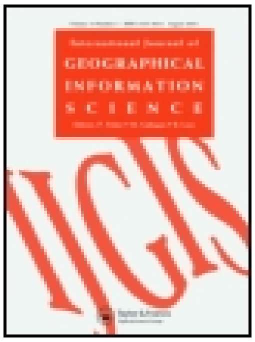


International Journal of Geographical Information Science

ISSN: 1365-8816 (Print) 1362-3087 (Online) Journal homepage: https://www.tandfonline.com/loi/tgis20

Metrics for characterizing network structure and node importance in Spatial Social Networks

Dipto Sarkar, Clio Andris, Colin A. Chapman & Raja Sengupta

To cite this article: Dipto Sarkar, Clio Andris, Colin A. Chapman & Raja Sengupta (2019): Metrics for characterizing network structure and node importance in Spatial Social Networks, International Journal of Geographical Information Science, DOI: 10.1080/13658816.2019.1567736

To link to this article: https://doi.org/10.1080/13658816.2019.1567736

Published online: 18 Feb 2019.

INTERNATIONAL JOURNAL OF GEOGRAPHICAL INFORMATION SCIENCE
https://doi.org/10.1080/13658816.2019.1567736

Taylor & Francis
Taylor & Francis Group

Check for updates

RESEARCH ARTICLE

Metrics for characterizing network structure and node importance in Spatial Social Networks

Dipto Sarkar[a,b], Clio Andris[c], Colin A. Chapman[d,e,f,g] and Raja Sengupta[d,b]

[a]Department of Geography, National University of Singapore, Singapore; [b]Department of Geography, McGill University, Montreal, Canada; [c]Department of Geography, Pennsylvania State University, University Park, PA, USA; [d]School of Environment, McGill University, Montreal, Canada; [e]Department of Anthropology, McGill University, Montreal, Canada; [f]Shaanxi Key Laboratory for Animal Conservation, Northwest University, Xi'an, China; [g]School of Life Sciences, University of KwaZulu-Natal, Pietermaritzburg, South Africa

ABSTRACT

Social Network Analysis offers powerful tools to analyze the structure of relationships between a set of people. However, the addition of spatial information poses new challenges, as nodes are embedded simultaneously in network space and Euclidean space. While nearby nodes may not form social ties, ties may exist at a distance, a configuration ill-suited for traditional spatial metrics that assume adjacent objects are related. As such, there are relatively few metrics to describe these nuanced situations. We advance the burgeoning field of spatial social network analysis by introducing a set of new metrics. Specifically, we introduce the *spatial social network schema, tuning parameter* and the *flattening ratio*, each of which leverages the notion of 'distance' to augment insights obtained by relying on topology alone. These methods are used to answer the questions: What is the social and spatial structure of the network? Who are the key individuals at different spatial scales? We use two synthetic networks with properties mimicking the ones reported in the literature as validation datasets and a case study of employer–employee network. The methods characterize the employer–employee as spatially loose with predominantly local connections and identify key individuals responsible for keeping the network connected at different spatial scales.

ARTICLE HISTORY

Received 21 February 2018
Accepted 7 January 2019

KEYWORDS

Spatial Social Network; metrics; network structure; node importance; social relationships

Introduction

A social network (SN) can be defined as a collection of nodes connected by edges that represent formal 'ties' or evidence of interaction (Wasserman and Faust 1994) between people or institutions. SNs are a useful data model for studying the relationships between entities in personal, professional, kinship, political and romantic relationships (Boccaletti *et al.* 2006). Social network analysis (SNA) has traditionally focused on the topology of nodes and edges within the 'network' or 'feature' space, without environmental or geographic context. While metrics have helped characterize interpersonal and institutional systems for decades, traditional, aspatial SN representations can be limiting because

CONTACT Dipto Sarkar dipto.sarkar@mail.mcgill.ca

societies and relationships do not exist in isolation but coexist among geographic features that affect these ties. Geographic context is an important factor for social networks as it influences individual decisions and drives network dynamics (Adams *et al*. 2012). Currently, the study of spatial social networks (SSN), i.e. spatially embedded networks (Radil *et al*. 2010a, Sarkar, Sieber, *et al*. 2016), lacks sufficient metrics, software and frameworks for the simultaneous analysis of interpersonal connectivity and environmental attachment.

The study of SSNs fits into the larger scope of GIScience by challenging the notion of proximal effects as predicted by Tobler's First Law. The suite of spatial analysis techniques (e.g. spatial autocorrelation, kriging) is built on the premise that proximal entities interact more frequently. However, humans tend to interact across distance, requiring metrics that can privilege spatial non-adjacency. This fabric of interdependence renders a social systems' behaviour too complex to model in terms of just distance. Thus, we use networks to model dependence between non-adjacent entities – as these networks can accommodate both nearby and distant connections. In GIScience, there have been calls for analysis of social systems as affected by, as well as affecting, geographic space (Shaw and Sui 2018). There is a need for spatial information theory and representation for this emerging field within GIScience, and the examination of social network metrics within spatial environments helps us advance its theoretical agenda. Specifically, by modelling how connectivity may reframe the importance of places or people that seem peripheral due to their spatial location, central locations or people with high accessibility may be considered crucial for overall connectivity of the social fabric.

In this work, we create new metrics for SSN analysts looking to simultaneously value a node (i.e. person) for its network connections and its geographical location. We choose to focus on nodes because unlike more common spatial network analyses that focus on transportation and road networks (O'Sullivan 2014), the edges of SNs are neither planar nor well-suited for analysis within a GISystem. Each of the following metrics/techniques stem from a prior legacy of network analysis and graph theory (Nordbeck 1964, Haggett and Chorley 1969). We apply our metrics to two simulated networks and a real-world network. While the simulated networks mimic common SS properties reported in SN literature, the real-world network of economic benefits around Kibale National Park, Uganda, was collected as part of a conservation effort. We are interested to know what are the social and spatial dimensions correlate in the network? What is the spatial distribution of the social connections (e.g. dense connections over a small extent vs sparse connection dispersed in space)? And who are the key individuals across different spatial scales responsible for spreading the benefits and keeping the network connected?

The metrics were designed to answer these questions and can be described as follows:

First, we introduce a scatter plot-based visualization called an *SS network schema* that simultaneously plots each pair of nodes as a function of both their network distance (measured in network hops) and Euclidean distance (e.g. distance apart in geographic space) to explain the extent to which connections are nearby or dispersed. This is a graphical representation of the concept called *route factor* (Nordbeck 1964, O'Sullivan 2014), defined as the number of hops to reach two nodes separated in Euclidean space. The *SS network schema* is an improvement over the route factor equation because it simultaneously explores the network properties of all nodes, providing an understanding of the entire network. The *SS tuning parameter*(a) is used to describe the extent to which nodes favour nearby or distant contacts. This tuning parameter used in conjunction with topological centrality metrics illustrates that a node deemed topologically important merely by its ability to connect its neighbours may not be efficient in connecting clusters of nodes at a distance, thereby losing its importance when the spatial scale of analysis changes.

Third, we measure a network's spatial tightness with the *network flattening ratio*, the proportion of the sum of original network edge distances to the same sum of distances in a re-configured network optimized to create the closest possible ties while maintaining each node's degree. The purpose of this metric is to quantify how dispersed the network structure is over geographic space as compared to a perfect hypothetical 'compactness' realized by connecting all nodes with the nearest possible candidates while maintaining the number of neighbours (degree) of each node. This metric can be used to mark the dispersion or tightening of a network over time or across different networks.

Our results for the Kibale network show that while some node with fewer ties serves as the only connection between people in different villages, others have many local ties but do not connect farther nodes. In conservation scenarios, such as in Kibale, where equitable distribution of resources is desired, these individuals are particularly important because they help identify individuals responsible for dispersing economic benefits arising from a single source, preventing clusters of economic wealth from forming. However, existing metrics, relying only on topology, disregard spatial embedding of the network and miss the role people play at different spatial scales. Our metrics identify important individuals who play a role in sustaining network connections at different spatial scales. The success of conservation plans is contingent on the support of local communities (Adams 2004). Individuals responsible for spreading economic benefits are also likely to be influencers and can be targeted to spread the message of conservation.

Background

SNA relies on a variety of metrics to characterize the network at different topological scales, namely entity-level (single node or edge), meso-level (collection of nodes) and network-level (the holistic network) metrics. Meso-scale network structures, or communities, have received significant attention in SSNs due to its close conceptual ties with the concept of 'community' in geography and a host of techniques have been developed to detect them (Porter *et al*. 2009, Onnela *et al*. 2011, Croitoru *et al*. 2014, Sarzynska *et al*. 2016). Here, we focus on network-level and entity-level metrics which are described in the sub-sections below.

The primary challenge of creating metrics for SSNs stems from the discordance between the definitions of distance in SNA and geography, measured in hops incurred from moving along edges between nodes and Euclidean space measured by (x, y) coordinates, respectively. Thus, standard SNA metrics (e.g. centrality) are unable to provide any information about spatial aspects of an SSN.

Researchers have amassed a helpful set of theoretical knowledge pertaining to propinquity (Fischer 1982), which deals with network structure as a function of distance. The concept has been verified with larger social networks, typically found through social media. Social ties tend to be local, with the probability of ties reducing with distance (Liben-Nowell *et al*. 2005, Wong *et al*. 2006, Mok *et al*. 2010, Preciado *et al*. 2011). Nearby nodes tend to have similar socio, cultural or demographic properties (Hipp and Perrin 2009) and like Euclidean distance, social similarity is also an attractor, manifested in the fact that individuals commonly choose to associate with others of similar traits (e.g. race, language, location, wealth) (McPherson *et al*. 2001). Moreover, nodes that are central to the network tend to be clustered in Euclidean space (Onnela *et al*. 2011).

While the aforementioned studies include distance as a variable, fewer studies embed their systems in the context of multivariate geography, leveraged by geolocating households or activities in a GISystem to include contextual information about geographic space (Andris 2016). Emch *et al*. (2012) modelled a social network of individuals and their positions in geographic space to demonstrate that spatial closeness of two agents is a stronger determinant of disease spread than their level of interpersonal interaction. Models of disaggregate social systems in a geographic setting have revealed that urban gangs have rivalries in adjacent and non-adjacent turf – showing the importance of proximity in institutionalized violence (Radil *et al*. 2010b, Papachristos *et al*. 2013). Another study showed that family members tend to live closer to one another, compared to unrelated community members in rural Thailand, and that this closeness is exacerbated with tie strength (Verdery *et al*. 2012).

Additionally, visual analytics through node-and-edge diagrams (sociograms) are of limited use since, in the case of SSNs, the positions of nodes are fixed using Euclidean coordinates. Hence, graph layout algorithms (Gibson *et al*. 2013), designed to create aesthetically and analytically viable sociograms by moving the nodes around, are not as useful in a spatial setting. Moreover, the significant size of real-world networks create 'hairballs' (Krzywinski *et al*. 2012) and are often too dense to view in Euclidean space using (x, y) anchored sociograms (Luo and MacEachren 2014).

Network expanse

Network-level properties are calculated to analyse the population dynamics of the social network in its entirety. Defined in network space, these metrics are thus non-spatial and almost entirely non-geographic. However, metrics such as **Average Path Length** and **Network Diameter** provide intuitions about node-hop distance in a network, indicating how quickly one can get from one part of the network to another. While average path length refers to the mean inter-node distance, network diameter is the maximum distance between two of the farthest social connections, thus providing a measure of how 'big' the network is (Hanneman and Riddle 2005, p. 81), and how much it 'costs' to reach all nodes.

While considering social networks embedded in geographic space, it is important to characterize the spatial extent of the network. Social-connections tend to be local highlighting the importance of characterizing spatial and social separation between distant entities. Hence, a specification of network diameter length with a spatial extent is essential to capture the SS expanse of spatial social networks. In addition to average path length and network diameter, summary statistics of other entity-level metrics such as degree, betweenness measure network structure as well.

Important nodes

Entity-level metrics help identify important network actors. Such nodes are considered to be in the 'thick of things' (Freeman 1978) as a virtue of being centrally located in the network. Freeman (1978) defines degree, betweenness and closeness centrality as key metrics for assessing relative importance. Node **Degree** is the number of nodes to which a focal node (termed ego in social networks) connects. **Betweenness** centrality is the number of shortest paths between all pairs of nodes that pass through (i.e. use) the focal node for transitivity. **Closeness** centrality captures the average distance with which a node can reach all other nodes in the network (Borgatti 2005). These definitions are modified to accommodate edges with directionality and edge weights, which reflect the strength of relationships or magnitude of flows on an edge.

In case of directed networks, nodes have out-degree and in-degrees. In Freeman's (1978) centrality measures, the focus is on the number of connections regardless of send/receive directionality. The modified metrics for directed edges are generally referred to as prestige metrics since they distinguish between choices made by the node and the collective choices made by the other nodes toward the central node (Knoke and Burt 1983, Wasserman and Faust 1994, Borgatti *et al*. 2002). Hence, prestige is a more refined concept than centrality and can only be measured when incoming and outgoing edges are separated. Using edge weights, degree can be redefined as the sum of weights of all the edges incident on the focal node (Barrat *et al*. 2004), although this makes it hard to distinguish between, for example, a node with 10 edges of weight 1 and a node with 1 edge of weight 10 (Opsahl *et al*. 2010). In case of closeness and betweenness, the least cost path is used (Brandes 2001, Newman 2001) although this may ignore the relative importance of edge weight versus number of edges (Opsahl *et al*. 2010).

Methods

We introduce a set of metrics to characterize the SS structure of the network, efficiency of spatial connectivity, and to identify important nodes embedded in the SSN. We visualize the network using an *SS network schema* which allows users to explore the changing importance of each node as the distance becomes an important factor. We quantify this using the node-level *SS tuning parameter (a)*, which provides modified node centrality measures, namely, degree, closeness and betweenness. We also provide a network-level metric, the *flattening ratio*, which compares an actual network to its most 'flattened', spatially compact hypothetical configuration, to quantify how dispersed it is.

Let $G = (V, E)$ be an undirected unweighted connected graph where $V = \{v_1, v_2,\ v_3, \ldots,\ v_n\}$ is the set of nodes and $E = \{e_1, e_2,\ e_3, \ldots,\ e_n\}$ is the set of edges where each edge e_k is associated with an unordered pair of vertices $(i,\ j)$. Locational information in the form of *(x, y)* coordinates is associated with each node. The Euclidean distance between any two nodes $(p,q) \in V$ is represented as $|p, q|$ while the shortest path along the network is represented as $C(p, q)$. However, $|p, q|$ and $C(p, q)$ are not directly comparable as they are defined in different measurement spaces.

SS network schema (SS-NS) for rendering of network expanse

The *SS network schema* plots $C(p,q)$ against $|p,q|$ to detect patterns that have been consistently reported in spatial social network literature (Figure 1). The axes of the plot afford a measure of how 'big' the network is both socially and spatially (Hanneman and Riddle 2005). The range of the x-axis specifies the spatial extent of the network using Euclidean distance between the most distant nodes (a continuous variable). The y-axis specifies the network diameter (shortest path distance between the topologically farthest nodes) (a discrete variable) (Figure 1). Along the y-axis, the clustering of points denotes the number of nodes at increasing topological distances. An example line passes through y = 1 (shortest path = 1). Points along this line represent the frequency of distances of various first-degree friends. Most social network studies that incorporate distance only focus on these first-degree ties, i.e. at y = 1. However, examining distances of n^{th}-degree ties is also crucial to determine processes on the network (e.g. diffusion). Overall, this plot is a workaround for comparing the two distance notions (network and Euclidean). As mentioned, it extends the concept of *route factor* equation, from being a summation of the total distance to total hops to an exploration of the distance to hops ratio for each individual node. This disaggregation allows the analyst to examine the central tendencies of this distribution and pinpoint nodes that may be anomalies in the distance-connection distribution.

Flattening ratio for measuring spatial tightness

To define the flattening ratio, we first create a degree-constrained nearest neighbour network $\bar{G}$ from the given social network G by reconfiguration, such that each node i in

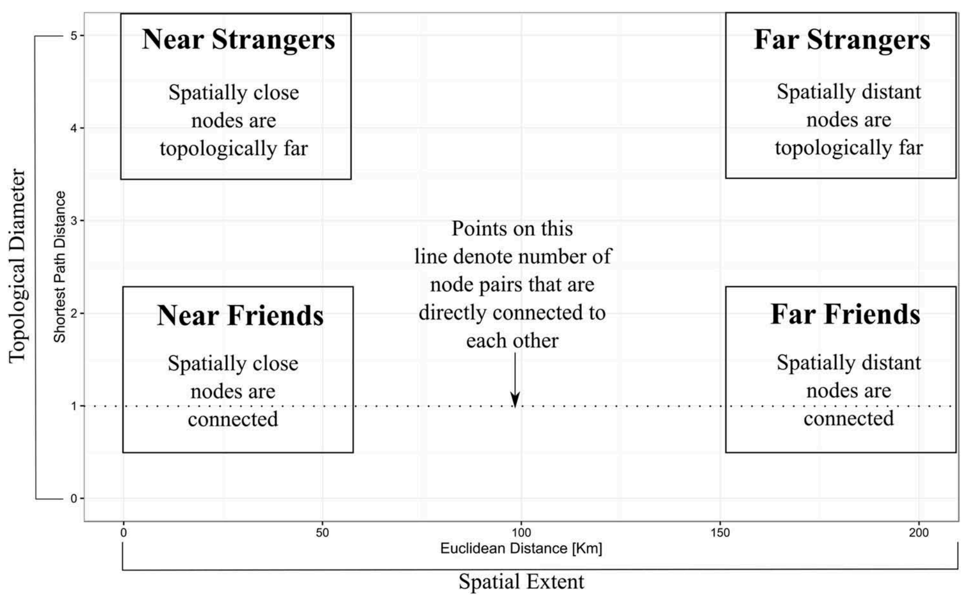


Figure 1. An SS network schema illustrates the different scenarios that any pair of nodes in a spatial social network could incur.

$\bar{G}$ with degree D connects to its nearest D neighbours in Euclidean space. The flattening ratio is the sum of the Euclidean distance between every pair of connected nodes in the network $\bar{G}$ compared to the sum of the Euclidean distance between every pair of connected nodes in G. Mathematically,

$$F_s = \frac{\sum \overline{|pq|}}{\sum |pq|} \tag{1}$$

where, as before, $|pq|$ denote the Euclidean distance between any two connected nodes in G and $\overline{|pq|}$ denote the Euclidean distance between any two nodes connected by an edge in $\bar{G}$. Since the nodes in $\bar{G}$ connect more efficiently to their spatially close neighbours, the overall distance $\sum \overline{|pq|}$ is expected to be less than $\sum |pq|$. Thus, F_s provides a measure of tightness with which nodes in G are connected to their closest spatial neighbours. The closer the ratio is to 1, the closer network G is to an ideal degree-constrained nearest neighbour network. Note that this is akin to the Erdös–Rényi configuration model (Erdős and Rényi 1960). Due to the degree constrained requirement, many possible resultant 'flattened' networks are possible. In other words, $\bar{G}$ is constructed stochastically and for a given G, several $\bar{G}$ are possible. Thus, to calculate F_s, we took the average of several iterations.

SS tuning parameter for measuring node importance through spatial centrality metrics

To define the important nodes in an SS context, we introduce the *SS tuning parameter a* to balance the importance attributed to near versus far social connections. We calibrate the value of a between 0 and 1. If $a = 0$, then farther connections are given higher weightage, and if $a = 1$, the nodes with nearby social connections are considered more important. If $a = 0.5$, both far and near connections in Euclidean space are given equal weightage, thus, nodes that either have intermediate connections, or a mixture of far and near connections, are deemed important by the *spatial centrality metrics* at $a = 0.5$. Thus, different values of a capture the changing importance of the nodes (calculated by *spatial centrality metrics*) at different spatial scales of analysis. The *spatial centrality metrics* are calculated as such:

Each edge $e \in E$, $|pq|$ can be considered as cost or benefit depending on whether near connections are preferred over farther connections. For example, network distance is considered a cost when looking to travel fast between people, but a benefit when trying to stop disease spread. Hence, when moving from one node (i) to another (j) by traversing along the edges, the total weight of the shortest path is calculated by Dijkstra's algorithm (Dijkstra 1959) as the total cost of travelling along the edges from node i to j:

$$d_{N_{i,j}} = min(|i,a| + |a,b| + \ldots + |x,j|) \tag{2}$$

Alternatively, when far connections are preferred, this distance is modified (Brandes 2001, Newman 2001):

$$d_{F_{i,j}} = min\left(\frac{1}{|i,a|} + \frac{1}{|a,b|} + \ldots + \frac{1}{|x,j|}\right) \tag{3}$$

The subscripts F and N in each case denote whether near or far connections are preferred and consequently whether the Euclidean Distance between the nodes was interpreted as benefit or cost.

The *Spatial Centrality* metric incorporates both the amassing of edge weights through traversal as a cost and benefit, is defined here as X_{S_i} for node i, where the tuning parameter a ranges from 0 to 1. X_s can, in turn, refer to the three measures of node centrality, namely, degree, betweenness and closeness. X_s is defined as:

$$X_{S_i} = \alpha \cdot \overline{X_{N_i}} + (1 - \alpha) \cdot \overline{X_{F_i}} \tag{4}$$

$$where,\ \overline{X_{N_i}} = \frac{X_{N_i} - \min(X_{N_i})}{\max(X_{N_i}) - \min(X_{N_i})},\ \overline{X_{F_i}} = \frac{X_{F_i} - \min(X_{F_i})}{\max(X_{F_i}) - \min(X_{F_i})} \tag{5}$$

Where X_s can take on three different values: *degree* (D_s), *closeness* (C_s) *and betweenness* (B_s);i is a focal node and j represents all other nodes. N is the total number of nodes in the network, and α is the tuning parameter such that $0 \leq \ \alpha \leq 1$.

For each case (degree, closeness and betweenness), X_F and X_N are defined as follows:

$$D_{N_i} = \ \sum\nolimits_j^N \frac{1}{|i,j|},\ D_{F_i} = \sum\nolimits_j^N |i,j| \tag{6}$$

$$B_{N_i} = \frac{g_{N_{x,y}}(i)}{g_{N_{x,y}}},\ B_{F_i} = \ \frac{g_{F_{x,y}}(i)}{g_{F_{x,y}}} \tag{7}$$

$$and,\ C_{N_i} = \ \sum\nolimits_j^N d_{N_{i,j}},\ C_{F_i} = \sum\nolimits_j^N d_{F_{i,j}} \tag{8}$$

where $g_{B_{x,y}}$ is the **total length** of the shortest paths between every pair of nodes $x, y \in V - \{i\}$ and $g_{F_{x,y}}(i)$ is the **total length** of the shortest paths that pass through the focal node i.

The values for $B_{F_i},\ B_{N_i},\ C_{F_i}, C_{N_i},\ D_{F_i},,\ D_{N_i}$are normalized to be in the range $[0, 1]$ by linear scaling. These centrality measures, like many other metrics in SNA, are better suited for providing a ranking of importance of nodes rather than for quantifying the difference in influence between nodes and should not be used to compare different networks by their scores (Bonacich 1987, Borgatti 2005). Consequently, the numeric values of metrics $D_{S_i},\ C_{S_i},$ and $B_{S_i}$are less important than the rankings of the nodes afforded by the values computed by the metrics.

Theoretically, the modified metrics provide a consolidation for network and Euclidean distance by treating distance as an intrinsic property of the network. Further, by considering Euclidean distance as either cost or benefit for social activities, the metrics leverage the topology of the network to quantify the spatial scale of influence of each node.

Data

Simulated data

We created two simulated datasets based on prior calibrations of inter-node distance (i.e. propinquity) and node distribution. These initiation metrics are derived from consistently reported accounts of inter-nodal distance (Festinger *et al*. 1950, Mok *et al*. 2010, Preciado *et al*. 2011) and spatial distribution of nodes (Fischer 1977, Butts *et al*. 2012) and act as test datasets for our methods. In the first synthetic **Poisson Network**, node location (x, y) is generated randomly using a Poisson process inside a bounded Euclidean space and the probability of forming an edge reduces exponentially as a function of the distance between the nodes, following the propinquity property (Figure 2(a)). In the second synthetic **Clustered Network**, nodes are clustered at different Euclidean distances inside a bounded space, and the probability of forming an edge reduces exponentially as a function of the distance between the nodes (Figure 2(b)), following findings that spatially clustered nodes tend to be well connected with relatively few links to other such clusters (Entwisle *et al*. 2007, Abizaid *et al*. 2016). Each network has 32 edges connected by 113 and 114 edges, respectively (see Supplementary Information for SS properties).

Kibale employment network

We also utilize a real-world dataset of an employment network near Kibale National Park (KNP), Uganda (hereafter Kibale), to complement the synthetic networks. The elevation in the study area ranges from 1445 to 1557 m (via a 30-m resolution digital elevation model) (USGS 2006). Although KNP is a mid-altitude moist evergreen forest, the area around the field-station is relatively flat and the primary means of transport are by foot, bicycles or motorbikes. The roads near the park are a mixture of paved and unpaved surfaces and people often take shortcuts through fields, rendering Euclidean distance a good approximation of the cost of travel and co-location. A lack of telecommunication infrastructure and the relatively high cost of keeping a mobile phone means most interaction is still carried out in person (Sarkar *et al*., 2016).

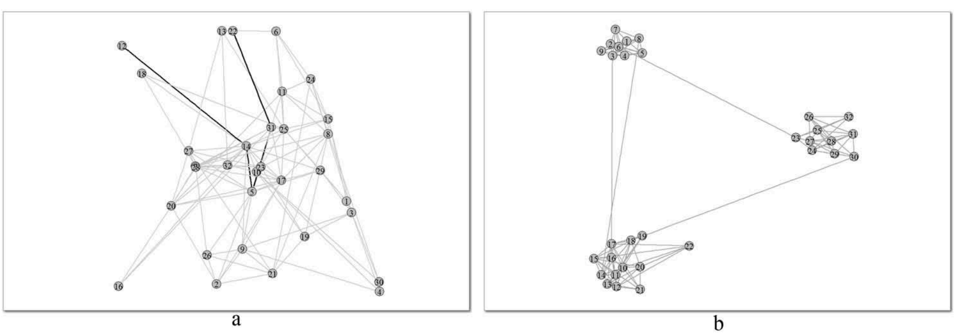


Figure 2. The simulated networks in arbitrary geographic space. The nodes are anchored to their corresponding (x, y) locations. (a) Poisson Network. The highlighted path shows two nodes that are spatially close but topologically far. (b) Clustered Network.

Employers and employees were interviewed in-person between January 2016-May 2017. Employees of Makerere University Biological Field Station (MUBFS) were interviewed first. Next, more participants were identified through a snowball sample, resulting in 209 participants from 21 villages, including people living at the field station. We contacted these employees and asked who they hired for agriculture and household work and for contact information for the people hired for these tasks. We learnt that employees hired for household and farm work by one person are sometimes hired by different employers. The agricultural and household workers informed us also of those they had hired. As many chains as possible were followed through the snowball sample until an individual on the chain did not hire anyone, could not be contacted, or lived more than 10 km from the field station by a motorable road. People who did not hire anyone were excluded. This data was used to develop a network where the location of everyone was geolocated to the village of their residence, and the edges represent employer–employee relationships. The resultant network is a network with 163 nodes and 155 edges in 21 connected components. We used the largest connected component (Figure 3), comprising 97 nodes and 106 edges, to demonstrate the effectiveness of the introduced metrics at providing insights on the spatial embedding of the network and social importance of the nodes at different spatial scales.

This network was collected with the aim of quantifying the percolation of economic benefits originating from employment at MUBFS through the community through employer–employee relationships (Sarkar *et al*. In Press). In this network, our hypothesis is that important individuals are those responsible for spreading benefits across different spatial scales by virtue of their SS position in the employment network. Thus, they are characterized as having a good mix of near-far connections, thus keeping the network connected at different spatial scales. MUBFS is one of the longest continuously running research field sites in Africa and provides a unique case study to understand the impacts of such an establishment on the livelihoods of the community living near the park. This dataset provides an ideal application scenario for the new metrics because its significant size makes it difficult to visualize the network using sociograms and challenging to identify important actors using traditional SNA metrics. Additionally, while the synthetic networks were generated programmatically to have the ideal SS properties, the properties of the real-world network are unknown and have more complexity and noise in their SS properties.

Results

SS network schema

The *SS-network schema* helps investigate the overall structure of the SSN. For the synthetic networks (Figure 4), the points have been aggregated at 5-km intervals and jittered vertically on the plot to provide visual cues as to how many points are at each x, y coordinate. The numbers provide a count of points at each geographic distance, giving an idea of distance dispersion of social connections. The extent of the x and y axes provides a notion of network size, both spatially and socially. The spatial extent of the network, as dictated by the distance of the between the farthest nodes in Euclidean space, is approximately 150 km for the **Poisson Network** and 100 km for the **Clustered**

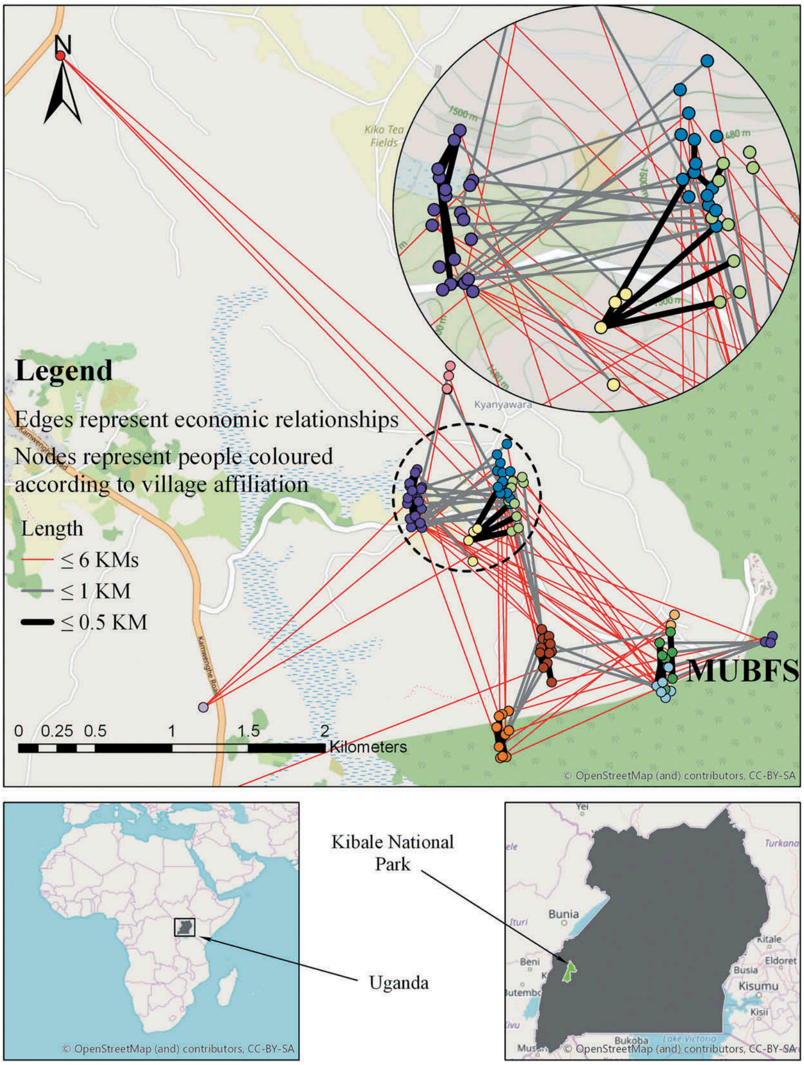


Figure 3. The Kibale network of economic benefits depicts workers and employers as nodes and their edges as relationships. There are a number of small villages that have internal connections (depicted mostly in black) while connections exist more frequently between each small village, indicating that there is an economic incentive to foster travel and communication between these areas. (The position of the nodes has been jittered to prevent overlap of nodes in the same village.)

Network (Figure 4). In terms of SNA, the diameters of the networks are 4 and 5, respectively, meaning that each node in the network can be reached from every other node by traversing relatively few edges. The y-axis entries along *Shortest Distance = 1* is akin to plotting a histogram of the distance between social connections, frequently used to assert distance-friendship patterns.

For **Poisson Network**, the generative process implies that geographically distant nodes are unlikely to be connected, and thus, most social connections should be local. Therefore, distant nodes will have many intermediaries for connection (larger

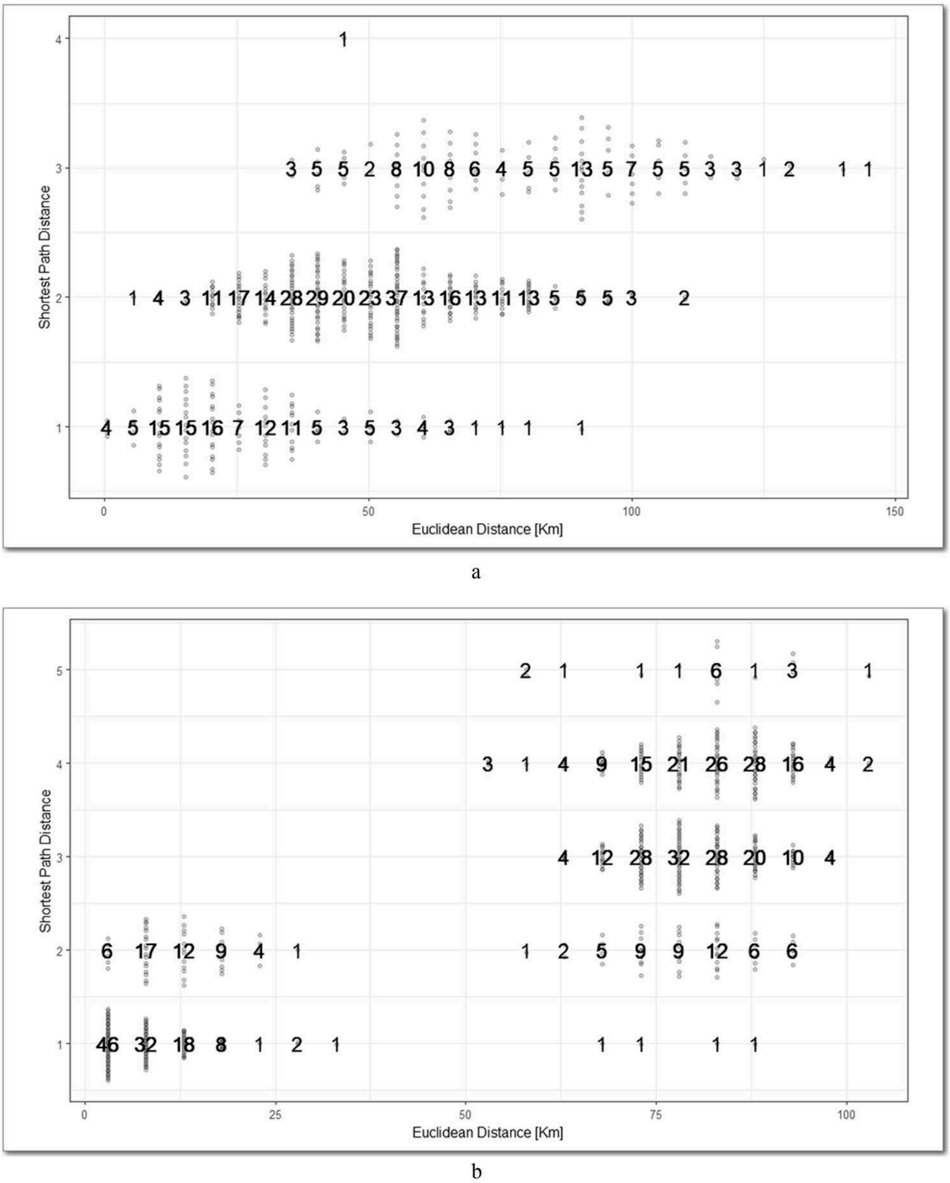


Figure 4. SS network schema for (a) Poisson and (b) Clustered simulated spatial social network. Points are aggregated at 5 km intervals.

shortest distance). The clustering of points in the 'Near Friends' zone (bottom right) of Figure 5(a) highlights the connections and points along the line *Shortest Distance = 2*, indicating that most nodes are reachable from each other within two steps. The farthest nodes in Euclidean space are separated by three hops in network space. Figure 4(a) also highlights an interesting outlier node at the top-left corner, which has spatially close neighbours but is separated from them by four hops. This four-hop path is highlighted in Figure 2. This case cannot be derived easily from standard sociogram methods of SNA

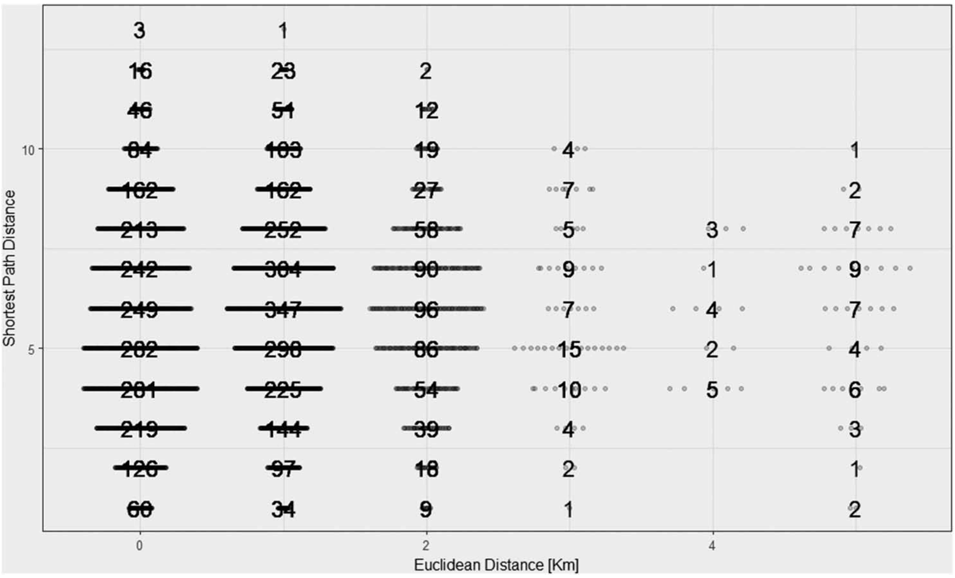


Figure 5. SS network schema for Kibale spatial social network of economic benefits. Points are aggregated at every 1 Km.

but is clear using this visual method. Thus, this node is an outlier with respect to the distance distribution of nodes in this network and would, hence, require further considerations for analysis.

For **Clustered Network**, the generative process implies that each spatial cluster will have many connections and relatively few direct connections between distant clusters. The highly clustered nature of node locations implies that the 'Near Friends' zone (Figure 4(b)) has many points, while there are very few 'Far Friends' (i.e. an absence of points in the lower right-hand quadrant). One can detect spatial clustering and distance-decay of tie formation from the large empty space in the middle of the chart and from the collection of many points towards extremities. Both the sociogram and the *SS-NS* illustrate that about 1/3rd of friends are nearby and 2/3^{rds} of friends are far away, yet the *SS-NS* confirms there are no third-degree ties in these tight clusters, wherein the sociogram may have visually concealed this predicament. A drawback of this plot is that it is unclear how many distinct (unique) nodes are participating in each X, Y plot point as nodes (egos) repeat for each combination with other network nodes (alter).

In the **Kibale Employment Network**, the average employer–employee distance is 0.9 km, and thus, the spatial resolution of the *SS-NS* has been set to 1 km to be able to capture the spatial variations in this network. The small numbers along *Shortest Path Distance* = 1, along with the relatively large diameter of five hints at a sparsely connected network. The spatial extent of the network is relatively small at approximately 6 km in diameter and most nodes are within 3.5 km of each other. The probability of employment decreases with Euclidean distance, and the sparsity of points beyond the 1.5-km mark along *Shortest Path Distance = 1* highlight the few employments that exist beyond 1.5 km. Interestingly, some of the farthest nodes have an edge between them, suggesting some connections are worth maintaining despite the challenges posed by sparse telecommunication and commuting infrastructure. The heavy clustering of points in the near friend region highlights that most hiring is local. However, some topologically farthest connections are spatially close, implying that a lack of opportunities in one's village may necessitate travel to find work. These two extreme cases point to potentially interesting employer–employee dynamics. The anchored sociogram in Figure 3 verifies distant connections are sustained by a single individual. It also points at the possibility that after 2 km, employment may be driven by the individual's reputation rather than his location, as all three long distance (>2 km) connections are sustained by a single node. Moreover, some node pairs are close together in Euclidean space may be far apart in network space, highlighting that the same person was rarely hired by two separate employers even if they lived close to each other. This is not surprising as most hiring is done for farm work, making it difficult for a person to be hired for the entire day at more than one farm.

The geographical network map (Figure 3) illustrates the topography in which the network is situated and the relative position of the nodes with respect to one another. Yet, this representation suffers from drawbacks owing to its 'hairball' like structure (Krzywinski *et al*. 2012) that would be increasingly pronounced with more nodes and edges. Alternatively, the *SS-NS* (Figure 5) provides discernible information about socio-spatial network properties such as size (both spatial and topological) and pattern of connections (i.e. how well nearby nodes are connected, how many nodes have distant connections).

Spatial network tightness (flattening ratio)

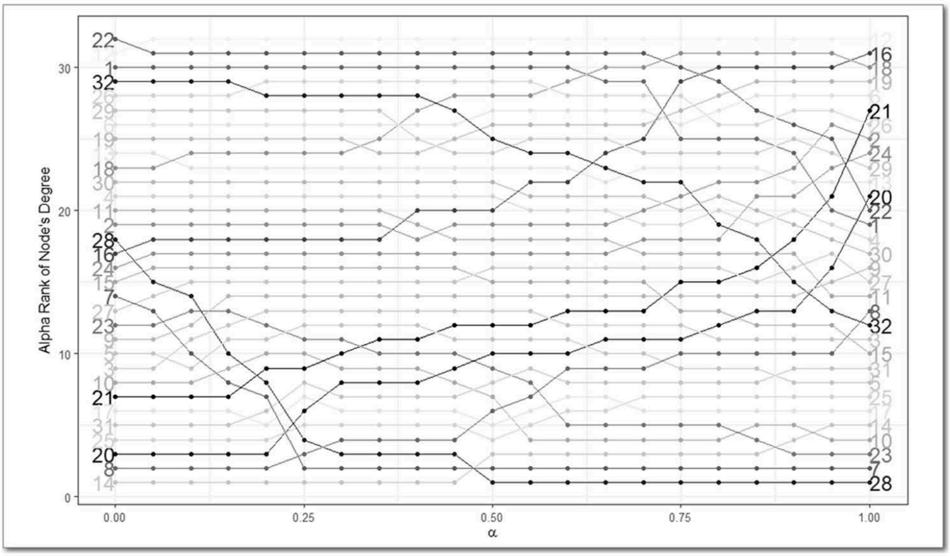
The original simulated networks are already spatially optimized as the probability of long connections was programmed to decrease exponentially. Thus, the flattening ratio for the **Poisson and Clustered Networks** are 0.797 and 0.923, respectively. For the **Kibale Employment Network**, the flattening ratio is 0.212, which implies that the original network is far from being spatially tight. This may be due to several hires from distant villages and the significant number of 'close strangers', which get optimized in the flattened networks (see Supplementary Information for flattened networks generated from the original networks).

We also experimented with elevation as a cost parameter on top of Euclidean distance in the Kibale flattened network since steep roads impede pedestrian mobility – a key mode of transportation in Kibale. Each edge was weighted by the maximum elevation change along the path. The sum of the change in elevation for all original edges was 1519.07 m and, for flattened networks, ranged between 1212.2 and 1512.3 m. Thus, the original network, with longer edges, also incurred more elevation change as can be expected.

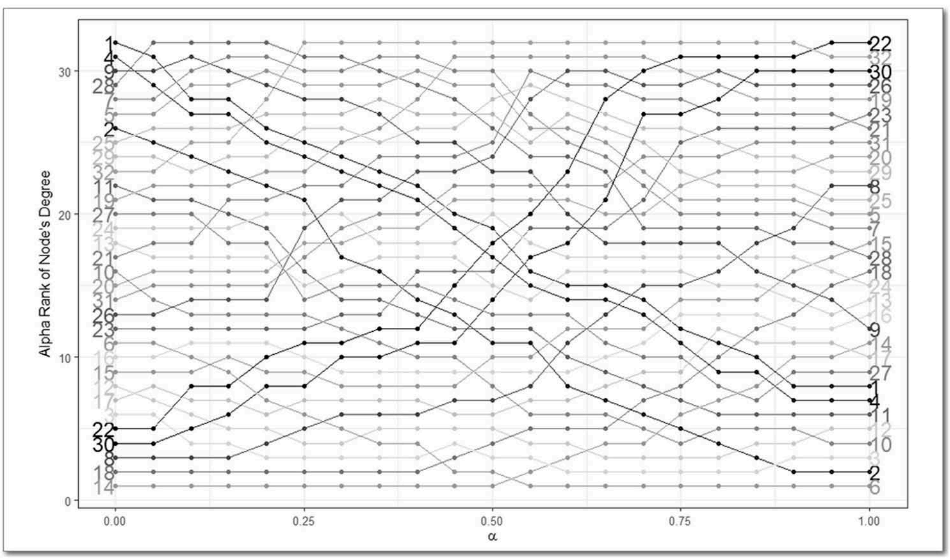
A combination of the two network-level metrics characterizes the structure of the Kibale network as being sparsely connected and spatially 'loose' even though most connections are local. This structural property can be explained by the fact that in an employment network, a person can mostly be hired by a single employer, albeit, there are exceptions. Thus, the number of interconnections between nodes are limited which, in turn, prevents a dense network structure and forces hiring from farther away when locals have already been engaged.

Identifying important nodes via SS tuning parameter

The plots allow the user to detect key nodes that shift, given a changing emphasis on having far versus near connections as the most valuable qualities of a node. Using **Poisson network**, node importance, as predicted by degree centrality, changes with different values of α (Figure 6(a)). Here, the x-axis shows the values of α, while the y-axis shows the rank of the node at different α's, beginning with the highest-ranking node at



a



b

Figure 6. Degree ranking of the nodes at different α values for simulated networks. Transparency of the lines depicts total absolute value change in rank between α = 0 and α = 1. (a) Poisson Network and (b) Clustered Network.

the top and incrementing to the lowest ranking node. The numbers correspond to node ID. The lines connecting the rank of a node at different values of α are more transparent if they change rank between $\alpha = 0$ and $\alpha = 1$ significantly. As expected, nodes with the farthest connections are considered important at $\alpha = 0$, while nodes with closer connections are given higher ranking at $\alpha = 1$. Nodes 20, 28 and 32 are affected most by α as they predominantly have further and nearer connections, respectively. On the contrary, node 14, as result of having connections at different Euclidean distances, remains relatively important at all values of α.

In **Clustered network**, node 22 and node 6 are remarkable in their significant rank variations (Figure 6(b)). Node 22 significantly loses its ranking between different α values, ending up in the last rank because of its position in Euclidean space, which is slightly away from a node cluster while still being well connected to five nodes in the nearest cluster. Its relatively intermediate distance connections make it significantly lose its ranking after $\alpha = 0.4$. However, its connections are geographically not close enough to make it important at $\alpha = 0.5$. On the contrary, node 6, being socially well-embedded in its spatial neighbourhood, becomes the most important node at $\alpha = 0.55$ and maintains its ranking throughout. Nodes 1, 2 and 4 gain in rank between $\alpha = 0$ *and* $\alpha = 1$ for the same reason as node 6.

In terms of identifying node importance, Poisson network is more stable than the clustered network given a shift in emphasis on distance vs nearness as an indicator of importance. This is highlighted by the many crossing lines in Figure 6(b) compared to Figure 6(a).

The considerable number of disconnected components (21) in the **Kibale Employment Network** begets evaluating the role of nodes in the largest components to keep the component connected across different socio-spatial scales. The graph of change in the betweenness values (Figure 7) highlights the shifting role of different nodes to keep the network connected across α values. Node 4 is arguably the most important as s/he connects six clusters (people in different villages) at different Euclidean distances and maintains a stable top rank. Node 23, on the other hand, significantly improves ranking between $\alpha = 0$ and $\alpha = 1$, serving as the connection between residents of the same village who are not connected by any other path. Node 18 only employs people in different villages that are quite far from their village of residence and serves as a crucial point of connection between the pair of villages when distant connections are preferred.

Directed graphs to distinguish between employer and employee can be used as a measure of 'prestige'. Thus, node 12 can be considered prestigious as s/he is hired by people from far villages although people (nodes) with similar experience (i.e. house making, carpentry) are spatially close. The high correlation ($r_s = 0.86$) between out-degree and degree importance at $\alpha = 0$ implies people usually tend to hire employees from far villages only when they have relatively saturated local connections. This also follows from the nature of hiring which tends to be for household and farm work, and thus, locals are preferred unless an alternative option is quite reputable or skilful, which is may be the case for the individual depicted by node 12.

Table 1 highlights the effectiveness of our metrics as compared to the traditional SNA ones. As expected, the variation in the top-ranked nodes in the modified metrics highlights the roles played by individuals at different spatial scales of analysis at keeping

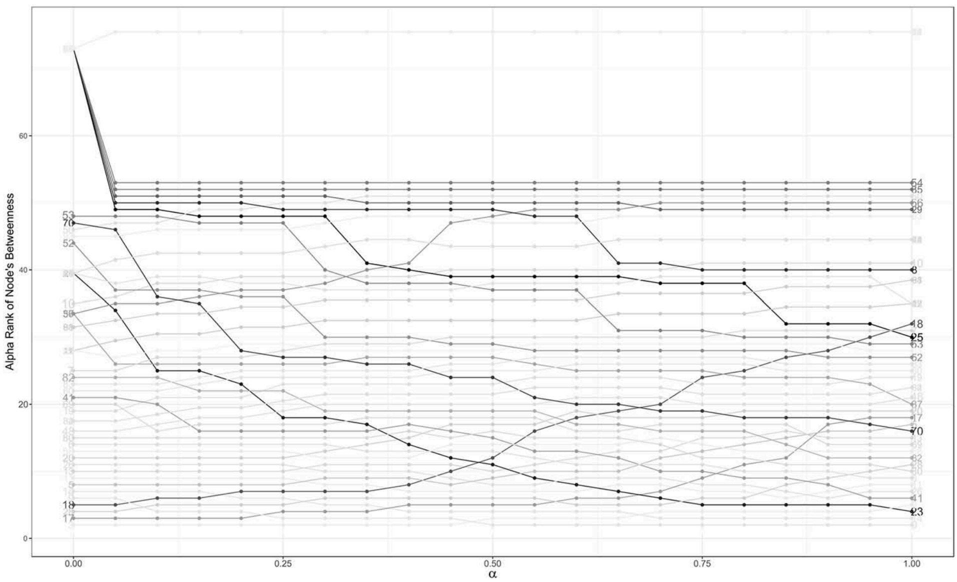


Figure 7. Betweenness ranking of the nodes at different α values in Kibale Employment Network. In case of a tie, all the tied nodes are given an average rank. Transparency of the lines depicts total absolute value change in rank between α = 0 and α = 1.

the network connected and thus help spread benefits far and wide across the landscape. Thus, as compared to traditional metrics, the modified metrics are better suited at eliciting the spatial roles of individuals in the network.

In SSNs, it important to look at both the number of connections and at the contribution of individuals in maintaining and spreading benefits over a network across the landscape. The use of the tuning parameter *α*, helps us characterize the importance of individuals on a landscape, particularly across different spatial scales by balancing the cost and benefits of distant connections. In the characterization of the spread of economic benefits, the absence of 'important' individuals will spatially condense the benefits originating from MUBFS. Since studies have shown that economic gains from the protected-areas impact a community's perception of conservation plans, it is important to ensure equitable percolation of the benefits to individuals living in communities surrounding the park. The combination of degree and betweenness of individuals, as reflected by the modified metrics, helps identify specific individuals who fulfil this important role.

Discussion and conclusion

We provide new methods to analyse SSNs that take both Euclidean space and network space into consideration and consequently inform how social systems are organized in and affected by geographic space. These methods highlight that the Kibale network is sparsely connected and not spatially compact although hiring is mostly local. Several key individuals act as brokers at keeping the network connected across larger spatial extents. In the context of this employer–employee network arising from a conservation effort, this 'loose' structure, with a core located near the field station is desirable as it ensures that

Table 1. Table showing the top three rankings obtained by the modified metrics compared to the standard definitions in the three networks. The numbers in the table denote the name (ID) of the nodes in the networks. The correlation row shows the overall Spearman's Rank Correlation in ranking between each of the modified metrics at different α values in relation to their standard definitions. Note that, even in cases where overall correlation is high, there is variation in the top-ranked nodes highlighting the spatial effects of the SSN being captured through the new metrics.

		Degree				Betweenness				Closeness			
	Ranking	Topological	Spatial at $\alpha = 0.0$	Spatial at $\alpha = 0.5$	Spatial at $\alpha = 1.0$	Topological	Spatial at $\alpha = 0.0$	Spatial at $\alpha = 0.5$	Spatial at $\alpha = 1.0$	Topological	Spatial at $\alpha = 0.0$	Spatial at $\alpha = 0.5$	Spatial at $\alpha = 1.0$
Poisson	1st	14	14	28	28	14	14	14	14	14	14	14	23
Network	2nd	17	8	7	7	25	8	25	25	17	8	17	10
	3rd	5 and 25	20	14	14	10	21	31	31	5, 10, 25 and 3	21	31	14
	Correlation	1.00	0.83	0.89	0.89	1.00	0.56	0.87	0.81	1.00	0.37	0.90	0.89
Clustered	1st	12	14	14	6	18	18	18	18	14 and 18	14	18	16
Network	2nd	17	18	6	2	30	30	30	30	8 and 12	18	14	17
	3rd	3, 6, 14, 15 and 18	8	3	3	6	14	16	16	3, 6. 10 and 17	8	17	10
	Correlation	1.00	0.34	0.81	0.80	1.00	0.59	0.88	0.71	1.00	0.61	0.86	0.67
Kibale	1st	4	12	4	41	4	4	4	4	4	4	4	4
Network	2nd	5 and 48	4	41	84	9	3	9	9	17	3	9	55
	3rd	9,14, 17, 41 and 60	11	12	26	17	17	3	14	9 and 3	17	6	9
	Correlation	1.00	0.61	0.65	0.61	1.00	0.95	0.99	0.98	1.00	0.77	0.82	0.69

economic benefits arising from the field station are accessible to local communities. The long and medium distant connections sustained by few individuals are important in ensuring that communities near the boundary of the park at various distances from the field station also derive some economic benefits from the field station. Together, these features of the network play a crucial role in engaging communities in conservation-related activities, a feature linked to successful conservation efforts.

In terms of methodologies, we introduced a scatterplot-based visualization approach providing an overview of the socio-spatial structure of SSNs by plotting the shortest network paths required to connect nodes at various Euclidean distances. We also designed a measure of socio-spatial network tightness called the flattening ratio. These two network-level methods characterize the spatial structuring of the social network in terms of distance distribution of connections. In general terms, these methods provide an intuition of where friends are located, how easy is it to meet n^{th} degree mutual friends and how fast information will percolate to distant places.

Furthermore, we proposed variations to the three commonly used centrality metrics, namely, degree, betweenness and closeness, to characterize the important roles played by the nodes in the network at different spatial scales. By using scaling parameter α, the metrics modify the interpretation of distance between nodes to interpret it as either beneficial or detrimental, thereby quantifying the importance at different spatial scales. The modified centrality metrics help distinguish between individuals who have many local friends versus those with few local friends but many long-distance friends. In the context of SSNs, both these classes of individuals are important depending on the spatial scale of interest.

We found instances where a node with many distant connections is interpreted as being important at $\alpha = 1$, thus, portraying its importance to be greater than a node with a few very close connections. Using the introduced metrics in conjunction with standard definitions helps weed out such outliers. Furthermore, one must be careful not to attach an excessive value to individual results of the SNA centrality metrics as node importance (as assigned by the centrality metrics) tend to be highly correlated (Valente *et al*. 2008, Li *et al*. 2011, 2015). Consequently, characterization of important nodes should be done as a combination of metric use, visual inspection and expert insight about the generative processes of the network.

The locational information in SSNs also provide opportunities to understand the network in conjunction with other geographic data known to influence tie formation, such as the nature of the built environment (Lund 2003, Hipp *et al*. 2014, Boessen *et al*. 2017). For example, in the Kibale network, the presence of numerous shortcuts through fields and unpaved roads makes it difficult to gauge travel distance. Thus, relying only on information on known roads would exaggerate the employer–employee distance and provide an incorrect proxy of the impact of the environment. The map also helps identify the location of the villages and the location of the spatial core close to the field station. Additionally, the lop-sided east-west orientation of the network is due to the barrier to the east posed by the national park and the presence of the roads that expedite movement. Furthermore, the concept of centrality in geography goes beyond topological connections and relative positions to include factors that may be social, economic, political, etc. For example, the concept of 'prestige' can be related to a social attribute of the node rather than its topological positioning (Abizaid *et al*. 2015, Entwisle *et al*. 2007). In Kibale, some long-distance hiring is explained by the requirement of people with specialized skills (Node 12), not available locally. Moreover, a node may be deemed important because of its geographic location (Fleming and Sorenson 2001, Owen-Smith and Powell 2004). Sometimes, geographical location becomes an important driver in tie formation, compensating for lack of a central position in a social network (Owen-Smith and Powell 2004) or, conversely, stop certain entities from forming alliances if they are geographically far even though socially central (Fleming and Sorenson 2001). Considering geography, as well as social networks, becomes indispensable for understanding the duality between the importance of location versus network (Castells 1996). This duality can be addressed by a combination of the standard centrality metrics along with the modified centrality metrics introduced here. While the standard metrics privilege network position, the modified metrics can distinguish between the nodes, playing a crucial role in local versus distant processes. Thus, the use of new socio-spatial techniques proposed herein, along with maps and traditional metrics, are required to identify the different important nodes according to their roles.[1]

Note

1. The metrics and visualization discussed in this article are available as an R library downloadable from https://github.com/diptosarkar/SpatNet/.

Acknowledgments

This research was supported by Rathlyn Fieldwork Award, Rathlyn GIS Award and the Graduate Mobility Award from McGill University. Colin Chapman was supported by the Humboldt Foundation, the Robert Koch Institute, Office for Academician Northwest University and an IDRC Grant while writing this paper. We would also like to thank Makerere University Biological Field Station and Uganda Wildlife Authority for their support.

Disclosure statement

No potential conflict of interest was reported by the authors.

Notes on contributors

Dipto Sarkar is a faculty in the Department of Geography at National University of Singapore. His research focuses on geographic information science, social networks, and computational social sciences. He is particularly interested in applying his methods to understand the human dimensions of biodiversity conservation.

Clio Andris is an assistant professor of geography at The Pennsylvania State University where she serves as director of the Friendly Cities Lab. Her research interests are geographic information science, social networks, interpersonal relationships, urban planning and spatial data mining.

Dr. Colin A. Chapman's research focuses on how the environment influences animal abundance and social organization and given their plight, he has applied his research to primate conservation. He has published 450+ articles, been cited 31000+ times, has a H factor of 97 and has received ~ $11 million in research funding and ~10 million in training grants. He has received a number of prestigious awards (Fellow of the Royal Society of Canada, Killam Fellow, Konrad Adenauer

Research Award from the Alexander von Humboldt Foundation, Anderson Teacher Scholar), was appointed as a Conservation Fellow to the Wildlife Conservation Society and as an advisor to National Geographic, and received the Velan Award for Humanitarian Service. He has conducted research in Kibale National Park in Uganda for 30 years and is interested in the roles of food abundance, disease, nutrition, and stress in determining primate abundance and how to best to conserve the world's biodiversity, where he focuses on primates and recently elephants because of their plight. During this time, he has not just been an academic, but has devoted great effort to promoting conservation by help the rural communities in the area he works.

Prof. Raja Sengupta is an Associate Professor in the Department of Geography and School of Environment at McGill University, Montreal, Quebec, Canada. His research interests in GIScience include Agent-Based Models (ABMs) of both human and primates, with his recent research focusing on behaviour and movement of red colobus monkeys in Kibale National Park, Uganda. He is also interested to understand how resource sites can be locations for transmissions of infectious diseases (as verified using ABMs), and is using network analysis to understand both spread of diseases and landscape-based factors that affect movement patterns of primates and other animals. He has published 35 journal papers, 10 book chapters, and one edited volume on related topics.

References

Abizaid, C., Coomes, O.T., and Perrault-Archambault, M., 2016. Seed sharing in amazonian indigenous rain forest communities: a social network analysis in three Achuar villages, Peru. *Human Ecology*, 44, 577–594. doi:10.1007/s10745-016-9852-7

Abizaid, C., *et al*., 2015. Social network analysis of peasant agriculture: cooperative labor as gendered relational networks. *The Professional Geographer*, 67 (3), 447–463. doi:10.1080/00330124.2015.1006562

Adams, J., Faust, K., and Lovasi, G.S., 2012. Capturing context: integrating spatial and social network analyses. *Social Networks*, 34 (1), 1–5. doi:10.1016/j.socnet.2011.10.007

Adams, W.M., 2004. Biodiversity conservation and the eradication of poverty. *Science*, 306 (5699), 1146–1149. doi:10.1126/science.306.5698.956a

Andris, C., 2016. Integrating social network data into GISystems. *International Journal of Geographical Information Science*, 8816 (March), 1–23. doi:10.1080/13658816.2016.1153103

Barrat, A., *et al*., 2004. The architecture of complex weighted networks. *Proceedings of the National Academy of Sciences*, 101 (11), 3747–3752. doi:10.1073/pnas.0400087101

Boccaletti, S., *et al*., 2006. Complex networks: structure and dynamics. *Physics Reports*, 424 (4–5), 175–308. doi:10.1016/j.physrep.2005.10.009

Boessen, A., *et al*., 2017. The built environment, spatial scale, and social networks: do land uses matter for personal network structure? *Environment and Planning B: Urban Analytics and City Science*, 45 (3), 239980831769015.

Bonacich, P., 1987. Power and centrality: a family of measures. *American Journal of Sociology*, 92 (5), 1170–1182. doi:10.1086/228631

Borgatti, S.P., 2005. Centrality and network flow. *Social Networks*, 27 (1), 55–71. doi:10.1016/j.socnet.2004.11.008

Borgatti, S.P., Everett, M.G., and Freeman, L.C., 2002. *UCINET 6 for windows: software for social network analysis*. doi:10.1044/1059-0889(2002/er01)

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

Butts, C.T., *et al*., 2012. Geographical variability and network structure. *Social Networks*, 34 (1), 82–100. doi:10.1016/j.socnet.2011.08.003

Castells, M., 1996. *The rise of the network society*. Oxford, UK: Blackwell.

Croitoru, A., *et al*., 2014. Linking cyber and physical spaces through community detection and clustering in social media feeds. *Computers, Environment and Urban Systems*, 53, 47–64. doi:10.1016/j.compenvurbsys.2014.11.002

Dijkstra, E.W., 1959. A note on two problems in connexion with graphs. *Numerische Mathematik*, 1 (1), 269–271. doi:10.1007/BF01386390

Emch, M., *et al*., 2012. Integration of spatial and social network analysis in disease transmission studies. *Annals of the Association of American Geographers. Association of American Geographers*, 105 (5), 1004–1015. doi:10.1080/00045608.2012.671129

Entwisle, B., Rindfuss, R.R., and Faust, K., 2007. Networks and contexts : variation in the structure of social ties 1. *American Journal of Sociology*, 112 (5), 1495–1533. doi:10.1086/511803

Erdős, P. and Rényi, A., 1960. On the evolution of random graphs. *Publication of the mathematical institute of the hungarian academy of sciences,* 5, 17–61.

Festinger, L., Back, K.W., and Schachter, S., 1950. *Social pressures in informal groups: a study of human factors in housing*. Stanford, CA: Stanford University Press.

Fischer, C.S., 1977. *Networks and places: social relations in the urban setting*. New York: Free Press.

Fischer, C.S., 1982. *To dwell among friends: personal networks in town and city*. Chicago, IL: Chicago University Press.

Fleming, L. and Sorenson, O., 2001. Technology as a complex adaptive system: evidence from patent data. *Research Policy*, 30 (7), 1019–1039. doi:10.1016/S0048-7333(00)00135-9

Freeman, L.C., 1978. Centrality in social networks conceptual clarification. *Social Networks*, 1 (3), 215–239. doi:10.1016/0378-8733(78)90021-7

Gibson, H., Faith, J., and Vickers, P., 2013. A survey of two-dimensional graph layout techniques for information visualisation. *Information Visualization*, 12 (3–4), 324–357. doi:10.1177/1473871612455749

Haggett, P. and Chorley, R.J., 1969. *Network analysis in geography*. London: Edward Arnold.

Hanneman, R.A. and Riddle, M., 2005. *Introduction to social network methods*. Riverside, CA: University of California, Riverside.

Hipp, J.R., *et al*., 2014. Examining the social porosity of environmental features on neighborhood sociability and attachment. *PloS one*, 9 (1), e84544. doi:10.1371/journal.pone.0084544

Hipp, J.R. and Perrin, A.J., 2009. The simultaneous effect of social distance and physical distance on the formation of neighborhood ties. *City & Community*, 8 (1), 5–25. doi:10.1111/j.1540-6040.2009.01267.x

Knoke, D. and Burt, R.S., 1983. Prominence. *In*: R.S. Burt and M.J. Mino, eds. *Applied network analysis*. Beverly Hills, CA: SAGE, 195–222.

Krzywinski, M., *et al*., 2012. Hive plots--rational approach to visualizing networks. *Briefings in Bioinformatics*, 13 (5), 627–644.

Li, C., *et al*., 2015. Correlation between centrality metrics and their application to the opinion model. *European Physical Journal B*, 88 (3), 1–13. doi:10.1140/epjb/e2015-50671-y

Li, C., *et al*., 2011. The correlation of metrics in complex networks with applications in functional brain networks. *Journal of Statistical Mechanics*, 2011 (11), P11018. doi:10.1088/1742-5468/2011/11/P11018

Liben-Nowell, D., *et al*., 2005. Geographic routing in social networks. *Proceedings of the National Academy of Sciences of the United States of America*, 102 (33), 11623–11628. doi:10.1073/pnas.0503018102

Lund, H., 2003. Testing the claims of new urbanism: local access, pedestrian travel, and neighboring behaviors. *Journal of the American Planning Association*, 69 (4), 414–429. doi:10.1080/01944360308976328

Luo, W. and MacEachren, A.M., 2014. Geo-social visual analytics. *Journal of Spatial Information Science*, 8 (8), 27–66.

McPherson, M., Smith-Lovin, L., and Cook, J.M., 2001. Birds of a feather: homophily in social networks. *Annual Review of Sociology*, 27, 415–444. doi:10.1146/annurev.soc.27.1.415

Mok, D., Wellman, B., and Carrasco, J., 2010. Does distance matter in the age of the internet? *Urban Studies*, 47 (13), 2747–2783. doi:10.1177/0042098010377363

Newman, M.E.J., 2001. Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality. *Physical Review E*, 64 (1), 016132. doi:10.1103/PhysRevE.64.016132

Nordbeck, S., 1964. Computing distances in road nets. *Papers of the Regional Science Association*, 12 (1), 207–220. doi:10.1007/BF01941253

O'Sullivan, D., 2014. Spatial network analysis. *In*: M. Fischer and P. Nijkamp, eds. *Handbook of regional science*. Berlin, Heidelberg: Springer Berlin Heidelberg, 1253–1273.

Onnela, J.-P., *et al*., 2011. Geographic constraints on social network groups. *PloS one*, 6 (4), e16939. doi:10.1371/journal.pone.0016939

Opsahl, T., Agneessens, F., and Skvoretz, J., 2010. Node centrality in weighted networks: generalizing degree and shortest paths. *Social Networks*, 32 (3), 245–251. doi:10.1016/j.socnet.2010.03.006

Owen-Smith, J. and Powell, W.W., 2004. Knowledge networks as channels and conduits: the effects of spillovers in the boston biotechnology community. *Organization Science*, 15, 5–21. doi:10.1287/orsc.1030.0054

Papachristos, A.V., Hureau, D.M., and Braga, A.A., 2013. The corner and the crew: the influence of geography and social networks on gang violence. *American Sociological Review*, 78 (3), 417–447. doi:10.1177/0003122413486800

Porter, M.A., Onnela, J.-P., and Mucha, P.J., 2009. Communities in networks. *Notices of the AMS*, 56 (9), 1082–1097.

Preciado, P., *et al*., 2011. Does proximity matter? Distance dependence of adolescent friendships. *Social Networks*, 34 (1), 18–31.

Radil, S.M., Flint, C., and Tita, G.E., 2010a. Spatializing social networks: using social network analysis to investigate geographies of gang rivalry, territoriality, and violence in los angeles. *Annals of the Association of American Geographers*, 100 (2), 307–326. doi:10.1080/00045600903550428

Radil, S.M., Flint, C., and Tita, G.E., 2010b. Spatializing social networks: using social network analysis to investigate geographies of gang rivalry, territoriality, and violence in los angeles. *Annals of the Association of American Geographers*, 100 (2), 307–326. doi:10.1080/00045600903550428

Sarkar, D., *et al*., 2016. Countering elephant raiding with short message service: challenges of deploying public participation-based systems in a setting with sparse information communication technologies resources. *The Canadian Geographer/Le Géographe Canadien*, 60 (4), 493–504. doi:10.1111/cag.12314

Sarkar, D., *et al*., In Press. A tiered analysis of community benefits and conservation engagement from the makerere university biological field station, Uganda. *Professional Geographer*. doi:10.1080/00330124.2018.1547976.

Sarkar, D., Sieber, R., and Sengupta, R., 2016. GIScience considerations in spatial social networks. *In*: J. Miller, D.O. Sullivan, and N. Wiegand, eds. *Geographic Information Science. GIScience*, *Lecture notes in computer science*, 9927. Cham: Springer International Publishing, 85–98.

Sarzynska, M., *et al*., 2016. Null models for community detection in spatially embedded, temporal networks. *Journal of Complex Networks*, 4 (3), 363–406. doi:10.1093/comnet/cnv027

Shaw, S.-L. and Sui, D., 2018. Editorial: GIScience for human dynamics research in a changing world. *Transactions in GIS*, 22 (4), 891–899. doi:10.1111/tgis.v22.4

USGS. 2006. *Shuttle radar topography mission*.

Valente, T.W., *et al*., 2008. How correlated are network centrality measures? *Connections*, 28 (1), 16–26.

Verdery, A.M., *et al*., 2012. Social and spatial networks: kinship distance and dwelling unit proximity in rural Thailand. *Social Networks*, 34 (1), 112–127. doi:10.1016/j.socnet.2011.04.003

Wasserman, S. and Faust, K., 1994. *Social network analysis: methods and applications*. Cambridge, UK: Cambridge University Press.

Wong, L.H., Pattison, P., and Robins, G., 2006. A spatial model for social networks. *Physica A: Statistical Mechanics and Its Applications*, 360 (1), 99–120. doi:10.1016/j.physa.2005.04.029