and London transport[4] or urban road networks in Germany[5]. All these examples have in common that they do not consider the impact of spatial configuration on centrality metrics, thus neglecting its potential bias.

Application-driven research proposed domain-specific concepts for weighting origin-destination pairs in computation of centrality, which may mitigate bias of spatial configuration. In urban analytics and transport modelling, weighting based on flow estimation is common. Spatial interaction between origins and destinations is modelled, resulting in an estimate for travel demand. Despite the long history of such methods, adequate modelling is highly complex and has been found not to meet real-world observed patterns in many cases.

Research Gap

While numerous studies applied betweenness centrality to spatial networks, effects of the spatial configuration of nodes on centrality have not been regarded systematically. Furthermore, we identified the lack of a simplistic null model of betweenness centrality for applications in spatial networks that avoids introducing complex (behavioural) models.

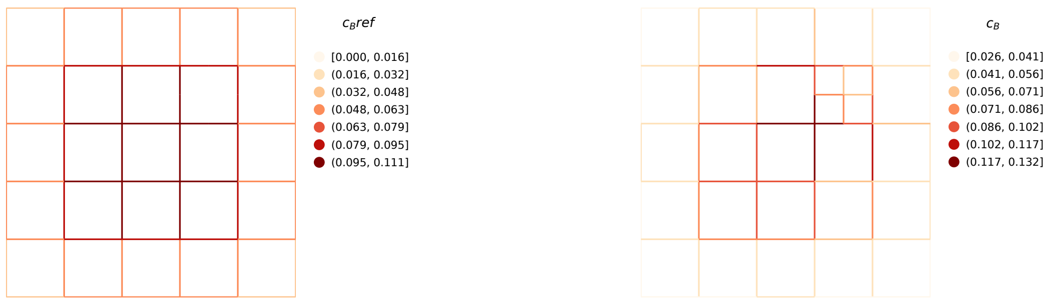
To fill this gap, we first assess the problem in more detail and then provide a method to compensate the influence of spatial node density. Motivated from the application domain of urban analytics and mobility, we focus on edge betweenness centrality as known key measure from which node betweenness centrality may easily be derived [2]. Where helpful, we motivate our theoretical considerations using examples of real-world transport networks.

2 Method

The Problem Illustrated

To illustrate the impact of node density on edge betweenness centrality (c_B), we use a simplistic reference case. In a network constructed as a regular grid, c_B is known to be highest in the spatial centre, as shown in figure 1 a). If we subdivide one grid cell by adding an additional node per edge and one node at the cell centre, we observe a shift in high c_B towards the newly subdivided area, visualised in figure 1 b). One may think of this as a city block to which access paths for pedestrians have been added. While the overall structure of this virtual residential area remained the same (i.e. no buildings have been added or removed), centrality shifted significantly. This can be explained by the fundamental definition of c_B. As we introduce new nodes - in the given case five nodes are added to a cell originally consisting of four nodes - each of these new nodes introduces an equally important origin and destination for all shortest paths computation. As a consequence, the influence of paths from and to this cell increases in relation to all paths within the given network.

To quantify this gain in influence, we can calculate the change in contribution of paths from and to the given cell relative to all paths within the network. Following the definition of c_B (see equation 1), it is more precisely the number of origin-destination relations that start or end within the given cell that we are interested in. As known from normalisation of c_B, the total number of origin-destination relations in a directed graph consisting of n nodes is $n(n-1)$. As one single node has the role as origin as well as destination for o-d relations to all other nodes, it contributes $2(n-1)$ o-d relations. Consequently, we can express the relative contribution of one node to all possible relations as $\frac{2(n-1)}{n(n-1)}$. In a more generic form, we can quantify the contributed relations of i nodes to the network beyond the given cell as $\frac{2i(n-1)-i(i-1)}{n(n-1)}$.



(a) Regular grid network.

(b) Same network with subdivision added.

Figure 1 The problem illustrated: Influence of node density on centrality.

Considering the case of adding subdivision to a regular grid, we can assess the contribution I of nodes forming a given cell C and the subdivided cell C' as follows:

$$I_C = \frac{2a(n-1) - a(a-1)}{n(n-1)} \text{ and } I_{C'} = \frac{2(m+a)(n+m-1) - (m+a)(m+a-1)}{(n+m)(n+m-1)}$$

where a denotes the number of nodes originally constituting the cell and m denotes the number of nodes added through subdivision.

The relative increase in influence of cell C can be assessed as $\frac{I_{C'}}{I_C}$ which for $a \ll n$ and $m \ll n$ can be approximated as: $n^2(a+m)\frac{-2n}{-2an^3} = \frac{a+m}{a}$. Thus, the influence of the given cell on centrality increases approximately proportional to the increase in node count for the same area for commonly large networks.

For our minimalistic example with $n = 36; a = 4; m = 5$, the cell's influence increases by $\frac{81}{205}/\frac{67}{315} \approx 1.858$. Due to the small network size and relatively high a and m, this value does not reach the approximate value for large networks of $\frac{4+5}{4} = 2.25$.

To summarise this section, we were able to show that the influence of spatial node density on betweenness centrality can be assumed to be proportional for commonly large networks. Consequently, spatial variation in node density has significant impact on betweenness centrality. While this might be intended in specific application cases, it appears unintended for generic, unbiased assessments and remains hard to control for in general.

Proposed Method: Spatial Betweenness Centrality

To mitigate the effect of varying node density on betweenness centrality, we propose a method for computing spatially normalised betweenness centrality. We refer to edge betweenness centrality as c_B and to our proposed spatial edge betweenness centrality as c_{SB}. The main idea behind c_{SB} is to weight all paths contributed to centrality per origin-destination pair relative to the area covered by their origin and destination nodes.

Our proposed method consists of two steps: 1) determining the spatial coverage per node, and 2) computing spatially weighted centrality based on node coverage.

Determining the spatial coverage per node. Spatial coverage of nodes can be determined using tessellation of the network space. While tessellation using Voronoi polygons is common e.g. for retrieving a network null model, it does not render suitable in the given case. Especially in networks with high variability in edge length, Voronoi polygons may intersect non-adjacent edges. Furthermore, motivated from the mobility domain, we assume that interaction is generated along edges rather than at nodes. Therefore, we propose utilising an edge-based tessellation such as the method described by Araldi and Fusco using proximity bands [1] or network-based Thiessen tessellation. The area of each polygon then describes the spatial coverage per edge. Each node's spatial coverage can thus be derived as the total spatial coverage of all edges adjacent to a node divided by two.

Compute spatially weighted centrality based on node coverage. Edge betweenness centrality $c_B(e)$ is defined for a graph $G(V, E)$, where V refers to the set of nodes and E refers to the set of edges as follows:

$$c_B(e) = \sum_{s,t \in V; s \neq t} \frac{\sigma_{s,t}(e)}{\sigma_{s,t}} \tag{1}$$

In this formula, for any pair of origin s and destination t nodes, $\sigma_{s,t}$ denotes the number of shortest paths from s to t, and $\sigma_{s,t}(e)$ refers to the quantity of paths that pass edge e.

In standard c_B, each origin-destination pair contributes equally to centrality. Consequently, the weight each o-d relation contributes equals to one: $w_{relation}(s, t) = 1$. For spatial normalisation we want to weight the paths contributed to centrality per o-d pair proportionally by their origin and destination node spatial influence (weights). Therefore, we propose a weight function that distributes an origin node's spatial influence (area covered) to all other nodes proportionally to their relative spatial influence as a destination:

$$w_{relation}(s, t) = w(s) \frac{w(t)}{\sum\limits_{u \in V; u \neq s} w(u)} \text{ for } s, t \in V; s \neq t$$

where $w(v)$ refers to the weight of a node v, respectively its spatial coverage. The full definition of our proposed spatial betweenness centrality metric consequently reads as:

$$c_{SB}(e) = \sum_{s,t \in V; s \neq t} \frac{\sigma_{s,t}(e)}{\sigma_{s,t}} w(s) \frac{w(t)}{\sum\limits_{u \in V; u \neq s} w(u)} \tag{2}$$

In order to obtain normalised values for c_{SB}, the absolute values are divided by the total area covered by origin nodes: $c_{SBnorm}(e) = c_{SB}(e) / \sum_{v \in V} w(v)$.

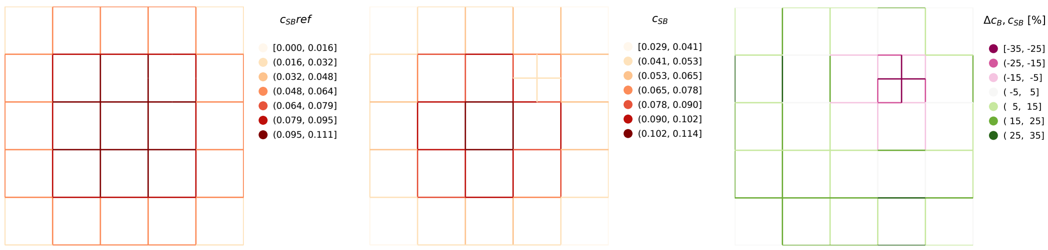
We propose an implementation based on spatial interaction incorporated betweenness centrality (SIBC)[7]. It builds upon Brandes algorithm [3] and adds a weight function $f(s, t)$, which represents a measure of spatial interaction - known flow or estimated flow based on a gravity model [7]. If one pre-computes the o-d weight matrix, it can be employed as spatial interaction matrix in the SIBC method. For applicability in large networks, we suggest computing o-d weights stepwise per origin, alongside solving the single-source shortest path (SSSP) problem.

3 Results

In this section we provide centrality assessments for different networks using both, standard edge betweenness centrality c_B and our proposed spatial variant c_{SB}.

The artificial case: regular grid network. As first example we assess the network that we used to illustrate the problem in section 2.

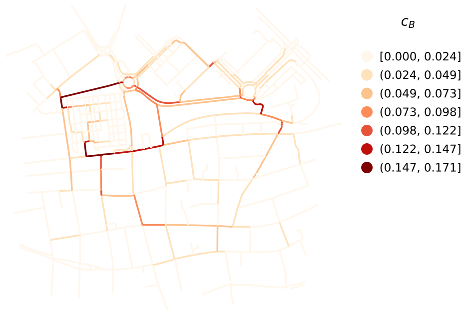
In figure 1 we can observe that c_B shows a shift of high centrality towards the subdivided cell, whereas such a shift is not present in c_{SB} shown in figure 2 a) and b). The differences between c_{SB} for the subdivision case and c_B for the regular grid case are relatively small. In contrast, figure 2 c) highlights the mitigated shift of high centrality when applying c_{SB}.



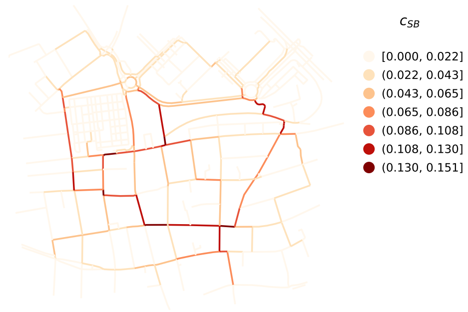
(a) Regular grid network. **(b)** Network with subdivision. **(c)** Difference compared to c_B.

Figure 2 Spatial betweenness centrality applied to the original problem case.

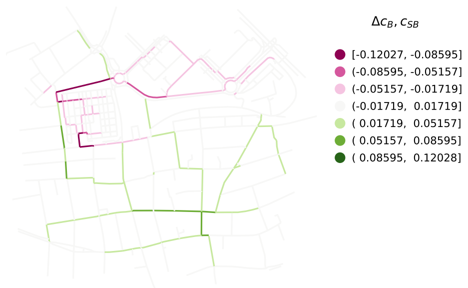
The real-world case: street networks. Additionally, we computed c_B and c_{SB} for several extracts of real-world road networks of varying form. For brevity, we only present one example here and provide more cases online at `https://doi.org/10.5281/zenodo.8125632`.



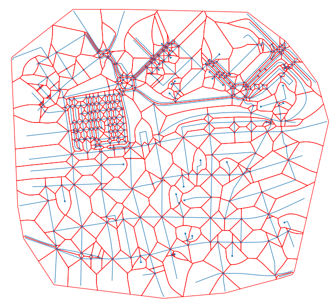
(a) Betweenness Centrality c_B.



(b) Spatial Betweenness Centrality c_{SB}.



(c) Difference: $c_{SB} - c_B$.



(d) Tessellated input network.

Figure 3 Betweenness centrality for a subset of a real-world street network (Stuttgart, Germany).

Results for the real-world case using a subset of Stuttgart, Germany are presented in figure 3. High node density is present in the North, whereas lower node density is prevalent in the centre and South, which is visible in subfigure d). Accordingly, the size of tessellation polygons decreases with higher node density. When comparing subfigures a) and b) or assessing the differences in subfigure c), c_B puts a clear emphasis on routes linking the high-density areas. For c_{SB}, part of these links also show above-average centrality. However, additional links in the centre and South are more pronounced in c_{SB}.

For all networks assessed, we can observe a tendency of higher centrality values for c_B in proximity of areas with higher node density compared to spatially normalised centrality c_{SB}.

4 Discussion and Outlook

We showed that spatial variation in node density has significant influence on betweenness centrality for spatial networks. Unless node density is an intended indicator to consider in a specific application case, we regard this as systematic bias that needs to be addressed. E.g. for applications in mobility, standard betweenness centrality c_B may only render suitable results, if node density spatially correlates with population density.

With the concept of spatial betweenness centrality c_{SB} we propose a generic solution that utilises spatial normalisation to weight the contribution of individual relations. Results applying c_{SB} show a clear mitigation of bias introduced through variations in node density in c_B. Spatial betweenness centrality c_{SB} can therefore provide a generic null model of betweenness centrality in spatial networks.

For practical application of c_{SB}, edge effects need to be considered. One may utilise a network covering larger extent than the area of interest for assessment to allow shortest paths on edges outside the area of interest and to avoid edge effects in tessellation. Future research should shed more light on specific edge effects of c_B and c_{SB}.

Depending on the domain-specific application case, additional factors may be integrated into c_{SB} assessments. Non-uniform weight may be applied to areas of e.g. different land use. This also allows for excluding certain areas from contributing to centrality computation as orign and destination. Furthermore, combination e.g. with population data can open new application scenarios.

We see great potential in the use of spatial betweenness centrality c_{SB} for unbiased, generic assessment of spatial networks. It combines both, morphological properties with spatial embedment of the network. However, it may depend on the specific domain application, whether c_{SB} or an advanced domain-specific modelling approach is preferable.

References

1 Alessandro Araldi and Giovanni Fusco. From the street to the metropolitan region: Pedestrian perspective in urban fabric analysis. *Environment and Planning B: Urban Analytics and City Science*, 46(7):1243–1263, September 2019. doi:10.1177/2399808319832612.

2 Marc Barthelemy. *Morphogenesis of Spatial Networks*. Lecture Notes in Morphogenesis. Springer International Publishing, Cham, 2018. doi:10.1007/978-3-319-20565-6.

3 Ulrik Brandes. A faster algorithm for betweenness centrality. *The Journal of Mathematical Sociology*, 25(2):163–177, June 2001. doi:10.1080/0022250X.2001.9990249.

4 L. da F. Costa, B. A. N. Travençolo, M. P. Viana, and E. Strano. On the efficiency of transportation systems in large cities. *Europhysics Letters*, 91(1):18003, July 2010. doi:10.1209/0295-5075/91/18003.

5 Stefan Lämmer, Björn Gehlsen, and Dirk Helbing. Scaling laws in the spatial structure of urban road networks. *Physica A: Statistical Mechanics and its Applications*, 363(1):89–95, April 2006. doi:10.1016/j.physa.2006.01.051.

6 Parongama Sen, Subinay Dasgupta, Arnab Chatterjee, P. A. Sreeram, G. Mukherjee, and S. S. Manna. Small-world properties of the Indian railway network. *Physical Review E*, 67(3):036106, March 2003. doi:10.1103/PhysRevE.67.036106.

7 Xiaohuan Wu, Wenpu Cao, Jianying Wang, Yi Zhang, Weijun Yang, and Yu Liu. A spatial interaction incorporated betweenness centrality measure. *PLOS ONE*, 17(5):e0268203, May 2022. doi:10.1371/journal.pone.0268203.