Spatial Social Networks: Exploring theoretical and methodological challenges

by

Dipto Sarkar

Department of Geography

McGill University

Montreal, Quebec, Canada

A thesis to the McGill University in partial fulfillment of the requirements of the degree of Doctor of Philosophy.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

BRIC	Brazil, Russia, India, and China
CR	Control Respondent
G7	Canada, France, Germany, Italy, Japan, the United Kingdom, and the United States
GWR	Geographically Weighted Regression
KHCC	Kibale Health and Conservation Centre
KNP	Kibale National Park
KSRP	Kibale Snare Removal Program
LDA	Latent Dirichlet Allocation
MHC	Mobile Health Clinic
MUBFS	Makerere University Biological Field Station
NGS	National Geographic Society
NTFP	Non-Timber Forest Product
PA	Protected Area
PI	Principle Investigator
SE	Standard Error
SNA	Social Network Analysis
SS	Socio-spatial
SSN	Spatial Social Networks
UWA	Uganda Wildlife Authority
WESP	World Economic Situation and Prospects

Abstract

Social Network Analysis (SNA) relies on a network structure composed of nodes connected via edges to represent entities (e.g., people as nodes) and relationships (e.g., friendships as edges) between them. The nodes and edges can be augmented with multiple attribute information (e.g. age, sex, relationship type) to enhance the insights available from SNA. Even though societies are embedded in geographic space which impact the formation, maintenance, and dissolving of social ties, hitherto adding spatial considerations in SNA has received relatively less critical attention because geographic embedding reshapes the structure of network and its processes, and thus cannot be treated akin to other attribute information. This dissertation looks at the challenges and opportunities of incorporating spatial information in SNA and introduces new methodological approaches to leverage socio-spatial properties of such networks, hitherto termed Spatial Social Networks (SSNs).

Introducing spatial information in SNA comes with its challenges. The first is deciding on the sophistication of the incorporation of spatial information in the social network. In response, we create a typology of existing research focused on the integration of geography and SNA. Additionally, although there is a long tradition of network analysis in Geography, SSNs require a new perspective for understanding social networks in the context of GIScience. For example, distance, community, and scale are three concepts that resonate in both fields and offers potential opportunities for understanding the socio-spatial properties that are modelled through SSNs.

In SNA, networks are abstract representations of a system which model conceptual relationships (e.g. friendship, collaborations) between entities (e.g. people, organisations). Thus, unlike road networks, where both the nodes and edges have explicit spatiality, a SSN can incorporate spatial information in different ways in its node and edge structure. Thus, SSNs are not constrained to its most common manifestation of incorporating spatial information into only the nodes in the form of nominal location or (x, y) co-ordinates. We create three conceptualizations of SSNs from a single National Geographic grants dataset that incorporate spatial information differently to highlight the different ways in which spatial information can be incorporated in the node-and-edge network structure. The three-different SSN highlight varied

spatial relationships latent in the dataset, and analysing them provides new insights into global and regional trends of research collaborations.

Further, SNA relies on metrics to extract meaningful information about network structure from underlying topological node-and-edge structure. In SSNs with geolocated nodes, non-spatial metrics provide limited insight into the socio-spatial structure of the networks. We introduce a new set of metrics which can be used to identify important nodes in a socio-spatial context. We prove the efficacy of the new metrics on two simulated networks as well as on a real-world network of economic benefits.

Finally, while SSNs are a unique way of understanding society, it provides a single dimensional view of a multi-faceted social system as it over-privileges connections above other social dimensions. Thus, SNA should be complimented with qualitative and quantitative analysis to provide complete understanding of the system under study. While use of the new metrics on the network of economic relationships originating from the research field station located at Kibale National Park helps understand the network structure and identify important individuals responsible for spreading the economic benefits through the community, additional analysis helps understand the role of the research field station in shaping the community-park relationship across space that could not be captured by only modelling the flow of economic benefits through social connections.

Résumé

L'analyse de Réseau Sociale (ARS) compte sur une structure de réseau de nœuds et de liens pour représenter des entités (par ex., les gens) et des relations (par ex., des amitiés). Les nœuds et les bords peuvent être enrichis avec des informations complémentaires (par ex., l'âge, le sexe, le type de relation). Bien que les sociétés se développent dans un contexte géographique qui influence la formation, le maintien et la dissolution de liens sociaux, l'ajout de considérations spatiales à ARS a reçu peu d'attention parce que l'ancrage géographique réorganise la structure du réseau et de ses processus et doit donc être traité différemment. Cette dissertation examine les défis et les opportunités d'incorporer des informations spatiales dans l'ARS et présente des nouvelles méthodologie afin de profiter des propriétés socio-spatiales d'un tel réseau, dès lors nommé Réseaux Sociaux Spatiaux (RSSs).

Inclure des informations spatiales dans l'ARS présente un ensemble unique de défis. Le premier défi est associé au choix du niveau de sophistication d'informations spatiales à incorporer. En réponse, nous avons créé une typologie de la recherche existante qui porte sur l'intégration de géographie et de l'ARS selon le niveau de sophistication. Malgré une longue tradition d'analyse de réseau en géographie des phénomènes humains et physiques, la réapparition d'ARS dans des champs divers exige une nouvelle perspective pour comprendre les réseaux sociaux dans le contexte de la Science de l'Information Géographique. Par exemple, la distance, la communauté et l'échelle sont trois concepts pertinents aux deux disciplines et offre l'opportunité de comprendre les propriétés socio-spatiales qui sont modelées par des RSS.

Les réseaux sont des représentations abstraites d'un système qui modélisent des relations conceptuelles entre les entités. Ainsi, un réseau social peut incorporer des informations spatiales de différentes façons dans ses nœuds et la structure des liens. Les RSSs peuvent aller au delà de l'intégrer de l'information spatiale dans nœuds sous forme de localisation nominale ou (x, y) ou de coordonnées. En utilisant une grande base de données de subventions de National Géographic, nous créons trois conceptualisations de RSS qui incorporent des informations spatiales différémments. Ces trois RSS mettent en évidence diverses relations spatiales latentes dans l'ensemble de données et leur analyse fournit de nouveaux aperçus des tendances mondiales et régionales dans la collaboration en recherche.

En outre, l’ARS compte sur des métriques pour extraire des informations significatives du graphique relationnelle sous-jacent. Dans l’ARS, la métrique non-spatiale fournit un aperçu limité de la structure socio-spatiale du réseau. Nous fournissons de nouveaux métriques pour les réseaux sociaux spatiaux qui fournissent une compréhension des propriétés socio-spatiales du réseau et l’identification des nœuds importants dans un contexte socio-spatial. Nous prouvons l'efficacité des nouveaux métriques sur deux réseaux simulés et un réseau réel.

Finalement, quoique unidimensionnel, les RSS offrent une façon unique de mieux comprendre un système social, en privilégiant les liens géographiques. Pour une vision globale du système à l’étude, l’ARS devrait être complété par une analyse qualitative et qualitative. L'utilisation de nouveaux métriques sur le réseau social de la station de recherche du Parc nationale Kibale facilite la compréhension de la structure du réseau et identifie les individus importants responsables du partage des avantages économiques au sein de la communauté. De plus, les analyses permettent de mieux comprendre comment la station de recherche, en fournissant une gamme de services supplémentaires, influence la relation entre le parc et la communauté. L’effet des ses services ne peut pas être perçu par une simple modélisation des flux d'avantages économiques permis par les connections sociales.

Acknowledgements

Behind every achievement is a team of people that help build the scaffoldings to make success possible. Here are some of the people without whom this dissertation would not be possible.

I would like to express the deepest appreciation to my advisors Dr. Raja Sengupta and Dr. Colin Chapman. They continually and convincingly conveyed a spirit of adventure in regard to research and scholarship, and an excitement with regards to teaching. Without their guidance, encouragement, support, and persistent help this dissertation would not have been possible.

In addition, I would like to thank Dr. Clio Andris for her willing participation and enthusiasm in guiding and collaborating on projects. Her ideas on the social implications captured by spatial social networks were very insightful and thought provoking. I would also like to extend my gratitude to Dr. Renee Sieber. She motivated me to find the relevant literature to help situate spatial social networks in the larger context of the discipline of geography.

In this dissertation, I primarily use two different datasets. I am grateful to National Geographic Society, specifically Dr. Christopher Thornton, Mr. Michael Jourdan, Dr. Peter Raven, and Ms. Angie Sanders for providing access to the dataset, helping with the analysis, and confirming the findings. The dataset in Kibale National Park, Uganda could not have been collected without the help of Mr. Jimmy Ogwang, Dr. Patrick Omeja, Mr. Wilson Kagoro, and Ms. Scarlet C. Angom.

The Department of Geography at McGill University also played a very important role in providing a congenial and vibrant workspace. I would like to thank Dr. Nigel Roulet, Dr. Sebastian Breau, and Dr. Oliver Coomes for providing financial and institutional support through the course of this dissertation. I would like to extend my gratitude to Mr. Warren and Ms. Cowie for there generous support through the Rathlyn Awards and Alexandra Irwin Cowie Fellowship.

Last but not the least, the work centric life of the last four years would have been very difficult without the constant understanding, help, and encouragement of friends and family. This list is too long to enumerate exhaustively but Ms. Chandrabarna Saha, Mrs. Shikha Sarkar, Mr. Dipon Sarkar, Dr. Kim Valenta, Ms. Shannon Swikert, Mr. Basil Vandervort-Charbonneau, and Ms. Gabriella Fanous deserve special mention.

Contribution of Authors

The four chapters directly presenting the results of this thesis are written as journal manuscripts that have been published or are in review in different journals. The contributions of the co-authors are as follows:

GIScience Considerations in Spatial Social Networks

Published in J.A. Miller, D.O. Sullivan, and N. Wiegand, eds. *Lecture Notes in Computer Science*. Springer International Publishing, 85–98.

Author	Contribution
Dipto Sarkar	Literature review, idea development, and manuscript writing
Renee Sieber	Idea development and manuscript writing
Raja Sengupta	Guided manuscript writing and idea development

Understanding Research Collaborations and Connectivity through Spatial Social Networks: Analysis of 126 years of Grantmaking

In review at the *Annals of the American Association of Geographers.*

Author	Contribution
Dipto Sarkar	Idea development, data analysis, and manuscript writing
Colin A. Chapman	Data acquisition
Raja Sengupta	Guided manuscript writing and idea development

Metrics for Characterizing Network Structure and Node Importance in Spatial Social Networks

In review at the *International Journal of Geographic Information Science.*

Author	Contribution
Dipto Sarkar	Idea development, statistical methodology, data analysis, and manuscript writing
Clio Andris	Idea development, and manuscript writing
Colin A. Chapman	Data acquisition
Raja Sengupta	Guided manuscript writing, idea development and spatial analysis

Research Stations as Conservation Instruments Provide Long Term Community Benefits Through Social Connections

Accepted subject to minor revisions at *The Professional Geographer.*

Author	Contribution
Dipto Sarkar	Idea development, data analysis, and manuscript writing
Colin A. Chapman	Data acquisition, and manuscript writing
Kim Valenta	Idea development, and manuscript writing
Scarlet C. Angom	Data acquisition, and transcription
Wilson Kagoro	Data acquisition

Raja Sengupta	Guided spatial analysis, manuscript writing, and idea development

1. Introduction

1.1. Background

1.1.1. The Networks Paradigm and SNA

The idea of modelling systems as networks originated in the 18th century (Alexanderson 2006, Barabási 2015, sec. 2.1), as a means of modelling a uniquely geographical problem, namely, whether it was possible to visit the four land masses of the European city of Königsberg by traversing the seven bridges over the Pregel River exactly once. Euler abstracted the complex cityscape into a simplified node- and edge-based structure (Figure 1.1), which he referred to a graph and conclusively demonstrated a solution to this conundrum using the topological properties of this node and edge structure. In this specific case, the landmasses were represented as nodes while the bridges connecting the landmasses consisted of the edges of the graph. Subsequent growth of interest in such node and edge topological structures lead to the emergence of the field of graph theory, including how to characterize connections within graphs, identify important nodes, and characterize emergent properties (Clark and Holton 1991, Barabási 2015).

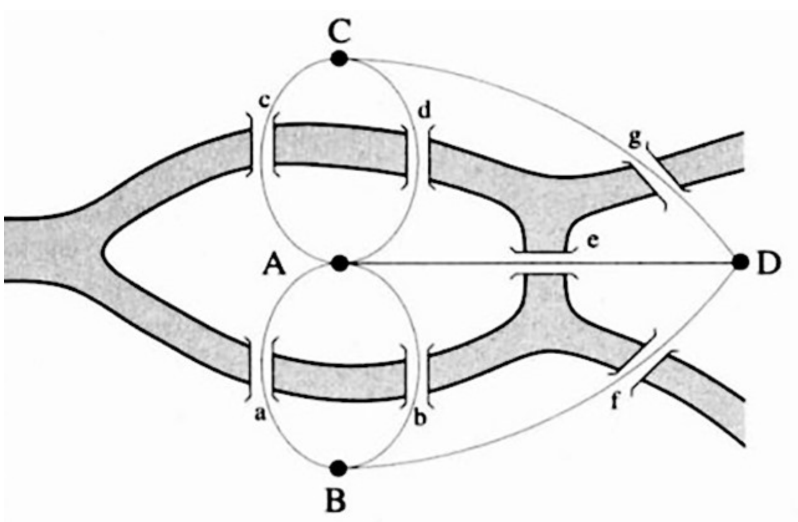


Figure 1.1.1: The seven bridges of Könisberg and its corresponding graph based representation with bridges shown as 'lines' labelled with small letters, connecting landmasses represented as 'dots' and labelled with capital letters.[1]

---

[1] Image source: http://math.stackexchange.com/questions/1173328/eulers-solution-of-seven-bridges-of-k%C3%B6nigsberg-in-layman-terms Uploaded by user 'blackened'. License: Creative Commons Attribution-ShareAlike (CC-BY-SA). Accessed: 02 May 2018

The application of graph theory concepts to empirical real-world networks is known as network science (Barabási 2015, sec. 2.2). Networks have been used to represent a variety of systems including the human brain (Greicius *et al.* 2003, Eguíluz *et al.* 2005, Iturria-Medina *et al.* 2008), to cosmological phenomenon (Krioukov *et al.* 2012), in addition to systems that are easier to conceptualize as networks, such as, communication infrastructure (e.g. internet, mobile phone) (Barabasi and Albert 1999, Barabási *et al.* 1999, Karsai *et al.* 2011) and transport infrastructure (e.g. road and railway networks) (Sen *et al.* 2003, Jiang 2007, Eppstein and Goodrich 2008, Chen and Hu 2013). Social network analysis (SNA) is a subset of this field, and primarily deals with social connections between entities. However, the term SNA is also used loosely to refer to systems that have been modelled as networks in which the edges represent conceptual, rather than physical, entities, e.g. friendship, kinship, trade relations (Wasserman 1994, Castells 1996, Fleming and Sorenson 2001, Durland and Fredericks 2005, Serrat 2017). While some scholars describe the SNA approach as a set of tools for "theory development" or "strategic approach", others have argued that it should be considered a paradigm in its own right (Kadushin 2012, Prell 2012, Buch-Hansen 2014, Colchester 2016), as it provides a set of methods and assumptions that can constitute a fundamentally different perspective of the world.

1.1.1. Adoption of SNA in the Social Sciences

Unlike Euler's original 18th century network (where the graphs and edges represented physical infrastructure like bridges), social networks describe conceptual relationships between entities using a node and edge structure. The entities represented by nodes can be people, countries, and organizations; while the relationship depicted by the edges is usually a conceptual connection (e.g., friendship, kinship, trade relation), depending on the system and the particular aspect of the system under study (Wasserman 1994, Durland and Fredericks 2005, Borgatti *et al.* 2009b, Serrat 2017).

The basic premise here is to define a society as a group of entities with persistent interactions. In this perspective, the social network is conceptually represented as a graph, and consequently as collection of nodes and edges. Specifically, a network is usually expressed as a non-directed graph $G = (V, E)$ where the set $V = v_1, v_2, v_3, \dots, v_n$ represents the set of nodes and the set $E = e_1, e_2, e_3, \dots, e_n$ is the set of edges. Each edge e_k is associated with a pair of vertices *(i, j)*. Consequently, this graph based structure of a social network is represented by a *n x n* matrix called an *adjacency matrix (A)*, where the entries ($A_{i,j}$) represent known connections between a pair of nodes (*i and j*). The binary values in A represent presence or absence of interaction whereas non-binary entries signify some measure of intensity of interaction, for example, number of years as a surrogate for acquaintance strength or number of trade agreements between countries as a measure of trade relationships. The edges in the graph can also incorporate directionality. In a non-directional graph, relationships are considered to be reciprocal. Thus, an edge between two nodes denotes a two-way relationship. In some applications, the edges between the nodes may be non-reciprocal and hence, the graph is directional. In a directional graph, each edge represents a non-reciprocal relationship and consequently a pair of edges are required between a pair of nodes to denote a two-way relationship. The main focus of a social network is, however, on topology, that is, the relationships between the entities as captured by the nodes and edges (Wasserman 1994, Lazer *et al.* 2009), and consequently in this dissertation, I primarily focus on non-directed graphs.

SNAs have found a natural home in the social sciences for many reasons, particularly as a quantitative toolkit. First, as stated above, the term SNA has fit well with the need to define a collection of discrete entities depicted as nodes representing people, countries, and organizations that are strung together by edges that represent relationship such as, friendship, kinship, trade, sharing of ideas (Wasserman 1994, Durland and Fredericks 2005, Borgatti *et al.* 2009b, Serrat 2017). Thus, SNA as a concept to model complex fabric of interactions is intuitive and easy to grasp with its standard conceptualization of entities as nodes and relationships as edges. The emphasis on relationships afforded by the network data model (Wasserman 1994, Lazer *et al.* 2009) aides in shifting the focus of analysis beyond the individual entities to the pattern of interactions that make the system 'other than the sum of its parts'. This gives the ability to analyse the fabric of entities and relationships that constitutes the social system, and provides an understanding of how local processes drive group-level properties by considering the different social environments experienced by each individual (Strandburg-Peshkin *et al.* 2013, Farine and Whitehead 2015).

Second, the flexibility of the SNA technique lies in the high level of abstraction provided by singling out only the relationship aspect from the rest of the data (Newman 2010). The abstraction facilitates scalability, making the techniques applicable to networks of all sizes from small groups to global systems (Kadushin 2004). Most SNA techniques rely solely on the topological structure of the data as expressed connections between the nodes and edges. However, enriching the relational data with additional attributes can aide in gaining further insight into the processes of the social network, like "attribute based homophily" (e.g. people with similar tastes tend to know each other) (McPherson *et al.* 2001).

Third, the rise of interest in SNA is concomitant with the exponential increase in relevant data available through Web 2.0 technologies like social media sites and location sharing services (Freeman 2004, Fu *et al.* 2008, Lewis *et al.* 2008, Borgatti *et al.* 2009a, Bughin and Chu 2010). Significant interest from several disciplines (such as Computer Science, Sociology, Biology) have helped SNA transition from a simple representation-based analysis method to a comprehensive toolset with facilities for visualization, characterization with metrics, as well as complex algorithms for analysis of intricate processes and prediction (Otte and Rousseau 2002, Freeman 2004, Borgatti *et al.* 2009a, Scott 2017).

Because any analysis using SNAs is significantly dependent on the representational form adopted to highlight entities and their relationships, it therefore becomes important to understand precisely how to define what an edge, or connection, consists of, and what it represents. If two entities can be distinguished from one another, and have some kind of connection, then this "relationship" can be modelled using network analysis. However, in any given set of entities there can be multiple relationships between different entities, and these relationships can take on a multiplicity of forms. Thus, there are several potential ways in which edges can be defined, and each different form of edge employed in a given network analysis will result in a different network (Zuckerman 2008). Thus, the flexibility of the network model provides opportunities for representing multiple relationships embedded in a social system using different networks. However, this requires careful consideration about what information each of the networks capture and the overlap between them (Perkins *et al.* 2009, Ansari *et al.* 2011, Magnani *et al.* 2013, Boccaletti *et al.* 2014).

1.2. Research Motivation

Currently, a considerable amount of social science data have spatial identifiers (Goodchild *et al.* 2000, Goodchild and Janelle 2004, Ward and Gleditsch 2008). This geographical component of an ever-growing trove of data has important ramifications for analysis and the understanding the processes under study. Considerations for incorporating the spatial component in analysis has usually taken the form of using 'spatial lag' functions, i.e. using distance decay to account for the diminishing influence of distant values, or by specifying a stochastic spatial process to account for the error term (Anselin 1988, Ward and Gleditsch 2008, Radil 2011). However, the basic premise of these models is that nearby entities interact more, with the intensity of interaction reducing with distance as succinctly expressed by *the first law of geography* (Tobler 1970). In terms of statistical modelling, the interactions are formalized by using a $n \times n$ called the matrix of *spatial weight* or *connectivity matrix*. Each entry of the matrix represents interaction between the entries i and j in terms representing either contiguity or proximity. In case of contiguity, the entries are usually binary indicating physical connections between spaces, while in case of proximity based models each entry of W each entry represents the interaction between i and j based on the measure of a certain variable scaled according to the distance between them (as described in Radil 2011).

Social networks usually do not assume interaction amongst nearby entities. The focus is typically on known relationships or interactions between nodes irrespective of their relative positions in geographic space (Kossinets 2006, Borgatti *et al.* 2009a). Thus, there is an exclusive focus on the topology of the node and edge structure without environmental or geographic context in which the entities are situated. This representation can be problematic because societies and their relationships do not exist in isolation, but coexist among geographic features that affect these ties. Although technologically challenging, the geographic or spatial context has recently been recognized as an important factor of social networks (Adams *et al.* 2012) that influences individual decisions that drive network level dynamics (Strandburg-Peshkin *et al.* 2013, Farine and Whitehead 2015). Geographic applications of SNA to understand specific systems in recent years include epidemiology (Moore 2010, Giebultowicz *et al.* 2011a, 2011b, Keeling *et al.* 2011, Emch *et al.* 2012a), criminology (Radil *et al.* 2010a), trade networks (Castells 1996, Fleming and Sorenson 2001, Owen-Smith and Powell 2004), as well as cooperative agriculture practices (Entwisle *et al.* 2007, Abizaid *et al.* 2015, 2016), to highlight a few.

In the context of GIScience, the simplest method of incorporating spatial information in a social network has been to geolocate the nodes by incorporating the spatial information as an attribute of the node (Andris 2016). Focusing on the topology (that is, node and edge connectivity) as well as geographic embedding of the network provides insights into spatial considerations shaping social interactions. The simplest geographic determinant of social ties is the Euclidean distance between entities, and several studies have reiterated the existence of distance decay via the exponential decline of number of social ties with increasing distance (Wong *et al.* 2006a, Barthélemy 2010, Mok *et al.* 2010, Scellato *et al.* 2010, Preciado *et al.* 2011, Tranos and Nijkamp 2012).

However, only considering Euclidean distance is a rather simplistic model of geography. For example, studies have shown that modelling travel distance is more effective at explaining in-person interaction (Salonen *et al.* 2012). In addition, spatial layout of the environment (Eagle *et al.* 2009, Sevtsuk *et al.* 2009, Sailer and McCulloh 2012, Hirschi 2013, Boessen *et al.* 2017), distribution of resources (Lund 2003, Wineman *et al.* 2009, Hipp *et al.* 2014), and distribution of entities (Entwisle *et al.* 2007, Butts *et al.* 2012, Verdery *et al.* 2012, Kowald *et al.* 2015, Boessen *et al.* 2017) also impact how social connections form, persists and dissolve.

SNA's explicit focus on known interactions also provides insights into the notion of communities which in geography encompasses the concepts of dense network social interactions amongst entities co-located in a bounded space (Hillery 1955, Clark 1973, Frug 1996, Expert *et al.* 2011, Onnela *et al.* 2011, Daraganova *et al.* 2012). The detection of these meso-scale well-connected components in a spatially embedded social network also revisits the concept of regions as being well defined geographic areas with persistent interaction between its inhabitants. These regionalization approaches often match administrative delineations (Ratti *et al.* 2010, Calabrese *et al.* 2011, Rinzivillo *et al.* 2012, Sobolevsky *et al.* 2013, Coscia *et al.* 2014, Hawelka *et al.* 2014), but also provides insights into socio-cultural (Blondel *et al.* 2010, 2011, McMenamin 2017) and political (Mossa *et al.* 2005, Ratti *et al.* 2010) influences shaping the sub-structures within the fabric of the larger system.

The integration of spatial social networks comes with its own set of challenges. The linking of spatial and social information represents a "system of systems", where the spatial information becomes a part of a complex set of underlying topological relationships (Andris 2011). Thus, spatial social networks enable representations of spatial systems where interactions are not merely a function of distance. Since spatial distance alone is a misleading metric, traditional spatial statistical analysis techniques which rely on modelling proximity may not be sufficient. On the other hand, graph theoretical techniques that only leverage the topological connections are also insufficient for understanding the system that is modelled as a spatial social network as it disregards the nuances of the spatial aspects of the system. This is exemplified by Dalton's statement, *"While they [graph theorists] have enabled much progress to be made, they are not entirely suitable to geography…"* (Dalton 1973, p. 1), because the topological structure of the graph disregards the spatial characteristics of the system, such as, distance of connection and spatial arrangement of nodes. In addition to such systemic challenges, researchers have also pointed out that the node and edge structure of SNA provides an overly simplistic representation of a system which primarily privileges connections above other factors of a complex social system (Mizruchi 1994, Newman 2010, Buch-Hansen 2014) and consequently the need to supplement SNA with other qualitative and quantitative analysis (Crossley 2010, Edwards 2010, Schipper and Spekkink 2015).

In the light of the current status of SNA research, and the lacunae of its implementation with respect to the spatial domain, the aim of this dissertation is to build upon the rising interest of using SNA to understand spatial processes, and in creating a novel framework for addressing social network and spatial analysis together. I build upon the existing literature in geography and SNA to create new theories for integrating spatial information as an integral part of social network analysis. I have also provided new methodological approaches to dealing with spatial information in SNA. In regard to terminology, social networks that incorporate spatial information has been referred to by various terms, such as location based social networks (Ye *et al.* 2010, Scellato *et al.* 2011a), geo-social (Scellato *et al.* 2010, Luo and MacEachren 2014), anthrospaces (Andris 2016) and spatial social networks (Radil *et al.* 2010). This dissertation uses the term "*spatial social networks*" (or SSNs for short), as it highlights the social connections, and recognises the embeddedness of the interactions in geographic space, thus not confining geographic notions to just a pair of coordinate locations.

For the integration of spatial and social information, however, several theoretical and methodological considerations are required. I attempt to answer three specific questions in this dissertation: How can spatial information be incorporated into the structure of the network? What metrics can be used to describe the socio-spatial properties of the networks? And finally, what complementary existing methods of analysis are available to support spatial social network analysis?

1.3. Thesis Outline

This thesis builds upon the existing literature on SSNs with the goal of adding new theoretical and methodological perspectives. The chapters of this dissertation are based on articles that are in various stages of the publication process. Each chapter is prefaced with a summary and linking statement that positions the contribution of the chapter towards the overall goal of the thesis. Chapter 2 of the dissertation entitled *"GIScience considerations in Spatial Social Networks"* is a published literature review that looks for congruence in concepts between the fields of social network analysis (SNA) and geography. Further, this chapter also identifies several forward-looking concepts to better situate the concept of Social Networks in Geography. Specifically, it identifies two lacunae. First, that Geography is almost always represented by giving x, y coordinates to the nodes. This does not always have to be the case. Second, that there are potential opportunities for reconciling network and geographic concepts as means of providing a better understandings of SSNs. Chapter 2A provides additional bridging literature in support of these two arguments as a means of advancing the discussion on SSNs. Chapter 3 entitled *"Understanding Research Collaborations and Connectivity through Spatial Social Networks: Analysis of 126 years of Grantmaking by the National Geographic Society"*, moves away from the restrictive embedding rules of physical networks and allows social networks to incorporate spatial information in different forms in the network structure to model different relationships embedded in the data. Chapter 4, entitled "Metrics for characterizing network structure and node importance in spatial social networks" presents new metrics specifically for analyzing SSNs that consider both spatial and network characteristics. Chapter 5, entitled *"Research stations as conservation instruments provide long term community benefits through social connections"* highlights how quantitative and qualitative analysis can compliment SSN analysis, with a view to identifying and capturing other elements of the system under study that were missed by using a pure network approach.

2. GIScience Considerations in Spatial Social Networks

Citation: Sarkar, D., Sieber, R., and Sengupta, R., 2016. GIScience Considerations in Spatial Social Networks. *In*: J.A. Miller, D.O. Sullivan, and N. Wiegand, eds. *Lecture Notes in Computer Science*. Springer International Publishing, 85–98.

Linking statement: This chapter acts as as a foundation for this dissertation with a literature review that specifically looked for congruence in concepts between the fields of social network analysis (SNA) and geography. The concepts of distance, community, and scale resonate in both the fields and offer avenues for assimilating SNA and GIScience. The review also revealed that there exist three levels of sophistication with which spatial information has been incorporated as part of social network structure. In its simplest form, this involves attaching location information to the entities; which is sufficient for inferring distance decay of relationships. However, more sophisticated forms are required to consider the embedding of the network in geographic space. These insights address some of the shortcomings in the SSN literature essential for situating social networks in the context of geography and GIScience. Chapter 2A compliments Chapter 2 with background information pertaining to the network model, the various metrics used in SNA, and the major themes of research utilizing SNA and geography.

Summary: There has been a proliferation of literature that incorporates social network analysis (SNA) to study geographic phenomena. We argue that these incorporations have mostly been superficial. What is needed is a stronger interrogation of the challenges and possibilities of a tight coupling of spatial and social network concepts, which take advantage of the strengths of each methodology. In this paper, we create a typology of existing research focused on the integration of geography into SNA: nodal, topographic and spatial. We then describe three core concepts that co-exist in the two fields but are not necessarily complementary: distance, communities, and scale. We consider how they can be appropriated and how they can be more tightly coupled into spatial social networks. We argue that the only way we can move beyond a superficial integration is to holistically identify the challenges and consider new methods to address the complexities of integration.

2.1. Introduction

In recent years, Social Network Analysis (SNA) has generated considerable attention due to the distinctive ways in which it characterizes and prioritizes the relationships among entities. The diagrammatic approach of social networks serves as a starting point for visual exploratory analysis. Social Network Analysis's foundation in graph theory provides a strong backbone for deriving metrics to analyze network patterns. The basic premise of a social network is to define a society as a group of entities with persistent interactions and shared attributes. A society can refer to a group of people sharing the same territory, subject to the same laws, interested in same activities (forming clubs), and belonging to the same economic or social status. It is the common attribute(s) between the individuals that gives rise to social interactions. In geography, the concept of a society is formed on the common attribute, irrespective of interactions between the entities. In SNA, it is not presumed that similar entities will interact. Thus, the focus is on the explicit interactions. This provides an interesting avenue to uncover patterns, discover important individuals and reveal interesting facts about the society, which a procedure starting with the assumption that everyone with similar qualities interact with each other may not provide. Thus, SNA provides a complimentary approach, focusing on studying individuals, groups and ultimately the society by concentrating on the known interactions that exist between the entities.

Both methods of enquiry, spatial and social analysis, can however benefit each other by a coupling of knowledge.

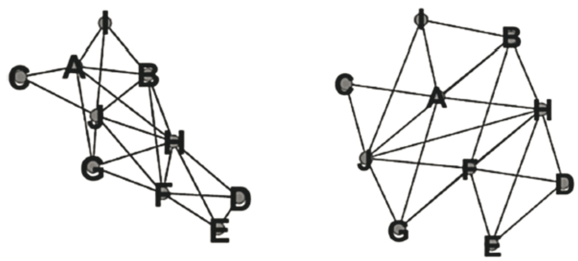
In this paper, we explore some of the requirements for tighter integration of SNA with geography. Although Barthélemy (2011) highlighted the long tradition of network analysis in Geography, whereby both human and physical phenomenon have been modeled as networks (Haggett and Chorley 1969), the resurgence of SNA in various fields compels revisiting the tradition and offer new perspectives for understanding social networks in the context of GIScience. Despite a proliferation of research that integrates geographic aspects in SNA, the current literature lacks a framework for classifying the different methods by which the integrations have been accomplished. We introduce a typology of integrating geography and SNA in the current literature. We then highlight three concepts commonly used in SNA and geography but warrant deeper understanding of what they mean in either context, highlighting the problems as a way to form working definitions required in the realm of spatial social networks. We hope to offer additional interesting avenues for exploring interconnectivity and interactions between people and also between people and their surroundings. Referred to by various terms, such as location based social networks (Ye *et al.* 2010, Scellato *et al.* 2011a), geo-social (Scellato *et al.* 2010), and spatial social networks (Radil *et al.* 2010a), we prefer the term Spatial Social Networks as it highlights the social connections, recognises the embeddedness of the interactions in geographic space, not confining geographic notions to just a pair of coordinate locations.

2.2. Social Network Analysis (SNA)

Social Network Analysis represents relationships between connected entities such as individuals, organizations, and groups. In SNA, a social network is computationally represented as a collection of nodes and edges. Specifically, a network is usually expressed as a non-directed graph defined G = (V, E) where V = $v_1, v_2, v_3, \ldots, v_n$ represents the set of nodes and E = $e_1, e_2, e_3, \ldots, e_n$ is the set of edges. Each edge e_k is associated with an unordered pair of vertices (i, j). In some applications, the edges between the nodes are non-reciprocate and hence the graph is directed, where each edge e_k is associated with an ordered pair of vertices (u, v). The main focus of the social network is on the edges (or ties), that is, the relationships that exist amongst the nodes (Wasserman 1994). Using graphs to represent social networks limits the possibility of self-loops, as in terms of social relationships, the concept of a person being a friend with themselves does not make little sense from the modelling aspect.

Most social networks are unweighted graphs. The presence of an edge between two nodes is binary, indicating whether there exists a relationship amongst the two nodes or not. In unweighted graphs, the edges do not convey any other information besides connectivity of nodes. Hence, navigation in the network space is only possible by moving along existing edges from node to node (like navigation on a road network). A sociogram is a visualization of the social network. In the sociogram, the widths, or the lengths of the edges are arbitrary. The nodes in a sociogram are located with the attempt to show interconnected nodes close to each other (Krzywinski *et al.* 2012). The position of the nodes in the layout is not directly interpretable on its own on a Cartesian plane and only has meaning in relation to other nodes (Jacomy *et al.* 2014). Any scaling or rotation of the sociogram does not change the underlying information (Hanneman and Riddle 2005).

Various graph layouts have been developed to produce aesthetically pleasing drawings by modifying the position of the nodes and edges and by changing the length of the edges (Battista *et al.* 1994). The drawings should not be confused with the graph itself; very different layouts can correspond to the same graph (Battista *et al.* 1994). Figure 2.1a and b shows the same graph represented with two different layout algorithms applied; the adjacency matrix of nodes and edges is shown in Fig. 1c. In the adjacency matrix (Fig. 2.1c), each non-diagonal entry, aij, is the existence of an edge connecting node i to node j. Usually the entries aij are binary and denote the existence or non-existence of an edge between the two nodes i and j. Unlike sociograms, there exists a unique adjacency matrix for each graph (up to permutations of rows and columns) (Garrido *et al.* 2009).



	A	B	C	D	E	F	G	H	I	J
A	0	1	1	0	0	0	1	1	1	0
B	0	0	0	0	0	1	0	1	1	1
C	0	0	0	0	0	0	0	0	0	1
D	0	0	0	0	1	1	0	1	0	0
E	0	0	0	0	0	1	0	1	0	0
F	0	0	0	0	0	0	1	1	0	1
G	0	0	0	0	0	0	0	1	0	1
H	0	0	0	0	0	0	0	0	0	1
I	0	0	0	0	0	0	0	0	0	1
J	0	0	0	0	0	0	0	0	0	0

(c)

Figure 2.1: Various representations of the same social network. (a) Sociogram with ForceAtlas2 layout (Jacomy *et al.* 2014) *(b) sociogram with Fruchterman–Reingold layout* (Fruchterman and Reingold 1991) *(c) adjacency matrix.*

Increasing amounts of data can now be used to nuance social relations. In many cases, an entity can be thought of as a cluster of structured or unstructured attributes, distinguished by a unique identifier. The emergence of big data has spawned new perspectives on social network structures (Watts and Strogatz 1998, Barabasi and Albert 1999, Albert and Barabasi 2001), energized the development of new metrics (Newman 2002, Opsahl 2009), and increased the availability of software and libraries (Borgatti *et al.* 2002, Hagberg *et al.* 2008, Bastian *et al.* 2009). Geography is increasingly playing a role in characterizing social networks. However, geography often tends to be treated similarly to other attributes and has, until recently, merited little critical attention about how it can be coupled with SNA to exploit spatial embeddedness of the network. In this paper, we use the term Geography rather broadly to refer to the field itself; a term synonymous with describing “the earth’s surface from a standpoint of distributions and interactions” (De Geer 1923).

2.3. Existing Methods of Coupling Geography and SNA

Numerous articles have discussed spatial social networks. We characterize the literature into three main types.

First, in its simplest form, articles treat geography as a nodal attribute (Entwisle 2007, Crandall *et al.* 2010, Cranshaw *et al.* 2010, Pelechrinis and Krishnamurthy 2015). This approach has location information of the entities stored in terms of nominal location (e.g., gazetted placenames). The location is often treated similarly to other nodal attributes (e.g., age, gender). The location provides information about similarity among different entities. This allows us to infer location based homophily (i.e. propinquity) or to consider places as promoters of social tie formation and maintenance. The main benefits of this type are that it is easier to treat one attribute like any other and does not demand the geometric transformation of a nominal location to a feature type. The analysis methods are usually rooted in SNA, as opposed to spatial analysis, and locational information merely adds more context to aid the SNA.

Second, researchers may treat location of the entities as a topographic attribute (Nag 2009, Batty *et al.* 2012, Comber *et al.* 2012, Koylu *et al.* 2014). This is a more sophisticated way of integrating spatial information into social networks by associating x, y locations with nodes, edges, or both. In this type, edges can take on two meanings: the social connection between two entities or the physical path between the entities. Having x, y locations aids representation of the social network on a Cartesian space enabling use of visual as well as spatial analysis techniques to understand the spatial characteristics. This type of research is typically accompanied by basemaps to visualize the social network. More importantly, these methods fit social relationships on to a Cartesian space. The fitting of the social network to Cartesian space makes spatial analysis, such as kernel densities, a more vital component than in the nodal attribute treatment. Nonetheless, reducing geography to be a mere nodal attribute or simply to a x, y pair to be rendered on or analyzed against a map disregards the nuanced effects on actors and tie formation in the social network.

Third, is the treatment of geography as a spatial property of the network (Kwan 2007, Radil *et al.* 2010a, Expert *et al.* 2011, Onnela *et al.* 2011, Butts *et al.* 2012, Daraganova *et al.* 2012, Doreian and Conti 2012, Luo and MacEachren 2014, Andris 2016). This not only considers the geographic locations of the nodes and/or edges but also exploits spatial properties and patterns to infer spatio-temporal characteristics of the network. Common ways of considering spatial aspects include not only Euclidean distance but also social distance, contiguity in terms of geography and in terms of social relationships, to name a few. This alleviates the handling of spatial information by recognizing it as more than x, y co-ordinate pairs, identifying and consequently exploiting different means of incorporating spatial information, not only as attributes but as a fundamental aspect of entities and relations embedded in a geographic space. The primary challenge of integration lies in the creation of a geo-social space that embodies characteristics of Cartesian space as well as network space. By recognizing the spatial properties of the social network, this type moves closest to the definition of the geo-social space in which spatial social networks are embedded.

In the following section, we encompass all three types of spatial social networks to discuss the terminological chaos that ensue when talking simultaneously about geography and social network analysis. Creating a typology helps understand the various levels of sophistications of integration, creating a baseline for further deliberation.

2.4. Different Expressions and Challenges

Table 2.1 shows three specific concepts that occur parallelly in the literature of SNA and geography literature that hold potential for avenues of reconciliation for a closer integration of spatial information in SNA. We discuss the terms as they appear in the two different contexts and move on to highlighting the challenges as well as the importance of creating solidarity of the terms for spatial social networks.

Table 2.1: Parallel concepts of SNA and geography literature and examples of how they tend to be expressed

Concept	Expressed in SNA	Expressed in Geography	Coupling Problem
Distance	Counts of edges; connectivity; shortest path; degrees of centrality; weighting	Measures in Euclidean space; shortest path; homophily of non-geographic attributes in Cartesian space; impedances; distance decays	Incongruent spatial metaphors
Community	Shared attribute; areas; number of social interactions; homophily	Static measures (jurisdictions); Dynamic measures (spatial distribution and clustering)	Semantics
Scale	Number of nodes and edges; completeness of	Resolution of collection and representation; spatial extent;	Reconciling and integrating the many

	capture; characteristic nodes	edge effect	interpretations

2.4.1. Distance

In social networks, distance is measured by movement from one node to another node, travelling along the edges. Nodes connected by an edge are said to be adjacent. Two nodes i and j are considered reachable, if there is a sequence of one or more edges that connects the said nodes. The sequence of edges between i and j is called a path. The number of links one traverses to reach another node equals the distance on the graph. More specifically, the geodesic distance, d(i, j), between two nodes is defined as the shortest path between them (Freeman 1978). It is also possible to create social networks with unconnected nodes, for example, separate groups of friends with no common friend between the groups. In SNA, the groups themselves are called connected components. If there is no path connecting the two nodes, that is, if they belong to different connected components then the distance between them is conventionally defined as infinite. Hence, two nodes belonging to two different connected components are unreachable from each other.

In SNA, the simplest way to characterize importance of nodes is by looking at the number of edges incident on it. Thus, an important node has many adjacent nodes by virtue of having a high degree centrality (Freeman 1978). In a social network, being friends with an important person is always beneficial because one potentially becomes closer to many other people in the network. It is important to note that being 'close' considers the geodesic distance on the social network and not physical distance in a Cartesian Space. Thus, even without having a high degree centrality, by virtue of having a few important friends, a node may have highly efficient paths connecting it to most other nodes in the network. These nodes are said to have a high closeness (Freeman 1978).

In a weighted graph, one cannot just as easily traverse one path as another. In certain scenarios, edges may be associated with weights to represent factors like strength of a tie, probability of forming a tie, or in case of spatial social networks, geographic distance between the nodes. Weights add a new property coaxing geodesic distance calculations to account for the different weights of the edges. Links may also have directionality (non-reciprocal relations). If the graph is directional then distance d(i, j) and distance d(j, i) are not symmetrical. The concept of adding weights to the edges is similar to geographical impedances along a road or stream network. (In addition to the discussion here, Andris (2011) provides a comprehensive account of social distance as measured by movement from none to node along edges and how this has been used in social networks with geographic considerations.)

Similar to associating weights with links, attribute information (e.g. age, gender) can be affixed to nodes. Each attribute can be considered a dimension and projected on to axes creating an n-dimensional attribute space[2]. Each node is represented as a point in this attribute space. The position is determined by the particular set of values of the node's attributes. The locations of the nodes are no longer arbitrary and the distance between them is interpretable (Hanneman and Riddle 2005). This information in the attribute space determine the similarity of the nodes with respect to their attributes. For example, people with similar incomes and similar age are closer together in the attribute space.

If the geographic location of each node is stored as attribute information, then longitude and latitude may be used to characterize the X and Y axes of the attribute space. The distance between the two nodes in the attribute space represents the distance in geographic space. Since there may be a variety of attribute information collected about the nodes, the specific attributes selected to characterize the axes in the attribute space may produce different sociograms for the same social network.

Incorporating location into a node's attributes allows exploration into the geographic properties of networks. One way geographic effects may be considered is by calculating the distance between nodes with connections. Tobler's First Law of Geography states that near features are more alike than distant features (Tobler 1970). Nodes located in proximity to one other in geographic space thus have properties that are similar to each other. In social networks, it is known that similar nodes tend to form connections (i.e., homophily) (McPherson *et al.* 2001). Thus, geographically closer nodes are more likely to have an edge than nodes that are further apart. Propinquity has been acknowledged to play a role in forming social relationships (Festinger *et al.* 1950). Many social processes are considered to be an outcome or affected by

[2] Machine learning uses the term 'feature' to refer to each attribute used to characterize an entity. Consequently, the n-dimensional space where the features live is referred to as a feature space. Here we use the term attribute space to avoid confusion with geographic features.

spatial proximity (Downey 2006). Milgram's (Milgram 1967) landmark work on "small world networks", which led to the famous concept of "Six degrees of separation", contained a geographic component as the letters were posted from the different cities to reach their final destination. This concept caught the attention of researchers in exploring the relationship between geographic distance and social ties.

Despite the telecommunication revolution and the fabled "death of distance" (Cairncross 1997, Sempsey 1998), researchers reiterate that relationships are often geographically local with the probability of forming long distance ties diminishing exponentially with increase in distance between the actors (Liben-Nowell *et al.* 2005, Wong *et al.* 2006b, Preciado *et al.* 2011). The dependence on geographic distance to form social ties can be exploited to form generative models for social networks, which mimic the properties found in real world networks (Kleinberg 2000, Watts *et al.* 2002).

The geographic distance-friendship relationship is recognized as having important consequences on the structures and processes of the network (Kleinberg 2000, Watts *et al.* 2002, Wong *et al.* 2006b). While studying sparsely-connected social networks, where despite the low density of links between the nodes, all nodes are reachable from each other via only a few steps (i.e. small world networks), Kleinberg (2000) and Watts *et al.* (2002) concluded that if the probability of linking two individuals is inversely proportional to the geographic distance separating them. In addition, a simple greedy strategy (i.e., searching by making the locally optimal choice at each step) based on geography is able to find a short path to a target in $(\ln N)^2$ time. The authors also point out that if a network is not structured like this, it is impossible to find the target using a simple greedy strategy in a poly-logarithmic time, making searches computationally expensive. A model proposed by (Watts *et al.* 2002) to explain the 'searchable' nature of small world networks considers individuals to belong to groups, which in turn are embedded hierarchically inside larger groups. The group can refer to any attribute, for example profession or geography. In this model too, searching using only local information (i.e., selecting a neighboring node of the current node that has the same attribute as the target) was successful only when the probability of acquaintance between two individuals was inversely related to distance.

A distance metric is an abstract notion, appropriated by different fields in various ways to describe what 'near' and 'far' means in the subject's realm. Geography usually uses measures in Cartesian space, which can be measured in different ways (e.g., Euclidean distance, Manhattan distance). Social networks, on the other hand, use network space where distance is measured as edge sequences between nodes. While studying social networks that are situated in geography, the distance of nodes on the surface of the earth is a strong determinant of social relationships and hence affects geodesic distance in the social network. In spatial social networks, metrics can be developed to leverage the different distance conceptualizations to characterize the nodes as well as the entire network. This duality of network space and cartesian space as highlighted here and mentioned in Andris (2016) can however can provide a starting point for developing new distance measures will require a conceptualization of a geo-social space in which spatial social networks are embedded.

2.4.2. Communities

Sporadic debates in geography and SNA have recommended various definitions of communities. The simplest differences between communities and societies are in terms of size and interactions. In a social network, parts of the network may be highly connected to each other. These sub-structures inside the social network are referred to as clusters, communities, cohesive groups, or modules (Palla *et al.* 2005). The principal elements defining a community in geography are usually identified as social ties, social interactions and area (Hillery 1955, Clark 1973). However, this definition is not all encompassing, nor are all the elements described above a necessary condition for a community as described in geography (Frug 1996, Smith 1999). Professional societies (e.g., the scientific community, community of lawyers) may not satisfy the requirement of sharing common geographic territory, or may not even interact with one another, yet form a community based on homogeneity of profession. Communities hence have three primary dimensions determining them, namely, shared area, social interactions, and homophily. The intersection of geography and social network helps explore communities as a function of both shared area and social connections, integrating social as well as spatial communities. Thus, a working definition for communities for spatial social networks adheres to the old-school definition encompassing shared area and based on social ties, yet is flexible to account for spatially discontinuous communities if the social ties between entities are known.

The mere fact that nodes are geographically co-located is insufficient to firmly combine spatial concepts of community to SNA ideas of community. Hence spatial cluster detection methods (e.g., Getis-Ord, global Moran's I and Ripley's K) and even Tobler's Law may not be informative in detecting social network communities. Even non-spatial topological community detection algorithms normally used in SNA, like clique or modularity based approaches, may be insufficient to find communities in spatial social networks. In the case of spatial SNA, the spatial arrangement of nodes as well as the nature of ties must be factored in to extract information from the network topology. Thus, for spatial social networks, it is important to consider not only the Euclidean distance between the nodes, but also the social distance between them, for most socio-spatial analyses including cluster and community detection (Crampton *et al.* 2013).

Modularity (Newman 2006) is a metric that provides a measure of the quality of graph partition. Modularity calculation for detection of social communities in spatial social networks must control for the spatiality of the network. Hence, community detection in spatial social networks needs to be perceptive of both spatial and network auto-correlation to distill social and geographic determinants of community formation. The modularity calculation can be modified to factor in the location of the node to find communities that are firmly determined by geographic factors (Onnela *et al.* 2011). However, researchers have argued that this form of approach to community detection provides little information about the underlying forces actually shaping the topology of the network, and have proposed a modularity measure that can factor out the effects of space, thus finding clusters of nodes that are similar but not just because of their location relative to one another (Expert *et al.* 2011). Interesting community patterns can also be extracted from spatial social networks by applying standard modularity based community detection approaches coupled with innovative ways of visualizing the community. One such visualization approach plots the communities on maps and uses Kernel Density Estimate to characterize the relative occurrence of a user in any given community in any given location (Batty *et al.* 2012, Comber *et al.* 2012).

Despite the nihility of a widely accepted unique definition of communities in geography and SNA, the existence of smaller connected structures within the larger society is a signature of the hierarchical nature of the complex social structure (Palla *et al.* 2005). Identification of topological clusters moves the focus from quantifying the importance of individual nodes to identifying important sub-structures in the network, representing a jump in the entity of analysis, that is, instead of just one node we look at a cluster of nodes. Coupling SNA and geography provides avenues for consolidating the various conceptualizations of communities, opening up opportunities to compare and contrast the various definitions of communities and their corresponding usefulness in revealing socio-spatial patterns and processes.

2.4.3. Scale

Scale has been a central notion in geography and also a particularly confusing one depending on the context (Quattrochi and Goodchild 1997, Atkinson and Tate 2000, Wu and Li 2009, Goodchild 2011). Scale in geography has consequently been separated prmarily into resolution or granularity of the data and extent covered by the dataset. Thus, it is imperative to reconcile concepts of spatial social networks with different meanings of geographic scale.

The observation scale or the measurement unit (Wu and Li 2009) needs two specifications, one for the geography and another for social network elements. Whereas the specifications in geography include the smallest object discernable and the smallest measurable units, social networks need to include disclaimers about what resolution of data is collected about the nodes and the edges. Details about nodes include not only a list of the attributes collected, but also metadata about the attributes. For example, age is denoted as a specific number or as a range. In terms of geography, it is vital that location resolution of the node is known. Moreover, if the nodes denote people, it is important to know how they are located and assess the implications of the locations for the study. For example, is the location of the person's home recorded or is it the location of work? A person, unlike a house, is not stationary in space. Thus, it is imperative to reflect on how the recorded location(s) affect the inferences from the spatial social networks and how qualifying the observation scale serves as the entry point for understanding the simplified model of reality at which the study is implemented (Atkinson and Tate 2000).

The geographic scale or the spatial extent refers to the area on the surface of the earth spanned by the social network under study (Wu and Li 2009). Thus, analysis on data from Facebook may have a geographic extent spanning the entire earth. Depending on the phenomenon under study, only a subset of the network may be used. In terms of analysis boundaries, social networks pose a two-fold problem, finding the entire population and then determining the links between the entities. When resorting to sampling, decisions need to be made to limit the population, or the links, or both. If geographic constraints are used, then the geographic boundaries used to subset the network itself define the geographic scale. Conversely, the social network itself might dictate the geographic scale that needs to be considered. For example, an experiment that requires the creation of the social network by following the connections of a person, the geographic extent determined by how far from the original person their connections live. Someone residing in Montreal may have friends only in Montreal in which case the geographic extent will be small, or may have friends residing all over the world requiring a very large geographic extent of study. When resorting to sampling, all relationships that lie beyond the sampling boundaries are ignored. As network algorithms are fundamentally relational, the results obtained from these algorithms will be erroneous as a result of the edge effect (Gil 2016).

When social networks are studied in the context of geography, the spatial extent at which the various social network metrics are reported may convey interesting information. For example, it is known that most of our social connections are local, with only a few long distance links. Thus, it may be interesting to classify the degree of a node with varying spatial extents. A person who has more long distance links than average may be of more interest in connecting disparate spatial locations. Similarly, when studying real world social networks, people living in small towns or villages often know each other, forming closely knit social networks. However, emergence of communities at a larger spatial extent with similar population density may be more interesting because of the lower probability of such an event.

Additionally, scale can be studied in terms of its phenomenology. The argument about Stommel diagrams (Stommel 1963, Holling 1992) in which geographic features only have meaning when observed in space-time (e.g., a flood, an oak forest), can be extended to social networks. The recognition of the fact that social networks are a spatio-temporal process is highlighted by the adoption of check-ins and timelines by almost all social media sites. Thus, in spatial social networks, a coupling of social networks requires recognition of the fact that social networks are not only contextually based on the conceptualization of edges, but also contextual in space-time.

Perhaps the most confusing use of the term ‘scale’ in social networks is when referring to the existence of characteristic nodes. In a uniform network, every node has an approximately equal number of edges. The degree distribution of a uniform network has a sharp peak with a very small standard deviation. Hence, a node with the mean number of edges is considered to be representative of all the nodes in the network. However, most social networks do not have an egalitarian degree distribution. Few nodes have disproportionately more edges compared to the majority of the nodes in the network. In terms of social networks, these nodes play a vital role in keeping the network connected. Thus, the degree distribution is skewed. These networks are considered to be scale-free because there is no characteristic node to represent all the other nodes (Barabási 2002). The closest equivalence in geography is regarding aggregating and rescaling data (Openshaw and Rao 1995). Data is said to be rescaled to a lower resolution by combining smaller regions into larger ones, aggregating the values based on some central tendency. This process is often used to aggregate data from county level up to the provincial level. Though the commonly used term for this operation is ‘rescaling’ or ‘upscaling’ as it involves a change in the observational spatial resolution, the idea is similar to ‘scale’ in networks where a large population is said to be represented by a single entity. The use of scale in social networks to refer to existence of a characteristic node is an incongruence of terminology between social networks and the different meanings of scale used in geography.

2.5. Conclusion

Spatial social networks have gained traction in the literature as a method of incorporating geographical information into SNA. In this paper, we have highlighted some of the inconsistencies that require closer deliberation for a tighter coupling of spatial information in SNA. We proposed a typology of the current literature of spatial SNA. We also highlighted some of the parallel concepts that exist in spatial and social network analysis. SNA provides an interesting perspective on explicitly studying relationships between entities. However, the incorporations of geography in social networks have been rudimentary thus far, and critical introspection is essential to incorporate concepts and concerns from the perspective of GIScience. We need to draw upon the long tradition of geography in working with non-Cartesian notions of space, moving towards a definition of geo-social space to succinctly reflect the subtleties of spatial social networks.

When using SNA, analysts must remain cognizant of the fact that a network is usually a snapshot of a social system in both time and conceptualization. Depending on the conceptualization of the relationships, multiple social networks can be created. For example, if explicitly declared friendships are used, it results in a particular social network that is different from the one when some common attribute between people is used to conceptualize the edges. Society is multi-faceted and the different conceptions highlight different aspects. Situating social networks in geography not only allows several new conceptualizations of relationships between entities, but has the potential to enrich analytic capabilities even when the network ties are based on non-geographic factors. In this paper, we have highlighted some of the considerations for progressing socio-spatial analytics utilizing spatial social networks. Investigators of new metrics, analysis techniques, and algorithms designed to leverage spatial and social networks should remain vigilant about unifying spatial and social concepts to reveal interesting phenomena, possible only through more deeply interrogating both spatiality and sociality together.

2A. Additional Literature Pertaining to Spatial Social Networks and this Dissertation

This chapter supplements the literature in Chapter 2, which includes a typology for SSNs and potential avenues for reconciliation of SNA in GIScience with additional literature pertinent to SSNs in general and to this dissertation specifically. While Chapter 2 provides an over-arching framework that provides an overview of the various ways in which spatial information has been incorporated in social networks and highlights the congruent use of terms in SNA and geography literature as avenues for a tighter integration, this chapter specifically highlights the background literature which provides the rational for Chapters 3, 4, and 5.

2A.1. Relational Multiplicity and the Network Model

As highlighted in Chapter 1.1.1, social networks describe conceptual relationships between entities. Provided that two entities can be distinguished from one another, and have some kind of relationship, this can be modelled using network analysis. However, for any given set of entities there can be multiple relationships between different entities, and these relationships can take on a multiplicity of forms. For example, given a set of people, there might be multiple relationships that exists between the entities (e.g. friendship, kinship, co-worker). Depending on the research question some or all of these relationships between the entities may be of importance. Thus, there are several potential ways in which edges can be defined, and each different form of edge employed in a given network analysis will result in a different network (Zuckerman 2008). Consequently, it becomes relevant to understand precisely how to define what an edge, or connection, consists of, and what the edge represents in a social network.

As described in Section 2.3, and Andris (2016), the dominant way of handling spatial information in SSNs has been to incorporate spatial information as an attribute of the nodes of a network. Thus, the relationships as captured by edges do not represent spatial relationships. However, given a set of entities, since multiple relationships can be present in any given network, it is possible to create multiple network-based representations out of the same dataset that capture spatial information in various ways. If the relationship between the nodes is a spatial one, then it is possible to encode spatial information in the edges of the network too. In addition to the multiple ways in which edges can be conceived and expressed within a single dataset, nodes themselves can also be conceived of in multiple ways. Thus, multiple networks can be created from a single dataset in which the nodes and edges represent various relationships embedded in the dataset. This means spatial information does not have to be restricted to be a nodal attribute (Table 1.1), and the edges of the network themselves can incorporate spatial information to represent spatial relationships. Thus, these networks created out of the same dataset can provide different views of the various latent spatial relationships.

Table 2A.1: Overview of the how spatial information can be incorporated into the nodes and edges of a social network. Type 1 networks with spatial information incorporated as part of the nodes are the most dominant form of SSNs in literature.

Type	**Node spatial?**	**Edge Spatial?**	**Example**
0	No	No	Non-spatial social network
1	**Yes**	No	Spatial social networks with nodes pinned to location and edges representing social connectivity
2	**Yes**	**Yes**	Spatial social network with both nodes and edges incorporating spatial information, but unlike a road network, the edges may not represent physical connections
3	No	**Yes**	Spatial social network that have spatial information in the edges, but not in the nodes. For example, people connected together by edges representing their country of origin.

2A.2. Metrics in SNA

SNA draws upon the principles of graph theory, representing entities and connections between them as nodes and edges respectively. Thus, in its simplest form, a social network is a collection of nodes connected together by edges, and consequently the primary focus is on the topology, or the relationships between the entities (Wasserman 1994, Lazer *et al.* 2009).

Based on the node-and-edge topological structure, SNA relies on a series of metrics to characterize the network at different topological scales – these can be subdivided into three categories, namely Entity level, Community level/Meso-scale, and network scale metrics. The following sub-sections provide an overview of the three categories of metrics used in SNA and explains how they have thus far been applied to SSNs.

2A.2.1. Entity Level Metrics

These metrics are primarily used to identify and characterize important entities, i.e. nodes, that are embedded in the network structure. The three most commonly used entity level metrics are degree, betweenness, and closeness, which are collectively referred to as centrality measures. The central entities in the network are considered to be in the “thick of things” (Freeman 1978) as a virtue of being more centrally located in the network than the other entities. Figure 1 exemplifies the simplest measures of centrality for unweighted connected graphs as defined by Freeman (1978), that is, degree, betweenness and closeness. *Degree* of a node is the number of other nodes the focal node is connected to. Hence, the three nodes A and C are important as they are connected to four other nodes. However, node B arguably plays a more important role than A, or C, in keeping the network connected. Being the only point of connection between parts of the network, which would otherwise have been disconnected, node B has a crucial brokerage advantage. Mathematically, this advantage of acting as a bridge is captured by the number of shortest paths that between all pairs of nodes in a network that pass through the focal node and is termed as *betweenness*. The importance of node D however, is not captured by the degree or betweenness. Due to node D’s connection to node B (which plays an important role in keeping otherwise disparate parts of node connected), node D despite having only a single connection can reach all other parts of the network in lesser number of steps compared to all other nodes in the network except for B. Thus, *closeness* centrality captures the efficiency with which a node can reach all other nodes in the network (Borgatti 2005).

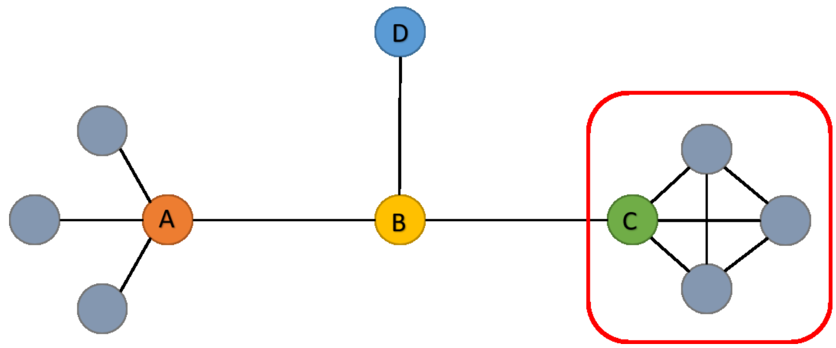


Figure 2A.1: A simple diagrammatic representation for demonstrating node importance and cliques. The colored nodes represent important nodes according to the centrality measures of degree, betweenness, and closeness. The highlighted part of the network shows a clique.

The basic definition of degree, betweenness and closeness has been modified to accommodate more nuanced definitions of networks where the relationships may further incorporate directionality or weights (Figure 2A.1). For instance, the relationships in the network may not be implicitly reciprocal, making the edges directional. Moreover, edges may have weights reflecting the strength of relationships, creating a network conceptualization which goes beyond dichotomous edge conceptualization. In either case, the definitions of the metrics have been adapted to reflect the network characteristics. In case of directed graphs, the challenge is circumvented by choosing only the appropriate edges according to the direction while traversing the graph, and the centrality measures are referred to as *Prestige* measures since they differentiate between choices made by the node and the collective choices made by the others toward the node (Knoke and Burt 1983, Wasserman and Faust 1999, Borgatti *et al.* 2002). However, if the edges have associated weights, several different interpretations for the metrics have been proposed which can collectively be termed as *Modified Centrality Metrics*. In the simplest form, the degree can be redefined as the sum of the weights of all the edges incident on the focal node (Barrat *et al.* 2004). However, in this interpretation, there is no distinction between the reporting of degree for a node with 10 edges of weight 1, and a node with 1 edge of weight 10 (Opsahl *et al.* 2010). In the case of closeness and betweenness, the edge weights are factored in the shortest path calculations by mutating the definition of shortest path to be the least costly path (Brandes 2001, Newman 2001). However, Opsahl et al. (2010) argue that these interpretations confound the importance of the number of edges incident on a graph with the consequences of edge weight. They have hence proposed a tuning parameter to modulate the relative importance of edge weight versus number of edges (Opsahl *et al.* 2010).

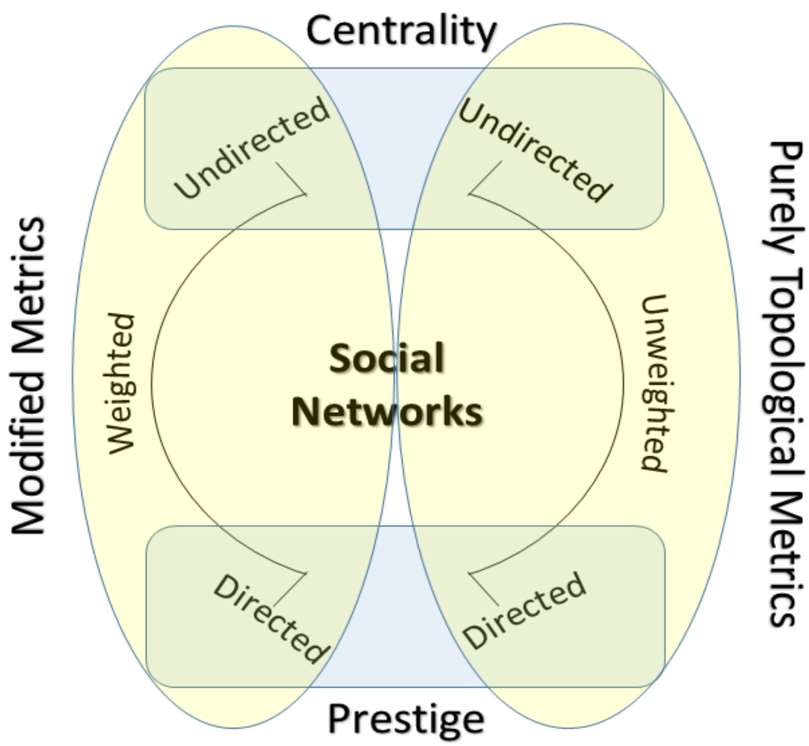


Figure 2A.2: Various metrics of centrality in SNA classified according to the type of networks being modelled according to edge reciprocity and weights.

In cases of spatial social networks, another layer of intricacy is added to the network specification. Enriching the nodes with locational information contextualizes the entities in terms of their relative position to each other on the earth's surface (Type 1 networks in Table 2A.1). The locational information associated with the nodes enables interpreting the nodes' position in Cartesian space in addition to network space. This raises the crucial question of using social network metrics to characterize the important nodes in the network in terms of spatial scales (Chapter 2.4.1). A person (say Alice) may have lots of local friends and thus be considered important in her region, but someone (say Bob) who has just moved into the locality may have an equal number of friends at his previous location. Thus, in the context of spatial social networks, both these entities are important but their importance varies depending on the spatial scale of interest. While Alice may be more efficient in spreading information locally, Bob may be more effective in spreading information far and wide. The same argument can be extended to the betweenness measure where Alice and Bob may have differing brokerage agencies at different spatial scales (Bronfenbrenner 1977, Boessen *et al.* 2017). Following the same line of argument, closeness centrality in case of spatial social networks must reflect how efficiently a node connects to local nodes as compared to far nodes. In the case of entity-level metrics, there is currently a lack of literature on how to capture both spatial and social importance of entities simultaneously.

2A.2.2. Community Level/Mesoscale Metrics

As described in Chapter 2.4.2, communities in social networks are meso-scale structures embedded in networks, within which nodes are highly connected to each other (Palla *et al.* 2005). As evident from the definition, the ideal topological form of a community are sub-structures embedded in the network in which all the nodes are connected to each other with minimal connection to the rest of the network (Figure 2.4.1 highlighted portion). These completely connected sub-structures in a network are known as *cliques*. However, in real world networks, such completely connected sub-structures are rare, but there do exist parts of the network where nodes are better connected to each other than to the rest of the network. Here, as with entity level metrics, a plethora of algorithms exist to detect these meso-scale structures. Notable examples of these algorithms include minimum cut method (Newman 2004a), hierarchical clustering based methods (Mann *et al.* 2008, Alvarez *et al.* 2015), Girvan-Newman algorithm (Girvan and Newman 2002), and modularity maximization-based methods (Newman 2004b, Blondel *et al.* 2008, Bader *et al.* 2013). In SSNs, these community structures are often the manifestation of latent effects of the underlying space. Consequently, there is considerable debate about how to detect community structures in SSNs, and algorithms have been developed that addresses the effect of space on community formation (see Chapter 2.4.2 for discussion).

As discussed in Chapter 2.4.2, the idea of community is also a central notion in geography and encompasses the concepts of social interactions amongst entities co-located in a bounded space along with having attribute similarities (Hillery 1955, Clark 1973, Frug 1996). Thus, in the context of spatial social networks, the idea of community can be explored as a function of both shared area, and shared social connections. The most common way of exploring this has been to identify community structures within the topological structure of networks, and to plot the resultant communities on maps. For example, Batty *et al.* (2012) and Comber *et al.* (2012) have shown that communities that are detected within social networks often form spatially cohesive structures when plotted on maps, re-enforcing the idea that members of a community often co-exist in a bounded space and have social mutual connections.

The spatial distribution of the communities detected within social networks also provides interesting insight into the idea of spatial regions as loci of homogeneity. The large corpus of data on human mobility and interactions allows re-visiting the concept of regions from a bottom-up approach, such that regions are delineated from observed interactions rather than focusing on interactions bounded in geographic space (e.g. political boundaries). Based on social and interaction data, previous work has seen regions segmented on the basis of similarities (used to form a region) and differences (used to split regions) in surnames (Cheshire *et al.* 2010), commuters (Nielsen and Hovgesen 2008, Coscia *et al.* 2012, Rinzivillo *et al.* 2012), currency circulation (Brockmann *et al.* 2006, Thiemann *et al.* 2010), and telephone calls (Ratti *et al.* 2010, Calabrese *et al.* 2011, Sobolevsky *et al.* 2013, Hawelka *et al.* 2014). The partitioning of human populations based on networks of communication from multiple countries has been shown to reflect linguistic and cultural borders of geographical space, and sometimes follows administrative boundaries closely (Calabrese *et al.* 2011, Sobolevsky *et al.* 2013). Even at a global scale, communities detected using a Twitter mobility network formed spatially cohesive regions reflecting the regional division of the world (Hawelka *et al.* 2014). However, it should be noted that here, as elsewhere, exist issues of scale (Gao *et al.* 2013, Coscia *et al.* 2014). Spatially, finer resolutions create over-detailed networks, while smaller components can be associated to several small clusters. At larger scales, models can generate an excessive aggregation of local movements. While in some instances, partitioning of geographic space based on human interaction data closely matches existing administrative boundaries, in other cases the data might suggest a completely different segregation. A study on the dating site OkCupid showed that while looking for romantic partners, the geographic search boundaries conceived by users are quite different from major administrative regions, and in some cases also transcend international boundaries (McMenamin 2017).

A caveat for regionalization studies based on interaction data is that the interactions themselves may be impacted and shaped by administrative boundaries and their associated laws. Regionalization on data from telecommunication in the United Kingdom have been found to closely match administrative boundaries, but the authors of this study note that this may be partially due to the fact that telecom regions also closely match historical and important administrative regions (Ratti *et al.* 2010). However, the study also points out that based on the pattern of communications, the effects of a possible secession of Wales from Great Britain would be more disruptive for the human network than that of Scotland. Similar studies based on mobile phone communication in Brussels have shown discovered regions to be disparate from the administrative boundaries, yet to be consistent with linguistic borders, thus providing clear traces of the evolution of Belgium as having its own unique administrative and geopolitical history (Blondel *et al.* 2010, 2011).

The abundance of human mobility and interaction data that exist can thus be useful in examining the meaningfulness of administrative boundaries, with the caveat that some administrative boundaries delineate resource availability in addition to meaningful divisions amongst human populations. Thus, the algorithmic partitions based on human interaction is interesting but may obscure historical, geographic, economic and political factors used in the creation of administrative boundaries.

2A.2.3. Network level metrics

Network level metrics go beyond individual entities and mesoscale structures to characterize the structure of the entire network. The simplest network level metric can be conceived of as the *size of the network*, e.g. number of nodes and edges. Additionally, given the number of nodes and the type of network, e.g. directed vs. undirected, the maximum number of edges required to connect each pair of nodes in a network is bounded. Thus, by comparing the number of edges that exist in a network, to the theoretical maximum number of edges required to fully connect a network, can provide insights into the density of edges present in the network. This metric is known as *network density* (Wasserman and Faust 1999).

With network level metrics, network properties are calculated to analyze the population dynamics of the social network as a whole. Average path length and network diameter provide information about how topologically close nodes are to each other, and thus indicate how quickly one can get from one part of the network to another. While *average path length* refers to the mean internode distance, *network diameter* is the maximum of internode shortest distances. Being a measure of the maximum distance between two of the furthest social connections, diameter provides a measure of how 'big' the network is (Hanneman & Riddle 2005, p.81). When considering social networks embedded in geographic space, it is also important to characterize the spatial extent of the network. The fact that social-connections tends to be local with the probability of connection diminishing exponentially as a function of distance (Liben-Nowell *et al.* 2005, Wong *et al.* 2006b, Mok *et al.* 2010, Preciado *et al.* 2011) highlights the importance of characterizing how far the most distant entities are, both spatially and socially. Hence, a specification of network diameter along with spatial extent is essential to capture the socio-spatial expanse of spatial social networks. However, similar to entity-level metrics, there is a dearth of measures that capture the socio-spatial structure of an SSN.

Moreover, in addition to using the aforementioned metrics, another way of understanding a network structure is to use node- and edge-based visual representations known as sociograms (Moreno 1934) (Figure 2A.1 is a simple sociogram). Various layout algorithms employ heuristic methods to position nodes and edges of a network in a sociogram (such as force-directed layout algorithms (Jacomy *et al.* 2014)), in order to provide nuanced information through visual analytics, and to create aesthetic layouts. This free-form way of positioning nodes and edges are possible as in a network, the focus is on the topology, and the position of the nodes and edges are not directly interpretable (Chapter 2.2). Despite the fact that the topology of a network can be exploited to create isomorphic structures, the size of a network poses a challenge to create interpretable visual representations that balance form and function (Krzywinski *et al.* 2012). Additionally, in SSNs, the problem of plotting the network is exacerbated by the fact that if nodes are anchored to (x, y) geographical locations, then the ability to employ graph layout algorithms that can create various isomorphic forms of the network by moving the nodes and edges to create either more aesthetic or easier to interpret visualizations are compromised. The challenge of characterization of the network structure and visualization of the socio-spatial properties of a network are areas of SSN that requires further research.

2A.3. Applications of SNA in "Spatially Integrated Social Sciences"

"One cannot understand social life without understanding the arrangements of particular social actors in particular social times and places." (Abbott 1997)

In the wake of "spatially integrated social science" (Goodchild *et al.* 2000, Goodchild and Janelle 2004), significant focus has been placed on understanding human interactions and space as being mutually established. The paradigm shift towards quantitative social science has seen a rise in interest in the inclusion of geographical considerations (Anselin 1999, Goodchild *et al.* 2000). As noted in Chapter 1.1.1, SNA has been embraced in social sciences as a quantitative toolkit to model relationships between entities. The growing interest in SNA from various disciplines coincides with the growing availability of data available through Web 2.0 technology, including social media and location sharing applications (Freeman 2004, Fu *et al.* 2008, Lewis *et al.* 2008, Borgatti *et al.* 2009a, Bughin and Chu 2010). The multidisciplinary applications of SNA have resulted in toolsets capable of visualization, characterization, and the inclusion of complex algorithms for the analysis of processes and the generation of predictions (Otte and Rousseau 2002, Freeman 2004, Borgatti *et al.* 2009a, Scott 2017).

The use of SNAs is pronounced in the fields of agriculture, economic geography, epidemiology, and criminology. Within agricultural networks, particular attention has been paid to the role of key actors that control agricultural resources and drive the formation of networks. For example, in the Peruvian Amazon, Abizaid et al. (2016) found that certain nodes representing households are disproportionately responsible for the distribution of resources in seed sharing networks, and that reciprocal seed sharing is rare. The households responsible for most seed sharing are those with higher relative community prestige, that can be associated with their role as healers, or as true members of the community. In addition to the importance of the role of community standing in structuring some social networks, several other important attributes of nodes within networks have been identified, notably kinship. Kinship has been found to be an important driver of network formation, and even a possible indicator of prestige. The importance of kinship in structuring social networks has been identified in labour sharing networks in Thailand (Entwisle *et al.* 2007), and in the Peruvian Amazon (Abizaid *et al.* 2015). Multi-scale networks differentiated on the basis of individuals, households, and communities not only shape spatial distribution of seeds, but also show gender based differences in the roles of actors (Zimmerer 2003). In the context of agricultural networks, the importance of networks of farmers in exchanging seeds have been identified as a key driver supplementing formal seed distribution networks responsible for circulation of seeds and promotion of diversity in agriculture (Zimmerer 2010, Coomes *et al.* 2015).

In economic geography, SNA has offered insight into the network of ties amongst entities in a geographic region. The work of Glückler (2007), highlights many nuances of SNA, including preferential attachments and brokering, as well as the role of properties like local ties. His work demonstrates the potential importance of both proximity and place. Further work has shown that in some cases, however, the geographical location of a node can also be an important driver in network and edge formation, and can go so far as to compensate for a lack of topological centrality within a social network (Owen-Smith and Powell 2004), or, conversely, preclude the formation of edges amongst spatially distant nodes, irrespective of their topological centrality (Fleming and Sorenson 2001). The inclusion of SNA and geography in economic geography has arguably been critical to the elucidation of the importance of location versus network ties, which is one of the fundamental concepts of this field (Castells 1996). In the global context of trade networks, the scale effects of network have been notable at highlighting how domestic networks influence international trade (Rauch 2001).

Epidemiology is another field which has utilized SNA as a tool for exploration and prediction of the patterns and processes of disease, particularly infectious diseases. In this context, the network paradigm has been successful in aiding the modeling of the relational concept of place (Cummins *et al.* 2007) in addition to modelling interactions between individuals (Valente 2010), both of which are crucial for the understanding of spread of disease and health outcomes. In short, infectious diseases can spread through populations based on environmental circumstances and interpersonal interactions within social networks. While environmental and interpersonal factors have traditionally been studied separately, recent attempts have been made to consider them together. For example, research on cholera in Bangladesh has shown the importance of environmental conditions in the transmission and persistence of the disease, but has also highlighted the role that social connections can play in disease transmission (Giebultowicz *et al.* 2011b, Emch *et al.* 2012b), and that the comparison of both spatial and social clustering is essential to understand the mechanism of disease transmission (Giebultowicz *et al.* 2011a, Emch *et al.* 2012a). Additional research corroborates the importance of the interaction between space and social ties, e.g. the importance of social meeting locations, and how these can be used to understand disease transmission between socially unrelated individuals (Wylie *et al.* 2005, Ziersch *et al.* 2005). Even in the case of non-transmissible diseases, such as cardiovascular diseases, geospatial clustering of social disadvantage maybe causally to the disease (Daniel *et al.* 2008), and this is not surprising considering the links between obesity and cardiovascular diseases, and the spatial (Pouliou and Elliott 2009, Wen *et al.* 2010, Huang *et al.* 2015) and social clustering (Christakis and Fowler 2007, Leroux *et al.* 2013) of obese individuals due to a variety of genetic (Barabási 2007, Frayling *et al.* 2007) and environmental factors (Cummins *et al.* 2005, Jeffery *et al.* 2006, King *et al.* 2006, Day and Pearce 2011).

SNA has also been used in the field of criminology, by studies that have researched the social and built environments in which gang form, and how gangs can influence local rates of violence. Indeed, multiple inductive methods have been developed to model the factors involved in spatial and temporal patterns of crime (Tita and Radil 2011), and how and why criminals partition geographic space into territories as a function of resource competition (Brantingham *et al.* 2012). Simultaneously, in the field of computational criminology SNA-based techniques have been employed to understand the role of key nodes (criminals) in the structures of criminal social networks (Brantingham 2011, Gallupe 2016). Recently, a more deductive approach has started to gain traction in which the theory of influence is used to model influence in geographic space. This approach considers the relative location of gangs in geographic space, while also considering the position of gangs within networks of gang rivalries. While still relatively recent, the application of SNA to the patterns and processes of crime have demonstrated that the spatial distribution of gang violence is strongly associated with socio-spatial dimensions of gang rivalries (which include geographic proximity, history of rivalry, organizational memory, and additional group processes) (Tita and Radil 2011, Papachristos *et al.* 2013). These applications also demonstrate how it is possible to simultaneously evaluate the way in which social positions in networks along with geographic space can constrain and shape criminal outcomes (Radil *et al.* 2010a).

2A.3.1. Shortcomings of the Network Paradigm for "Spatially Integrated Social Sciences"

SNA can provide interesting insight into patterns of interactions inherent in social systems, as described above. The node and edge structure used in network science has been shown to be highly scalable, and capable of modelling a variety of systems (from cells in the human body to stars in a galaxy) (Kadushin 2004, Havlin *et al.* 2012, Krioukov *et al.* 2012, Sullivan 2014, Andris 2016) and has thus been labeled as non-reductionist (Kadushin 2004). Like all models and theories, the network paradigm is not perfect and enable understanding some aspects of a system while inhibiting others as a consequence of the underlying set of assumptions (Colchester 2016). Consequently, SNA has been derided by some as simplistic, as it abstracts complex systems into simplified representations of sets of points connected by edges that can capture only the basic patterns of connectedness, and little else (Mizruchi 1994, Newman 2010, Buch-Hansen 2014) and as a result over privileging the nodes' perspective (Mejias 2006, 2010). Attempts to bridge the theoretical disconnect between the utility of SNA and its shortcomings can be informed by the addition of spatial dimensions into social networks to introduce geographical nuance into the structure and function of the networks. Spatial information is particularly important because spatial considerations, like spatial layout of the environment (Eagle *et al.* 2009, Sevtsuk *et al.* 2009, Sailer and McCulloh 2012, Hirschi 2013, Boessen *et al.* 2017), distribution of resources (Lund 2003, Wineman *et al.* 2009, Hipp *et al.* 2014), and distribution of entities (Entwisle *et al.* 2007, Butts *et al.* 2012, Verdery *et al.* 2012, Kowald *et al.* 2015, Boessen *et al.* 2017) have all been shown to impact the formation of ties and the nature and consequences of interactions. Thus, while studying social systems it is important to remember that these are multifaceted systems, in which interactions are only one of the components. To get a better understanding of the multifaceted nature of a social system it is important that SNA is complemented by qualitative and quantitative methods (Crossley 2010, Edwards 2010, Herz *et al.* 2015, Schipper and Spekkink 2015).

In this dissertation, I develop techniques for spatial social network analysis and apply these to a novel system(Chapter 4), to understand how community-park relationships are mediated by the presence of a long-term research field station. Like other complex social systems, interaction is a key component, but not the only component, and spatial social networks help identify how economic benefits that originate in the research field station percolate through the community, specifically with the help of a few key individuals. Economic benefits are an important, but not sole, component of the research field station, that impacts community-park relationships. Thus, I further supplement the spatial social network analysis in Chapter 4 with additional quantitative and qualitative methods to identify the various additional factors which relate to the research field station and impact community-park relationships (Chapter 5). By applying a multiplicity of methods to understand the same system, I attempt to discover the nuanced complexity of social systems, which are often abstracted out by the simplistic tendencies of SNA analysis to reduce a system under study into nodes and edges.

3. Understanding Research Collaborations and Connectivity through Spatial Social Networks: Analysis of 126 years of Grantmaking by National Geographic Society

Citation: Sarkar, D., Chapman, C.A., and Sengupta, R. Understanding Research Collaborations and Connectivity through Spatial Social Networks: Analysis of 126 years of Grantmaking by National Geographic Society. In Review at the *Annals of the American Association of Geographers.*

Linking statement: In this chapter, we move away from the restrictive embedding rules of physical networks that require both nodes and edges to be anchored to geographic space, and allow different network conceptualizations to be used to study various forms of spatial relationships embedded in the data. Thus, this chapter builds upon the literature covered in Chapter 2A.1 which highlights the flexibility of the network model to represent different relationships embedded in the same dataset. We use the National Geographic Society (NGS) grants database consisting of more than 12,000 grants given between 1890 and early 2016 as our case study, and create three different network realizations that embed spatial information in multiple ways as part of the node and edge representation of social networks. The nature of researchers supported by NGS supports the presence of spatial identifiers that are related to fieldwork locations, in addition to the information about the location of the applicants of the grants, make this particular dataset rich in terms of number of fields with spatial information, rendering it ideally suited for the aforementioned analysis. The three different networks incorporate spatial information in various ways in the network structure to create different realizations of spatial social networks from the same dataset. Each of these networks highlight different aspects of spatial relationships latent in the dataset, and the analysis of the networks using SNA techniques provide interesting insights into international and regional trends in research collaborations.

Non-Disclosure Agreement: signed by Mr. Sarkar with NGS allows scientific publications.

Summary: Global collaborations are a key factor defining modern academic excellence. We utilize Social Network Analysis, which provides a lucrative set of tools to analysis connections between entities, to obtain a unique view of research collaborations between countries. To do so, we incorporate spatial information in various ways in the network structure to create different Spatial Social Networks. Namely, we incorporate the spatial information as part of nodes, edges, or both. We use the National Geographic Society grants database consisting of grants given between 1890 and February 2016 as our case study and create three different network realizations that embed spatial information into the network in the aforementioned ways. Each of these networks highlight different aspects of connectivity latent in the dataset, and along with spatial information, provides insights into international and regional trends in research collaborations. Importantly, it enables us to answer specific questions about global collaborations using visual and analytical frameworks. Additionally, we deliberate on the abstraction afforded by Social Network Analysis, along with its rich toolset, which helps capture spatial relationships differently using the node and edge structure, and how these alternate networks compare to traditional network realizations used in GIScience.

3.1. Introduction

In the age of globalization, the question of international and interdisciplinary research is one that frequently dominates discussions around academic endeavors (Batty 2003). For example, global university rankings, such as the Times Higher Education (THE), World University Ranking (Times Higher Education 2017) and QS World University Ranking (QS World University Rankings 2017) measure "internationalization" by using either demographics (e.g., percentage of international staff and students) or publications. When using publications, the standard mechanisms for discussing research collaborations both between countries and disciplines is to use journal authorships and citation indexes (Hood and Wilson 2001, Ponds 2009, Adams 2013). For example, Tijssen & Winnink (2017) use the contribution of cited research to patents as a measure of research productivity and knowledge translation, and rank various countries standardized by their research expenditures.

Such methods frequently ignore the role of geography in the manifestation and production of knowledge. Malecki (2010, p. 493)argues that "geography of knowledge presupposes that knowledge is not uniformly but, rather, unevenly distributed across the landscape". Relying on citations alone is arguably inadequate, as it is an aggregated and lopsided view of scientific collaborations since highly cited publications are dominated by researchers from a few academic institutions, and therefore, countries. Notably, the US and EU15 (Western Europe) dominate top 1 percent most frequently cited publications, whereas China is present in top 10%, mostly through internationally co-authored publications (Leydesdorff *et al.* 2014). Further, universities have recently mobilized diasporic academics to create "global knowledge networks" which are also overlooked by citations to publications alone (Larner 2015). Therefore, there is a need to capture the different mechanisms by which collaborations happen, which involves going beyond analysis of citation counts.

Other forms of potential collaborations such as researchers working at a field site outside their country of residence are possible. In order to model such collaborations, it is necessary to move beyond summary statistics of demographics and known connections solely through co-authorship and citations. This requires a data model that is flexible enough to model different types of connections and can exclusively focus on these newer connections for analysis. The node and edge structure of social networks provides an adaptable method of representing research collaborations. Indeed, co-authorship and citation databases have been modelled as social networks (Kretschmer 1997, 2004, Otte and Rousseau 2002, Wagner and Leydesdorff 2005, Hou *et al.* 2008, Leydesdorff *et al.* 2018) but the flexibility of the social network data structure has not been exploited to model different conceptualizations of collaborations embedded in a datasets. We propose conceptualizations that go beyond explicitly declared connections and instead rely upon geographical considerations, and hence spatial information is a crucial aspect of both the derived social networks and the relationship the network models.

In this paper, we use an international database of research grants (i.e., more than 12,000 records representing grants given by the National Geographic Society between 1890-early 2016) to emphasize different spatially-explicit conceptualizations of research collaborations. Our conceptualizations of the network use geographical considerations as a central notion defining collaborations. We go beyond explicitly declared connections as commonly used in co-authorship and citation based studies. We subsequently use the rich literature on SNA to analyze connectivity as reflected in the different conceptualizations in which spatial information plays a crucial role. This also allows exploration beyond the first degree connections to understand where each individual entity fits into the larger fabric of a network. The spatial aspects of social networks have recently seen a great deal of interest (Barthélemy 2010, 2011, Andris 2016, Sarkar, Sieber, *et al.* 2016). Various studies have re-iterated the probability of social ties decreasing exponentially with increase of distance (Liben-Nowell *et al.* 2005, Wong *et al.* 2006a, Preciado *et al.* 2011, Scellato *et al.* 2011b). Geographic applications of SNA include epidemiology (Moore 2010, Giebultowicz *et al.* 2011a, 2011b, Keeling *et al.* 2011, Emch *et al.* 2012a), criminology (Radil *et al.* 2010a), trade networks (Castells 1996, Fleming and Sorenson 2001, Owen-Smith and Powell 2004), as well as co-operative agriculture practices (Entwisle *et al.* 2007, Abizaid *et al.* 2015, 2016), to highlight a few. The aforementioned examples all consider network representations where the spatial information is part of the nodes. The goal of this paper is to demonstrate how different Spatial Social Networks (SSN) can be created from a single dataset that embeds spatial information in different forms. We use the NGS database as our case study and consequently demonstrate how the varied SSNs conceptualizations provide different insights into collaborations.

3.2. SNA in GIScience

A social network is represented as a graph with nodes representing discrete entities (e.g. people, companies) connected together by edges that represent relationships. The relationships in a social network are conceptual entities and are commonly used to depict social connections like kinship and friendship, and also to represent other types of associations such as trade relationships and collaborations.

SNA has a lucrative set of tools to define and explore connections, and “connectedness” or “topology” and thus is of considerable interest in geography and Geographic Information Science (GIScience). SNAs enable operationalization of topological connections beyond spatial relationships in Euclidean space (Egenhofer and Franzosa 1991), a realization that remains difficult in Geographic Information Systems (GISystems). In the realm of GIScience, SNA is one of the many techniques explored for making sense of large, complex datasets with important topological implications. For example, Multi-Dimensional Scaling (MDS) and Self Organizing Maps (SOM) (Skupin and Fabrikant 2003, Skupin and Agarwal 2008) are notable examples of methods which rely on relational data to create visual structures from data that may not be explicitly spatial to highlight underlying relationships. SNA is particularly attractive amongst the alternatives principally for the following reasons.

First, SNA is an intuitive concept to model the complex fabric of relationships with its standard conceptualization of entities as nodes and relationships as edges. The emphasis on the relationships afforded by the network data model (Wasserman 1994) aides in shifting the focus of analysis beyond the individual entities to the pattern of interactions that make the system '*other* than the sum of its parts'.

Second, the flexibility of the SNA lies in the high level of abstraction provided by filtering out the relationship aspect from the rest of the data (Newman 2010). The abstraction facilitates scalability, making the techniques applicable to networks of all sizes from small groups to global systems (Kadushin 2004).

Finally, there has been a rise in interest SNA concomitant with the exponential increase in relevant data available through Web 2.0 technologies like social media sites and location sharing services (Freeman 2004, Fu *et al.* 2008, Borgatti *et al.* 2009a, Bughin and Chu 2010). Significant interests from several disciplines have evolved SNA from being a simple representation-based analysis method to a comprehensive toolset with facilities for visualization, characterization with metrics, as well as complex algorithms for analysis of intricate processes and prediction (Otte and Rousseau 2002, Freeman 2004, Borgatti *et al.* 2009a, Scott 2017). This avid interest in SNA has spawned a great wealth of knowledge, as well as readily available analysis frameworks that have helped researchers easily incorporate SNA in their projects (Borgatti *et al.* 2002, Csardi and Nepusz 2006, Hagberg *et al.* 2008, Bastian *et al.* 2009).

We primarily focus on the topological structure of datasets. We emphasize associating only spatial information with the nodes and edges and will not be enriching the nodes and edges with additional attribute data needed for sophisticated data mining. However, the concepts we present remain valid even when additional attribute information is used. We use the National

Geographic Society Grants Database to create three different social networks that have spatial information incorporated in them in different ways. We analyze these networks to demonstrate how the different social networks provide different answers based on their conceptualization and incorporation of spatial information.

3.3. Understanding spatiality of nodes and edges of a network

It is valuable to highlight two aspects of networks in the context of geography; the topological and the topographical. Whereas graph theory provides the metrics and algorithms to describe the topological aspects of the system, geography provides understanding of the spatial distribution of the components of the system. For example, road networks explicitly represent the location of the various streets, as well as the connectivity between the streets, and thus have topological and topographical information. The connection between the nodes and edges forming the network itself captures the topology; situating the network on a geographic plane enforces the topography. Exploiting both topological and topographical information has been shown to provide insights into the dynamics of urban structure (Mossa *et al.* 2005, Porta *et al.* 2006, Zhong *et al.* 2014), whereas comparing topological signatures of road networks across the world have been shown to display remarkable quantitative similarities (Jiang 2007).

Non-spatial networks like social networks usually have their focus only on the topological connections that exists amongst the entities. To make the social networks spatial, in most cases the spatial information is associated to the nodes, whereas the edges representing conceptual relationships between the spatial entities are devoid of spatial information (Sarkar, Sieber, *et al.* 2016). This representation has proved useful in understanding the spatial pattern of social connections that re-enforce the friendship-distance relationship (Liben-Nowell *et al.* 2005, Wong *et al.* 2006a, Preciado *et al.* 2011, Scellato *et al.* 2011b). SNA has provided insights into a number of spatial processes, like crime (Radil *et al.* 2010a) and disease transmission (Bian and Liebner 2007, Giebultowicz *et al.* 2011b, 2011a, Emch *et al.* 2012a, Gao and Bian 2016). However, SSN representations thus far mainly been limited to only the nodes having spatial information, with the primary spatial aspect of the edges being the distance of connection which is a consequence of the spatially located nodes. On the other hand, schematic transit representations, like the London Underground Transit Map, exemplify graphs in which the topology of the nodes and links are central even though the system they represent are both topological and topographical (like road or utility networks). This map seeks to convey the connectivity information of the different underground metro-rail stations, but abstracts away much of topographical information leaving only a few features, like the simplified shape of the Thames River, to highlight the north-south divide of London. Despite this abstraction, the simplification afforded by the schematic representation has lead even long-time residents to use it as not only an interface to the subway system, but also as an interface to the city overall (Vertesi 2008).

The different spatial embeddedness considered in physical networks (like road networks), spatial social networks, and the London Underground Transit Map highlight how varied spatial information can be incorporated as part of a network and begets the question: how can spatial information be incorporated and represented as part of the node and edge structure of social networks? And, whether different network based conceptualizations from the same dataset that model nodes and edge differently to capture multiple latent relationships provide different insights into the system under study.

3.4. Network Conceptualizations of the National Geographic Database of Grants

The database consists of more than 12,000 records representing grants given by The National Geographic Society between 1890 and early 2016. The number of grants given out annually are few in the early years (3 in the first decade) and have increased significantly over the years (more than 400 in 2015). The fields used from the records consist of fieldwork location, grant discipline and country of grantee (Table 3.1), amongst other fields. Table 3.2 provides an overview of the various networks developed from the National Geographic Society (NGS) grants database depending on how spatial information was incorporated, and the question about research collaborations that it primarily addresses. Specifically, we present three social networks with spatial information incorporated in different ways, along with a discussion of the questions that can be best answered using particular social networks, and the implications of incorporating spatial information in the particular form to make the network spatial. It is worth noting from this table that for any spatial social networks, attaching x, y co-ordinates is not the only way of incorporating spatial information in social networks. In fact, for a dataset of this size, plotting the social network as a sociogram or on a map creates visual clutter and does not provide any discernible information.

In the following sub-sections, we describe the creation of each type of network, followed by the results obtained through social network analysis to address the question of interest pertaining to each network.

Table 3.1: Fields of interest in the original National Geographic Society grants database

Field Name	**Description**
Contact Primary Address Country	The country of the researcher's primary affiliation.
Grant Discipline(s)	The discipline(s) of research. For most grants (70.43 percent), multiple disciplines are listed.
Fieldwork Location – Country	The country(s) where field work was carried out. Several grants (22.3 percent) have multiple entries for fieldwork country, implying that there were multiple field sites involved.
Fieldwork Location – Continent	The continent(s) where the field work was carried out. When the fieldwork location was located in different countries, the countries in question may be in different continents. The continents listed in the database used for the analysis in this paper are Africa (2894 instances), Asia (3348 instances), Europe (1198 instances), Oceania (1140 instances), North America (3448 instances), South America (2464 instances), and Middle America (2231 instances). Other listings include Oceans (138 instances), Ocean Islands (34 instances), Space (24 instances), Worldwide (33 instances), and Laboratory/Archival Research Only (31 instances).

Table 3.2: Overviews of the three different spatial social networks created

Type	Node	Node spatial?	Edge	Edge Spatial?	Question
1	Fieldwork Location (Country/Continent)	Yes	Same Grant	No	What is the "research connectivity" between countries?
2	PI Country	Yes	Fieldwork Country	Yes	How "internationalized" is the research of the various PIs?
3	Discipline	No	Fieldwork Country	Yes	What is the multi-disciplinary research potential of different countries?

3.4.1. Network Type 1: What is the "research connectivity" between countries?

In the Fieldwork Country-Grant Network, nodes represent the field work locations taken from the column that lists the continent and country where the project was carried out. Thus, the aggregation level for the fieldwork location is considered to be at the country level. Each record in the database refers to a grant and each grant may have multiple fieldwork locations listed. In some grants, more than 10 locations are listed as the fieldwork country, however, for each row a maximum of first 10 entries were considered. The first 10 entries were used only a few records had more than a few countries and given the long history of the dataset, some of the later entries may not have been relevant as the locations were listed according to priority. The entries related to non-country or non-continent entities like 'Space', 'Laboratory' and 'Oceans' were discarded. Two locations are considered to be connected if they are mentioned simultaneously in grants as fieldwork locations. A graph was consequently created with the nodes representing fieldwork countries connected together by edges representing grants. The edges have weights that denote the strength of connection between two nodes. Thus, an edge weight of 2 means that two grants have mentioned the same two countries as fieldwork location. The graph was not fully connected as entries like "Cape Evans" and "Estonia" only appeared in single grants, and these isolated nodes were discarded. The connected part of the graph consists of 193 nodes and 1763 edges without any self-loops or parallel edges, that is, two edges between the same set of nodes. Instead, if there were parallel edges between nodes, then the edges were collapsed into a single edge and the weights of the edges were summed. We also need to account for changes to countries over time given the long time frame of the dataset. For example, here are entries that may relate to the same or different countries over time, for example "Serbia" and "Serbia and Montenegro". These were left as they were presented in the database. The primary question that can be answered using this network is what are the relationships between countries in terms of research connectedness? In other words, how are countries connected to each other by research grant?

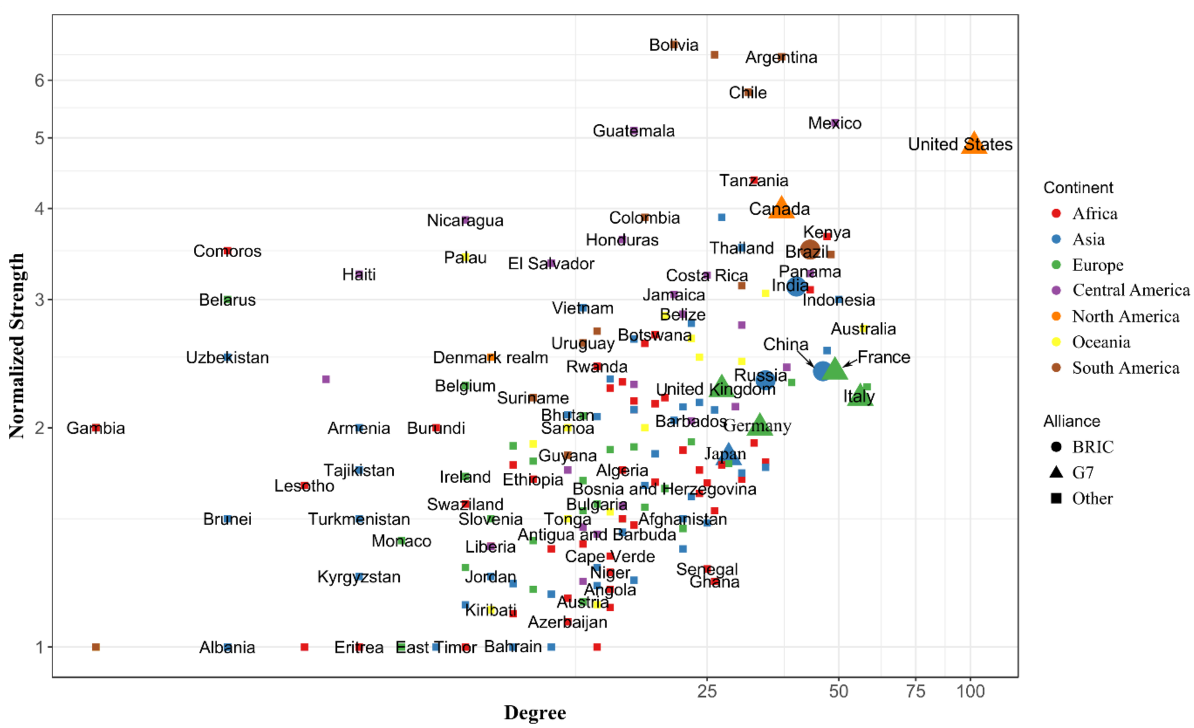


Figure 3.1: Scatterplot of Degree versus Normalized Strength in Type 1 network on a log-log graph. Each point represents a country where fieldwork was carried out. The BRIC and G7 Alliances are shown using different symbols. The color of the points represents the continent of the fieldwork country.

To elicit regional trends of "research connectivity" from the network, Figure 3.1 shows the scatterplot of the degree (number of topological neighbors of a node) versus normalized strength (sum of edge weights/degree) for each node in the graph. Each node represents a country colored according to the continent where the country is situated. The shape of each point denotes whether it is in the G7 (Canada, France, Germany, Italy, Japan, the United Kingdom and the United States), or BRIC (Brazil, Russia, India, China) alliance. The color and the shape of the points in the plot help reveal spatial and economic trends in "research connectivity". This analysis indicates that while a few countries have very high degrees of "research connectivity", led by the United States, most countries have low degrees. G7 and BRIC countries have been highlighted to show their relative high degrees and normalized strength ($p<0.01$ on comparing linear model with degree and normalized strength as dependent variable and economic status, that is, G7, BRIC, and others as predictors against a null model using Analysis of Variance), indicating many strong collaborations formed by their researchers. This probably indicates that economic factors play an important role in forming research connections. Interestingly, there also appears to be regional trends of "research connectivity" between low degree countries. The scatterplot of normalized strength against each node's degree reveal that even though a few countries may only be connected to a few other countries, these pairings between countries are strong with several grants listing the pairs together. Thus, countries like Tajikistan and Uzbekistan have high normalized strength even if they have low degrees. This is logical as countries near to each other often have similar geographic features and hence are often listed together in grants which study those features. This explains the clustering of the Central Asian countries of Comoros and Belarus near low degrees but high strength; as well as South and Central American Countries of Bolivia, Argentina, Mexico, Guatemala, and Chile towards high degree and strength.

The geographical component of the nodal information is also hierarchical in nature, i.e., country locations can be grouped into continents thus providing a depiction of connectedness at a coarser spatial scale. The aggregated view of the data at the continent level reveals that most studies take place within multiple countries in the same continent, but there are grants in which countries in different continents are involved. Figure 3.2 reveals that there are grants that connect each pair of continents with North America having the strongest ties to other continents (exemplified by the large number of points between North America and other parts of the chord diagram in Figure 3.2). In terms of SNA, aggregation might be done by looking for nodes in the graph that are more densely connected with each other, than to other nodes. Detecting topological communities using the walk trap algorithm (Pons and Latapy 2005) also reveals a similar story. The walk trap algorithm uses the idea of short random walks to find well-connected sub-structures called communities. Since there are few links that connect different topological communities, random walks are more likely to be trapped within the same community. Figure 3.3 highlights that there is one major topological community spanning all three North American countries along with considerable number of Asian, Oceania, and some African Countries. This is not surprising considering the high degree centrality of the United States. Asian countries share communities with countries from every other continent, while South and Central America belong almost exclusively in Community 4. The homogeneity of topological communities consisting of multiple countries detected in each continent shows that there are dense connections between counties in the same continent. Asian and African countries form several communities because of the large size of the continents and the large number of countries in the continents. The large number of chords which start and end in Africa and its lack of links to other continents in the chord diagram (Figure 3.2) is reflected in the bar chart (Figure 3.3) as the largest community (Community 1) containing only African countries. Interestingly, the 5th Community consists of 9 countries that are in Middle East (Bahrain, Oman, Kuwait, Qatar, Saudi Arabia, UAE, and Yemen) and the Horn of Africa (Eritrea and Djibouti), regions which are spatially proximate to each other. Thus Figure 3.2 and Figure 3.3 together help explain research collaborations in terms of spatial locations and topological connections as expressed through this spatial social network.

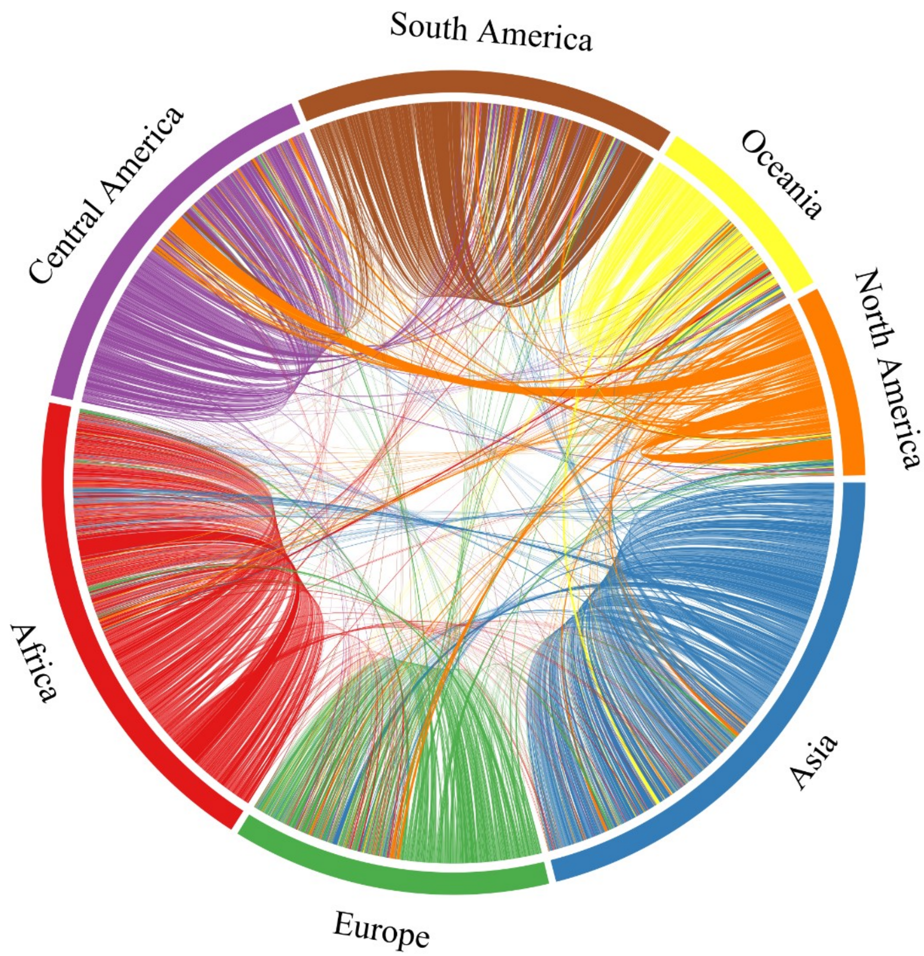


Figure 3.2: Chord diagram showing fieldwork collaborations across continents in Type 1 network. Each chord in the diagram represents a grant and connects two countries where the fieldwork was carried out. All countries in the same continent are grouped and have the same color. The minor chords along with the color provides a multi-scale representation of connections between fieldwork locations.

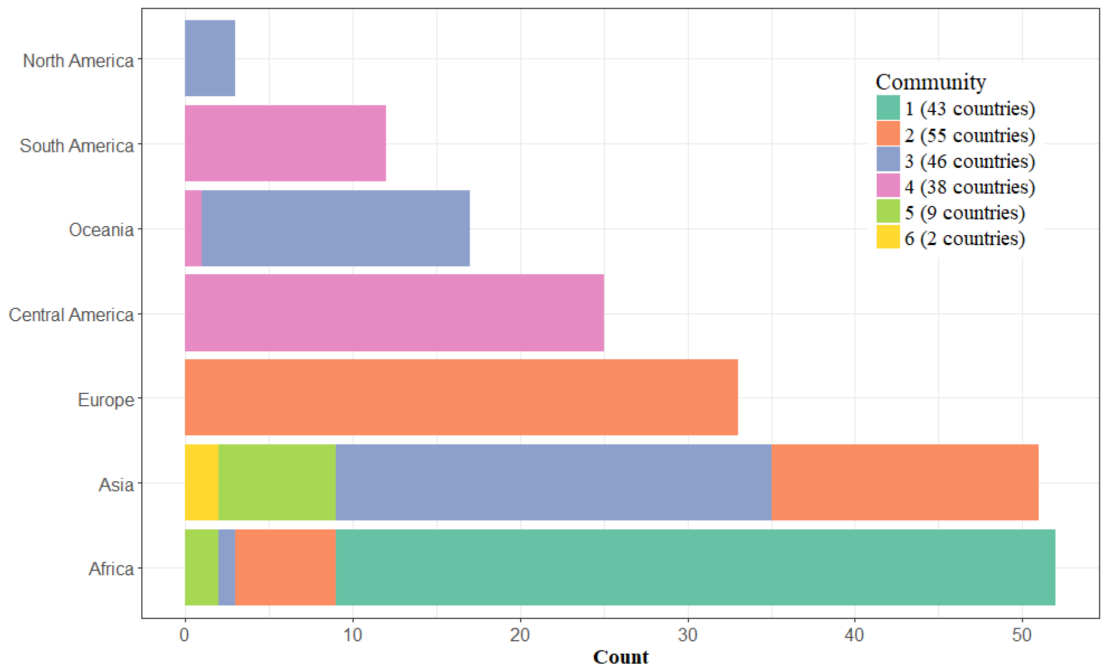


Figure 3.3: Stacked bar chart showing number of countries in each community per continent detected using the walk-trap algorithm on the Type 1 network. The legend also shows the number of countries in each community. The communities were detected using the walk-trap algorithm.

3.4.2. Network Type 2: How "internationalized" is the research of various the PIs?

In the second type of network, both the nodes and the edges have explicit spatial information attached to them. In the PI Country-Fieldwork Country network, the country of residence of the PI are connected by Fieldwork countries. Thus, if two PIs from the same country have a fieldwork country in common, then the two countries are connected by a third country, that is., the Fieldwork country. This network has 123 countries representing PI Countries and 197 places representing Fieldwork countries. The resulting network consists of 123 nodes and 1595 edges excluding self-loops and both the nodes and edges in this network represent countries. Self-loops can be of importance for this network conceptualization, as it signifies research interest in the same country as the PIs location of residence; however, we ignore self-loops as we are focusing on the connections between countries. The primary question of interest from this network then is: how "internationalized" is the research of the various PIs?

The resulting network becomes an instrument which can be used to predict possible collaborations between countries based on the patterns of PIs interests in the same field locations. As often reported in literature, in this network too, there is a high correlation between the degree and betweenness metrics ($p<0.01$) (Valente *et al.* 2008, Li *et al.* 2011, 2015), thus PI countries with a lot of connections also tend to act as bridges connecting different parts of the network. Here, the metrics are also correlated with the number of grants that were allocated to PIs from each country. The number of grants obtained by PIs of different countries is radically different with the PIs from a few countries, namely, United States (69 percent of all grants), United Kingdom (4 percent of all grants), and Canada (3 percent of all grants) have the most number of grants. The correlations imply that these few countries are also the ones that have high degree and betweenness. Consequently, it can be assumed that these countries play an important role in keeping the network connected. Even if the United States is removed, the graph still remains connected (with 122 nodes and 1473). It is also worth remembering that 46 PI country-fieldwork country pairs appear only once in the 126 years long dataset. Thus, the number of such tenuous connections were high in the database. However, the connectivity even in this network with tenuous connections are not broken if the United States is removed. This emphasizes the highly connected nature of the graph, indicating an ease of forming research collaboration almost anywhere in the world based on common fieldwork locations.

We switch from examining the country of only the PI to countries representing the location of all the co-applicant of the grant. In this case, the nodes represent not only the countries of the PI, but also that of the co-applicants listed in the grants. The grant co-applicants are often researchers from the country where the fieldwork was carried out. Thus, this network highlights the collaborations taking place specially as a consequence of fieldwork based research. Note that this represents a network that is a superset of the network in which the edges represented only the PI countries. This network thus consists of 179 nodes representing country of the researcher (PI and co-applicants) and 4426 edges representing the fieldwork countries. The significant increase in the size of this graph (number of nodes and edges) compared to the previous one highlights that co-applicants are often from different countries and including them highlights relationships between researcher countries not apparent before. Figure 3.4 shows a heatmap of researcher's countries against fieldwork countries grouped together by their World Economic Situation and Prospects (WESP) development status (WESP 2014) and then by continent. The x and y axis of the heatmap are symmetrical as they follow the same ordering of countries. The developed countries are grouped towards the origin of the plot. Japan is the only Asian country marked as developed according to WESP and it is plotted closest to the origin to the plot along both x and y-axis. Along the y-axis (Fieldwork Country) the grouping of countries according WESP development status is made clear by using labels. Along the x-axis (Country of Researcher) United States is the country separating the two classes of countries according to WESP development status. Since PIs from the United States have worked in almost every country (193 out of the 197 fieldwork countries), there is a clear line making this division apparent. For countries such as Somalia, Haiti, and Guyana, the potential for collaboration is limited to researchers from United States, United Kingdom, and Canada coming to do research. In fact, for most countries in the database, researchers from developing countries form ties to researchers from developed countries for work in their home countries and/or its neighbors, and not abroad. This is highlighted by the clustering of points along the diagonal for the developing countries, implying that for developing countries researchers, they either work in their own countries, or in nearby countries. However, PIs from developed countries tend to work everywhere, often in collaboration with co-PIs in the host or nearby countries, something the National Geographic Society explicitly encourages , and which had initially manifested itself in this second network (having both PI and co-applicants) in the form of increased network size.

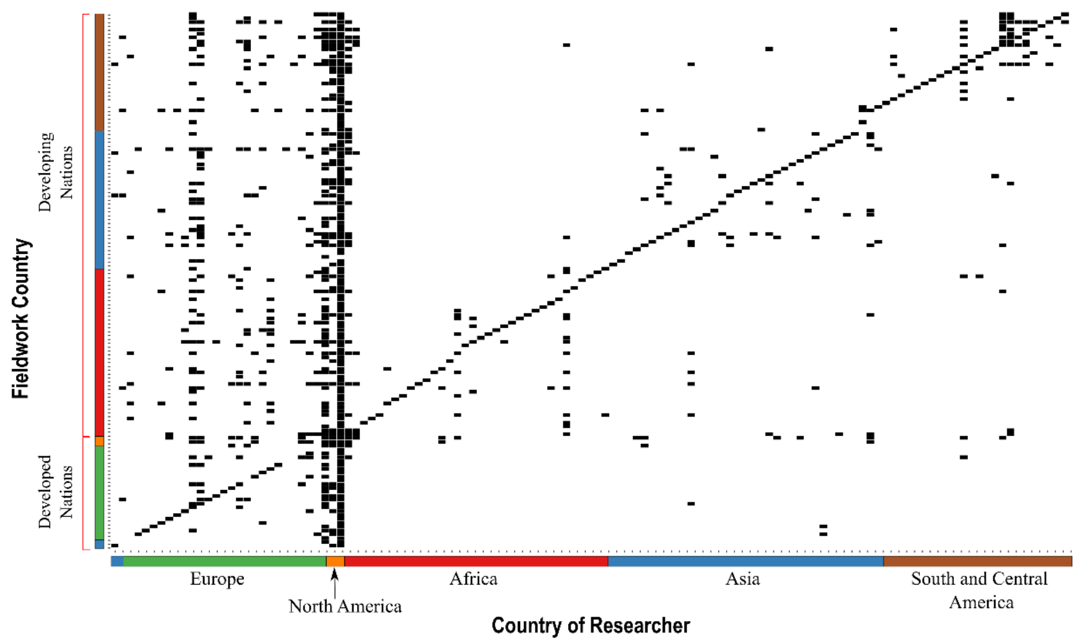


Figure 3.4: Heatmap showing the researcher's country of affiliation along the x-axis and the country where the fieldwork was carried out on the y-axis. The countries are grouped by their WESP development status and then by continents. The order in which the countries are plotted are the same along the x and y axes with North America serving as the line separating the developed and the developing nations along the x-axis. The colors denote the continent.

3.4.3. Network Type 3: What is the multi-disciplinary research potential of different countries?

In this network, nodes represent grant disciplines that are connected by the countries where the fieldwork took place. Thus, two grant disciplines were connected by an edge representing a country if both disciplines had at least one fieldwork country in common (see illustrative example in Table 3.3). Here, spatial information is associated with the edges but not with the nodes. The first two fieldwork countries and grant disciplines listed in each grant were used to create this network. Note that parallel edges between a pair of nodes were not merged to a single edge. Thus, there may be multiple edges between two nodes, one for each country that have both the disciplines in common. (For example, the nodes B and C in Table 3.3B have two edges representing Canada and India, respectively). Edge weight is associated with each edge and captures the number of instances of the same country that connect the two nodes. Thus, if multiple edges between the same pair of nodes represent different countries, they were kept as parallel edges. But, if multiple edges between a pair of nodes represented the same country, they were collapsed to a single edge and the weight of the new edge represents the number of original edges that were collapsed into the final edge. Note that in Table 3.3B, the nodes A and C have a weight of 4 on the edge denoting Canada because the disciplines A and/or C appeared with Canada as the fieldwork country 4 times in Table 3.3A. Creating the network in this manner ensures that the exhaustive set of relationships between disciplines and fieldwork country was captured. We excluded edges with edge weights below 5 (the bottom 10 percent), assuming that these associations may not be relevant considering the large size of the database. The resultant network has 80 research disciplines connected together by 20,017 edges representing 170 Countries and contain parallel edges between nodes depicting different countries.

Table 3.3: Example illustrating how the nodes, edges, and edge weights were created for Type 3 network. Table A depicts organization of entries in the database, Table B shows the weighted edge list created to make the Type 3 network.

(A)

Disc1	**Disc2**	**Disc3**	**Country**
A	E	B	Canada
C	B		India
C			Canada
B			Ireland
	A		India
A	C	F	Canada

(B)

Node1	**Node2**	**Country**	**Weight**
A	C	Canada	4

A	E	Canada	2
A	B	Canada	2
A	F	Canada	2
C	E	Canada	2
B	C	Canada	2
C	F	Canada	2
B	E	Canada	1
E	F	Canada	1
B	F	Canada	1
B	C	India	2
A	C	India	2
A	B	India	1

This network-based conceptualization, where the node is liberated from spatial constraints, proves instrumental in answering geographically relevant questions such as: Which countries are particularly conducive for certain research disciplines? Is there spatial correlation amongst countries and disciplines, i.e., do neighboring countries have similarity in the fields of research they choose to pursue? It also allows us to identify the strength of a field of research across countries, and answer additional questions such as, which disciplines have been explored in many countries? Are there disciplines on which research has been conducted in only a few countries? Does the discipline-country network have patterns where some disciplines are tightly knit together by countries?

Querying the graph for the stated questions regarding multi-disciplinary research potential achieved in different countries yields interesting results. Despite the great variety of research funded by NGS to researchers from different countries, most grants have historically awarded to United States-based researchers. Hence, according to our analysis, research on most combinations of disciplines has been conducted in the United States, accounting for almost 10 percent of the edges. Further, although the network is very well connected (the presence of several edges makes the graph resistant to be decomposed into disconnected components by the removal of a few edges), four nations, namely, United States, China, Mexico, and Canada, account for 26 percent of all edges, highlighting the predominance and achieved potential for multi-disciplinary research in those countries (19 percent accounts for G7 and 14 percent for BRIC countries). However, since our network is created based on grants that have hitherto been given out by NGS, the lack of edge connecting two countries does not imply that research in these two disciplines is not possible in that country. On the contrary, the non-existence of such links potentially points to research opportunities that may yet be untapped. For example, among European nations, Switzerland is in the bottom 10 percent of total number of edges, which suggests low level interdisciplinary research (for example, connecting Biology, Geography, Paleontology, Invertebrate, and Geomorphology). Thus, the lack of edges representing countries with low degrees represents an opportunity for more multi-disciplinary research in these countries, especially considering that their neighboring countries (i.e., Italy, France, Austria, and Germany) connect 33 disciplines, which include Geology, Geography, Anthropology and Archeology (Figure 3.5). For a well-connected network such as this one, the non-existence of edges is often more interesting than the existence of one.

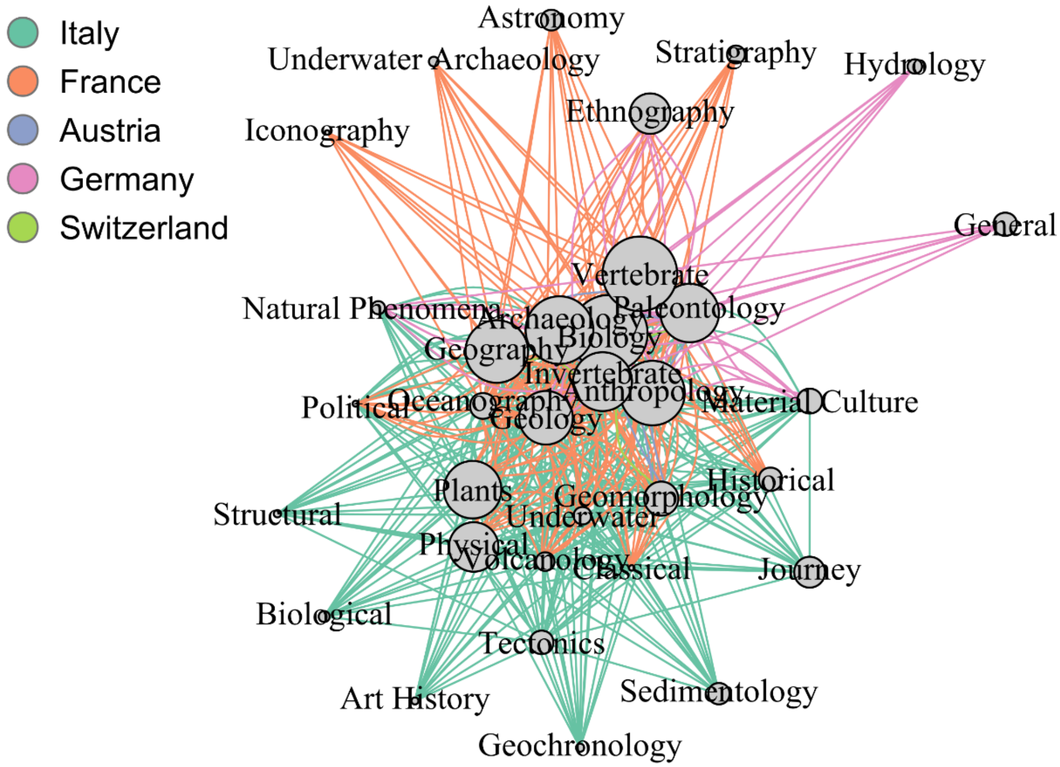


Figure 3.5: Graph of neighboring countries of Switzerland highlights potential for research using Type 3 network. Edges colored according to Country. Vertex size represents total degree of the vertex.

Shifting the focus of analysis from the edges to the nodes puts emphasis on the disciplines. Using this network, we can highlight the research disciplines that can potentially be researched in only a few countries, if the disciplines are to be explored in conjunction with each other. This implies that an edge (representing a country) must exist between the two disciplines of interest. The degree of the different nodes is a good indication of the number of countries that can be the field location for research. If one is interested in finding the potential of multi-disciplinary research, it is necessary that an edge links the two nodes. Filtering the graph to only keep nodes that have their degrees in the bottom 28 percent of the degree distribution curve (the threshold at which most number of edges can be removed without disconnecting the network) reveals "Pollution" as a theme is researched in conjunction with others, such as "Freshwater" and "Climate Change" in both United States and China (Figure 6). The threshold was selected at 28 percent because beyond threshold, the sub-graph produced contained disconnected components. The sub-graph consists of 21 nodes and 91 edges.

However, this approach of filtering the original graph to create a sub-graph also produces a significant number of false-negatives, i.e., combinations of research topics that are likely incompatible, such as “Marine” and “Terrestrial”. It is therefore up to the discretion of the user to find sensible combinations. These potentially incompatible connections are formed as a result of the discipline classifications in the database and the sheer number of connections formed in the original networks as a result of its combinatorial nature of construction. It is also worth noting that, even in this case, when trying to find a combination for a novel research topic, the absence of edges between two nodes may be more informative than the presence of one. We did not create a complementary graph (a graph with all the edges representing all the countries connecting nodes that are not connected in the primary graph) as the aim of this article was to focus on the different ways in which a social network can be made spatial, and providing examples for each type. One may also go beyond selecting just two disciplines and look for longer chains connected by edges representing the same country to find novel research topics, such as, “Innovation and Technology” in “Transportation and Communication” to reduce “Pollution” with the hope of countering “Climate Change” in China (highlighted path in Figure 3.6).

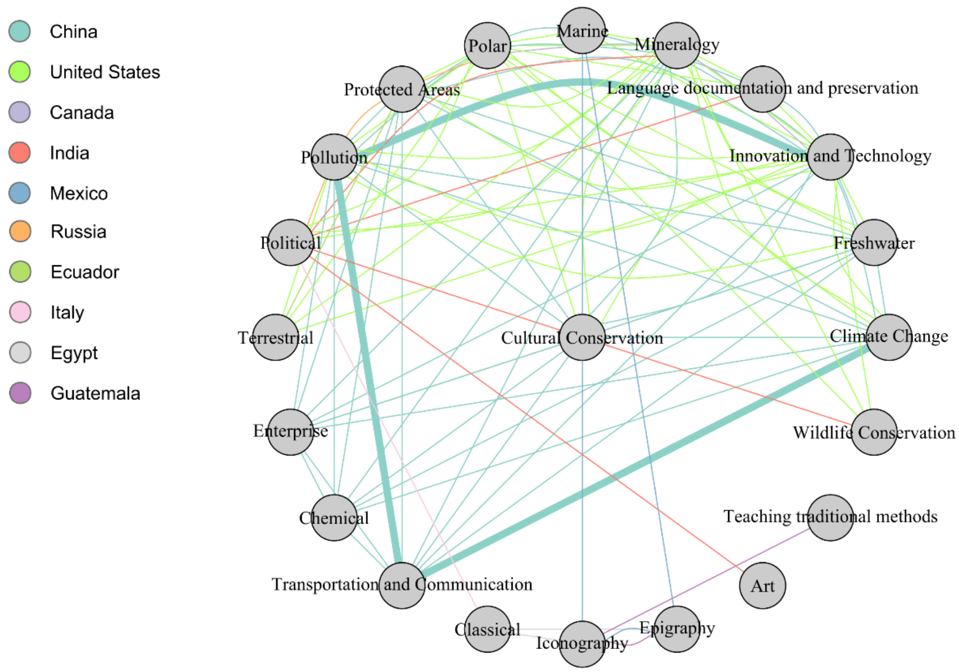


Figure 3.6: Nodes with degrees in the bottom 28 percent of the degree distribution in the Type 3 network. Edges are colored according to country. The bold edges show a path connecting "Pollution", "Climate Change", "Innovation and Technology, and "Transportation and Communication" highlighting the potential for a new research topic.

3.5. Discussion

Scientific collaborations are increasingly global, as our analysis reveals. The NGS dataset, when analyzed through the lens of Spatial Social Networks (SSN) provides a multi-faceted understanding of research collaborations.

The three networks created in this study incorporate spatial information in different ways. Thus, it is important to discuss what is shown in the network structures, how they compare to network realizations used in GIScience; and what insights the analysis provides. Type 1 networks with edges depicting countries where the fieldwork was carried out and the edges represent grants depict the most common representation of spatial social networks where the spatial information is part of the nodes. Even though we did not plot the sociogram on a map using the (x, y) of the country centroids to represent the nodes and the edges drawn as graphical artifacts, we were able to elicit regional patterns of fieldwork locations. Visualizing the entirety of the dataset would have produced a “hairball” (Krzywinski *et al.* 2012) sociogram making it difficult for visual interpretation. Type 1 networks (with spatial nodes) are probably the easiest to conceptualize and, even using this method with the NGS database, several different networks could have been made. For example, the nodes could have represented the country of researcher connected together if the same grant had researchers as co-PIs from two countries. Such a network is similar to co-authorship based networks and would not highlight the unique possibilities of using fieldwork locations as a collaboration opportunity offered by the NGS database. In addition, NGS being a US based organization, the majority of the grants went to PIs from the US. Consequently, the analysis of the alternate Type 1 network could not provide interesting insights about global collaborations in addition to that provided by large co-authorship networks. Because of the types of researchers supported by NGS, the network we created was able to highlight interesting regional trends about the groups of countries where research is often carried out in tandem as highlighted by Figures 3.1 and 3.2.

The Type 2 network is similar to a road network in how spatial information is incorporated in the structure of the graph (i.e., as part of both the nodes and the edges). However, there are a few fundamental differences between this conceptualization and a road or utility network. First, as opposed to a road or utility network, these edges (even though they have spatiality) connect nodes conceptually and not physically. Second, in this network, the edges are not linear features acting as conduits. Rather these edges are conceptual entities and hence do not have impedances to hinder connections (i.e., even though the edges have weights associated with them, the weights do not represent impedance; they signify existence of multiple connections between two nodes). If an analogy is to be drawn with road or utility networks, then an edge weight of two represents two alternate paths between the nodes, which have been represented by a single edge. Collapsing multiple edges into a single edge without loss of information is possible. Thus, higher edge weights, unlike impedance, in fact, makes connection easier as it denotes the strength of the connection (Opsahl *et al.* 2010). Third, as with impedance, other restrictions such as turns are not valid in our network. All nodes are treated equal and the movement from node to node along edges by algorithms for calculating metrics (e.g., betweenness, shortest route) is only dependent on the existence of edges and not on any other rules imposed on nodes. Fourth, in a road network, edges meet each other at nodes, enforcing topology. Edges can only overlap when they have different elevation (z-values). Such edges may be drawn on top of each other without having any physical connection between them to represent fly-overs and underpasses. A social network, even when rendered on a map, show edges only as a graphical artifact portraying connections. Even if edges overlap each other, they do not imply connectedness. This network when compared to a citation or co-authorship based network displays similarities because both primarily put emphasis on first degree connections. In a citation network, just because there is a path between two nodes may mean that the two papers are in similar, but not necessarily, in the same field. The network itself highlights the connected inter-disciplinary nature of science (Börner *et al.* 2003, Boyack *et al.* 2005, Leydesdorff and Schank 2008, Porter and Rafols 2009). Similarly, our network highlights that researchers in different countries may be interested in the same fieldwork location even if they are not working on the same topics, and consequently portrays the highly connected nature of field based research. Unlike a citation based network, this network is of limited use to understand potential collaborations beyond the first degree. Creating a finer resolution version of this network may emphasis interests among researchers in specific domains in particular field sites, and consequently potential collaborations.

In Type 3 network, spatial information is associated with the edges but not with the nodes. Understandably, this conceptualization of network spatiality is feasible only in case of social networks where the edges represent conceptual connections. In case of physical networks like roads, the physicality of the edges also enforces spatiality on the nodes. Liberating the spatiality constraints from the nodes enables the edges to be free from having length as one of the attributes, a necessary condition in case of road networks. Having a network based conceptualization where the nodes represent grant disciplines which are inherently non-spatial is difficult to represent in a GIS system. In a GIS, a topological relationship between the nodes is difficult to enforce with non-spatial nodes. The closest analogy in a GIS based system would be to add the disciplines researched in each country as attribute information and resort to attribute based queries to highlight combinations of disciplines researched in countries. Such a GIS based model will require significant data manipulation to focus on the connections between various disciplines, especially when creating larger chains of country discipline pairings to suggest potential research collaborations (Figure 3.6 highlighted path). GIS based representation, while not ideally suitable for “connectivity” based analysis, would be better suited to reveal spatial patterns in disciplinary research.

There are also limitations specific to the NGS dataset used as a case-study here. The long-time frame of the grants database implies that the countries reported change over time. We have maintained the names as reported in the database. Most grants have country-level reporting, which does not distinguish between locations within the parts of the same country. This is particularly an issue for large countries like The United States. In addition, although this database is remarkable for its geographical coverage, NGS being a U.S.-based organization has over-representation its home country. Traditionally, most grants have been awarded to PIs based in the U.S.. Whereas this database provides an interesting look at collaborations; it does not equally represent all countries. It is worth re-iterating that the primary aim of this paper was to emphasis different various ways of creating spatial social networks by using the NGS database as a case study. More sophisticated analysis which include enriching the nodes and edges with additional attribute information along with complimentary co-authorship and citation based studies should provide a more concrete picture of global collaborations.

3.6. Conclusion

Wagner, Park, and Leydesdorff (2015) claim that the growth in research output of scientific papers, which has more than doubled in the past 20 years, is primarily driven by growth in internationally-coauthored scientific papers. However, co-authorship of scientific papers is just one method of evaluating international collaborations. We highlight an innovative use of social spatial network terminology, methods and visualization techniques to emphasize the potential for answering varied geographic questions related to this growth from different perspectives using the NGS’s grants database. The attempt to incorporate spatial information into edges can be seen as an alternate means for extracting information from a network with spatial attributes. We assert that there are many forms of representation by which spatial information can be incorporated into the node and edge structure. These altered perspective yields unique insights on the nature of the spatial social network and consequently can be applied to most standard databases with spatial identifiers.

These varied network conceptualizations of the same database capture geographic information differently and provide insights into the myriad perspectives latent in a dataset. Each network representation allowed us to answer a different question. For example, Type 1 networks indicated that few countries have high degree of research connectivity and there are significant regional trends. This is akin to previous research presented by Adams et al. (2014), who identified that collaborations patterns in Africa were far from universal; instead they were affected by regional geographic factors such as history, culture, and language. We may be observing similar trends here. Type 2 networks identified that international research often means collaborating with local researchers. Additionally, researchers from developed countries work everywhere; whereas researchers from developing countries work mostly in their own countries or in nearby countries. Binka (2005) terms this 'scientific colonialism', although there are welcome signs of concentrated efforts by researchers and granting agencies to turn these into more equitable partnerships. Analyzing Type 3 networks helped in understanding and identifying potential for interdisciplinary research in different countries, which has been shown to lead to a convergence between applied and theoretical fields supporting the evolution of scientific disciplines (Coccia and Wang 2016).

Taken together, the analysis allowed new interpretations of collaborations and collaborative potential from a standard grants database. Since globalization of research is emerging as an important measure of research excellence, we presented new ideas and techniques towards analyzing large, spatially-explicit datasets. The nature of research supported by the NGS implies that considerable fieldwork is involved and consequently geographically constrained silos of knowledge production (Latour 1987, Zaggl 2017) are broken down through international collaborations that defy the standard assumptions of geographic proximity in forming scientific associations (Wuestman *et al.* 2018). The NGS database, when analyzed using the network conceptualization, provided new insight into the phenomenon of quantifying "internationalization" that goes beyond the use and limitations of demographics (e.g., percent international students and staff) and citation indices. This allows a more refined evaluation of the concept of "internationalization" in research, a modern cornerstone of defining academic excellence. There are constraints to this growth of international collaborations, as fears arise of knowledge spillovers from domestically funded research beyond national boundaries. In some cases, this may result in a slowdown in such research collaborations (Ponds 2009). Just how these fears play out as 'techno-nationalism' and eventually change the shape of collaborative networks remain to be seen, but the methods presented herein can become useful measures of the change.

3.7. Supplementary Information

The communities detected in the Type 1 network were depicted as a stacked bar chart in Figure 3.3. Figure 3.7 shows a map of the communities. This map provides a representation of the spatial location of the six communities detected using the walk-trap algorithm. Given the long 126-year history of the database all the countries represented in the database are not reflected in the map. For example, Serbia and Montenegro was constituted in 1992 and dissolved in 2006. Other countries like Greenland do not appear in the NGS database. Because of these discrepancies, some of the countries are shown in grey. The number of countries represented with community information in the map based representation (178 countries) is less than the number of countries in Figure 3.3 (193 countries).

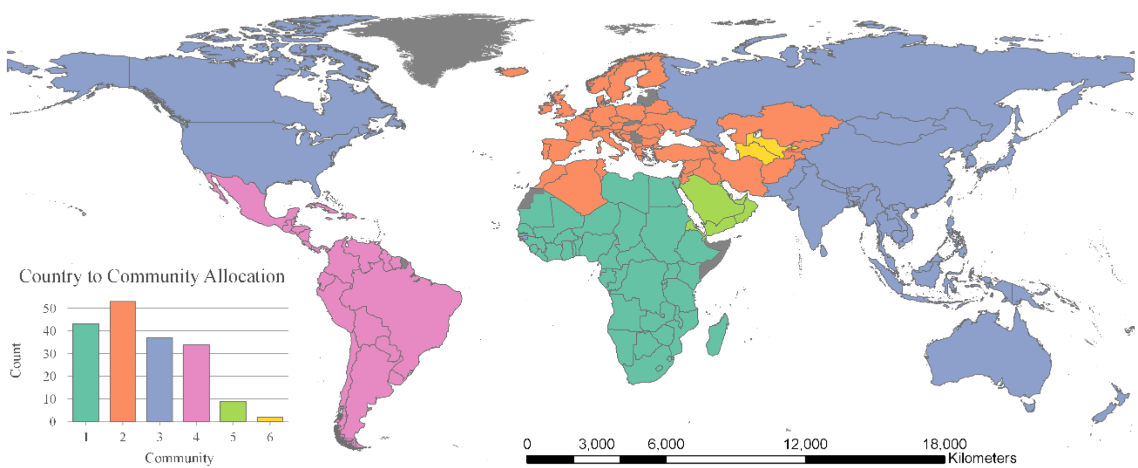


Figure 3.7: Map based visualization of the communities detected in Type 1 networks and portrayed as a stacked bar graph in Figure 3.3. The countries denoted in grey do not have community membership associated with them.

4. Metrics for Characterizing Network Structure and Node Importance in Spatial Social Networks

Citation: Sarkar, D., Andris C., Chapman, C.A., and Sengupta, R. Metrics for Characterizing Network Structure and Node Importance in Spatial Social Networks. In review at the *International Journal of Geographic Information Science.*

Linking statement: SNA relies on metrics and visualization to analyse complex network structures. However, when spatial information is incorporated in the network, there is currently a dearth of methods that can simultaneously leverage the spatial and social aspects of the network. These are needed to provide insights which go beyond the social aspect as captured by the topology of the network, and to include spatial characteristics (Chapter 2A.2). In this chapter, we present new metrics for SSNs. We present two sets of metrics for characterisation of the networks structure, and for identifying important nodes embedded in a network. The metrics rely on measurement of distances between nodes within the context of possible extremes, and provide information about the socio-spatial expanse of the network, as well as importance of nodes across different spatial scales of analysis. Thus, these metrics utilize the topological and Euclidean distances between nodes of SSNs (Chapter 2.4.1) and provide a measure of the extent of the network. They also characterize important nodes across different spatial scales (Chapter 2.4.3). The efficacy of these new metrics has been first demonstrated using simulated datasets that mimic socio-spatial properties previously reported in the literature. The primary reason for initially using simulated networks is because the properties of these networks are known, thus the outputs of the metrics can be verified as they returned the expected information. Additionally, the metrics are applied to a network of employer-employee relationships in Uganda's Kibale National Park. The snowball sampled dataset provides an ideal application scenario for the new metrics because of its significant size (number of nodes and edges) which make it difficult to visualize the network using sociograms and also challenging to identify important actors using traditional social network metrics. These metrics applied to the dataset obtained from the employer-employee network help visualize the socio-spatial structure of the network, and to identify individuals who are responsible for keeping the network connected at different spatial distances, and those who help percolate the economic benefits originating from the research station across multiple spatial scales.

Ethics Clearance: The research conducted as part of this chapter adheres to the ethics guidelines REB II as set up by the Tri-Council Policy Statement, TCPS 2 (2014) and has been reviewed and cleared by the McGill University Ethics Review Board (File No: 251-1215).

Summary: Social network analysis (SNA) offers powerful tools for revealing the structure of relationships between a set of people. These structures include basic node and edge components enriched with attribute information (such as node type or feature). These variables have been handled well in SN studies. However, associating nodes with spatial information such as location poses new challenges, as nodes are embedded simultaneously in a feature space network and a Euclidean space. We advance the analysis of spatial social (SS) networks by introducing set of new metrics for spatial social network analysis: the SS network schema, SS tuning parameter, and the flattening ratio, each of which measure edge distances between nodes within the context of possible extremes. We apply these metrics to a case study network of employer-employee relationships in Uganda's Kibale National Park, as well as two synthetic networks. These metrics applied to the dataset obtained from the employer-employee network helps identify individuals who are responsible for keeping the network connected and help percolate the economic benefits originating from the research station across different spatial scales.

4.1. Introduction

Social networks are a useful data model for studying relationships between entities that constitute a larger system. In its simplest form, a social network is a collection of nodes connected together by edges that can represent the existence of a formal "tie" or evidence of interaction (Wasserman 1994). Under the heading of social network analysis (SNA), a variety of metrics have been developed to understand the role of the nodes vis-a-vis the larger configuration of the surrounding edges, rendering the network 'other than the sum of its parts' (Wasserman 1994). As a result, certain nodes can leverage different "social" advantages due to their network position (Freeman 1978, Wasserman 1994, Borgatti and Everett 2006). Broadly, these nodes have various roles in transmitting information through the network, and serving as liaisons between other nodes (Hristova *et al.* 2015).

Much of SNA is focused on human societal relationships, such as, kinship (Dunbar and Spoors 1995, Micheli 2000), friendship (Lewis *et al.* 2008, Cho *et al.* 2011, Preciado *et al.* 2011), economic relationships (Zimmerer 2003, Abizaid *et al.* 2015, 2016, Coomes *et al.* 2015). SNA largely focuses on the topology of the node and edge structure within a "feature space" without environmental or geographic context (Sarkar, Sieber, *et al.* 2016). This schematic representation can be problematic because societies and relationships do not exist in isolation, but coexist among geographic features that affect these ties. Although technologically challenging, the geographic or spatial context has recently been recognized as an important factor of social networks (Adams *et al.* 2012) that influences individual decisions that drive network level dynamics (Strandburg-Peshkin *et al.* 2013, Farine and Whitehead 2015). Consequently, social network metrics have been conceived to characterise the societal dynamics of the entire population, embedded sub-groups, as well as to identify important nodes and edges by leveraging the topology of the network.

While there is mounting evidence to support modeling the social network as embedded in the context of the geographic environment, appropriate metrics that both characterize the socio-spatial (SS) properties of the network and identify key actors embedded in the system are currently lacking. A major discordance between social network representation and geographic space is that social network "distance" is measured in hops incurred from moving along edges between nodes, while geographic properties are described in continuous Euclidean space measured by (x, y) coordinates (Sarkar, Sieber, *et al.* 2016). Even visualization is problematic: geospatial networks create 'hairballs' (Krzywinski *et al.* 2012) and are often too dense to view in Euclidean space (Luo and MacEachren 2014). Additionally, these visualizations give the illusion that the length of the connectors reflect the geographic distance between the pair of entities they connect.

Instead, researchers have amassed a helpful set of theoretical knowledge regarding social network structure as a function of distance. First, social ties tend to be local with the probability of ties reducing with distance (propinquity) (Liben-Nowell *et al.* 2005, Wong *et al.* 2006b, Mok *et al.* 2010, Preciado *et al.* 2011). Individuals commonly choose to associate with others of similar age, nationality, location, race, income, educational level, religion, or language (McPherson *et al.* 2001); nearby nodes tend to have similar socio, cultural or demographic properties (Hipp and Perrin 2009); and social similarity is as much of an attractor as a certain Euclidean distance. Moreover, nodes that are central in the network tend to be clustered in Euclidean space (Onnela *et al.* 2011).

Whereas 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) simultaneously model a social network of individuals and their positions in geographic space to determine that the spatial closeness of two agents is a stronger cause of disease spread than their level of interpersonal interaction. Other successful models of disaggregate (e.g. agent-based) social systems in a geographic setting include those of urban gangs (Radil *et al.* 2010a, Papachristos *et al.* 2013), that ties exist between non-adjacent turf. 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 confounded with strength of relationships (Verdery *et al.* 2012).

Thus, while we have obtained significant insight on socio-spatial structures of networks, there is still a dearth of metrics that leverages this knowledge and makes available tools that can quickly identify such socio-spatial properties of the network. We extend geographically-embedded network analysis by creating new metrics under a common, burgeoning framework of SS metrics that simultaneously values a node for its network connections as well as its geographic location (Andris 2016, Sarkar, Sieber, *et al.* 2016). First, we experiment with a new scatter plot-based visualization called a *SS network schema* that preserves edge distance dispersion and network size. This approach plots each pair of nodes as a function of their network distance (e.g. one network hop, two network hops, etc.) and Euclidean distance (e.g. distance apart in geographic space) concurrently. Second, we apply a *SS tuning parameter* (α) to describe the extent to which system nodes favor nearby or distant contacts. This metric illustrates that a node deemed topologically-important merely by its ability to connect its neighbors may not be efficient in connecting clusters of nodes located at a distance, thereby losing its importance when the spatial scale of analysis changes. Third, we measure a network's spatial efficiency with the *network flattening ratio*, the proportion of the sum of network edge distances to the sum of network edge distances in a re-configured network optimized to create the closest possible ties while maintaining each node's degree.

We apply these methods on two simulated networks and a real-world network. The simulated networks mimic SS properties consistently reported in literature and serve as important benchmarks for verifying the introduced methods. On the other hand, the real-world network of economic benefits around Kibale National Park, Uganda, collected as part of a conservation effort provide a case-study highlighting the value of the metrics to provide unique insights.

We find that some nodes have relatively few connections but serve as the only connection between people in different villages, whereas others have many local ties (but not connecting villages). The modified centrality metrics using the *SS tuning parameter* helps identify both these classes of important individuals, as they play different roles in keeping the network connected at different spatial scales. The two network level parameters on the other hand provide information about the SS structure of the network, namely with regards to the extent of the network and how spatially efficient the connections in the network are in comparison to its ideal alternative rendition.

4.2. Background

Social networks are measured in many different ways. Four major types of metrics include those that characterize: the entire network, each individual node, network groups, and edges, for their systemic properties. In this work, we focus on the first two types of metrics, specifically, holistic network expanse and node importance.

4.2.1. Network Expanse

Network structural properties are calculated to analyse the population dynamics of the social network in its entirety. These metrics are traditionally non-spatial and almost entirely non-geographic, and are calculated in a feature space. Metrics like **Average Path Length** and **Network Diameter** provide intuitions about how 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 internode distance, network diameter is the maximum amongst the internode shortest distances. As a measure of the maximum distance between two of the farthest social connections, diameter provides a measure of how 'big' the network is (Hanneman and Riddle 2005, p. 81), and how much it 'costs' to reach all nodes in the network.

When considering social networks embedded in geographic space, it is also important to characterize the spatial extent of the network. Social-connections tends to be local with the probability diminishing exponentially as a function of distance, (Liben-Nowell *et al.* 2005, Wong *et al.* 2006b, Mok *et al.* 2010, Preciado *et al.* 2011) highlighting the importance of characterizing the spatial and social separation between the most 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 micro and meso-scale metrics such as degree, betweenness measure network structure as well.

4.2.2. Important Nodes

Social network metrics also help identify important network actors or entities (nodes). Such important nodes are considered to be in the "thick of things" (Freeman 1978) as a virtue of being more centrally-located in the feature space of the network than other nodes. Freeman (1978) defines degree, betweenness centrality, and closeness centrality as key metrics for assessing relative importance. Node **Degree** is the number of other nodes to which a focal node (called an 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 can be modified to accommodate edges with directionality, (ex. A connects to be B, but B does not connect to A, so the edge is non-reciprocal) 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 generally 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 others toward the node (Knoke and Burt, 1983; Wasserman and Faust, 1999; Borgatti, Everett and Freeman, 2002). Prestige is hence a more refined concept than centrality and can only be measured when incoming and outgoing edges can be separated. Using edge weights, degree can be redefined as the sum of the 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 node 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).

Given that Euclidean space and network space each reflect the different scales of social network which contribute to a nodes importance (Bronfenbrenner 1977, Boessen *et al.* 2017), how can social network metrics adapt to the influx of spatial information? We next propose modifications to the centrality metrics that identify these nodes as being important.

4.3. Methods

We introduce a set of metrics to characterise the SS structure of the network, efficiency of spatial connectivity, and to identify important nodes embedded in the spatial social network. We visualize the network using a *SS network schema*. We then provide a network-level metrics, the *flattening ratio* and node level *SS tuning parameter (α)* which provides modified node centrality measures, namely, degree, closeness, and betweenness.

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)* co-ordinates is associated with each node. For any two nodes $(p, q) \in V$ the Euclidean distance between them is represented as $|p, q|$ while the shortest path along the network is represented as $C(p, q)$. As mentioned earlier, $|p, q|$ and $C(p, q)$ are not directly comparable to each other by virtue of being defined in different measurement spaces. However, categorizing social connections at different Euclidean distances provides insight into the structure of the spatial social network.

4.3.1. SS Network Schema for Rendering of Network Expanse

The *SS network schema* plots $C(p,q)$ against $|p,q|$ to quickly and efficiently detect patterns that have been consistently reported in spatial social network literature (Figure 4.1). 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 reference line passes through y=1 (shortest path=1), where 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. this example y = 1. However, a geographical perspective also requires examining distances of second, third, etc. degree ties, which can be achieved partially by using different centrality measures.

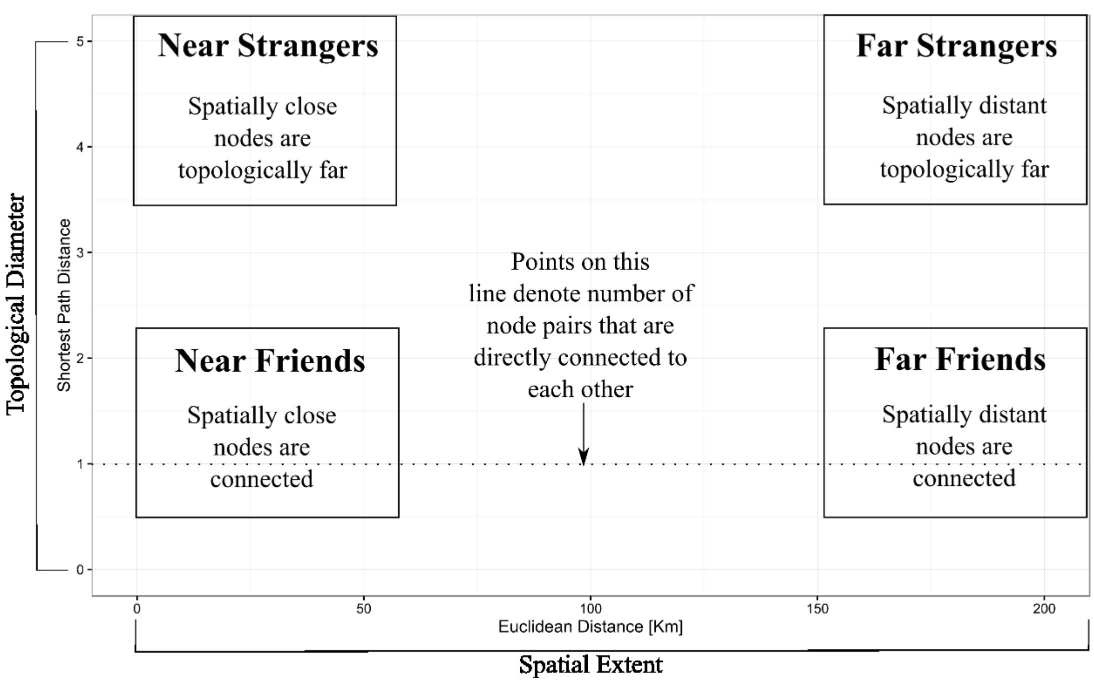


Figure 4.1: A SS network schema illustrates the different scenarios that any pair of nodes in a spatial social network could incur.

4.3.2. Flattening Ratio for Measuring Spatial Efficiency

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 $\bar{G}$ with degree D connects to its nearest D neighbors 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,

$$F_s = \frac{\sum \overline{|pq|}}{\sum |pq|} \ \ldots \text{(Eq. 1)}$$

Where as before $|pq|$ denote the Euclidean distance between any two connected node in G, while $\overline{|pq|}$ 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 neighbours (i.e. via shorter connections), the overall distance $\sum \overline{|pq|}$ is expected to be less than $\sum |pq|$. Thus, F_s provides a measure of the efficiency with which nodes in G connected to their closest spatial neighbours. The closer the ratio is to 1, the more efficient network G is to an ideal degree constrained nearest neighbour network. Note that, like the Erdős-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 thus for a given G, several $\bar{G}$ are possible. Thus, to calculate F_s, we took the average of several iterations.

4.3.3. SS Tuning Parameter for Measuring Node Importance

To reflect the important nodes in the social spatial context, we introduce the *SS tuning parameter* α to balance the importance of near versus far social connections. We calibrate the value of this parameter between 0-1 depending on whether geographically-farther or closer social connections are preferred: if $\alpha = 0$, then farther connections are preferred, while if $\alpha = 1$, the nodes with nearby social connections are considered more important. If $\alpha = 0.5$, both far and near connections in Euclidean space are given equal importance. Thus, nodes that either have intermediate connections, or a mixture of far and near connections are considered important at $\alpha = 0.5$. Thus, different values of α captures the importance of the nodes at different spatial scales of analysis. It is 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 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 (amongst all the alternate paths between i to *j*) is calculated by Dijkstra's algorithm (Dijkstra 1959) as the total cost of travelling along the edges from node i to *j*:

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

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

$$d_{F_{i,j}} = min(\frac{1}{|i,a|} + \frac{1}{|a,b|} + \cdots + \frac{1}{|x,j|}) \ldots \text{(Eq. 3)}$$

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.

We introduce the *Spatial Centrality* metric, defined here as X_s for node i, where the tuning parameter α ranging from 0-1 is applied. X_s can in turn refer to the three measure of node centrality, namely, degree, betweenness, and closeness. X_s is defined as:

$$X_{S_i} = \alpha \cdot \overline{X_{N_i}} + (1-\alpha) \cdot \overline{X_{F_i}} \ \ldots \text{(Eq. 4)}$$

$$where,\ \overline{X_{N_i}} = \frac{X_{N_i} - \min(X_{N_i})}{\max(X_{N_i}) - \min(X_{N_i})},\ \overline{X_{F_i}} = \frac{X_{F_i} - \min(X_{F_i})}{\max(X_{F_i}) - \min(X_{F_i})} \ldots \text{(Eq. 5)}$$

Where X_s can take on three different values: *degree* (D_s), *closeness* (C_s), *and betweenness* (B_s), i is the focal node and j represents all other nodes. T 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:

$$D_{N_i} = \sum_j^T \frac{1}{|i,j|}\ , \ \ D_{F_i} = \sum_j^T |i,j| \ \ \text{... (Eq. 6)}$$

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

$$and,\ C_{N_i} = \sum_j^T d_{N_{i,j}}\ , \ C_{F_i} = \sum_j^T d_{F_{i,j}} \ \text{... (Eq. 8)}$$

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.

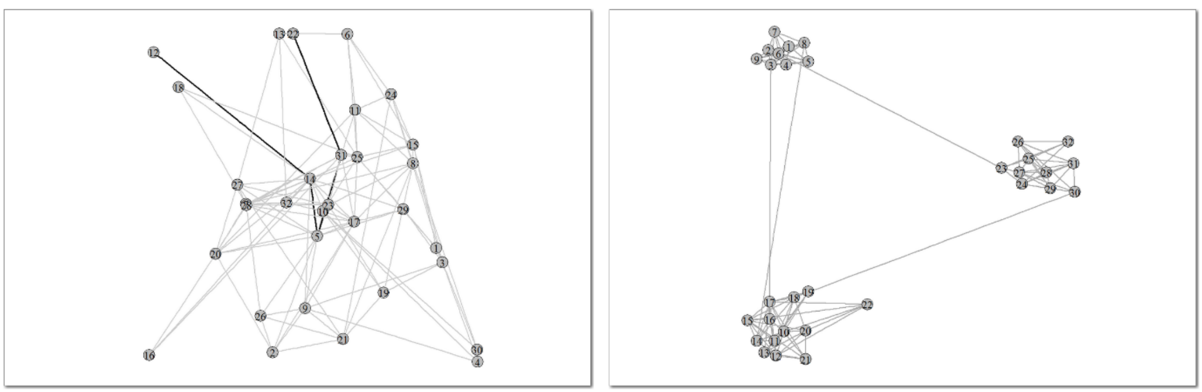
The values for $B_{F_i}, B_{N_i}, C_{F_i}, C_{N_i}, D_{F_i}, D_{N_i}$ are normalized to be in the range $[0,1]$ by linear scaling because as with many other metrics in SNA, these metrics are better suited to provide a ranking of importance of node rather than quantify the difference in influence between the nodes as given by the raw unnormalized values (Bonacich 1987, Borgatti 2005). In this study, 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 from the metrics. These centrality measures, like many other metrics in SNA, by design, are better suited for providing a ranking of importance of node rather than for quantifying the difference in influence between nodes (Bonacich 1987, Borgatti 2005). Thus, the metrics should not be used to compare different networks by comparing their scores.

4.4. Data

4.4.1. Simulated Data

We created two simulated datasets based on prior calibrations of inter-node distance (i.e. propinquity) and node distribution. These datasets were created because the results provided by the proposed methods, when applied to them, are predictable as a consequence of the properties of the simulated networks being known. Thus, the results obtained can be verified to see whether the methods perform as expected before applying them to real-world datasets with unknown socio-spatial properties. The initiation metrics for the simulated data are derived from consistently-reported accounts of distances between nodes (Festinger *et al.* 1950, Mok *et al.*

2010, Preciado *et al.* 2011) and spatial distribution of nodes (Fischer 1977, Butts *et al.* 2012). In a 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 4.2A). Second, in a synthetic **Clustered Network**, nodes are clustered at different Euclidean distance inside a bounded feature space, and the probability of forming an edge reduces exponentially as a function of the distance between the nodes (Figure 4.2B), 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 verification of spatial and social properties).



A B

Figure 4.2: 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.

4.4.2. Kibale Employment Network

We also focus on a real-world dataset on an employment network near Kibale National Park, Uganda (hereafter Kibale) to complement the results obtained from the synthetic networks. However, there is an obvious lack of predictability of results for a real-world network. Nevertheless, the use of this dataset allows for exploration of real-world spatial patterns for a better link to the geographic space in which they are embedded. Employers and employees were interviewed in-person between January 2016-May 2017. Participants were initially identified through from employee records furnished from Makerere University Biological Field Station (MUBFS). Next, more participants were identified through a snowball sample, resulting in 209 participants from 21 villages (including people who lived at the field station itself). 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. Most of the hiring is done for household and farm work with the same person sometimes hired by different employers. We contacted the agricultural and household workers who provided us with information for those they also had hired. As many chains as possible were followed, until an individual on the chain did not hire anyone, a person could not be contacted, or lived more than approximately 10 kilometers distance from the field station by motorable road (mixture of paved and unpaved surface). People who did not hire anyone were excluded. This data was subsequently used to develop a network where the location of each individual was geolocated to the village of their residence, and the edges represents employer-employee relationships. The resultant network is a network with 163 nodes and 155 edges in 21 connected components. The largest connected component (Figure 4.3), consisting of 97 nodes connected by 106 edges was used to demonstrate the effectiveness of the metrics introduced in this paper.

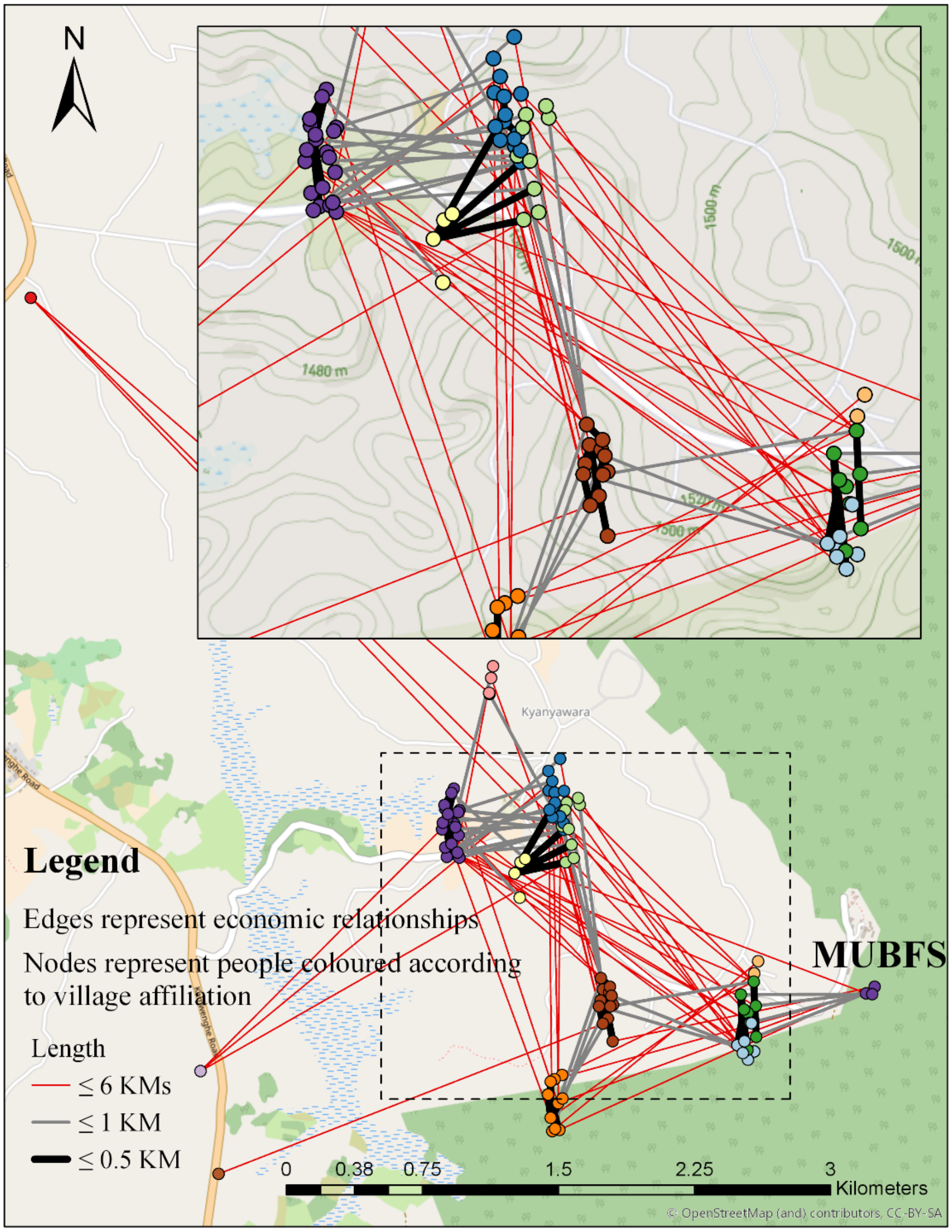


Figure 4.3: The Kibale network of economic benefits depicts workers and employers as nodes and their edges as relationships. There are a number of small villages that have internal connections (depicted mostly in black) while connections exist more frequently between these areas. (The position of the nodes has been jittered to prevent overlap of nodes in the same village.)

This network was collected with the aim of quantifying the percolation of benefits originating from the research field station located in the park (Sarkar et al. submitted) and how these benefits diffuse from the research field station through employer-employee relationships. The MUBFS is one of the longest continuously running research field sites in Africa and thus provides a unique case study to understand the impacts of such an establishment on the livelihoods of the community living near the park. In this network we hypothesis that important individuals are the ones who are responsible for spreading the benefits across different spatial scales. Thus, they are characterised as having a good mix of near-far connections, thus keeping the network connected at different spatial scales.

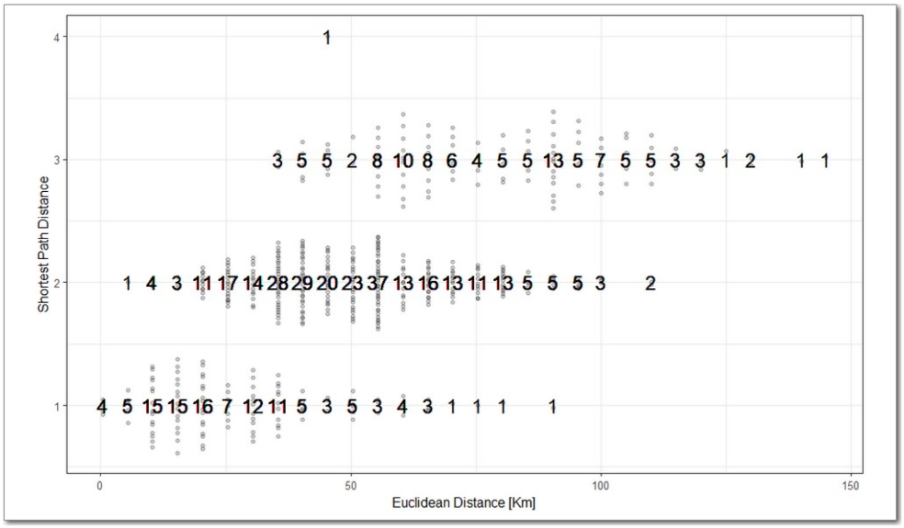
Contextually, the lack of telecommunication infrastructure and the relatively high cost of keeping a mobile phone means that most of the interaction is still carried out in-person (Sarkar, Chapman, *et al.* 2016). Although Kibale is a mid-altitude moist evergreen forest, the area around the field-station is relatively flat and the primary means of transport are by foot, bicycles, or motorbikes. The roads near the park are unpaved and people often take shortcuts through fields making travel distance difficult to estimate. Thus, we use the Euclidean distance between the villages as the measure of distance.

4.5. Results

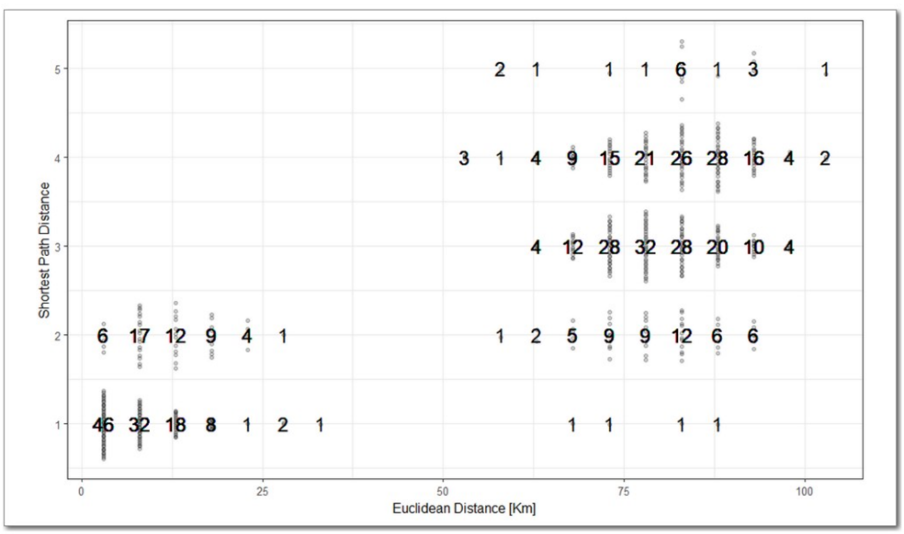
4.5.1. SS Network Schema

The *SS network schema* was utilized to investigate the overall structure of the spatial social network. For the synthetic networks (Figure 4.4), the points have been aggregated at 5-kilometer intervals and jittered vertically on the plot to give a visual cue as to how many points are at each x, y coordinate. The numbers provide a count of the number of points at each geographic coordinate. 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 nodes that are farthest apart in Euclidean space is approximately 150 kilometers for the **Poisson Network** and 100 kilometers for the **Clustered Network** (Figure 4.4). In terms of SNA, the diameter of the networks are 4 and 5, respectively, meaning that all nodes in the network can be reached from each other by traversing a relatively few edges The Y axis with only the value

Shortest Distance = 1 is akin to plotting a histogram of the distance between social connections, which have been frequently used to assert distance-friendship patterns.



A



B

Figure 4.4: SS Network Schema for (A) Poisson and (B) Clustered simulated spatial social network. Points are aggregated at 5 km intervals.

For the **Poisson Network**, the generative process implies that if nodes are far apart spatially, then they are unlikely to be connected and thus, most social connections should be "local". Thus, distant nodes will have many intermediaries for connection (larger shortest distance). The clustering of points in the "Near Friends" zone (bottom right) of Figure 4.4A, highlight the local nature of connections and large number of points along the line *Shortest Distance = 2*, indicates that most nodes are reachable from each other within 2 steps. The farthest nodes in Euclidean space are separated by 3 hops in network space. Figure 4.4A also highlights an interesting outlier node, which only has one connection (which is unascertainable from the plot) but this connection is four hops away, despite its relative nearby distance of 45 kilometers. This 4-hop path is highlighted in figure 2 for clarification. This case cannot be derived easily from the standard sociogram method, but is clearly evident using this visual method.

For the **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 4.4B) has many points while, there are very few "Far Friends" (i.e. an absence of points in the lower right hand quadrant). Without prior knowledge of the generative process, one can detect the spatial clustering and exponential decay of tie formation from the large empty space in the middle of the chart and the large number of points towards the extremities. Both the sociogram and the *SS network schema* illustrate that about one-third of friends are nearby and two-third of friends are far away), yet the *SS network schema* confirms that 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 possible combination with each other network node (alter).

In the **Kibale Employment Network**, the average distance between the employer and the employee is 0.9 Km and thus, we set the resolution of the *SS network schema* has been set to 1 Km so that the local connections are not aggregated to a courser spatial scale. The small numbers along *Shortest Path Distance*=1, along with the relatively large diameter of 5 hints at a sparsely connected network. The spatial extent of the network is also relatively small at approximately 6 kilometers in diameter, although most nodes are located 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 in the network have an edge between them. The heavy clustering of points in the near friend region highlights that most hiring is local. However, some of the topologically farthest connections are spatially close, implying that a lack of opportunities in one's village may necessitate travel to find work. These 2 extreme cases point to potentially interesting employer-employee dynamics. The anchored sociogram in Figure 4.3 verifies that each distant connection is sustained by a single individual. It also points at the possibility that after 2 km, employment may to be driven by the individual's reputation rather than his location, as all three long distance (>2 kilometer) connections are sustained by a single node. Moreover, some node pairs are close together in Euclidean space, they may be far apart in network space. This highlights the fact 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 of the hiring is done for farm work, making it difficult for a person to be hired for the entire day at more than one farm.

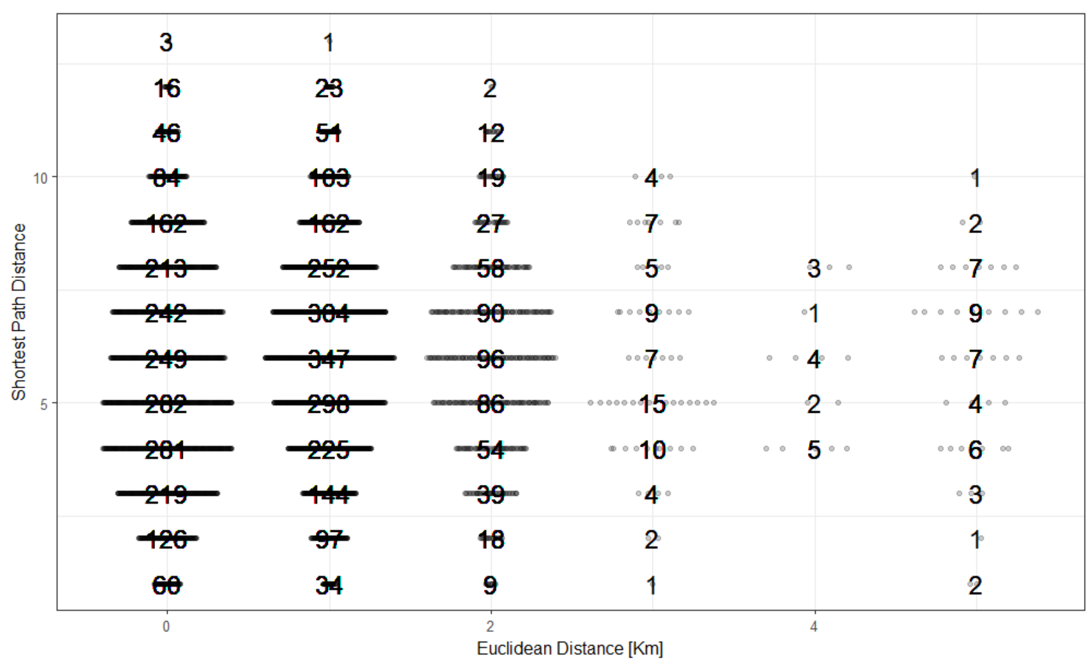


Figure 4.5: SS Network Schema for Kibale spatial social network of economic benefits. Points are aggregated at every 1 Km.

The geographical network map (Figure 4.3) 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 network schema* (Figure 4.5) provides discernable 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).

4.5.2. Spatial Network Efficiency (Flattening Ratio)

In case of the simulated networks, the original networks are highly spatially optimized as the probability of long connections decreased exponentially. Thus, the flattening ratio for the **Poisson and Clustered Networks** are 0.797 and 0.923 respectively by taking the average of 10 trials which of which created a different flattened network. For the **Kibale Employment Network**, the flattening ratio is 0.212, implying that the original network is far from being spatially efficient. This may be due to several hires from distant villages, and to the significant number of “close strangers”, which get optimized in the flattened networks. (See Supplementary Information for different flattened networks generated from the original networks).

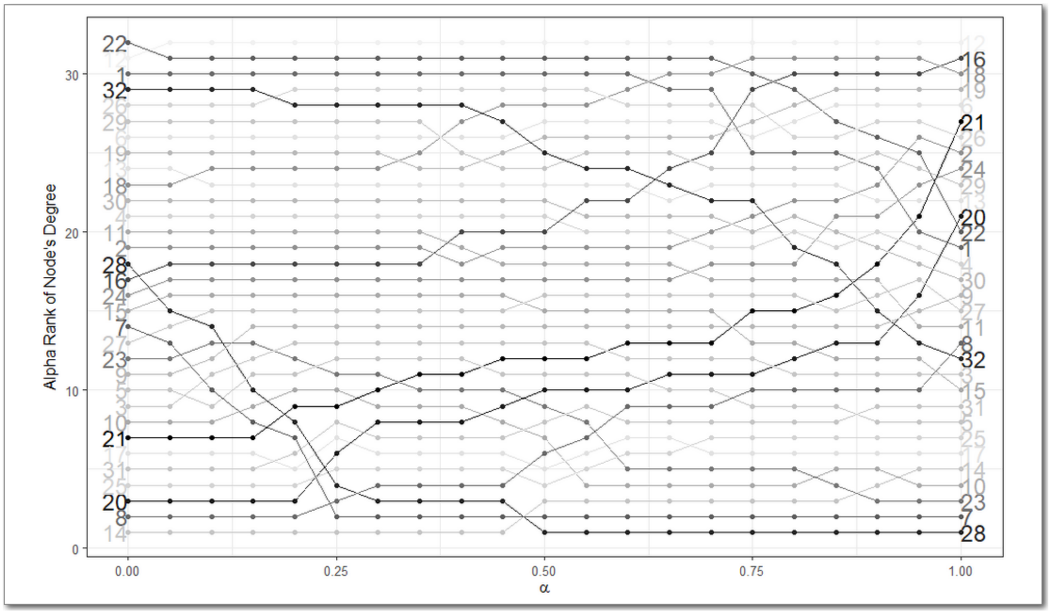
We also experimented with elevation as a cost parameter, on top of Euclidean distance, in the Kibale flattened network, since steep roads impede pedestrian mobility—a key mode of transportation in Kibale. The elevation in the study area ranges from 1,445-1,557 meters (via a 30-meter resolution digital elevation model) (USGS 2006). Each edge was given a cost weight equivalent to the range, that is, the maximum elevation of its path minus the minimum elevation on its path. The sum of the change in elevation for all original edges was 1519.066 meters, and for flattened networks, ranged from 1212.2-1512.3 meters. Thus, the original network, with longer edges, also incurred more elevation change, as can be expected.

4.5.3. Identifying Important Nodes via SS Tuning Parameter

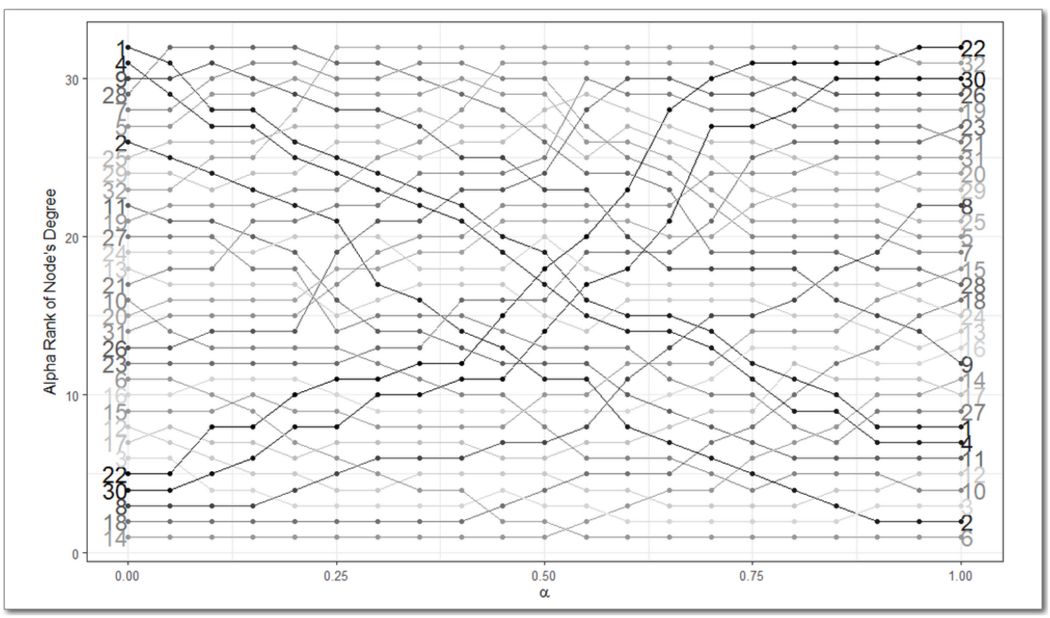
The following plots allow the user to detect key nodes that shift, given a changing emphasis on spatial nearness versus network nearness as the most valuable qualities of a node. Using the **Poisson network**, node importance, as predicted by degree centrality (D_S), changes with different values of α (Figure 4.6A). Here, the X axis shows the values of α, while the Y axis shows the rank of the node at a particular α, beginning with the highest-ranking node (rank 1) and incrementing to the lowest ranking node. The numbers correspond to node ID. The lines depicting the rank of a node at different values of α have been provided a transparency gradient according to their total change of rank between $\alpha=0$ and $\alpha=1$. As expected, nodes with the longest 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 distance, remains relatively important at all values of α.

In the **Clustered network**, node 22 and node 6 are remarkable in their significant variation in ranks as calculated by degree centrality (D_S) (Figure 4.6B). 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 5 nodes in the nearest cluster. Its relatively intermediate distance connections make significantly lose its ranking after $\alpha=0.4$. However, its connections are not near enough to make it important at $\alpha=0.5$. The classification of the term "intermediate distance connections" is however network dependent dictated by the distribution of the connection distances. 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 through-out. 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, the Poisson network is more stable than the clustered network given a shift in emphasis on social vs spatial as an indicator if importance. This can be seen by the many crossing lines in Figure 4.6B compared to Figure 4.6A.



A



B

Figure 4.6: Degree Ranking of the nodes at different α values for simulated networks. Transparency of the lines depict total absolute value change in rank between α=0 and α=1. (A) Poisson Network and (B) Clustered Network.

The considerable number of disconnected components (21) in the **Kibale Employment Network** begets evaluating the importance of the nodes in the largest components in terms of their role in keeping the component connected across different socio-spatial scales. The graph of change in the betweenness values (B_S) (Figure 4.7) highlights the shifting influences of the different nodes to keep the network connected across α values. Node 4 is arguably the most important as s/he connects 6 clusters (people in different villages) at different Euclidean distances and hence maintains a stable top rank. Node 23, on the other hand, significantly improves ranking between *α=0* and *α=1*, serving as connection between residents of the same village who are not connected by any other path. Node 18 only employs people located in different villages that are quite far from their village of residence and hence serves as an crucial point of connection between the pair of villages when distant connections are preferred.

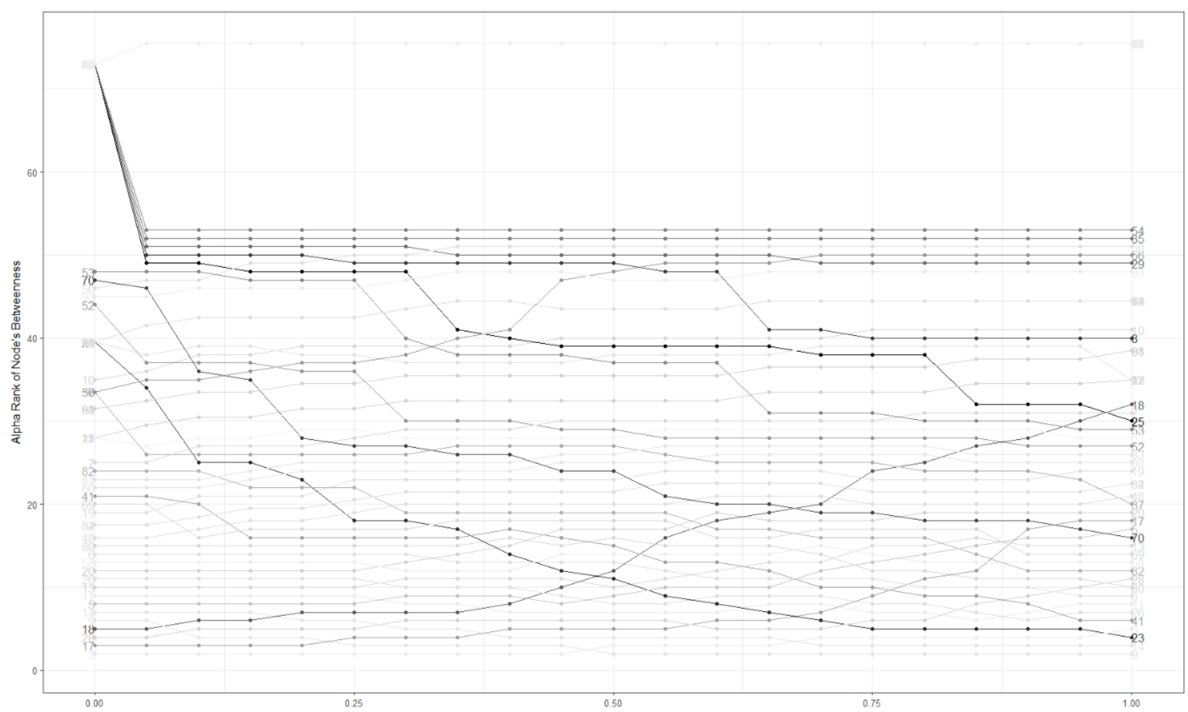


Figure 4.7: Betweenness ranking of the nodes at different α values in Kibale Employment Network. In case of a tie, all the tied nodes are given an average rank. Transparency of the lines depict total absolute value change in rank between α=0 and α=1.

The use of 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 that 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 skillful, which is may be the case for the individual depicted by node 12.

In a spatial social network, it is not just important to look at the number of connections, but also to look at the contribution of individuals in both maintaining and spreading the 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 characterisation of the spread of economic benefits, the absence of "important" individuals will spatially condense the economic benefits originating from the research field station. Since studies have shown that economic gains from protected-areas impact communities perception of conservation plans, it is important to ensure equitable percolation of the benefits in communities surrounding the park. The combination of degree and betweenness importance of individuals as reflected by the modified metrics, helps identify specific individuals who fulfill this important role.

4.6. Discussion and Conclusion

Currently, SNA relies on a battery of metrics to provide insights about social networks, but which do not readily consider space. We provide new methods to analyze a spatial social network that takes both Euclidean distance and underlying social structure into consideration. In essence, we extend and introduce new metrics to accommodate the spatial embedding of the network, and show its utility through the use of two simulated and one real-world network.

We introduce a scatterplot based visualization approach which provides an overview of the socio-spatial structure of spatial networks. By plotting the shortest network paths required to connect nodes at various Euclidean distances, the visualization contextualizes the distribution of the nodes and connections in Euclidean space and provides the network diameter and the spatial extent. Furthermore, we propose simple variations to the three commonly used centrality metrics, namely, degree, betweenness, and closeness, to characterize the importance of nodes in the network when embedded at different spatial scales. By using a 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. In addition, the network flattening ratio characterizes the spatial efficiency of networks by comparing it to its ideal most spatially efficient counterpart.

Whereas the modified centrality metrics affords insights to the socio-spatial importance of the nodes, they may not be used independent of the standard definitions of the corresponding metrics as defined by Freeman (1978). The metrics rely primarily on the distance of the connections rather than the number of connections of a node. Moreover, there may be instances where a node with a significant number of relatively far connections is interpreted as being important at $\alpha=1$, if all those small values add up to be significantly large, thus, portraying its importance to be greater than a node with a few very close connections. Using the metrics introduced in this paper in conjunction with standard definitions helps identify such outliers. Furthermore, one must be careful not to attach excessive value to the results of the SNA centrality metrics. Importance to nodes as assigned by the centrality metrics tend to be highly correlated (Valente *et al.* 2008, Li *et al.* 2011, 2015) (See supplementary information for correlation matrix for the Kibale network). Interesting information might be gleaned from the outliers which have not been marked important by the metrics. Thus, 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.

While the scatterplot visualization of network structure we introduced provides a summary of the spatial social network, the resolution or the scale at which points should be integrated require careful deliberation. In case of the Kibale social network, a resolution of 1 km was considered appropriate as several of the connections are local and thus a resolution of 1Km was able reflect the variations of the distance in employment relationships (Sarkar *et al.* n.d.). In case of the two simulated networks, the network with clustered nodes was simulated first with the median distance between the connections as approximately 10 km, thus resolution for visualization was 5 km. The network with the uniform distribution of nodes was created with approximately the same network density and spatial extent as the one with clustered nodes to facilitate comparison. As a general guide, it is recommended to set the resolution of the visualization to be half the distance of the average connection. In practise, it may be worth setting the resolution to reflect the generative processes of the network, or to match the aim of the study.

The node's locational information afforded by spatial social networks also provide opportunities to understand the network in conjunction with other geographic data which are known to influence tie formation, such as the nature of the built environment (Lund 2003, Hipp *et al.* 2014, Boessen *et al.* 2017). Thus, the metrics and visualization used here should be used to complement standard metrics and sociograms. 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 (Entwisle *et al.* 2007, Abizaid *et al.* 2015). 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 important driver in tie formation, compensating for lack of a central position in a social network (Owen-Smith and Powell 2004), or, conversely, may stop certain entities from forming alliances if they are located geographically far, even though socially central (Fleming and Sorenson 2001). Considering geography, as well as social networks, becomes indispensable to understanding the duality between the importance of location versus network (Castells 1996). Thus, use of new SS techniques as proposed herein, along with maps, are required to identify important nodes.

4.7. Supplementary Information

4.7.1. Simulated networks

Poisson Network

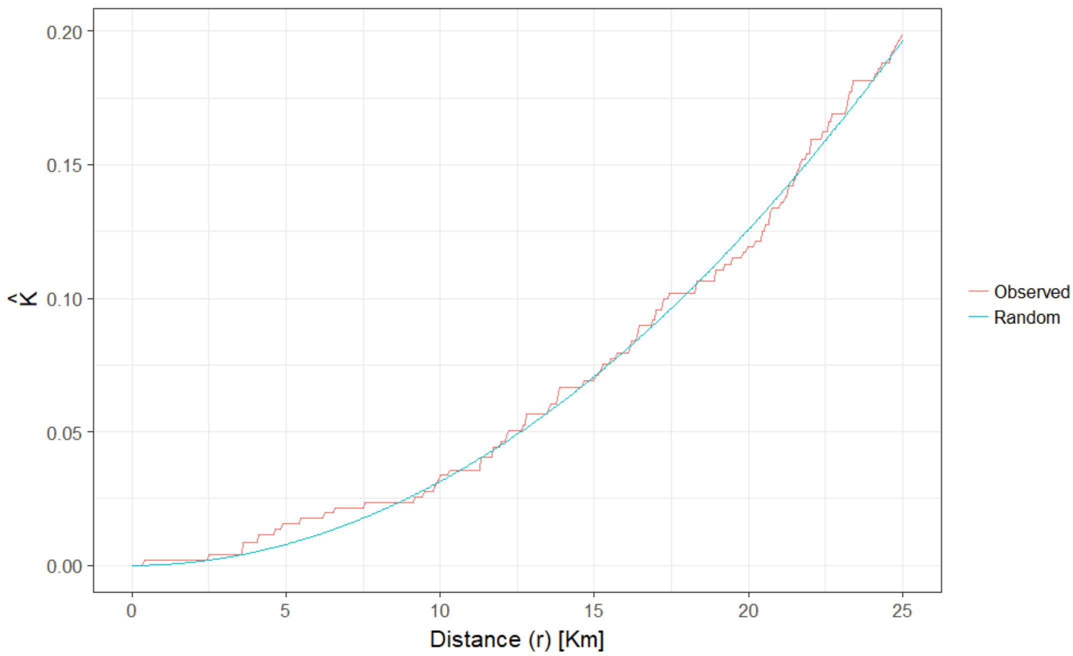


Figure 4.8: Multi-distance spatial cluster analysis showing uniform distribution of nodes in Euclidean space. The blue line shows the expected number of points for a perfect Poisson distrisbution and the red line shows the observed number of points in the dataset.

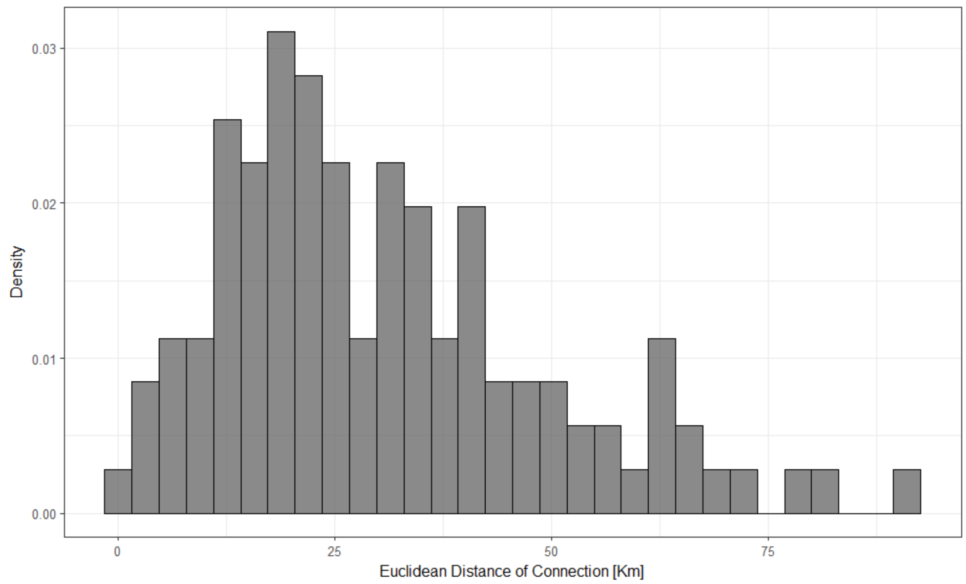


Figure 4.9: Histogram of connection distances showing exponentially decreasing probability of connection with distance.

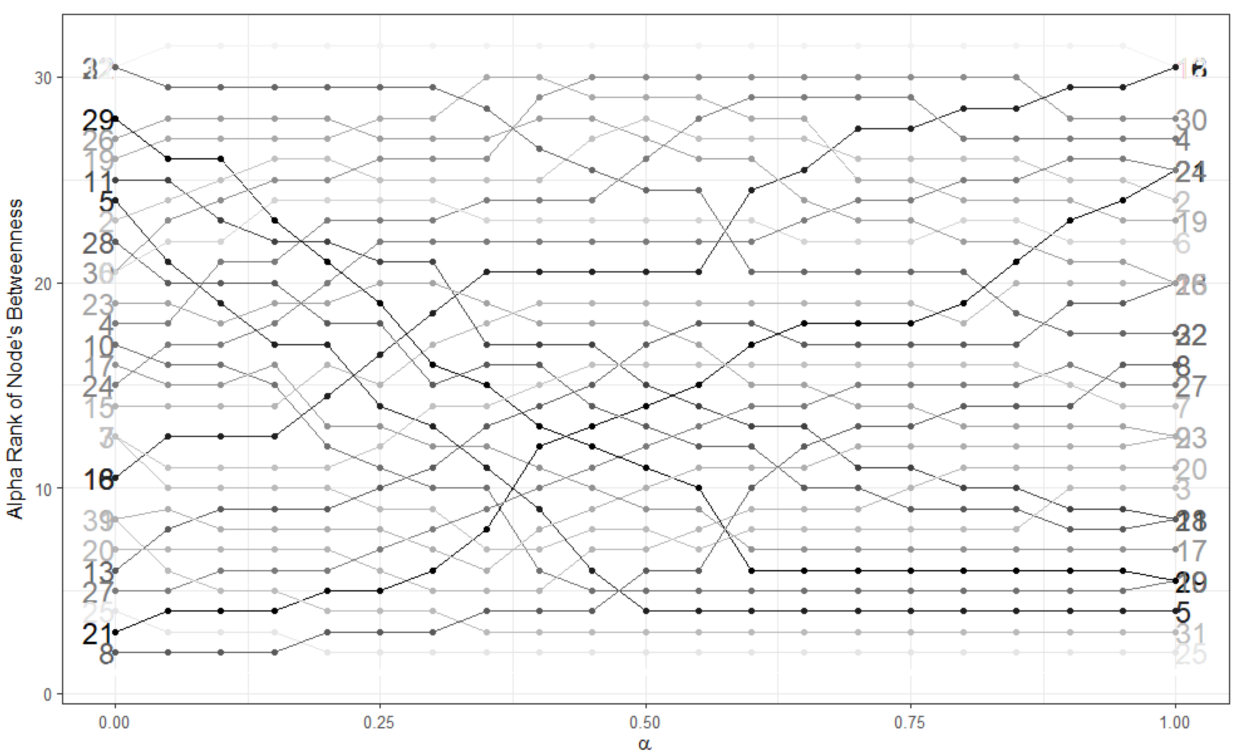


Figure 4.10: Betweenness Ranking of the nodes at different α values for the Poisson Network. Nodes which are tied for a rank are given the average rank.

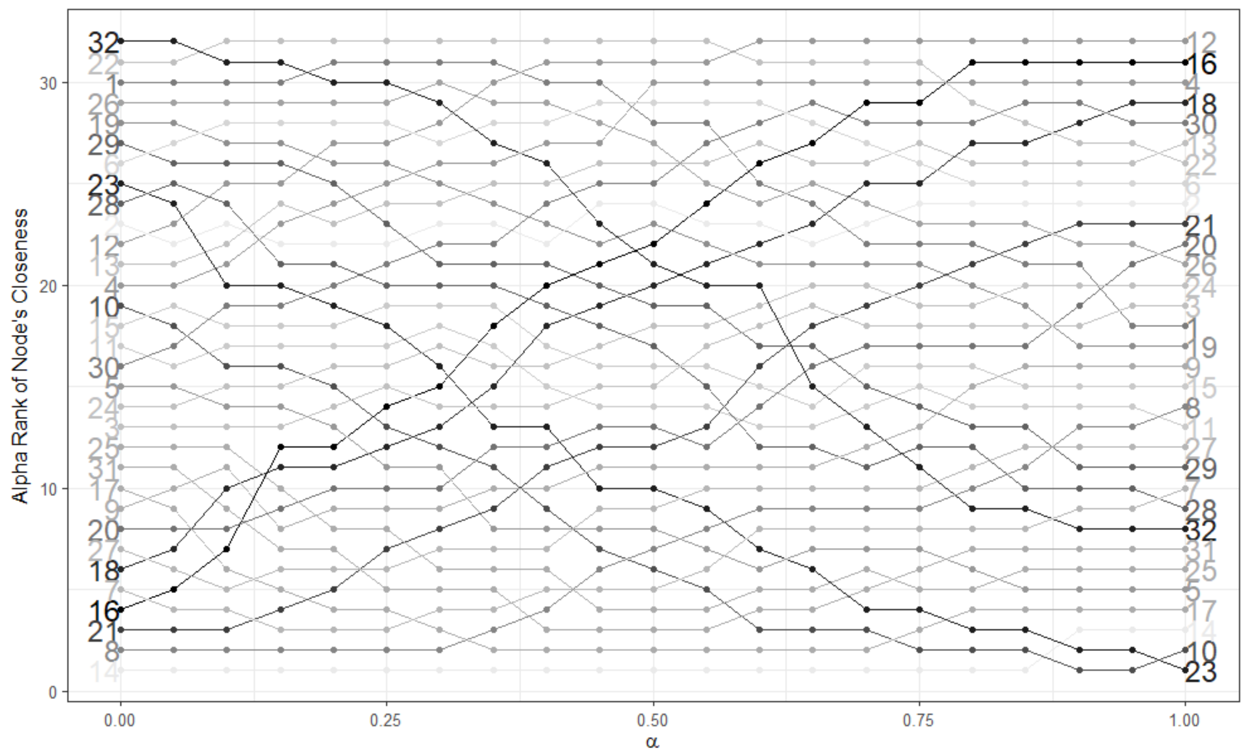


Figure 4.11: Closeness Ranking of the nodes at different α values for the Poisson Network. Nodes which are tied for a rank are given the average rank.

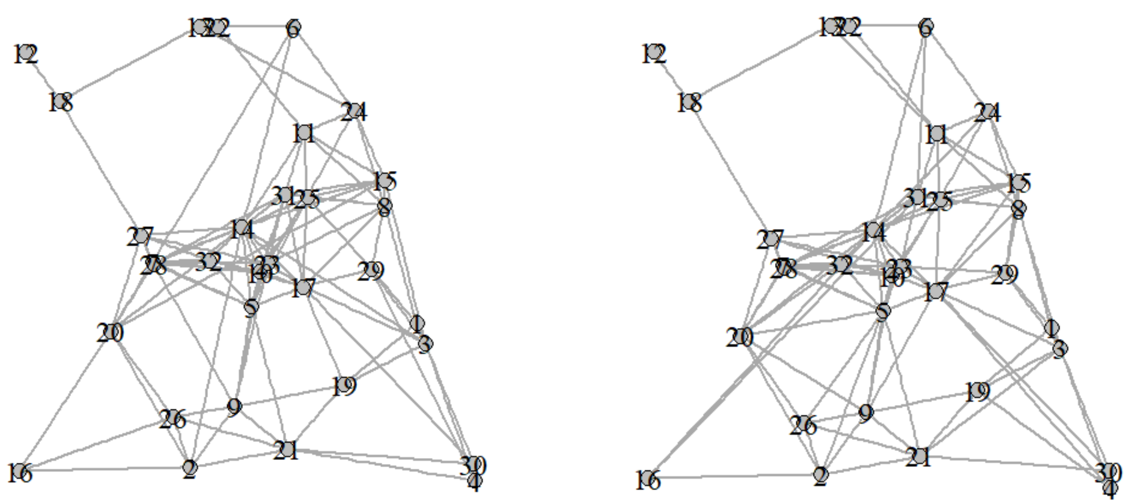


Figure 4.12: Two different flattened network created from the Poisson network.

Clustered Network

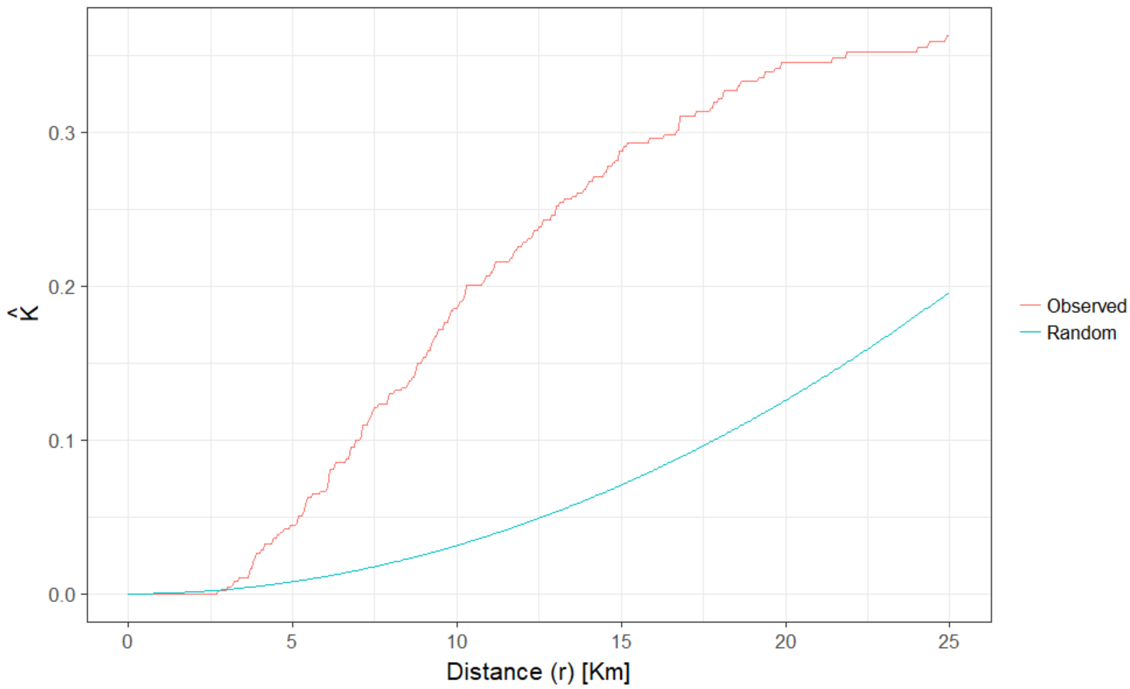


Figure 4.13: Multi-distance spatial cluster analysis showing clustered distribution of nodes in Euclidean space. The blue line shows the expected number of points for a perfect Poisson distribution and the red line shows the observed number of points in the dataset

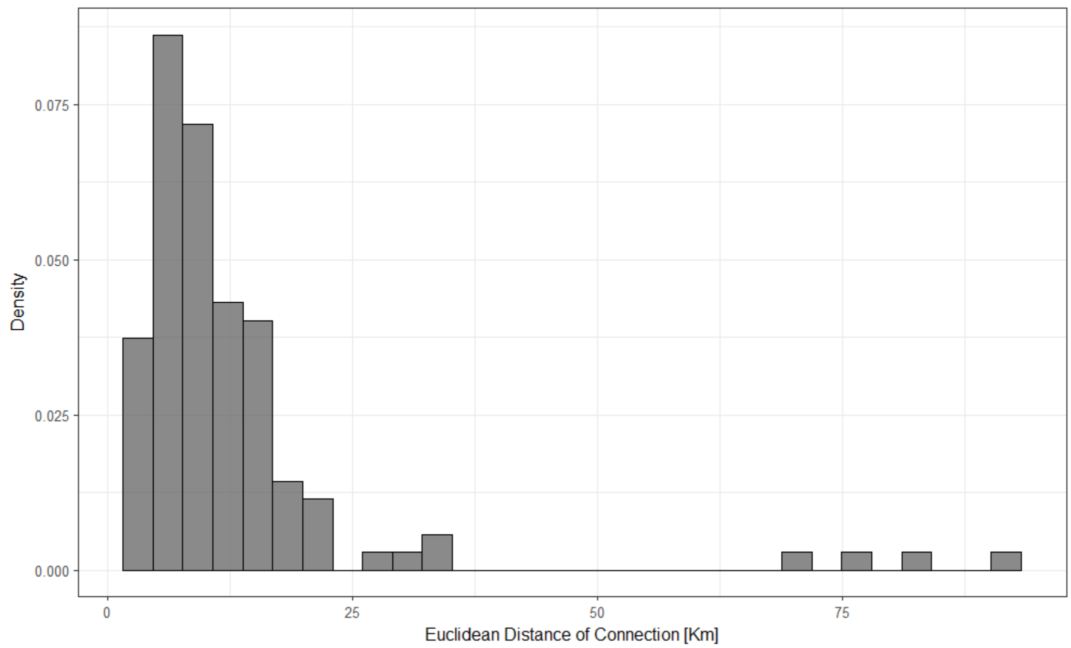


Figure 4.14: Histogram of connection distances showing exponentially decreasing probability of connection with distance.

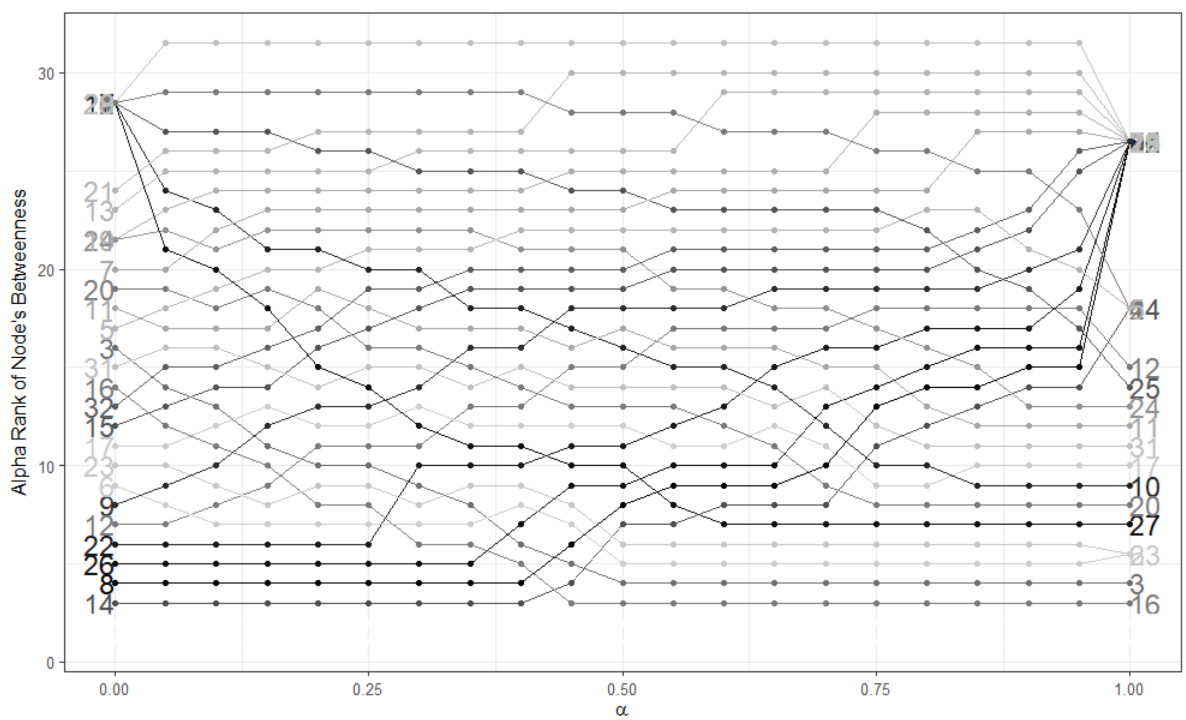


Figure 4.15: Betweenness Ranking of the nodes at different α values for the Clustered Network. Nodes which are tied for a rank are given the average rank.

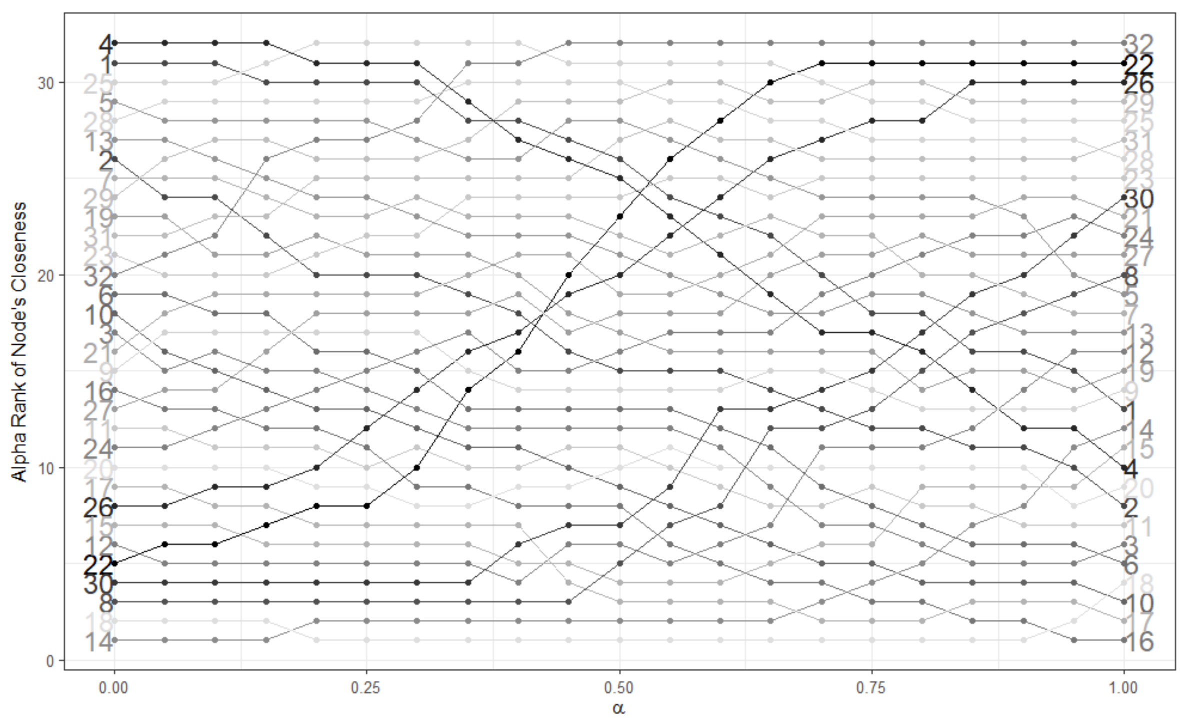


Figure 4.16: Closeness Ranking of the nodes at different α values for the Clustered Network. Nodes which are tied for a rank are given the average rank.

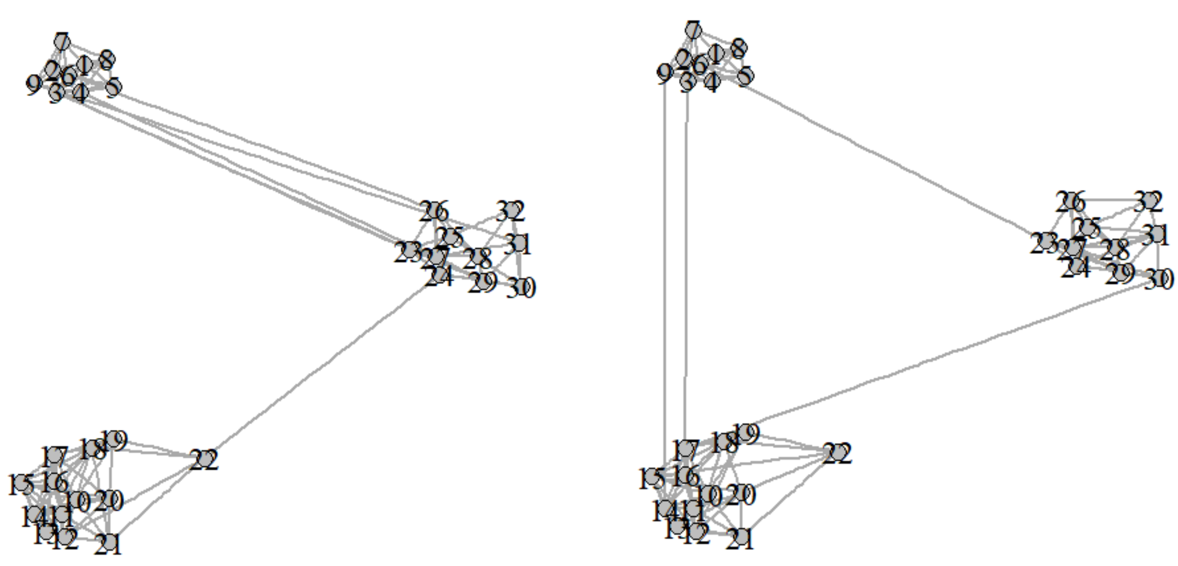


Figure 4.17: Two different flattened network created from the Clustered network.

4.7.2. Kibale Employment Network

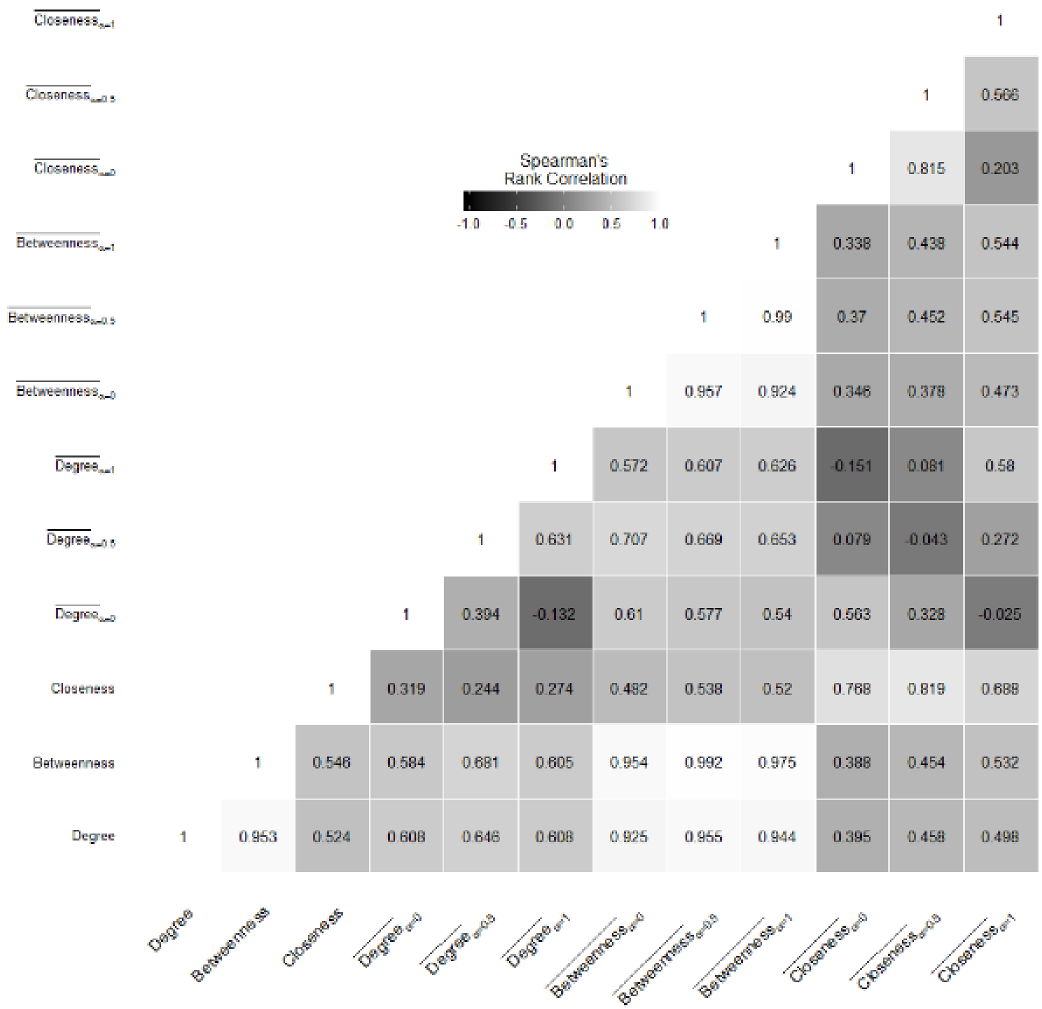


Figure 4.18: Spearman's Rank Correlation between the different metrics for the Kibale network. The dashes over the name of the centrality metrics denote the metrics modified with SS Tuning Parameter at α=0, α=0.5, and α=1 respectively.

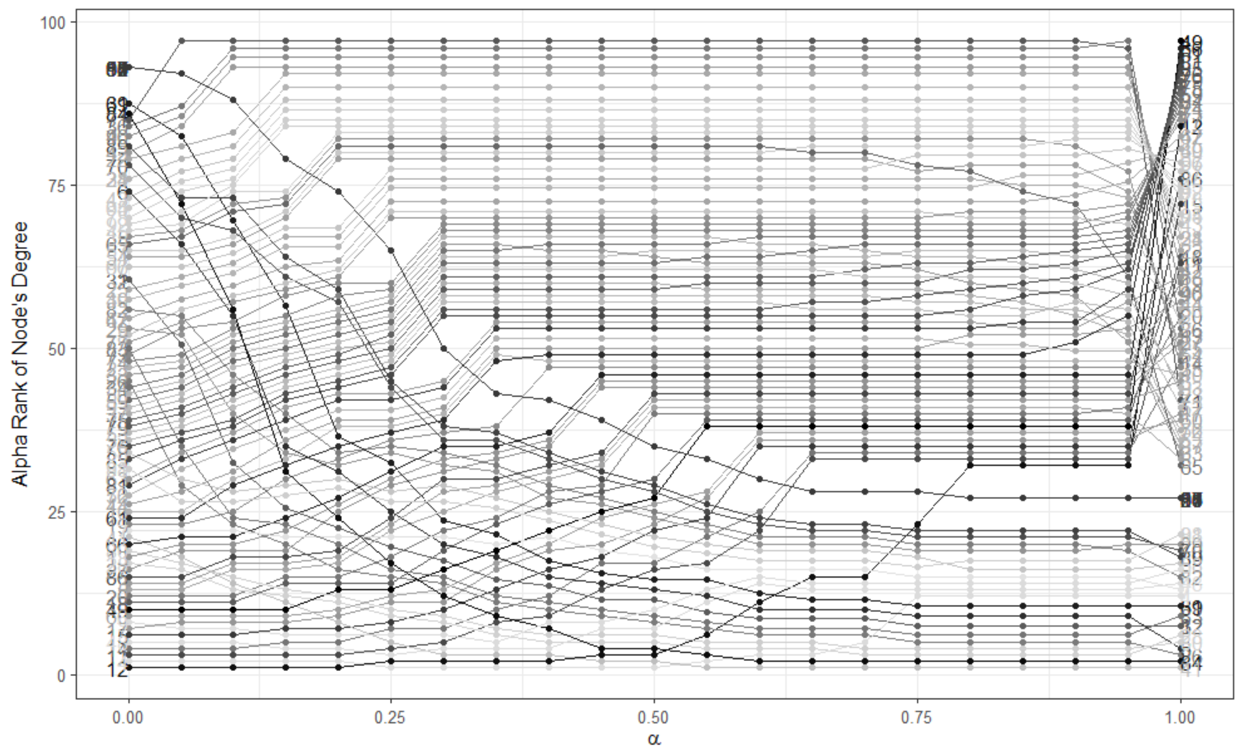


Figure 4.19: Degree Ranking of the nodes at different α values for the Kibale Network. Nodes which are tied for a rank are given the average rank.

Figure 4.15, the degree values of most nodes tend to be very low at $\alpha=0$ as most nodes in the network have mostly connections located in the same village. Thus, having $D_s \approx 0$ when $\alpha=0$, apart from the first few ranks, the rest are of limited use. Thus nodes 12 and 11 are considered the most important as they only have far connections. On the other hand, node 4 is also considered important as their hirings are mostly from nearby villages but not from their own village. Thus, by virtue of having comparatively intermediate distant connections, Node 4 becomes the number 1 ranked node at $\alpha=0.5$ before losing its ranking. When some importace is given to near connections by setting α to be close to 0.5, the ranks change significantly, as the large number of very near connections most nodes have start to add up to significant values. After $\alpha=0.5$, the top ranks remain relatively stable as there is a high correlation between nodes that have large number of connections (high degree) and nodes that have a good mixture of near and far connections. This is not surprising, as people who hire more people often have large farm lands in addition to having other endeavours like cattle and brick making, thus, requiring more people with varies skill sets. Hence they tend to have connections at all distances. Unlike the top ranked nodes, the other nodes keep changing their importance as more weight is afforded to near connections and slight variations to the combination of number of near connections and how far the connections are keep the rankings fluctuating. After $\alpha=0.9$, when only near connections are given importance, significant reshuffling occurs amongst the lower ranking nodes as the ranks become more correlated to the number of connections they have.

5. Research Stations as Conservation Instruments Provide Long Term Community Benefits Through Social Connections

Citation: Sarkar, D., Chapman, C.A., Valenta K., Angom S.C., Kagoro W., and Sengupta, R., xxxx. Research Stations as Conservation Instruments Provide Long Term Community Benefits Through Social Connections. Accepted subject to minor revisions at *The Professional Geographer.*

Linking statement: In this chapter, we highlight how quantitative and qualitative analysis compliments social spatial network analysis to provide a holistic picture of the system under study by overcoming the bias of networks to over-privilege “connections” above other aspects of a multi-faceted system (Chapter 2A.3.1). In this case, the system under study is the research field station located at Kibale National Park, Uganda and the socio-economic benefits it provides to the community around the park. This study relies on the same snowball sampled employer-employee network used in Chapter 4 to understand the percolation of benefits through the communities. However, this chapter uses quantitative and qualitative analysis to understand how research stations shape community-park relationships by providing not only economic opportunities (as modelled in Chapter 4), but also additional support through education and health care initiatives. This exemplifies the need to go beyond the sole use of social network analysis, which reduces a system to a collection of nodes and edges to capture the dynamics mediating park-community relationships, of which economic benefits and its flow through the community is one aspect. Thus, a combination of Chapter 4 and Chapter 5 provide a holistic understanding of the system under study, which aims to understand the benefits of a research field station, as these benefits spread out across the landscape (of which the social connections and importance of certain actors are an important, but a singular aspect of the picture).

Ethics Clearance: The research conducted as part of this chapter adheres to the ethics guidelines REB II as set up by the Tri-Council Policy Statement, TCPS 2 (2014) and has been reviewed and cleared by the McGill University Ethics Review Board (File No: 251-1215).

Summary: Conservation plans have evolved beyond biodiversity protection to include the welfare of the communities surrounding protected areas. Local community engagement initiatives include development of ecotourism, revenue sharing arrangements, and resource access agreements. Though research stations are common in African National Parks, their contribution to biodiversity protection and community benefits have seldom featured in the literature. In this article, we consider whether community benefits accruing from field research stations are effective and how they may promote community-park relationship. We employ a mixed-methods approach to understand the impacts on the local community of a field station located in Kibale National Park, Uganda. We find that the presence of a research station in Kibale National Park provides long-term direct employment for 52 people, and indirect, cascading benefits for up to 720 people several kilometers away. Additionally, the research field station is associated with other important community benefits, primarily healthcare and education. While benefits of the research station do not eliminate community-park conflict, the long-term presence of researchers and the gains to local people associated with them is an underappreciated and important means to better integrate the goals of biodiversity protection and local community investment. Most notably, the health and education related benefits reported by the participants can primarily be attributed to the initiatives taken by researchers.

5.1. Introduction

The rhetoric on management plans for Protected Areas (PAs) has focused extensively on the costs and benefits of delineating large areas for biodiversity protection, while accounting for the welfare of local communities. While PAs have generally been effective at biodiversity protection (Struhsaker *et al.* 2005, Laurance *et al.* 2012), they have often been accused of exacerbating poverty by disenfranchising neighboring communities (Ninan, 2006; Nyhus et al., 2005). Over the years, expectations of conservation plans have broadened to encompass the welfare of people living around the PAs (Robinson 1993, Daniels and Bassett 2002, Miteva *et al.* 2012, Martin *et al.* 2013). These community welfare initiatives have taken various forms including ecotourism, revenue sharing, resource access agreements, health care, and education (Ferraro *et al.* 2011, Child 2013, Chapman *et al.* 2015). Focused and sustained efforts to improve the livelihood of nearby communities have generally assuaged the reputation of PAs as poverty traps (Ferraro *et al.* 2011, Naughton-Treves *et al.* 2011, Mackenzie 2012); however, the best methods for managing PAs in light of local community needs remain contentious (Miteva *et al.* 2012).

Persistent threats to Earth's biodiversity intensifies the urgency to set aside land for protection. Since 1992, PAs have grown steadily, and as of 2006, they covered 24 million km^2, in 133,000 designated areas (Butchart *et al.* 2010, Rands *et al.* 2010). This surge in the size and number of PAs raises the concomitant challenge of ensuring that PAs do not disenfranchise local people. Creation of new PAs to maximise conservation without doing so at the expense of the socioeconomic well-being of adjacent communities requires treading a fine balance between the often conflicting requirements of biodiversity conservation, human rights, and development goals (Robinson 1993, Zimmerer 1994, 2006, Wilshusen *et al.* 2002, Brockington *et al.* 2006). Most biodiversity hotspots are located in the world's poorest countries (Cincotta *et al.* 2000, Fisher and Christopher 2007) where attempts to conserve land by excluding local people can impact their tenuous livelihoods (Guha 1989, DeFries *et al.* 2005). In addition to opportunity costs in the form of reduced access to natural resources, there are also issues such as increased crop raiding by park-protected animals (Naughton-Treves 1998, Mackenzie and Ahabyona 2012), eviction (Brockington *et al.* 2006, Karanth 2007), and threats to personal safety from park-protected animals (Packer *et al.* 2005, Inskip *et al.* 2013). In some cases the development of ecotourism has been effective in off-setting these costs, developing a vibrant ecotourism industry is not always a viable option and funds from tourists often do not diffuse to local communities, thus proving inadequate at providing substantial economic incentives for the local communities to protect biodiversity (Krüger 2005, Child 2013). For example, through gorilla ecotourism in Bwindi Impenetrable National Park, the local community received approximately $400,000 annually via revenue sharing agreements, yet the high population density around this park renders benefits to households negligible (Sandbrook and Adams 2012). Whether this small amount of money can alter the perception of the local communities is questionable (Karanth *et al.* 2012). Community-based conservation efforts developed partly in reaction to state-run exclusionary conservation (Wilshusen *et al.* 2002), have achieved mixed success in linking biological conservation objectives with local development endeavours (Campbell and Vainio-

Mattila 2003) and have managed to provide only modest supplements to local livelihoods (Kiss 2004). Thus, many PAs are left with limited opportunities to fulfill the mandate of ensuring community welfare alongside biodiversity protection.

We present data on a rarely discussed means of enhancing community welfare, while protecting biodiversity – the promotion of long-term research stations. Like eco-tourism, a research field station provides opportunities for economic gains. But unlike eco-tourism sites, research stations are not constraint to PAs with attractions such as large mammals. Moreover, researchers often spend significant amount of time in the PAs with research stations and return for several seasons, thus building a relationship with the place and the people. A large scale study to find correlates of conservation success revealed that 37.5 percent of PAs in Africa have research stations (Struhsaker *et al.* 2005) and that PAs with a research station were better able to evaluate success and management of PAs, even when research activities usually covered only 2–3 percent of the PA's area. Here, we focus on Makerere University Biological Field Station (MUBFS), a long-term research station in Kibale National Park, Uganda. Our research is the first to attempt to isolate the role of the long-term research field station (MUBFS) on the community. In this paper, we have used snowball interviews to understand how the MUBFS impacts livelihoods and consequently how the community perceives MUBFS. This study provides unique insights into how research stations shape community-park relationships by providing not only economic opportunities, but also additional support through education and health care. Some of the multi-faceted benefits from the field station are a result of initiatives started by long-term researchers and compliment the goals and initiatives of both the research station and conservation plans in general.

5.2. Study Area

Located in Western Uganda, Kibale National Park (Kibale) is a 795 km^2 mid-altitude moist evergreen forest (Figure 5.1). At the start of the 19th century the area was a large forest inhabited by approximately 40 households of agriculturists (Naughton-Treves 1998, 1999). In 1932, Kibale became officially recognized as a forest reserve with the goal of providing sustained hardwood timber production (Struhsaker 1999, Chapman and Lambert 2000). In the 1920s the Uganda

Game Department was tasked with confining wild animals to parks, to enable growth of agricultural land without hindrance of wild animals (Karanth et al. 2013; Naughton-Treves 1999a; Naughton-Treves 1998). The focus of Kibale shifted to biodiversity conservation in 1993, when Kibale's status was changed from forest reserve to national park under the stewardship of the Uganda Wildlife Authority (UWA). UWA manages PAs in Uganda using the 'Park and Neighbour' strategy (Jones 2006), where conservation research, community education and outreach, resource access agreements, and revenue sharing are vital components of management (Mugisha and Jacobson 2004, Mackenzie *et al.* 2015).

Research intensified in the area around 1970 and became localized in the Kanyawara area with the work of Dr. Thomas Struhsaker from the Wildlife Conservation Society (WCS). In 1987, the operations of the growing field station were handed over to Makerere University and the site was named Makerere University Biological Field Station (MUBFS). The field station infrastructure grew substantially in the early 1990s through international funding from WCS, European Union (EU), and The United States Agency for International Development (USAID)[3]. Currently, MUBFS consists of three research sites, Kanyawara, Ngogo and a newly developed site at Sebatoli, but the major site remains Kanyawara (Figure 1). Kibale's history of human impacts in terms of commercial logging/agricultural clearing and its location in an area conducive to working with the local community made the site a focal point for conservation research. Today, MUBFS is considered by some to be Africa's leading tropical forest research, conservation, and training site (Callahan 1997, Box *et al.* 2008).

[3] Objectives of MUBFS: http://caes.mak.ac.ug/mubfs/about-us.html

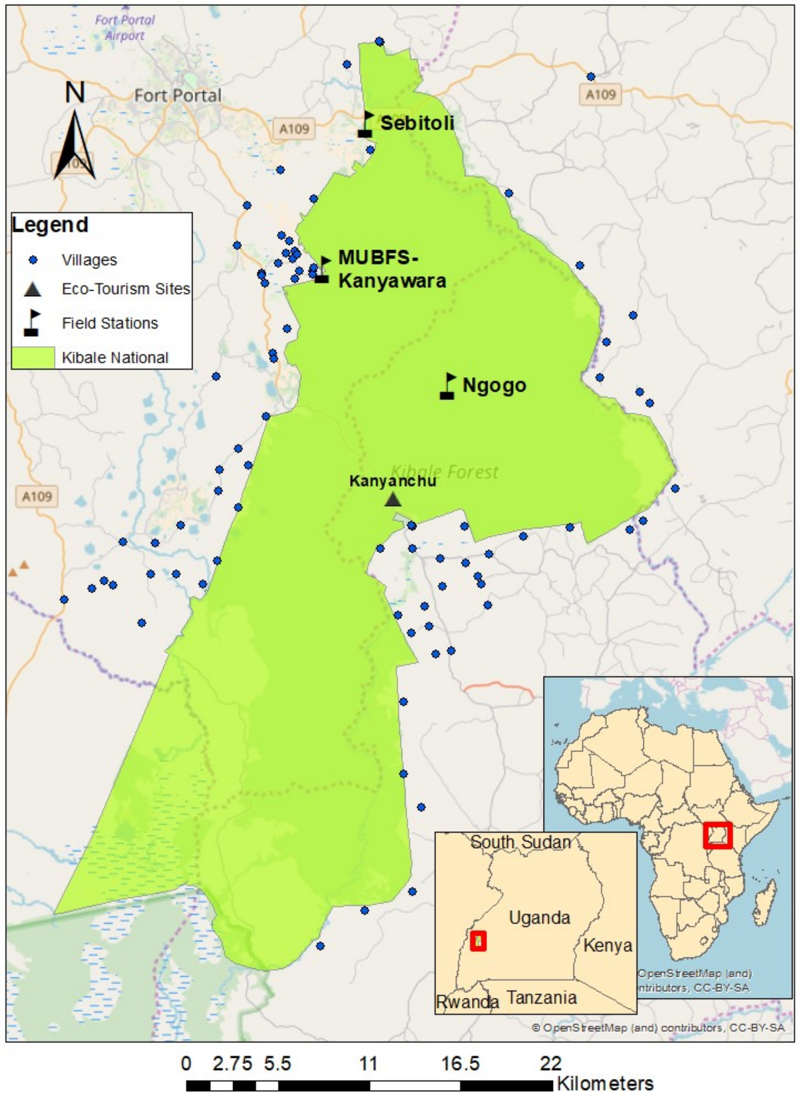


Figure 5.1: Map showing Kibale National Park, Uganda, villages around the park, the research field sites and the eco-tourism site at Kanyanchu

We focus on Kanyawara, which is the primary site housing administrative offices, lodging, classrooms, library, a mess hall and kitchen, laboratory space, and medical facilities (Figure 5.2). Kanyawara is located approximately 320 kms from Entebbe International Airport and is accessible via the Fort Portal - Kampala Road. Kanyawara currently employs 52 people of which 78.8 percent are from nearby villages. The number of people reported as employed by MUBFS are the ones in their payroll as administrative and support staff (primarily house keeper and trail cutters). People are also engaged on a temporary basis in capacities of cooks and housekeepers for researchers and field courses. In addition to the jobs offered directly by MUBFS, researchers hire Field Assistants (FAs) and long-term projects often hire several FAs with yearly renewable contracts. FAs tend to be from nearby villages, typically living within 6 Km of the field station.

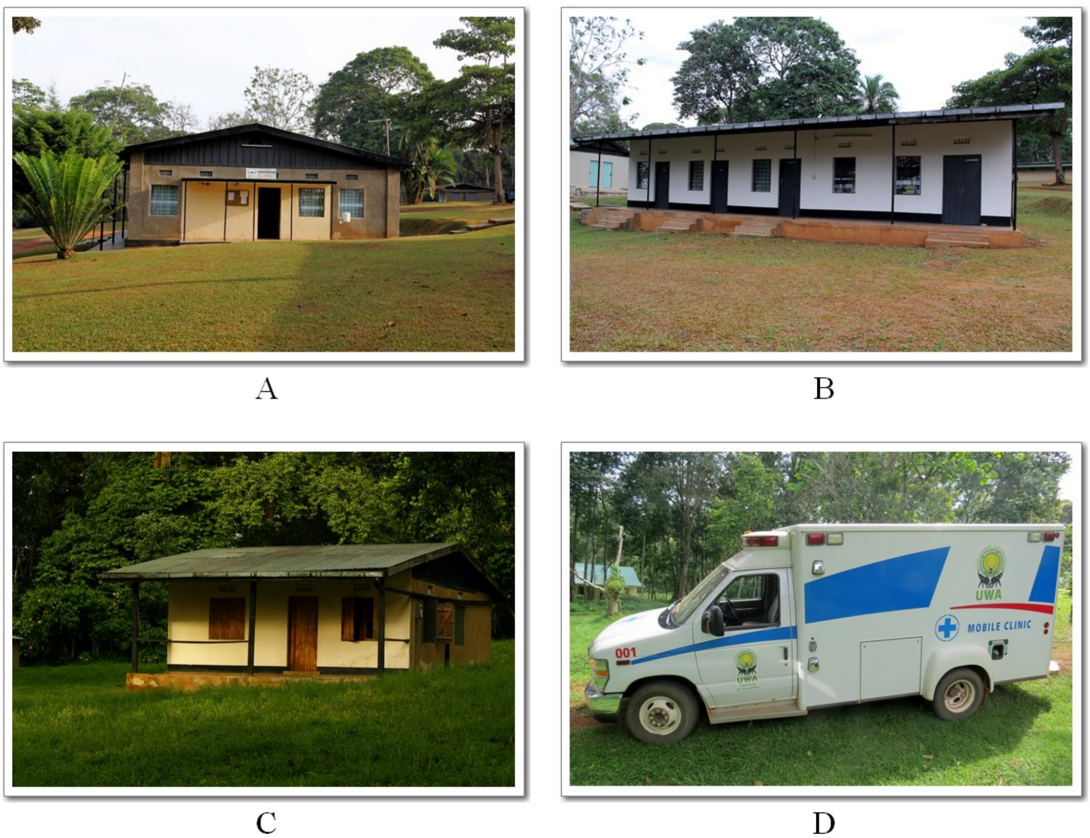


Figure 5.2: Pictures of MUBFS. (A) The main administration building, (B) Library and some classroom facilities, (C) Sample housing available on site, and (D) The Kibale Mobile Clinic.

Several outreach and development projects have been started near the field station by researchers working in Kibale (Table 5.1). The Kasiisi Project was set up by researchers from Harvard University in 1997 to enhance local education by improving school infrastructure. Even though Uganda has implemented free universal primary education (Oketch and Rolleston 2007), public schools are often overcrowded, understaffed, and lack basic infrastructure (Deininger 2003). In addition to lacking amenities such as school uniforms, school supplies, and books, children are faced with the trade-off of going to school versus helping with agricultural activities including guarding crops against the park's animals (Ross 2013, Mackenzie *et al.* 2015, MacKenzie, Moffatt, *et al.* 2017). This project has assisted over 10,000 children in 14 schools (Ross 2013) and has carried out several outreach activities (Project 2016).

In 2007, researchers and students from McGill University established the Kibale Health and Conservation Centre (KHCC) to provide medical services to the community (Bunting 2008, Chapman *et al.* 2015). The primary goal of the centre was to provide free consultation and at-cost medication to the local community. The center is located at one of the park's entrances, and employs a full-time nurse and a medical advisor whose responsibilities include outreach activities for disease prevention. From September 2008-September 2012, the centre provided care to 7,200 people and its outreach programme extended to 4,500 schoolchildren each year (Chapman *et al.* 2015). Since 2012, the addition of a Mobile Health Clinic (MHC) (Figure 2D) has expanded the reach of some of the health services provided by KHCC as well as conservation education outreach to areas surrounding the whole of the national park. The MHC caters to the plight of remote villages for basic health care, family planning, HIV/AIDS treatment, counselling, and vaccinations.

Table 5.1: Descriptive information about the staff and conservation initiatives in MUBFS (Struhsaker 1999, updated in this study)

Founded in	1970	
Founded by	Dr. Thomas Struhsaker	
Maintained by	Makerere University since 1987	
Number of employees (excluding Field Assistants)	Men: 39	Women: 13
	Total: 52	
Number of employees from nearby villages (Excluding Field Assistants)	Total: 41	
Conservation, health and education related	Kibale Health and Conservation Centre (Health)	

projects conducted using MUBFS as its original base and initiated by researchers at the field station,	Mobile Health Clinic (Health)
	The Kasiisi Project (Education)
	Kibale Snare Removal Program (KSRP) (Conservation)
	Herbarium (Education)
	Kibale Fuel Wood Project (Conservation)
Long-term research projects at MUBFS	Kibale Chimpanzee Project
	Kibale Monkey Project
	Kibale Fish Project
	Kibale EcoHealth Project
	Primate Ecology and Nutrition Project

5.3. Data and Analysis Methods

In this study, we use a mixed method approach comprised of household surveys and focus groups to capture local community perceptions of the field station and to understand the extent to which socio-economic life is affected by the research station. The survey primarily comprised of open-ended questions inviting descriptive answers and a few multiple-choice questions. It was administered to 70 employees of the field station representing residents of 13 surrounding villages and the field station between January 2016 and May 2017. To evaluate how the benefits from the station percolated through the community, each employee and FA hired by researchers were asked to list people they hired for various household and farming related activities, who were in turn interviewed and again asked who they employed. Thus, a snowball sampling approach was used which started with administering the surveys to people working at MUBFS and then consequently expanding it to people who are hired by them for household and farm work and so on and so forth. Since one aim of our study is to understand the community's perception of the benefits of a field station, the snowball sample allows us a closer look into how the perception of the people change as they are further removed from directly economically benefiting from MUBFS. To control for the self-selecting bias of the snow-ball sampling method, in addition to interviewing people who directly or indirectly benefit from the field station, the survey was also administered to 27 randomly selected respondents who were not a part of the snowball chain. These 27 respondents will be referred to as the Control Respondents (CR). The respondents from the field station and FAs are referred to as Tier 1 responders and each consequent referral is denoted as Tier 2, Tier 3, and so forth. The surveys were administered between January 2016-May 2017 by a Ugandan field assistant to minimize the bias associated with the respondents interacting directly with researchers. As many chains (connected Tiers) as possible were followed (i.e. until an individual on the chain did not hire a person to help them). The primary limitation to following some chains to completion was the unavailability or unwillingness of an individual to participate in the survey and if an individual lived more than 10 kms from the field station.

Most of the survey questions were open-ended responses to elicit as many themes as possible. Specifically, the respondents were asked to list what they perceive as positive and negative impacts of the research station. The descriptive survey questions along with the large number of survey respondents meant that a considerable amount of data was gathered as free form text. This text was analysed using Latent Dirichlet Allocation (LDA) (Blei *et al.* 2012) fitted with Gibbs sampling to automatically elicit topics latent in the discussions. LDA was chosen as the topic modelling algorithm as it is a unsupervised classification technique and thus could be run relatively easily and quickly in the field. The optimum number of topics latent in the corpus was determined by running the algorithm repeatedly to discover three to ten topics. The result of the algorithm where it was instructed to extract 5 topics was determined to be optimum, producing the most logical topic to term allocation. The extracted topics were further clarified and discussed by conducting six focus groups. Each focus group was comprised of three to four employees of the field station. The participants were given the set of topics discovered thus far and were specifically asked if there were topics which were more topics that may still be missing. After about 150 interviews, which included four tiers of participants, the focus groups agreed that persistent points had been covered and new information was unlikely to emerge. However, a few more surveys were administered to participants in Tier 5, 6, 7, and to the CRs to double check the completeness of information received.

Upon returning from the field, qualitative analysis was performed where by the interviews were manually coded and themes elicited. The extracted themes from the coding closely represented the ones found through LDA. To extrapolate the results of the survey and to get an understanding of how many secondary jobs were created through the employments at the field station, Geographically Weighted Regression (GWR) was used with the number of people from each village employed at the field station and the distance of the village from the field station as explanatory variables. The predicted output from GWR was used to create a surface depicting the secondary job potential using Inverse Distance Weighting (IDW).

5.4. Results

The snowball sampled respondents consisted of 209 people from 21 villages (including people who lived at the field station itself) within 10 km by road from the field station. Of these, 148 were men and 61 were women. There was a male bias in Tier 1, with 90 percent of the people interviewed being male, while only 61 percent of the people in subsequent tiers were male. Out of the 70 people interviewed in Tier 1, 40 were FAs of researchers and the remaining 30 people were administrators, cooks, and trail cutters. The people in the remaining Tiers (Tier 2 = 118 people, 3=8, 4=5, 5=5, 6=2, 7=1) primarily subsist with small scale agriculture supplemented by payments from the employees of the field station who hire them. These individuals were employed to assist with agriculture and cattle grazing (63 percent), helping around the household (13 percent), construction (12 percent), and in brickmaking (12 percent). There were 27 CRs from 9 villages consisting of 12 men and 15 women, 66 percent of whom were farmers. Thus, in total (snowball respondents and CRs) there were 236 respondents from 21 villages including the field station.

Overall demographics of the interviews is presented in Table 5.2. It is worth noting the varying number of respondents in each tier impacts the demographic results. Tier 1 respondents overall demonstrated better indicators of wealth by having larger number of livestock, and greater proportion of households had eucalyptus and cash crop plantations compared to all other Tiers and CRs. Tier 1 respondents also reported higher education levels. Across the various education and wealth indicators, the CRs are comparable to Tier 2 respondents.

Table 5.2: Demographic information obtained through the survey.

	Tier 1	Tier 2	Tier 3	Tier 4	Tier 5	Tier 6	Tier 7	CR
Number of respondents	70	118	8	5	5	2	1	27
Average household size	4.97	4.34	3.25	5.4	3.2	6	3	4.85
Percentage of adults who have completed primary but not secondary education	13.97	9.25	15.62	0	0	0	0	9.92
Percentage of adults who have completed secondary education	21.87	4.15	9.37	5	0	0	0	4.58
Average number of livestock (Cow, Goat, Pig, Chicken) per household	8.66	5	5.87	6	5	9	9	5
% of households with Eucalyptus plantations	55.71	29.66	12.5	0	0	0	0	29.63
% of households with Cash Crop (Tea, Coffee, Sugarcane) plantations	28.57	9.32	12.5	0	20	0	0	14.81

The topics revealed by LDA provided a snapshot into the themes that were later confirmed by interview coding. The following sub-sections discuss the themes in context of four over-arching motifs: economic benefits, crop raiding, resource access, and community-park relationship. These themes are consistent with the broad topics that have been discussed in the literature focusing on the community around KNP. While the studies focus primarily on the impacts of conservation plans on the community, the aim of our study was to understand how the research field station has affected the communities and consequently, the responses received provides insight on how MUBFS has contributed to the previously discussed themes.

5.4.1. Economic benefits

The most apparent community benefit of the field station was employment. The field station typically employs 52 people (Table 5.1), and 88.6 percent of Tier 1 respondents subsequently hired people to work in their household or farms. Most Tier 1 respondents (62.2 percent) stated that they would/could not have hired additional labour if they did not have a job at the field station. These findings explain why the most commonly cited benefit of the field station was employment, which was also recognized by 21.5 percent of the people not employed at the field station (including CRs).

CRs also iterated that employment in MUBFS is mostly available to the educated, *"Some inhabitants have been employed in the park but only those who are educated."* (Respondent CR15), and recognised the trickle-down effects of the employment benefits. Tiered Respondents also acknowledge that the economic benefits from MUBFS flow primarily through social connections (friends and family), *"...others get jobs from friends who are working in the park"* (Respondent 4.2).

Each person hired created on average 2.3 (S.E.=0.11) additional job opportunities for their community members (Figure 5.3). The majority (93 percent) of the hiring was from the villages < 3 km from the field station, but some employees traveled up to 8 km along dirt roads to get to work (Figure 5.4). In contrast, Tier 2 was mostly local. The mean distance between Tier 2 employees and their employers was 0.9 km. A Geographically Weighted Regression (GWR) with number of Tier 2 respondents at each location as the dependent variable of number of Tier 1 employees and the distance of the location from the field station was used to extrapolate the Tier 2 employment potential around the park ($R^2 = 0.617$) (Figure 5.5). Since, Tier 2 jobs tend to be local, they are concentrated around the villages with most people employed by the field station; however, as Tier 1 jobs were correlated with distance from the field station, the number of Tier 2 jobs also reduced with distance from the field station. Extrapolating the results of the survey to all employees living in nearby villages including the FAs, the field station helps support approximately 158 secondary employment opportunities in villages located within 5 km. The CRs on the other hand employ 1.7 (S.E.=0.29) additional people on average.

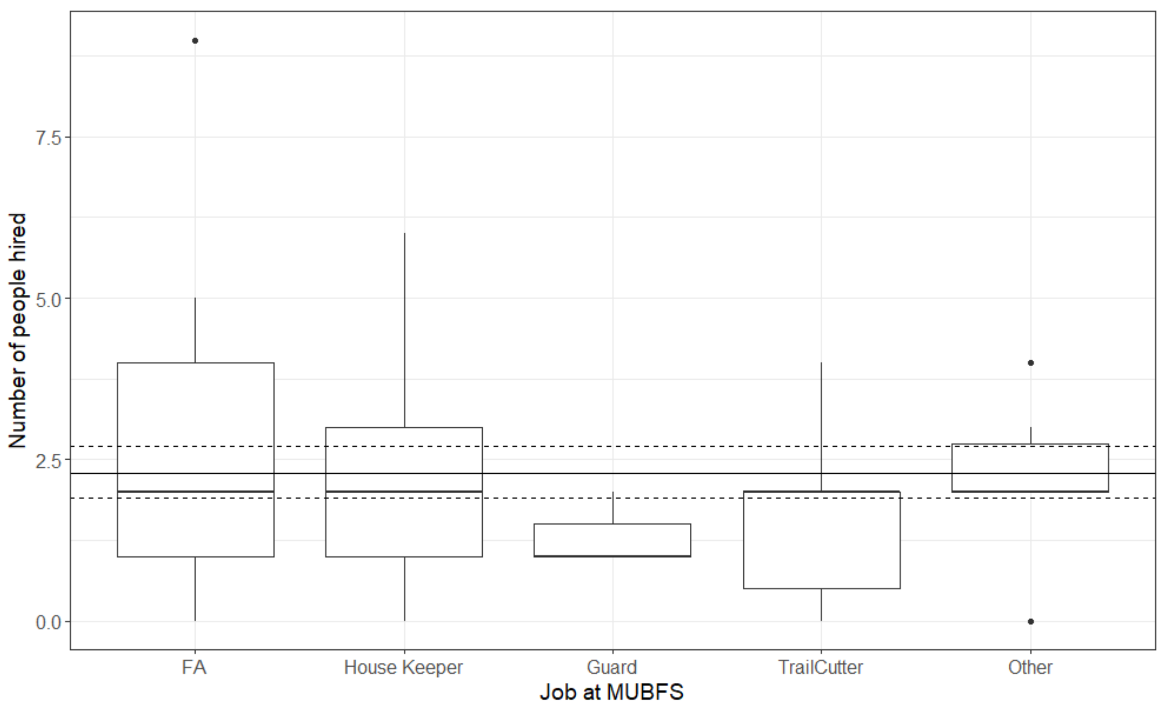


Figure 5.3: Number of people hired by employees of Makerere University Biological Field Station in Kibale National Park, Uganda according to their position at the field station. The thick solid line shows the sample mean along with thick dashed 95 percent confidence interval. Also displayed, as typical in whisker box plots is the median, the first and third quartile, and the maximum and minimum values.

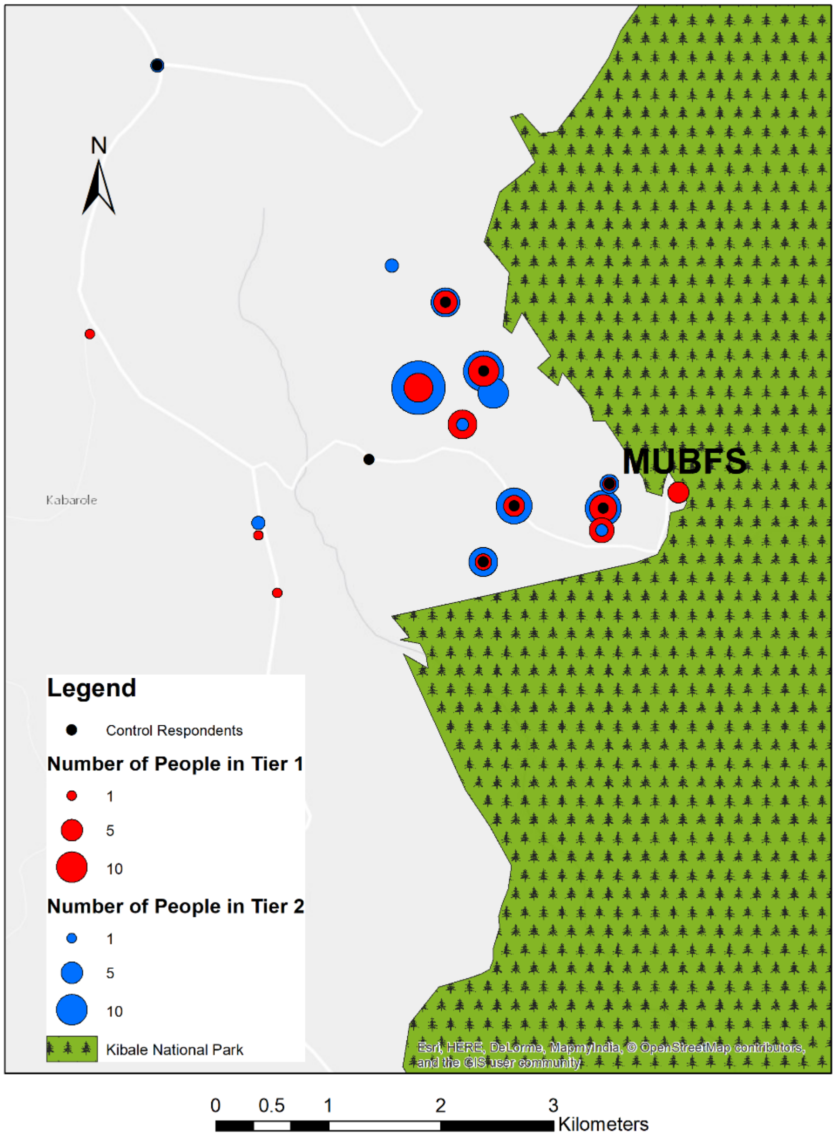


Figure 5.4: Map showing location of Tier 1, Tier 2 and Control Respondents of the survey conducted adjacent to the Makerere University Biological Field Station in Kibale National Park, Uganda.

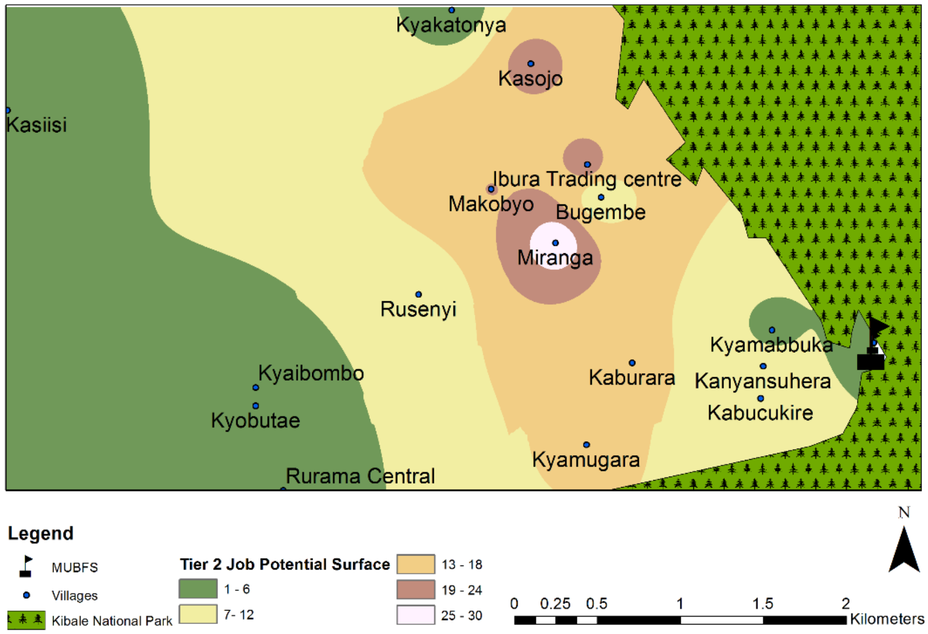


Figure 5.5: Inverse Distance Weighted surface showing predicted number of Tier 2 jobs within 5 km of Kibale National Park, Uganda.

Apart from employment opportunities, the field station also created economic opportunities akin to those created by ecotourism, such as sales of farm produce and crafts to researchers and students. Another recurrent economic benefit that emerged from the surveys was gifts received from researchers. When surveys are conducted, respondents stated that researchers reimbursed their time with small gifts, like soap, sugar, or mosquito nets. Moreover, unlike eco-tourism sites, researchers spend long periods in the field station, often coming back for several years. This often fosters congenial relationships between employees and researchers. Several people employed as FAs have identified benefits like help with their children's education from educational projects and direct payment from researchers. Indeed, 61 percent of people employed at the field station reported receiving advances or sponsorship from researchers for various reasons, most commonly children's education. Despite efforts such as the Kasiisi Project to make education accessible, the general cost of continuing education remains high and making getting sponsorship for children's education was one of the most cited benefits of working for the field station, as exemplified by the following comment of a Tier 2 respondent: *"Because one is working in the park, he has sponsors helping his child [for education]"* (Respondent 2.42).

5.4.2. Crop Raiding

Researchers have reiterated the miseries of the community due to crop-raiding animals, particularly baboons and elephants (Naughton-Treves 1998, Mackenzie and Ahabyona 2012). The same concerns were raised during our interviews and crop-raiding was reported as the most common negative effect of living next to a PA (63.98 percent including CRs). Locals consider wildlife as state property and believe that UWA should be accountable for animal's actions. However, the community recognises the efforts of the researchers to provide insights into how to tackle the crop-raiding problem. For example, the revenue sharing model employed by the park has been used to excavate and maintain trenches (MacKenzie 2012), and 11.02 percent of the respondents (including CRs) recognized trench excavation and maintenance as an useful expenditure. The plight of the community and the constant attention from researchers have made UWA more responsive to the night-time calls from villages seeking help to drive away elephants, and researchers and UWA have developed a protocol for villagers to report elephant raiding and gather data for understanding spatio-temporal patterns of raids (Sarkar, Chapman, *et al.* 2016).

5.4.3. Resource Access

Resource Access Agreements (RAAs) have been persistent points of contention in community-park relationships. RAA permits included keeping beehives in the park, collecting craft materials, and fishing. Previously the agreement allowed firewood and/or Non-Timber Forest Products (NTFPs) collection, but these activities were discontinued due to exhaustion of resources or due to non-compliance with regulations (Mackenzie *et al.* 2011). Nonetheless, the local dependence on firewood for cooking means that illegal fuelwood extraction remains a common practise. In the interviews, lack of legal access to firewood was a frequently voiced grievance (26.27. percent including CRs), and the presence of international researchers in the park was held responsible for the heightened vigilance of the community and UWA regarding illegal resource extraction. Efforts by the researchers (through the Kibale Fuel Wood Project) have helped raise awareness of planting eucalyptus for firewood and a handful of people said that they grow their own trees for firewood (34.74 percent including CRs, Tier 1 = 54.28 percent, 2 = 29.66 percent, 3 = 12.5 percent, 4, 5, 6 and 7 = 0 percent, CR = 29.63 percent) while a similar proportion (36.01 percent including CRs) of respondents were depended on friends' and neighbours' plantations or collected firewood from the park. Access to medicinal plants was another similar strain of complaints regarding restrictions on resource access. The blanket ban on extraction of NTFPs have also restricted the community from accessing the medicinal herbs and plants found in the park that are used in traditional medicine. The plight of the community resonates with the people employed at MUBFS (Tier 1 respondents) and when asked for suggestions for improving community park relationships, two Tier 1 respondents suggested holding more workshops to raise awareness on alternate fuels for cooking.

5.4.4. Community Park Relationship

The park and its animals are often viewed by community members as state property and of little use to locals. People's reaction to the park is often mixed with the positive perceptions attained from employment countered by the negative perceptions of being excluded from resources and the damage caused by crop raiding animals (Hartter and Goldman 2009). Researchers working on livelihood issues around Kibale have often worked as a liaison between UWA and the community, helping to change the perception of the park. In fact, one interviewee stated that, *"Before it [perception of the park] was bad. Rules were hard to follow, now things are better. Researchers help to solve problem. ... fees collected to help building trenches"* (Respondent 1.2).

Research field stations are uniquely positioned to enhance community engagement and conservation. Unlike eco-tourism sites, the people working in research sites are more invested in conservation by virtue of being involved in a set-up focused on enquiries on conservation. People employed in the park re-iterated that outreach workshops conducted by the field station and UWA were beneficial in spreading awareness about "the importance of a forest", "conservation and its use", and "how the park operates" and one employee stated, *"Community understand now the good in the park because of Mzungos (foreigners) coming from very far to make their studies [research] in the park"* (Respondent 1.46).

Overall, when asked about the community's perception of the park, the respondents in the snowball sample (Tiers) had a mixed response (29.66 percent positive, 4.66 percent neutral, 8.9 percent negative, 56.36 percent can't say), with the CRs were more agnostic with a high percentage of neutral responses (29.63 percent positive, 37.03 percent neutral, 22.22 percent negative, 11.11 percent can't say). Unsurprisingly, Tier 1 respondents were most positive (40 percent) about the community's perception of the park with the positive attitude petering out with Tiers (Tier 2= 27.11 percent, 3= 35 percent, 4, 5, 6 and 7= 0 percent). However, even the CRs (29.6 percent) noted that presence of researchers has helped raise awareness about the benefits of conservation while 37.03 percent mentioned the convenience of the KHC and the various education opportunities available.

5.5. Discussion and Conclusion

The community around KNP has been well studied to understand the impacts of conservation plans on people living around PAs (Naughton-Treves 1997, 1998, Lepp and Holland 2006, Hartter and Goldman 2009, Hartter and Southworth 2009, Naughton-Treves *et al.* 2011, Mackenzie and Ahabyona 2012, Mackenzie and Hartter 2013, MacKenzie, Moffatt, *et al.* 2017). Our research is the first to attempt to isolate the role of the long-term research field station (MUBFS) on the community. A large number of interviews were thus required to ensure that we captured the different perceptions about MUBFS in the community, along with understanding how MUBFS impacts livelihoods. The major themes elicited were consistent with the topics discussed in literature, but we provided new insights pertaining to the relationship between MUBFS and the community. We integrate previous findings with our own to identify themes that came out of our data that warrant further scholarship, in the context of the main findings of our research, in particular, the trickle-down effects of research station employment, economic status by tier, and how perceptions of the park vary in light of the research station.

Previous research around Kibale had highlighted that perceptions of the park vary in the community (Hartter and Goldman 2009, MacKenzie, Salerno, *et al.* 2017). The negative impacts of living next to a protected area were somewhat mitigated by policies for revenue sharing, resource access, and opportunities for employment and tourism. Our research highlights the fact that several of the positive measures can at least partially be attributed to the presence of a long-term research field station. The research station not only provides employment opportunities, but has also catalyzed the setting up of several community welfare projects in the area as a result of researchers conducting long-term research in the area and building connections to the place. Thus, MUBFS has not only acted as a platform for scientists worldwide to conduct their research in the Kibale area, but has also evolved over the years to provide services to the community, mainly through the efforts of researchers conducting long-term research in the area.

Our results indicate that the benefits from the field station spread out to the local community. The 52 people hired directly resulted in the hiring of 2.3 times that number, or 120 people. If these 120 people belonged to different households, an estimated 720 people obtain financial benefits from the station (average household contains six people (Mackenzie and Hartter 2013). These people purchased goods and services, spreading the benefits even further. Respondents indicated that the people providing these goods and services recognize they are benefitting from funds from the field station. The snowballed interviews allowed us to understand the dissemination of this benefit. The results revealed that although the economic benefits spread along social connections, people spatially and socially closer to MUBFS also have access to other benefits such as education subsidies and secondary employment opportunities. Thus, the perception of the field station and the park in general in concomitant with the association with MUBFS, thus Tier 2 and higher along with CRs had a more tempered reaction towards the park. In addition, the Tier 1 respondents as a dual effect of living close to MUBFS and being employed there not only view the park positively, but also show better indicators of wealth. The CRs showed wealth and education indicators that was comparable to the people hired by the Tier 1 respondents (Tier 2).

This spread of benefits appears to be substantial, but to guide planning, evaluating relative effectiveness of different approaches must be made with respect to at least four considerations. **First**, community benefits must be sufficient to alter the perception of the local community and thereby affect behaviour. If the density of the population neighboring a field station is low, even small amounts of financial gain from the project can translate to large individual gains. In Africa, the density of people outside parks is often high, and on average only 4 percent of the communities derive financial benefits from adjacent parks (Struhsaker *et al.*

2005). In Kibale, the population near the park's boundary is fairly high at around 300 people/km^2 (Hartter 2009). Thus, the impact of fund distribution to the 720 people considered to benefit from the field station is relatively small and needs careful evaluation.

Second, administrators should compare the benefits of the research station to other conservation schemes. In general, comparisons of different schemes are rare and difficult to make, stemming partly from the fact that the schemes operate at different temporal scales (e.g., cash benefits are immediate, but benefits of education may require decades to have an impact). Like the salary benefits of the field station, ecotourism provides financial benefits to people who are directly employed, or sell goods and services. Tourism in Kibale attracts approximately 7700 visitors each year, creating part-time employment for approximately 250 people and generating $271,000 in revenue (Mackenzie 2012). Given the large population around Kibale, tourism provides direct income to only 0.5 percent of the population. This income spread equally among community members bordering the park would translate into $1.08 per year *per capita*. Whether this amount is sufficient to alter community perceptions towards Kibale remains untested.

Third, different approaches to promoting conservation operate at different spatio-temporal scales. In our opinion, one important benefit of the field station is that it encouraged the development of other conservation programs. This was achieved as the field station provided a base for long-term research and promoted interaction with the local community and UWA. Several conservation programs arose from the field station, each operating on different spatial and temporal scales, with some operating only near the field station (e.g., Kibale Health Clinic) and others operating park-wide (e.g., Kasiisi Project, Kibale Mobile Clinic), some potentially influencing park-people relationships rapidly (e.g., health care), and others having an effect from the immediate to very long term (e.g. education).

Fourth, the employment opportunities arising from a field station have the potential of reinforcing the 'rich get richer' phenomenon as jobs in the field station are mostly offered to educated people. The secondary benefits also flow along social connections (relatives and friends) and thus create hubs of concentrated wealth. To mitigate the problem, careful steps to ensure that a wide spectrum of jobs are available should be taken, and secondary education initiatives are launched as part of the community outreach aspect of conservation plans. Otherwise, only the educated will get the benefits (including loans from the researchers for children's education) who in-turn can afford to educate their children, afford healthcare, and get a better quality of life, further marginalising the poor. At the field station, initiatives such as the Kasiisi Project and Kibale Health Clinic (KHC) have been instrumental in dispersing secondary benefits of the research station to the larger community. Thus, evaluating the success of conservation projects also needs to encompass associated secondary projects that affect the larger community.

Finally, the goal of conservation schemes is to protect biodiversity. Thus, planners should assess biodiversity and community perceptions before and after their implementation. There is evidence that the field station is protecting biodiversity by deterring poaching with a consequent there have been population increases in several species (Chapman *et al.* 2017). Researchers have initiated programs for patrolling and snare removal in addition to programs that improve people-park relations through the provision of health services and education.

Research Field stations have the potential to fill important conservation roles and can strike a balance between the duality of biodiversity protection and community enhancement. These stations help further scientific understanding, protect biodiversity, and can facilitate community welfare, involvement and development, even in the absence of ecotourism attractions. The multi-faceted benefits from the field station indicate that development agencies should consider investing in research stations, and governments should consider facilitating the establishment of research stations in collaboration with universities as part of conservation and community outreach mandates.

5.6. Supplementary Information

Sample survey administered to respondents from MUBFS

PART 1 – EMPLOYMENT DETAILS

1) Please list your position within the Field Station, Health Clinic, or assisting researchers with duration of employment.

Position	Duration Employed	Employer

2) Do you usually get a bonus at the end of a researcher's stay?

3) Has anyone in your family or extended family been sponsored through researchers? If yes, how many?

4) Do you ever get loans from researchers for school fees?

5) Do you ever get loans for other expenses? If yes, for what?

6) Do you have family working at the field station, clinic or with researchers?

Name	Gender	Relationship	Position	Duration Employed	Contact Details

PART 2 – PERSONAL DEMOGRAPHICS

7) Where do you live?

□ Village: ___________________ □ Parish: ______________________

8) Please list the people in your household, including gender, age, occupations and level of education. *Include yourself.*

Name	Position Within Family	Gender	Age	Job(s)	Level of Education	Contact Info
First and Last	□ Grandparent (GP) □ Parent (P) □ Child (C) □ Grandchild (GC) □ Other Relative (OR) □ House-Helper (HH)	□ Male (m) □ Female (f)	#	Please Specify:	□ None □ Some Primary (SP) □ Completed Primary (CP) □ Some Secondary (SS) □ Completed Secondary (CS) □ Vocational Training (V)	□ Phone Number □ Address

Name	□ Dependents (D)				□ College, University or Higher (CUH)	

9) How many of each type of livestock belong to your household?

	Cows	Goats	Chickens	Pigs	Other:
How many:					

10) What type of gardens do you have? *Check all that apply*

□ Tea □ Fruit □ Vegetables

□ Sugarcane □ Eucalyptus Trees

□ Other:

____

11) What produce do you sell? *Check all that apply*

□ Tea □ Fruit □ Vegetables

□ Sugarcane □ Eucalyptus Firewood

□ Bricks □ Other:

____

12) How often do your gardens get raided by elephants?

Almost Daily Weekly Every 2-4 weeks Monthly Every 2+ Months

13) How often do you gardens get raided by baboons?

Almost Daily Weekly Every 2-4 weeks Monthly Every 2+ Months

14) How do you protect your gardens against animal raids?

____

____

15) If a person acts as a guard, who are they?

Name	Gender	Age	Relationship	Contact Details

16) Do you own, rent, or borrow the land you are using? *Check all that apply*

□ Rent □ Borrow □ Own

17) How did you acquire the land you own? *Check all that apply*

□ Inherited □ Bought

18) What are some of the materials your home is constructed with? *Check all that apply*

□ Mud □ Animal Dung □ Elephant grass

□ Iron Sheets □ Cement □ Brick

□ Grass □ Timbers □ Other: ___________________________

19) Which of the following does your household have? *Check all that apply:*

□ Bicycle □ Radio □ Television

□ Mobile Telephone □ Motorcycle □ Car

20) What is your main method of communicating with people you know?

□ Cell phone –calling □ Cell phone – texting □ E-mail

□ Word of Mouth □ Other: ______________________________

PART 3 – ADDITIONAL LABOUR

21) Do you ever hire additional labor to help you tend to your gardens or property?

Name of	Relationship	Hired Job	Frequency of	Contact Details

Worker			Hire	
First and Last Name	□ Child (C) □ Grandchild (GC) □ Other Relative (OR) □ Friend (F) □ Other:_______	□ Household □ Tea □ Firewood □ Cattle □ Vegetables □ Fruit □ Brick □ Other:_______	□ Every week □ Every 2-4 weeks □ Every month □ Every 2-3 months □ Every 3-6 months □ Every 6+ mths	□ Phone number □ Address

22) Did/Would you hire additional labor if you were not employed at the park?

____

PART 4 – HEALTH PROFILE

23) How often do you use the local Kibale Clinic for you or your family?

Never Almost Never Sometimes Almost Always Always

24) How often do you use the Mobile Health Clinic for you or your family?

Never Almost Never Sometimes Almost Always Always

PART 5 – CULTURE, CONSERVATION & KIBALE

25) What animals are endangered (at risk of extinction) in the park?

____

____

26) Has the park changed your leadership role in the community? If so, how?

____

____

27) Have you run for or held any electoral position since working for the park?

____

____

28) Has the influx of mzungos into the park changed your community beliefs or traditions in any way? If yes, what changed?

____

____

____

29) How does your community view the park?

____

30) Could you please tell us what are some of the benefits this household receives from living close to the national park? (Write answers in table below)

Benefits of living next to the national park

31) Could you please tell us what are some of the negative effects this household receives from living close to the national park? (Write answers in table below)

Negative effects of living next to the national park

32) Do you have any ideas as to how to improve community-park relations?

33) Do you have any ideas as to how to improve conservation efforts in the park?

6. Conclusion and Future Research Plans

This dissertation has built upon the momentum of integrating spatial information in social sciences in the form of Spatial Social Networks (SSNs). As mentioned through the course of this dissertation, SSNs provide certain advantages of explicitly modelling social relationships and situating the social system in a spatial context providing avenues for understanding how space and society are mutually shaped. The four chapters comprising this dissertation make theoretical and methodological contributions to the field of GIScience and SSNs. Starting from providing theoretical avenues for integrating social network and spatial analysis, I have demonstrated the various ways in which social network can incorporate spatial information to highlight different spatial relationships embedded in a dataset, introduced metrics that built upon theoretical avenues not yet covered by SSN researchers, and situated how qualitative and quantitative analysis techniques can be used in tandem with SSN analysis to understand complex socio-spatial phenomenon.

In Chapter 1, I broadly introduce the background concepts explored in the dissertation. I state that relationships between entities are central in geography, and the development of SNA provides a unique way of modelling these relationships. Partly due to the increasing amount of spatial information available to researchers, spatial information has been increasingly integrated into various fields of social sciences. Consequently, SNA has also been used by several studies in ‘spatially integrated social sciences’ to model the interactions and relationships between the entities of the system. The chapter concludes by highlighting some of the shortcomings of SNA and SSNs that have motivated the research covered in this dissertation.

In Chapter 2, I provide a literature review pertaining to social networks, and discuss how social networks can be integrated with GIScience. I provide a typology enumerating the ways in which spatial information has been incorporated into SNA. Namely, there are three levels of sophistication of integrating spatial information into social networks; nodal, topographic, and as a property of a network. While the simplest form involves attaching location information to the entities, and is sufficient for inferring distance-friendship based relationships, more sophisticated forms are required to consider the embedding of the network in geographic space. In this chapter I also identify three potential avenues for integrating SNA in geography. The concepts of distance, scale, and community resonate in both fields and offers avenues for defining and interpreting SSNs. I also discuss (in Chapter 2A), literature related to spatial social networks, paying special attention to networks as models capable of representing various relationships amongst a given set of entities, and the various ways in which the application of geographic attributes to the nodes and edges within a network analysis can alter the results of models. I also provide an overview of the metrics commonly used in SNA, and explain how these have been adapted for SSNs. Further, this chapter also highlights how SNA has been applied to various problems within the social science literature with a spatial component.

Chapter 3 builds upon the concepts (introduced in Chapter 2A.1), of creating different networks from the same dataset by altering what the nodes and edges represent. Specifically, this chapter addresses the question regarding the various ways in which spatial information can be incorporated in the network structure. In the network paradigm where it is possible to represent a multitude of relationships among the same set of entities using different networks, it is important to contemplate the implications of creating these different network based representations from the same dataset, namely, what information can be gleaned from the network, and, how to handle the spatial components of the network. Specifically, I show that spatial information within networks can be incorporated as nodal attributes, edge attributes, or both. I do this using a grant dataset obtained from the National Geographic Society (NGS) spanning 126 years that includes spatial information in various fields, such as the location of grantees, grantee co-applicants (collaborators), and field work sites, in addition to non-spatial information, such as grantee discipline. I conclude that the different spatially explicit network-based representations that are created from the same dataset highlight different spatial relationships and provides insights into regional and international research trends. These spatial patterns of fieldwork based research provide new perspectives on one of the new standards of academic excellence, popularly known as “internationalization” of research, by going beyond the traditional ways in which this is measured, e.g. publication, citation, and demographics.

Chapter 4 introduces new metrics and visualization tools that leverage spatial information as well as social information as captured by SSN. This chapter specifically addresses the research objective of creating new methods for understanding socio-spatial properties of SNA. These new techniques provide socio-spatial information about the entire network structure and are used to identify important entities within a socio-spatial context. These metrics are first validated by applying them to simulated networks with known spatio-temporal properties, and then to a real-world dataset that I collected in Kibale National Park (KNP), Uganda. The KNP dataset was collected to model the flow of economic benefits originating from the research field station (MUBFS), through the local community. Economic benefits to local communities have been identified as crucial to the success of protected areas (Fiallo and Jacobson 1995, McClanahan *et al.* 2005, Lepp 2007, Mackenzie 2012), and these new metrics and visualization provide a better understanding of the structure of the network by identifying key individuals who are responsible for the percolation of benefits to the community at multiple spatial scales. The results show that most of the economic connections are local, as expected due to the nature of secondary jobs. However, some of the hiring is from further villages, highlighting that some skills may not be locally available. The combined effect of the employments at different distances help spread the economic benefits originating from MUBFS to as far as 5 kms from the source. However, in order to identify the important individuals responsible for spreading the benefits through the community, it is important to balance the role played by the individuals in providing local employment with the role of the individual who help disseminate the economic benefits over a larger area. This chapter addresses the dearth of methods pertaining to entity-level and network-level metrics and shows that distance between entities plays a crucial role in understanding the structure of a SSN and in determining which entities are important at different spatial scales.

Chapter 5 builds upon both the work of Chapter 4, and its dataset, and addresses a key problem with SNA that I identify and discuss in Chapter 2A and as part of the research objectives. In particular, the network data model used in SNA can be overly simplistic because the model relies on nodes and connections alone, and thus over-privileges interactions above other factors of multifaceted social systems. To address this shortcoming, I use the same dataset I used to build the SSN analysed in Chapter 4, but in this chapter, I utilize additional quantitative and qualitative methods to further understand how the presence of a field research station mediates community-park relationships. Even in this study, economic benefits come out as one of the primary benefits; however, this is not identified as the sole benefit of the field research station by community members. This illustrates that reliance on nodes and connections alone in modelling may overlook some important factors of a SSN. In this case, because the network constructed in Chapter 4 was focused on highlighting the economic benefits of the field station, the model de facto overlooked non-economic benefits, including important researcher contributions to the health and education of local communities.

6.1. Temporal SNA

In this dissertation, the focus was on providing methodological avenues for the integration of spatial information to social networks. In addition to incorporating spatial considerations, the other dimension that is of considerable interest is the dimension of time. In GISc, temporal topology and data models have generated considerable interest (Peuquet 1994, Peuquet and Duan 1995, Kwan 2002, 2004, Yuan et al. 2014). However, time-series analysis, change detection, and modeling evolution of individual entities is particularly difficult in social networks as the focus of social networks is on the topology, that is, the connection between the nodes and edges. In other words, in SNA, the focus is not on individual actors and their attributes but on the fabric of relationship between that exist in the system (Hanneman and Riddle 2005, chap. 1). Thus, when comparing networks over time, structural properties of the networks are more useful than reporting individual level statistics to model the evolution of each node (Barabási et al. 2002, Faust and Skvoretz 2008, Li and Yang 2009).

Given the same set of actors, statistical methods of comparing networks are well developed (Snijders 1996, Wasserman and Iacobucci 1998, Fowler and Christakis 2009). However, another set of challenges for comparing networks at different time points arises from the fact that often the set of nodes change over time. The challenge is thus with regards to detecting changes not only in the edges, but also with the need to account for the appearance and disappearance of actors (nodes) and the impact they have on the overall structure of the network. Given a significant timespan, this problem is akin to measuring the same relation on two or more different sets of actors. In order to compare the networks against each other, significant knowledge about the system under study including factors shaping the network is required beyond knowing just the topological network. Notable examples of such studies can be found in Resnick et al.'s (1997) work on longitudinal study on adolescent health, and in the series of works of on the networks of kinship and social and economic relations in 51 villages in Nang Rong District, Thailand (Rindfuss et al. 2004, Entwisle et al. 2007, Verdery et al. 2012). In recent years significant progress has been made to develop statistical techniques to compare networks at different time points that take into account stochastic nature of longitudinal social network data which can handle missing data with the help of probabilistic estimations (Faust and Skvoretz 2008, Krivitsky and Goodreau 2013, Dietrich 2017). Such new development offers potential to compare between different networks and provides interesting avenues for predicting link formation and network structure.

In this dissertation, comparing the network structure at various time points was a lucrative opportunity with the NGS database which covered a timespan of 126 years. In addition to the fact that the focus of the dissertation was to provide methodological avenues SSNs, the challenge of modelling the NGS network over time was several folds. First, the number of grants given by NGS over the years have significant variations. The large number of external socio-economic factors influenced how many grants were given, what type of research was supported, spatial variations of where the research was conducted, and who conducts the research. In order to make conclusions about the changes in network structure, significant research is required that can factor in the effect of the socio-economic developments over time. The second issue with modelling the network over time arises from the event-based nature of the socio-economic factors that have shaped the network. For example, it may be argued that increased affordability of commercial flights and the advent of the internet has had significant impact on where research is conducted and who the collaborators are. However, considering the large spatial and temporal aspect of the network, splicing the network at single time points to reflect such changes is difficult as the events did not occur simultaneously everywhere across the world. Finally, another challenge with this dataset is with regards to the geopolitical changes over time. While in social networks that model relationships between people, nodes may appear or disappear over time, in this network where countries are considered as entities of interest the challenge is exacerbated by the changing (shrinking, expanding, splitting, merging) political boundaries which require accounting for changes of location where researchers were based, and where research was conducted. Despite the challenges mentioned above, the dataset does provide unique opportunities to model evolution of different forms of collaboration as a result of field-based research. Consequently, handling such challenges provide new opportunities for theoretical research on time-based evolution of social networks, as well as for empirical research opportunities to understand the changing nature of scientific collaborations.

6.2. Future Research Plans

From the perspective of future research challenges, I would like to expand the spatial social networks to two specific application scenarios, namely, multi-modal communication networks and social-ecological networks.

In terms of multi-modal or multiplex networks, the challenges arise from the desire to model multiple parallel connections between entities which tend to be collapsed into single edges in SNA. Improvements in transport and telecommunication infrastructure have significantly altered our everyday interactions. While most researchers have studied social interaction as sustained by one of the methods, there is a lack of understanding how urban dwellers blend and supplement face-to-face interaction with virtual interactions. By using multi-modal SSNs that can capture urban social connections sustained through both physical transport infrastructure and virtual telecommunication interactions, I wish to expand the scope of my research to incorporate the social fabric of urban citizens. The multiple modes of interactions also engender questions regarding the agencies of real and virtual places in forming, sustaining, and dissolving social connections. In addition to the critical and theoretical contributions, this research can also be used to inform policies that drive urban planning and mass communication strategies in modern cities. For example, advancement in telecommunications has enabled people living outside the jurisdictional boundaries of cities to take part in several activities (e.g. jobs) contained in the city. Thus, the city as a 'place' extends beyond its boundaries and multi-modal networks provides a method to study the spill-over and plan telecommunication and transport infrastructure to prevent marginalisation. This study can be further enhanced by simultaneously modelling the representation of the city as a place which shifts the focus from interactions alone and hence may complement the network based findings.

The research on social-ecological networks is an extension of the work on social networks around Kibale National Park by including more components in the network. I wish to incorporate ecological components to understand the interdependence between the societies living around the park and the ecological services provided by the park. Beyond adding more components to the network, I plan to use concepts from complex systems (Bar-yam 1997, Chu *et al.* 2003) to understand the human-environment interactions as reflected in the network. Environmental systems have been modelled as complex systems (Nakamori and Sawaragi 2000, Müller and Li 2004) and networks of interaction between the components are generally regarded as the sources of complexity in complex systems (Bar-yam 1997, Hoffstein 2012). Hence, the use of complex systems as a lens to view and model these interactions can act as a bridge to better incorporate societal and environmental components in a single network. For example, the concept of scale is one of the overarching concept that recurs in both the complex system and geography literature. The congruent definitions of scale can be reconciled to couple the different parts of the network. In geography, scale represents a combination of three concepts, namely, spatial extent, spatial resolution, and scale of representation. Whereas scale in complex system refers to the existence of a large numbers of variables, structure of interconnected subsystems, and other features that introduce complications, such as non-linearity and uncertainty in behavior. The interdependence between scales of analysis is a critical component to understand interactions both in a spatial sense (global vs. local), as well as a system scale (micro vs. macro).

Primarily, I am interested in quantifying responses and resilience of socio-ecological systems to disturbances (anthropogenic and environmental). From the perspective of complex systems, the entire network will be resilient to shocks and remain at equilibrium, but a small disturbance in terms of complex system which only affects a small number of components in the network may not translate to only local implications from a spatial or social perspective. Although, social-ecological networks are an active area of research, the disparity in the spatial embeddedness of the various parts of the network has proven to be a barrier for modelling disturbances in the network. This research will provide a critical examination on how to model disparate components in the same spatial social network as well as explore the practical applications possible because of the incorporated spatiality.

7. References

Abbott, A., 1997. Of Time and Space: The Contemporary Relevance of the Chicago School. *Social Forces*, 75 (4), 1149–1182.

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*.

Abizaid, C., Coomes, O.T., Takasaki, Y., and Brisson, S., 2015. Social Network Analysis of Peasant Agriculture: Cooperative Labor as Gendered Relational Networks. *The Professional Geographer*, (March), 1–17.

Adams, J., 2013. The fourth age of research. *Nature*, 497 (7451), 557–560.

Adams, J., Faust, K., and Lovasi, G.S., 2012. Capturing context: Integrating spatial and social network analyses. *Social Networks*, 34 (1), 1–5.

Adams, J., Gurney, K., Hook, D., and Leydesdorff, L., 2014. International collaboration clusters in Africa. *Scientometrics*, 98 (1), 547–556.

Albert, R. and Barabasi, A.-L., 2001. Statistical mechanics of complex networks. *Reviews of modern physics*, 74 (1), 54.

Alexanderson, G.L., 2006. About the cover: Euler and Königsberg's Bridges: A historical view. *Bulletin of the American Mathematical Society*, 43 (04), 567–574.

Alvarez, A.J., Sanz-Rodríguez, C.E., and Cabrera, J.L., 2015. Weighting dissimilarities to detect communities in networks. *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences*, 373 (2056).

Andris, C., 2011. Metrics and Methods for Social Distance. MIT.

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

Ansari, A., Koenigsberg, O., and Stahl, F., 2011. Modeling Multiple Relationships in Social Networks. *Journal of Marketing Research*, 48 (4), 713–728.

Anselin, L., 1988. Spatial Econometrics: Methods and Models.

Anselin, L., 1999. The future of spatial analysis in the social sciences. *Geographic information sciences*, 5 (2), 67–76.

Atkinson, P.M. and Tate, N.J., 2000. Spatial Scale Problems and Geostatistical Solutions: A Review. *The Professional Geographer*, 52 (4), 607–623.

Bader, D., Meyerhenke, H., Sanders, P., and Wagner, D., eds., 2013. *Graph Partitioning and Graph Clustering*. Providence, Rhode Island: American Mathematical Society.

Bar-yam, Y., 1997. *Dynamics Of Complex Systems*. Westview Press.

Barabási, A.-L., 2002. *LINKED: The New Science of Networks*. 1st ed. Perseus Books Group.

Barabási, A.-L., 2007. Network Medicine — From Obesity to the "Diseasome". *New England Journal of Medicine*, 357 (4), 404–407.

Barabási, A.-L., 2015. *Network Science*.

Barabasi, A.-L. and Albert, R., 1999. Emergence of scaling in random networks. *Science*, 286 (5439), 509–512.

Barabási, A.-L., Albert, R., and Jeong, H., 1999. Internet: Diameter of the World-Wide Web. *Nature*, 401 (6749), 130–131.

Barabási, A.., Jeong, H., Néda, Z., Ravasz, E., Schubert, A., and Vicsek, T., 2002. Evolution of the social network of scientific collaborations. *Physica A: Statistical Mechanics and its Applications*, 311 (3–4), 590–614.

Barrat, A., Barthelemy, M., Pastor-Satorras, R., and Vespignani, A., 2004. The architecture of complex weighted networks. *Proceedings of the National Academy of Sciences*, 101 (11), 3747–3752.

Barthélemy, M., 2010. Spatial Networks. *Networks*, 86.

Barthélemy, M., 2011. Spatial networks. *Physics Reports*, 499 (1–3), 1–101.

Bastian, M., Heymann, S., and Jacomy, M., 2009. Gephi: An Open Source Software for Exploring and Manipulating Networks. *In*: *International AAAI Conference on Weblogs and*

Social Media.

Battista, G. Di, Eades, P., Tamassia, R., and Tollis, I.G., 1994. Algorithms for drawing graphs: an annotated bibliography. *Computational Geometry*, 4 (5), 235–282.

Batty, M., 2003. The Geography of Scientific Citation. *Environment and Planning A: Economy and Space*, 35 (5), 761–765.

Batty, M., Hudson-Smith, A., Neuhaus, F., and Gray, S., 2012. Geographic Analysis of Social Network Data. *In*: *Proceedings of the AGILE'2012 International Conference on Geographic Information Science,*. 24–27.

Bian, L. and Liebner, D., 2007. A Network Model for Dispersion of Communicable Diseases. *Transactions in GIS*, 11 (2), 155–173.

Binka, F., 2005. Editorial: North-South research collaborations: a move towards a true partnership? *Tropical Medicine and International Health*, 10 (3), 207–209.

Blei, D.M., Ng, A.Y., and Jordan, M.I., 2012. Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 3 (4–5), 993–1022.

Blondel, V., Deville, P., Morlot, F., Smoreda, Z., Dooren, P. Van, and Ziemlick, C., 2011. Voice on the border: Do cellphones redraw the maps? [online]. *Paris Innovation Reviews: Paris Science and Letters*. Available from: http://parisinnovationreview.com/2011/11/15/voice-border-cellphones-redraw-maps/ [Accessed 14 May 2017].

Blondel, V. V.D., Krings, G., and Thomas, I., 2010. Regions and borders of mobile telephony in Belgium and in the Brussels metropolitan zone. *Brussels Studies*, 32 (42), 0–12.

Blondel, V.D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E., 2008. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008 (10), P10008.

Boccaletti, S., Bianconi, G., Criado, R., del Genio, C.I., Gómez-Gardeñes, J., Romance, M., Sendiña-Nadal, I., Wang, Z., and Zanin, M., 2014. The structure and dynamics of multilayer networks. *Physics Reports*, 544 (1), 1–122.

Boessen, A., Hipp, J.R., Butts, C.T., Nagle, N.N., and Smith, E.J., 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*, 239980831769015.

Bonacich, P., 1987. Power and Centrality: A Family of Measures. *American Journal of Sociology*, 92 (5), 1170–1182.

Borgatti, S.P., 2005. Centrality and network flow. *Social Networks*, 27 (1), 55–71.

Borgatti, S.P. and Everett, M.G., 2006. A Graph-theoretic perspective on centrality. *Social Networks*, 28 (4), 466–484.

Borgatti, S.P., Everett, M.G., and Freeman, L.C., 2002. UCINET 6 For Windows: Software for Social Network Analysis.

Borgatti, S.P., Mehra, A., Brass, D.J., and Labianca, G., 2009a. Network analysis in the social sciences. *Science (New York, N.Y.)*, 323 (5916), 892–5.

Borgatti, S.P.S., Mehra, A., Brass, D.D.J., and Labianca, G., 2009b. Network analysis in the social sciences. *Science*, 323 (April), 892–5.

Börner, K., Chen, C., and Boyack, K.W., 2003. Visualizing knowledge domains. *Annual Review of Information Science and Technology*, 37 (1), 179–255.

Box, H., Butynski, T.M., Chapman, C.A., Lwanga, J.S., Oates, J.F., Olupot, W., Rudran, R., and Waser, P.M., 2008. Thomas T. Struhsaker: Recipient of the lifetime achievement award of the international primatological society 2006. *International Journal of Primatology*, 29 (1), 13–18.

Boyack, K.W., Klavans, R., and Börner, K., 2005. Mapping the backbone of science. *Scientometrics*, 64 (3), 351–374.

Brandes, U., 2001. A faster algorithm for betweenness centrality. *The Journal of Mathematical Sociology*, 25 (2), 163–177.

Brantingham, P.J., Tita, G.E., Short, M.B., and Reid, S.E., 2012. The ecology of gang territorial boundaries. *Criminology*, 50 (3), 851–885.

Brantingham, P.L., 2011. Computational Criminology. *In*: *2011 European Intelligence and*

Security Informatics Conference. IEEE, 3–3.

Brockington, D., Igoe, J., and Schmidt-Soltau, K., 2006. Conservation, Human Rights, and Poverty Reduction. *Conservation Biology*, 20 (1), 250–252.

Brockmann, D., Hufnagel, L., and Geisel, T., 2006. The scaling laws of human travel. *Nature*, 439 (7075), 462–465.

Bronfenbrenner, U., 1977. Toward an experimental ecology of human development. *American Psychologist*, 32 (7), 513–531.

Buch-Hansen, H., 2014. Social network analysis and critical realism. *Journal for the Theory of Social Behaviour*, 44 (3), 306–325.

Bughin, J. and Chu, M., 2010. The rise of the networked enterprise: Web 2.0 finds its payday. *McKinsey Global Institute*, December.

Bunting, M., 2008. Medical emergencies facing rural Uganda [online]. *The Guardian*. Available from: https://www.theguardian.com/katine/2008/sep/17/health.news.

Butchart, S.H.M., Walpole, M., Collen, B., van Strien, A., Scharlemann, J.P.W., Almond, R.E.A., Baillie, J.E.M., Bomhard, B., Brown, C., Bruno, J., Carpenter, K.E., Carr, G.M., Chanson, J., Chenery, A.M., Csirke, J., Davidson, N.C., Dentener, F., Foster, M., Galli, A., Galloway, J.N., Genovesi, P., Gregory, R.D., Hockings, M., Kapos, V., Lamarque, J.-F., Leverington, F., Loh, J., McGeoch, M.A., McRae, L., Minasyan, A., Morcillo, M.H., Oldfield, T.E.E., Pauly, D., Quader, S., Revenga, C., Sauer, J.R., Skolnik, B., Spear, D., Stanwell-Smith, D., Stuart, S.N., Symes, A., Tierney, M., Tyrrell, T.D., Vie, J.-C., and Watson, R., 2010. Global Biodiversity: Indicators of Recent Declines. *Science*, 328 (5982), 1164–1168.

Butts, C.T., Acton, R.M., Hipp, J.R., and Nagle, N.N., 2012. Geographical variability and network structure. *Social Networks*, 34 (1), 82–100.

Cairncross, F., 1997. *Death of Distance*. McGraw-Hill Ryerson Agency.

Calabrese, F., Dahlem, D., Gerber, A., Paul, D., Chen, X., Rowland, J., Rath, C., and Ratti, C., 2011. The connected states of America: Quantifying social radii of influence. *Proceedings -*

2011 IEEE International Conference on Privacy, Security, Risk and Trust and IEEE International Conference on Social Computing, PASSAT/SocialCom 2011, 223–230.

Callahan, J.T., 1997. Research Opportunites at the Makerere University Biological Field Station, Uganda. *Focus on Field Station: The Bulletin of the Ecological Society of America*, 78 (1), 80–84.

Campbell, L.M. and Vainio-Mattila, A., 2003. Participatory Development and Community-Based Conservation: Opportunities Missed for Lessons Learned? *Human Ecology*, 31 (3), 417–437.

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

Chapman, C.A., van Bavel, B., Boodman, C., Ghai, R.R., Gogarten, J.F., Hartter, J., Mechak, L.E., Omeja, P.A., Poonawala, S., Tuli, D., and Goldberg, T.L., 2015. Providing health care to improve community perceptions of protected areas. *Oryx*, 49 (04), 636–642.

Chapman, C.A., Bortolamiol, S., Matsuda, I., Omeja, P.A., Paim, F.P., Reyna-Hurtado, R., Sengupta, R., and Valenta, K., 2017. Primate population dynamics: variation in abundance over space and time. *Biodiversity and Conservation*.

Chapman, C.A. and Lambert, J.E., 2000. Habitat Alteration and the Conservation of African Primates : Case Study of Kibale National Park , Uganda, 185 (February 1999), 169–185.

Chen, H. and Hu, Y., 2013. Finding Community Structure and Evaluating Hub Road Section in Urban Traffic Network. *Procedia - Social and Behavioral Sciences*, 96, 1494–1501.

Cheshire, J.A., Longley, P.A., and Singleton, A.D., 2010. The Surname Regions of Great Britain. *Journal of Maps*, 6 (1), 401–409.

Child, B., 2013. *Parks in Transition:" Biodiversity, Rural Development and the Bottom Line*. Routledge.

Cho, E., Myers, S.A., and Leskovec, J., 2011. Friendship and mobility. *In*: *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '11*. New York, New York, USA: ACM Press, 1082.

Christakis, N.A. and Fowler, J.H., 2007. The Spread of Obesity in a Large Social Network over

32 Years. *New England Journal of Medicine*, 357 (4), 370–379.

Chu, D., Strand, R., and Fjelland, R., 2003. Theories of complexity. *Complexity*, 8 (3), 19–30.

Cincotta, R.P., Wisnewski, J., and Engelman, R., 2000. Human population in the biodiversity hotspots. *Nature*, 404 (6781), 990–992.

Clark, D.B., 1973. The Concept of Community: A Re-examination. *The Sociological Review*, 21 (3), 397–416.

Clark, J. and Holton, D.A., 1991. *A first look at graph theory*. World Scientific Publishing.

Coccia, M. and Wang, L., 2016. Evolution and convergence of the patterns of international scientific collaboration. *Proceedings of the National Academy of Sciences*, 113 (8), 2057–2061.

Colchester, J., 2016. Network Paradigm [online]. *Complexity Labs*. Available from: http://complexitylabs.io/network-paradigm/ [Accessed 31 Jan 2018].

Comber, A., Batty, M., and Brunsdon, C., 2012. Exploring the Geography of Communities in Social Networks. *In*: B. Rowlingson and D. Whyatt, eds. *Proceedings of the GIS Research UK 20th Annual Conference*. Lancaster, 33–37.

Coomes, O.T., McGuire, S.J., Garine, E., Caillon, S., McKey, D., Demeulenaere, E., Jarvis, D., Aistara, G., Barnaud, A., Clouvel, P., Emperaire, L., Louafi, S., Martin, P., Massol, F., Pautasso, M., Violon, C., and Wencélius, J., 2015. Farmer seed networks make a limited contribution to agriculture? Four common misconceptions. *Food Policy*, 56, 41–50.

Coscia, M., Giannotti, F., and Pedreschi, D., 2012. A Classification for Community Discovery Methods in Complex Networks.

Coscia, M., Rinzivillo, S., Giannotti, F., and Pedreschi, D., 2014. Spatial and temporal evaluation of network--based analysis of human mobility. *State of the Art Applications of Social Network Analysis*, 269–293.

Crampton, J.W., Graham, M., Poorthuis, A., Shelton, T., Stephens, M., Wilson, M.W., and Zook, M., 2013. Beyond the geotag: situating 'big data' and leveraging the potential of the geoweb. *Cartography and Geographic Information Science*, 40 (2), 130–139.

Crandall, D.J., Backstrom, L., Cosley, D., Suri, S., Huttenlocher, D., and Kleinberg, J., 2010. Inferring social ties from geographic coincidences. *Proceedings of the National Academy of Sciences of the United States of America*, 107 (52), 22436–41.

Cranshaw, J., Toch, E., Hong, J., Kittur, A., and Sadeh, N., 2010. Bridging the gap between physical location and online social networks. *In*: *Proceedings of the 12th ACM international conference on Ubiquitous computing - Ubicomp '10*. New York, New York, USA: ACM Press, 119.

Crossley, N., 2010. The Social World of the Network. *Sociologica*, 2010 (1).

Csardi, G. and Nepusz, T., 2006. The igraph software package for complex network research. *InterJournal*, Complex Sy, 1695.

Cummins, S., Curtis, S., Diez-Roux, A. V., and Macintyre, S., 2007. Understanding and representing 'place' in health research: A relational approach. *Social Science & Medicine*, 65 (9), 1825–1838.

Cummins, S., McKay, L., and Macintyre, S., 2005. McDonald's Restaurants and Neighborhood Deprivation in Scotland and England. *American Journal of Preventive Medicine*, 29 (4), 308–310.

Dalton, R., 1973. *Networks in geography : Modern techniques in geography*. G. Philip.

Daniel, M., Moore, S., and Kestens, Y., 2008. Framing the biosocial pathways underlying associations between place and cardiometabolic disease. *Health & Place*, 14 (2), 117–132.

Daniels, R. and Bassett, T.J., 2002. The Spaces of Conservation and Development around Lake Nakuru National Park, Kenya. *The Professional Geographer*, 54 (4), 481–490.

Daraganova, G., Pattison, P., Koskinen, J., Mitchell, B., Bill, A., Watts, M., and Baum, S., 2012. Networks and geography: Modelling community network structures as the outcome of both spatial and network processes. *Social Networks*, 34 (1), 6–17.

Day, P.L. and Pearce, J., 2011. Obesity-Promoting Food Environments and the Spatial Clustering of Food Outlets Around Schools. *American Journal of Preventive Medicine*, 40 (2), 113–121.

DeFries, R., Hansen, A., Newton, A.C., and Hansen, M.C., 2005. Increasing isolation of protected areas in tropical forests over the past twenty years. *Ecological Applications*, 15 (1), 19–26.

Deininger, K., 2003. Does cost of schooling affect enrollment by the poor? Universal primary education in Uganda. *Economics of Education Review*, 22 (3), 291–305.

Dietrich, E.A., 2017. Modeling Social Network Change over Time: A Comparison of Methods. University of Washington.

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

Doreian, P. and Conti, N., 2012. Social context, spatial structure and social network structure. *c*, 34 (1), 32–46.

Downey, L., 2006. Using Geographic Information Systems to Reconceptualize Spatial Relationships and Ecological Context. *American Journal of Sociology*, 112 (2), 567–612.

Dunbar, R.I.M. and Spoors, M., 1995. Social networks, support cliques, and kinship. *Human Nature*, 6 (3), 273–290.

Durland, M.M. and Fredericks, K.A., 2005. An introduction to social network analysis. *New Directions for Evaluation*, 2005 (107), 5–13.

Eagle, N., Pentland, A., and Lazer, D., 2009. Inferring friendship network structure by using mobile phone data. *Proceedings of the National Academy of Sciences*, 106 (36), 15274–15278.

Edwards, G., 2010. Methods Review paper Mixed-Method Approaches to Social Network Analysis. *ESRC national Centre for Research Methods*, NCRM.015 (January), 1–30.

Egenhofer, M.J. and Franzosa, R.D., 1991. Point-set topological spatial relations. *International journal of geographical information systems*, 5 (2), 161–174.

Eguíluz, V.M., Chialvo, D.R., Cecchi, G.A., Baliki, M., and Apkarian, A.V., 2005. Scale-Free Brain Functional Networks. *Physical Review Letters*, 94 (1), 018102.

Emch, M., Root, E.D., Giebultowicz, S., Ali, M., Perez-Heydrich, C., and Yunus, M., 2012a. 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.

Emch, M., Root, E.D., Giebultowicz, S., Ali, M., Perez-Heydrich, C., and Yunus, M., 2012b. 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.

Entwisle, B., 2007. Putting people into place. *Demography*, 44 (4), 687–703.

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.

Eppstein, D. and Goodrich, M.T., 2008. Studying Geometric Graph Properties of Road Networks Through an Algorithmic Lens. *arXiv.org*, cs.CG, 1–18.

Erdős, P. and Rényi, A., 1960. On the Evolution of Random Graphs. *In*: *PUBLICATION OF THE MATHEMATICAL INSTITUTE OF THE HUNGARIAN ACADEMY OF SCIENCES*.

Expert, P., Evans, T.S., Blondel, V.D., and Lambiotte, R., 2011. Uncovering space-independent communities in spatial networks. *Proceedings of the National Academy of Sciences of the United States of America*, 108 (19), 7663–8.

Farine, D.R. and Whitehead, H., 2015. Constructing, conducting and interpreting animal social network analysis. *Journal of Animal Ecology*, 84 (5), 1144–1163.

Faust, K. and Skvoretz, J., 2008. Space and Time , Size and Species. *Networks*, 32 (2002), 267–299.

Ferraro, P.J., Hanauer, M.M., and Sims, K.R.E., 2011. Conditions associated with protected area success in conservation and poverty reduction. *Proceedings of the National Academy of Sciences*, 108 (34), 13913–13918.

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

Fiallo, E.A. and Jacobson, S.K., 1995. Local Communities and Protected Areas: Attitudes of Rural Residents Towards Conservation and Machalilla National Park, Ecuador. *Environmental Conservation*, 22 (03), 241.

Fischer, C.S., 1977. *Networks and Places: Social Relations in the Urban Setting*. Free Press.

Fisher, B. and Christopher, T., 2007. Poverty and biodiversity: Measuring the overlap of human poverty and the biodiversity hotspots. *Ecological Economics*, 62 (1), 93–101.

Fleming, L. and Sorenson, O., 2001. Technology as a complex adaptive system: Evidence from patent data. *Research Policy*, 30 (7), 1019–1039.

Fowler, J.H. and Christakis, N.A., 2009. Dynamic spread of happiness in a large social network: Longitudinal analysis over 20 years in the Framingham Heart Study. *BMJ: British Medical Journal*, 338 (7685), 1–13.

Frayling, T.M., Timpson, N.J., Weedon, M.N., Zeggini, E., Freathy, R.M., Lindgren, C.M., Perry, J.R.B., Elliott, K.S., Lango, H., Rayner, N.W., Shields, B., Harries, L.W., Barrett, J.C., Ellard, S., Groves, C.J., Knight, B., Patch, A.-M., Ness, A.R., Ebrahim, S., Lawlor, D.A., Ring, S.M., Ben-Shlomo, Y., Jarvelin, M.-R., Sovio, U., Bennett, A.J., Melzer, D., Ferrucci, L., Loos, R.J.F., Barroso, I., Wareham, N.J., Karpe, F., Owen, K.R., Cardon, L.R., Walker, M., Hitman, G.A., Palmer, C.N.A., Doney, A.S.F., Morris, A.D., Smith, G.D., Hattersley, A.T., and McCarthy, M.I., 2007. A Common Variant in the FTO Gene Is Associated with Body Mass Index and Predisposes to Childhood and Adult Obesity. *Science*, 316 (5826), 889–894.

Freeman, L.C., 1978. Centrality in social networks conceptual clarification. *Social Networks*, 1 (3), 215–239.

Freeman, L.C., 2004. *The development of social network analysis*. Empirical Press Vancouver.

Fruchterman, T.M.J. and Reingold, E.M., 1991. Graph drawing by force-directed placement. *Software: Practice and Experience*, 21 (11), 1129–1164.

Frug, J., 1996. The Geography of Community. *Stanford Law Review*, 48 (5), 1047–1108.

Fu, F., Liu, L., and Wang, L., 2008. Empirical analysis of online social networks in the age of

Web 2.0. *Physica A: Statistical Mechanics and its Applications*, 387 (2–3), 675–684.

Gallupe, O., 2016. Network Analysis. *In*: *The Handbook of Measurement Issues in Criminology and Criminal Justice*. Hoboken, NJ: John Wiley & Sons, Inc, 555–575.

Gao, P. and Bian, L., 2016. Scale effects on spatially embedded contact networks. *Computers, Environment and Urban Systems*, 59, 142–151.

Gao, S., Liu, Y., Wang, Y., and Ma, X., 2013. Discovering spatial interaction communities from mobile phone data. *Transactions in GIS*, 17 (3), 463–481.

Garrido, A., Ciencias, F. De, and Uned, D., 2009. Graph properties and invariants , by their associate matrices. *Advanced Modeling and Optimization*, 11 (3), 337–348.

De Geer, S., 1923. On the Definition, Method and Classification of Geography. *Geografiska Annaler*, 5 (1923), 1–37.

Giebultowicz, S., Ali, M., Yunus, M., and Emch, M., 2011a. A comparison of spatial and social clustering of cholera in Matlab, Bangladesh. *Health & place*, 17 (2), 490–7.

Giebultowicz, S., Ali, M., Yunus, M., and Emch, M., 2011b. The simultaneous effects of spatial and social networks on cholera transmission. *Interdisciplinary perspectives on infectious diseases*, 2011, 604372.

Gil, J., 2016. Street network analysis 'edge effects': Examining the sensitivity of centrality measures to boundary conditions. *Environment and Planning B: Planning and Design*, 147:1-147:16.

Girvan, M. and Newman, M.E.J., 2002. Community structure in social and biological networks. *Proceedings of the National Academy of Sciences of the United States of America*, 99 (12), 7821–6.

Glückler, J., 2007. Economic geography and the evolution of networks. *Journal of Economic Geography*, 7 (5), 619–634.

Goodchild, M.F., 2011. Scale in GIS: An overview. *Geomorphology*, 130 (1–2), 5–9.

Goodchild, M.F., Anselin, L., Appelbaum, R.P., and Harthorn, B.H., 2000. Toward Spatially

Integrated Social Science. *International Regional Science Review*, 23 (2), 139–159.

Goodchild, M.F. and Janelle, D.G., 2004. Thinking spatially in social science.pdf. *Spatially integrated social science*, 3–22.

Greicius, M.D., Krasnow, B., Reiss, A.L., and Menon, V., 2003. Functional connectivity in the resting brain: A network analysis of the default mode hypothesis. *Proceedings of the National Academy of Sciences*, 100 (1), 253–258.

Guha, R., 1989. Radical American Environmentalism and Wilderness Preservation: A Third World Critique. *Environmental Ethics*, 11 (8), 71–83.

Hagberg, A.A., Schult, D.A., and Swart, P.J., 2008. Exploring network structure, dynamics, and function using {NetworkX}. *In*: *Proceedings of the 7th Python in Science Conference (SciPy2008)*. 11–15.

Haggett, P. and Chorley, R.J., 1969. *Network Analysis in Geography*. Hodder & Stoughton Educational.

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

Hartter, J., 2009. Attitudes of Rural Communities Toward Wetlands and Forest Fragments Around Kibale National Park, Uganda. *Human Dimensions of Wildlife*, 14 (6), 433–447.

Hartter, J. and Goldman, A.C., 2009. Life on the Edge: Balancing Biodiversity, Conservation, and Sustaining Rural Livelihoods around Kibale National Park, Uganda. *Focus on Geography*, 52 (1), 11–17.

Hartter, J. and Southworth, J., 2009. Dwindling resources and fragmentation of landscapes around parks: Wetlands and forest patches around Kibale National Park, Uganda. *Landscape Ecology*.

Havlin, S., Kenett, D.Y., Ben-Jacob, E., Bunde, A., Cohen, R., Hermann, H., Kantelhardt, J.W., Kertész, J., Kirkpatrick, S., Kurths, J., Portugali, J., and Solomon, S., 2012. Challenges in network science: Applications to infrastructures, climate, social systems and economics. *The European Physical Journal Special Topics*, 214 (1), 273–293.

Hawelka, B., Sitko, I., Beinat, E., Sobolevsky, S., Kazakopoulos, P., and Ratti, C., 2014. Geo-located Twitter as proxy for global mobility patterns. *Cartography and Geographic Information Science*, 41 (3), 260–271.

Herz, A., Peters, L., and Truschkat, I., 2015. How to do Qualitative Structural Analysis: The Qualitative Interpretation of Network Maps and Narrative Interviews. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 16 (1).

Hillery, G.A., 1955. Definitions of Community: Areas of Agreement. *Rural Sociology*, 20 (2), 111–123.

Hipp, J.R., Corcoran, J., Wickes, R., and Li, T., 2014. Examining the Social Porosity of Environmental Features on Neighborhood Sociability and Attachment. *PLoS ONE*, 9 (1), e84544.

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.

Hirschi, C., 2013. The Organization of Innovation- The History of an Obsession. *Angewandte Chemie International Edition*, 52 (52), 13874–13878.

Hoffstein, B., 2012. The Power of Networks: Fractals of Complexity [online]. *BigThink*. Available from: http://bigthink.com/think-tank/the-power-of-networks-fractals-of-complexity%0A.

Holling, C.S., 1992. Cross-Scale Morphology, Geometry, and Dynamics of Ecosystems. *Ecological Monographs*, 62 (4), 447.

Hood, W.W. and Wilson, C.S., 2001. The Literature of Bibliometrics, Scientometrics, and Informetrics. *Scientometrics*, 52 (2), 291–314.

Hou, H., Kretschmer, H., and Liu, Z., 2008. The structure of scientific collaboration networks in Scientometrics. *Scientometrics*, 75 (2), 189–202.

Hristova, D., Panzarasa, P., and Mascolo, C., 2015. Multilayer Brokerage in Geo-Social Networks. *International Conference on Web and Social Media*, (Granovetter 1973), 159–167.

Huang, R., Moudon, A. V., Cook, A.J., and Drewnowski, A., 2015. The spatial clustering of obesity: does the built environment matter? *Journal of Human Nutrition and Dietetics*, 28 (6), 604–612.

Inskip, C., Ridout, M., Fahad, Z., Tully, R., Barlow, A., Barlow, C.G., Islam, M.A., Roberts, T., and MacMillan, D., 2013. Human–Tiger Conflict in Context: Risks to Lives and Livelihoods in the Bangladesh Sundarbans. *Human Ecology*, 41 (2), 169–186.

Iturria-Medina, Y., Sotero, R.C., Canales-Rodríguez, E.J., Alemán-Gómez, Y., and Melie-García, L., 2008. Studying the human brain anatomical network via diffusion-weighted MRI and Graph Theory. *NeuroImage*, 40 (3), 1064–1076.

Jacomy, M., Venturini, T., Heymann, S., and Bastian, M., 2014. ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software. *PloS one*, 9 (6), e98679.

Jeffery, R., Baxter, J., McGuire, M., and Linde, J., 2006. Are fast food restaurants an environmental risk factor for obesity? *International Journal of Behavioral Nutrition and Physical Activity*, 3 (1), 2.

Jiang, B., 2007. A topological pattern of urban street networks: Universality and peculiarity. *Physica A: Statistical Mechanics and its Applications*, 384 (2), 647–655.

Jones, S., 2006. A political ecology of wildlife conservation in Africa. Review of African Political Economy. *Review of African Political Economy*, 33 (109), 483–495.

K. N. Ninan, 2006. *The Economics of Biodiversity Conservation: Valuation in Tropical Forest Ecosystems*. Routledge.

Kadushin, C., 2004. Introduction to Social Network Theory. *Networks*, 63, 60.

Kadushin, C., 2012. *Understanding social networks*. Oxford: Oxford University Press.

Karanth, K.K., 2007. Making resettlement work: The case of India's Bhadra Wildlife Sanctuary. *Biological Conservation*, 139 (3–4), 315–324.

Karanth, K.K., DeFries, R., Srivathsa, A., and Sankaraman, V., 2012. Wildlife tourists in India's emerging economy: potential for a conservation constituency? *Oryx*, 46 (03), 382–390.

Karsai, M., Kivelä, M., Pan, R.K., Kaski, K., Kertész, J., Barabási, A.-L., and Saramäki, J., 2011. Small but slow world: How network topology and burstiness slow down spreading. *Physical Review E*, 83 (2), 025102.

Keeling, M.J., Danon, L., Ford, A.P., House, T., Jewell, C.P., Roberts, G.O., Ross, J. V., and Vernon, M.C., 2011. Networks and the epidemiology of infectious disease. *Interdisciplinary Perspectives on Infectious Diseases*, 2011.

King, T., Kavanagh, A.M., Jolley, D., Turrell, G., and Crawford, D., 2006. Weight and place: a multilevel cross-sectional survey of area-level social disadvantage and overweight/obesity in Australia. *International Journal of Obesity*, 30 (2), 281–287.

Kiss, A., 2004. Is community-based ecotourism a good use of biodiversity conservation funds? *Trends in Ecology and Evolution*, 19 (5), 232–237.

Kleinberg, J., 2000. Navigation in a small world. *Nature*, 406 (6798), 845.

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

Kossinets, G., 2006. Empirical Analysis of an Evolving Social Network. *Science*, 311 (5757), 88–90.

Kowald, M., Arentze, T.A., and Axhausen, K.W., 2015. Individuals? spatial social network choice: model-based analysis of leisure-contact selection. *Environment and Planning B: Planning and Design*, 42 (5), 857–869.

Koylu, C., Guo, D., Kasakoff, A., and Adams, J.W., 2014. Mapping family connectedness across space and time. *Cartography and Geographic Information Science*, 41 (1), 14–26.

Kretschmer, H., 1997. Patterns of behaviour in coauthorship networks of invisible colleges. *Scientometrics*, 40 (3), 579–591.

Kretschmer, H., 2004. Author productivity and geodesic distance in bibliographic co-authorship networks, and visibility on the Web. *Scientometrics*, 60 (3), 409–420.

Krioukov, D., Kitsak, M., Sinkovits, R.S., Rideout, D., Meyer, D., and Boguñá, M., 2012. Network Cosmology. *Scientific Reports*, 2 (1), 793.

Krivitsky, P.N. and Goodreau, S.M., 2013. STERGM - Separable Temporal ERGMs for modeling discrete relational dynamics with statnet, 1–38.

Krüger, O., 2005. The role of ecotourism in conservation: panacea or Pandora's box? *Biodiversity and Conservation*, 14 (3), 579–600.

Krzywinski, M., Birol, I., Jones, S.J., and Marra, M.A., 2012. Hive plots--rational approach to visualizing networks. *Briefings in Bioinformatics*, 13 (5), 627–644.

Kwan, M., 2004. Gis Methods in Time-Geographic Research : Geocomputation and Geovisualization, 267–280.

Kwan, M.P., 2002. Feminist visualization: Re-envisioning GIS as a method in feminist geographic research. *Annals of the Association of American Geographers*, 92 (4), 645–661.

Kwan, M.P., 2007. Mobile communications, social networks, and urban travel: Hypertext as a new metaphor for conceptualizing spatial interaction. *Professional Geographer*, 59 (4), 434–446.

Larner, W., 2015. Globalising knowledge networks: Universities, diaspora strategies, and academic intermediaries. *Geoforum*, 59, 197–205.

Latour, B., 1987. *Science in Action: how to follow scientists and engineers through society*. Harvard University.

Laurance, W.F., Carolina Useche, D., Rendeiro, J., Kalka, M., Bradshaw, C.J.A., Sloan, S.P., Laurance, S.G., Campbell, M., Abernethy, K., Alvarez, P., Arroyo-Rodriguez, V., Ashton, P., Benítez-Malvido, J., Blom, A., Bobo, K.S., Cannon, C.H., Cao, M., Carroll, R., Chapman, C., Coates, R., Cords, M., Danielsen, F., De Dijn, B., Dinerstein, E., Donnelly, M.A., Edwards, D., Edwards, F., Farwig, N., Fashing, P., Forget, P.-M., Foster, M., Gale, G., Harris, D., Harrison, R., Hart, J., Karpanty, S., John Kress, W., Krishnaswamy, J., Logsdon, W., Lovett, J., Magnusson, W., Maisels, F., Marshall, A.R., McClearn, D., Mudappa, D., Nielsen, M.R., Pearson, R., Pitman, N., van der Ploeg, J., Plumptre, A., Poulsen, J., Quesada, M., Rainey, H., Robinson, D., Roetgers, C., Rovero, F., Scatena, F., Schulze, C., Sheil, D., Struhsaker, T., Terborgh, J., Thomas, D., Timm, R., Nicolas Urbina-Cardona, J., Vasudevan, K., Joseph Wright, S., Carlos Arias-G., J., Arroyo, L., Ashton, M., Auzel, P., Babaasa, D., Babweteera, F., Baker, P., Banki, O., Bass, M., Bila-Isia, I., Blake, S., Brockelman, W., Brokaw, N., Brühl, C.A., Bunyavejchewin, S., Chao, J.-T., Chave, J., Chellam, R., Clark, C.J., Clavijo, J., Congdon, R., Corlett, R., Dattaraja, H.S., Dave, C., Davies, G., de Mello Beisiegel, B., de Nazaré Paes da Silva, R., Di Fiore, A., Diesmos, A., Dirzo, R., Doran-Sheehy, D., Eaton, M., Emmons, L., Estrada, A., Ewango, C., Fedigan, L., Feer, F., Fruth, B., Giacalone Willis, J., Goodale, U., Goodman, S., Guix, J.C., Guthiga, P., Haber, W., Hamer, K., Herbinger, I., Hill, J., Huang, Z., Fang Sun, I., Ickes, K., Itoh, A., Ivanauskas, N., Jackes, B., Janovec, J., Janzen, D., Jiangming, M., Jin, C., Jones, T., Justiniano, H., Kalko, E., Kasangaki, A., Killeen, T., King, H., Klop, E., Knott, C., Koné, I., Kudavidanage, E., Lahoz da Silva Ribeiro, J., Lattke, J., Laval, R., Lawton, R., Leal, M., Leighton, M., Lentino, M., Leonel, C., Lindsell, J., Ling-Ling, L., Eduard Linsenmair, K., Losos, E., Lugo, A., Lwanga, J., Mack, A.L., Martins, M., Scott McGraw, W., McNab, R., Montag, L., Myers Thompson, J., Nabe-Nielsen, J., Nakagawa, M., Nepal, S., Norconk, M., Novotny, V., O'Donnell, S., Opiang, M., Ouboter, P., Parker, K., Parthasarathy, N., Pisciotta, K., Prawiradilaga, D., Pringle, C., Rajathurai, S., Reichard, U., Reinartz, G., Renton, K., Reynolds, G., Reynolds, V., Riley, E., Rödel, M.-O., Rothman, J., Round, P., Sakai, S., Sanaiotti, T., Savini, T., Schaab, G., Seidensticker, J., Siaka, A., Silman, M.R., Smith, T.B., de Almeida, S.S., Sodhi, N., Stanford, C., Stewart, K., Stokes, E., Stoner, K.E., Sukumar, R., Surbeck, M., Tobler, M., Tscharntke, T., Turkalo, A., Umapathy, G., van Weerd, M., Vega Rivera, J., Venkataraman, M., Venn, L., Verea, C., Volkmer de Castilho, C., Waltert, M., Wang, B., Watts, D., Weber, W., West, P., Whitacre, D., Whitney, K., Wilkie, D., Williams, S., Wright, D.D., Wright, P., Xiankai, L., Yonzon, P., and Zamzani, F., 2012. Averting biodiversity collapse in tropical forest protected areas. *Nature*, 489 (7415), 290–294.

Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabasi, A.-L., Brewer, D., Christakis, N., Contractor, N., Fowler, J., Gutmann, M., Jebara, T., King, G., Macy, M., Roy, D., and Van Alstyne, M., 2009. SOCIAL SCIENCE: Computational Social Science. *Science*, 323 (5915), 721–723.

Lepp, A., 2007. Residents' attitudes towards tourism in Bigodi village, Uganda. *Tourism Management*, 28 (3), 876–885.

Lepp, A. and Holland, S., 2006. A Comparison of Attitudes Toward State-Led Conservation and Community-Based Conservation in the Village of Bigodi , Uganda. *Society & Natural Resources*, (December 2004), 609–623.

Leroux, J.S., Moore, S., and Dubé, L., 2013. Beyond the "I" in the Obesity Epidemic: A Review of Social Relational and Network Interventions on Obesity. *Journal of Obesity*, 2013, 1–10.

Lewis, K., Kaufman, J., Gonzalez, M., Wimmer, A., and Christakis, N., 2008. Tastes, ties, and time: A new social network dataset using Facebook.com. *Social Networks*, 30 (4), 330–342.

Leydesdorff, L. and Schank, T., 2008. Dynamic animations of journal maps: Indicators of structural changes and interdisciplinary developments. *Journal of the American Society for Information Science and Technology*, 59 (11), 1810–1818.

Leydesdorff, L., Wagner, C.S., and Bornmann, L., 2014. The European Union, China, and the United States in the top-1% and top-10% layers of most-frequently cited publications: Competition and collaborations. *Journal of Informetrics*, 8 (3), 606–617.

Leydesdorff, L., Wagner, C.S., and Bornmann, L., 2018. Betweenness and diversity in journal citation networks as measures of interdisciplinarity—A tribute to Eugene Garfield. *Scientometrics*, 114 (2), 567–592.

Li, C., Li, Q., Van Mieghem, P., Stanley, H.E., and Wang, H., 2015. Correlation between centrality metrics and their application to the opinion model. *European Physical Journal B*, 88 (3), 1–13.

Li, C., Wang, H., de Haan, W., Stam, C.J., and Mieghem, P. van, 2011. The correlation of metrics in complex networks with applications in functional brain networks. *Journal of Statistical Mechanics*, 2011 (11), P11018.

Li, W. and Yang, J.-Y., 2009. Comparing Networks from a Data Analysis Perspective. *Lecture Notes in Computer Science*, 1907–1916.

Liben-Nowell, D., Novak, J., Kumar, R., Raghavan, P., and Tomkins, A., 2005. Geographic routing in social networks. *Proceedings of the National Academy of Sciences of the United States of America*, 102 (33), 11623–8.

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.

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

Mackenzie, C. a., Sengupta, R.R., and Kaoser, R., 2015. Chasing baboons or attending class: protected areas and childhood education in Uganda. *Environmental Conservation*, (2007), 1–11.

Mackenzie, C.A., 2012. Analysis Accruing benefit or loss from a protected area: Location matters. *Ecological Economics*, 76.

MacKenzie, C.A., 2012. Trenches like fences make good neighbours: Revenue sharing around Kibale National Park, Uganda. *Journal for Nature Conservation*, 20 (2), 92–100.

Mackenzie, C.A. and Ahabyona, P., 2012. Elephants in the garden: Financial and social costs of crop raiding. *Ecological Economics*.

Mackenzie, C.A., Chapman, C.A., and Sengupta, R., 2011. Spatial patterns of illegal resource extraction in Kibale National Park, Uganda. *Environmental Conservation*, 39 (01), 38–50.

Mackenzie, C.A. and Hartter, J., 2013. Demand and proximity: drivers of illegal forest resource extraction. *Oryx*, 47 (02), 288–297.

MacKenzie, C.A., Moffatt, S.P., Ogwang, J., Ahabyona, P., and Sengupta, R.R., 2017. Spatial and temporal patterns in primary school enrolment and exam achievement in Rural Uganda. *Children's Geographies*, 15 (3), 334–348.

MacKenzie, C.A., Salerno, J., Hartter, J., Chapman, C.A., Reyna, R., Tumusiime, D.M., and Drake, M., 2017. Changing perceptions of protected area benefits and problems around Kibale National Park, Uganda. *Journal of Environmental Management*, 200, 217–228.

Magnani, M., Micenkova, B., and Rossi, L., 2013. Combinatorial Analysis of Multiple Networks. *arXiv.org*.

Malecki, E.J., 2010. EVERYWHERE? THE GEOGRAPHY OF KNOWLEDGE. *Journal of Regional Science*, 50 (1), 493–513.

Mann, C.F., Matula, D.W., and Olinick, E. V., 2008. The use of sparsest cuts to reveal the hierarchical community structure of social networks. *Social Networks*, 30 (3), 223–234.

Martin, A., McGuire, S., and Sullivan, S., 2013. Global environmental justice and biodiversity conservation. *The Geographical Journal*, 179 (2), 122–131.

McClanahan, T., Davies, J., and Maina, J., 2005. Factors influencing resource users and managers' perceptions towards marine protected area management in Kenya. *Environmental Conservation*, 32 (1), 42–49.

McMenamin, B., 2017. The pitfalls of A/B testing in social networks [online]. *OkCupid*. Available from: https://tech.okcupid.com/the-pitfalls-of-a-b-testing-in-social-networks/ [Accessed 2 May 2017].

McPherson, M., Smith-Lovin, L., and Cook, J.M., 2001. Birds of a Feather: Homophily in Social Networks. *Annual Review of Sociology*.

Mejias, U.A., 2006. The tyranny of nodes: Towards a critique of social network theories [online]. *Blog Post*. Available from: https://blog.ulisesmejias.com/2006/10/09/the-tyranny-of-nodes-towards-a-critique-of-social-network-theories/ [Accessed 30 Oct 2017].

Mejias, U.A., 2010. The limits of networks as models for organizing the social. *New Media & Society*, 12 (4), 603–617.

Micheli, G.A., 2000. Kinship, family and social network. *Demographic Research*, 3.

Milgram, S., 1967. The small world problem. *Psychology today*, 2 (1), 60–67.

Miteva, D.A., Pattanayak, S.K., and Ferraro, P.J., 2012. Evaluation of biodiversity policy instruments: what works and what doesn't? *Oxford Review of Economic Policy*, 28 (1), 69–92.

Mizruchi, M.S., 1994. Social Network Analysis: Recent Achievements and Current Controversies. *Acta Sociologica*, 37 (4), 329–343.

Mok, D., Wellman, B., and Carrasco, J., 2010. Does Distance Matter in the Age of the Internet? *Urban Studies*, 47 (13), 2747–2783.

Moore, S., 2010. Social Networks, Social Capital and Obesity. *In*: *Obesity Prevention*. Elsevier, 673–685.

Moreno, J.L., 1934. Who shall survive?: A new approach to the problem of human interrelations. *Nervous and Mental Disease Publishing Co*.

Mossa, S., Turtschi, A., Amaral, L. a N., Guimera, R., Guimerà, R., Mossa, S., Turtschi, A., Amaral, L. a N., Guimera, R., Guimerà, R., Mossa, S., Turtschi, A., Amaral, L. a N., Guimera, R., Guimerà, R., Mossa, S., Turtschi, A., and Amaral, L. a N., 2005. The worldwide air transportation network: Anomalous centrality, community structure, and cities' global roles. *Proceedings of the National Academy of Sciences of the United States of America*, 102 (22), 7794–7799.

Mugisha, A. and Jacobson, S.K., 2004. Threat reduction assessment of conventional and community-based conservation approaches to managing protected areas in Uganda. *Environmental Conservation*, 31 (3), 233–241.

Müller, F. and Li, B.-L., 2004. Complex Systems Approaches To Study Human-Environmental Interactions: Issues And Problems. *In*: *Ecological Issues in a Changing World*. Dordrecht: Springer Netherlands, 31–46.

Nag, M., 2009. Mapping networks: A new method for integrating spatial and network data. *Princeton University*. Princeton University.

Nakamori, Y. and Sawaragi, Y., 2000. Complex systems analysis and environmental modeling. *European Journal of Operational Research*, 122 (2), 178–189.

Naughton-Treves, L., 1997. Farming the forest egde: vulnerable places and people around Kibale National Park, Uganda. *Geographical Review*, 87 (January), 27–46.

Naughton-Treves, L., 1998. Predicting patterns of crop damage by wildlife around Kibale National Park, Uganda. *Conservation Biology*, 12 (1), 156–168.

Naughton-Treves, L., 1999. Whose animals? A history of property rights to wildlife in Toro, western Uganda. *Land Degradation & Development*, 10 (4), 311–328.

Naughton-Treves, L., Alix-Garcia, J., and Chapman, C.A., 2011. Lessons about parks and poverty from a decade of forest loss and economic growth around Kibale National Park, Uganda. *Proceedings of the National Academy of Sciences of the United States of America*.

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

Newman, M.E.J., 2002. Assortative mixing in networks. *Physical review letters*, 89, 208701.

Newman, M.E.J., 2004a. Detecting community structure in networks. *The European Physical Journal B - Condensed Matter*, 38 (2), 321–330.

Newman, M.E.J., 2004b. Fast algorithm for detecting community structure in networks. *Physical Review E*, 69 (6), 066133.

Newman, M.E.J., 2006. Modularity and community structure in networks. *Proceedings of the National Academy of Sciences of the United States of America*, 103 (23), 8577–82.

Newman, M.E.J., 2010. *Networks: An Introduction*. Oxford University Press.

Nielsen, T.A.S. and Hovgesen, H.H., 2008. Exploratory mapping of commuter flows in England and Wales. *Journal of Transport Geography*, 16 (2), 90–99.

Nyhus, P.J., Osofsky, S.A., Ferraro, P., Madden, F., and Fischer, H., 2005. Bearing the costs of human-wildlife conflict: the challenges of compensation schemes. *CONSERVATION BIOLOGY SERIES-CAMBRIDGE*, 9, 107.

Oketch, M.O. and Rolleston, C.M., 2007. *Policies on Free Primary and Secondary Education in East Africa: A Review of the Literature*. CREATE Pathaways to Literature.

Onnela, J.-P., Arbesman, S., González, M.C., Barabási, A.-L., and Christakis, N. a, 2011. Geographic constraints on social network groups. *PloS one*, 6 (4), e16939.

Openshaw, S. and Rao, L., 1995. Algorithms for reengineering 1991 Census geography. *Environment and Planning A*, 27 (3), 425–446.

Opsahl, T., 2009. Structure and Evolution of Weighted Networks. University of London.

Opsahl, T., Agneessens, F., and Skvoretz, J., 2010. Node centrality in weighted networks: Generalizing degree and shortest paths. *Social Networks*, 32 (3), 245–251.

Otte, E. and Rousseau, R., 2002. Social network analysis: a powerful strategy, also for the information sciences. *Journal of Information Science*, 28 (6), 441–453.

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*.

Packer, C., Ikanda, D., Kissui, B., and Kushnir, H., 2005. Conservation biology: lion attacks on humans in Tanzania. *Nature*, 436 (7053), 927–928.

Palla, G., Derényi, I., Farkas, I., and Vicsek, T., 2005. Uncovering the overlapping community structure of complex networks in nature and society. *Nature*, 435 (7043), 814–818.

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.

Pelechrinis, K. and Krishnamurthy, P., 2015. Socio-spatial affiliation networks. *Computer Communications*, 000, 1–12.

Perkins, S.E., Cagnacci, F., Stradiotto, A., Arnoldi, D., and Hudson, P.J., 2009. Comparison of social networks derived from ecological data: implications for inferring infectious disease dynamics. *Journal of Animal Ecology*, 78 (5), 1015–1022.

Peuquet, D.J., 1994. It's About Time: A Conceptual Framework for the Representation of Temporal Dynamics in Geographic Information Systems. *Annals of the American Association of Geographers*, 84 (3), 441–461.

Peuquet, D.J. and Duan, N., 1995. An event-based spatiotemporal data model (ESTDM) for temporal analysis of geographical data. *International Journal of Geographical Information Science*, 9 (1), 7–24.

Ponds, R., 2009. The limits to internationalization of scientific research collaboration. *The Journal of Technology Transfer*, 34 (1), 76–94.

Pons, P. and Latapy, M., 2005. Computing Communities in Large Networks Using Random Walks. *In*: *Computer and Information Sciences - ISCIS 2005*. 284–293.

Porta, S., Crucitti, P., and Latora, V., 2006. The network analysis of urban streets: A dual approach. *Physica A: Statistical Mechanics and its Applications*, 369 (2), 853–866.

Porter, A.L. and Rafols, I., 2009. Is science becoming more interdisciplinary? Measuring and mapping six research fields over time. *Scientometrics*, 81 (3), 719–745.

Pouliou, T. and Elliott, S.J., 2009. An exploratory spatial analysis of overweight and obesity in Canada. *Preventive Medicine*, 48 (4), 362–367.

Preciado, P., Snijders, T. a. B., Burk, W.J., Stattin, H., and Kerr, M., 2011. Does proximity matter? Distance dependence of adolescent friendships. *Social Networks*, 34 (1), 18–31.

Prell, C., 2012. *Social network analysis*. London: SAGE Publications Ltd.

Project, K., 2016. *Kasiisi Project annual Report 2016*.

QS World University Rankings, 2017. QS World University Rankings Methodology [online]. *topuniversities.com*. Available from: https://www.topuniversities.com/qs-world-university-rankings/methodology [Accessed 19 Jan 2018].

Quattrochi, D.A. and Goodchild, M.F., 1997. *Scale in Remote Sensing and GIS*. Boca Raton, FL: CRC Press.

Radil, S.M., 2011. Spatializing Social Networks: Making Space for Theory in Spatial Analysis. *Dissertation*. University of Illinois at Urbana-Champaign.

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.

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.

Rands, M.R.W., Adams, W.M., Bennun, L., Butchart, S.H.M., Clements, A., Coomes, D., Entwistle, A., Hodge, I., Kapos, V., Scharlemann, J.P.W., Sutherland, W.J., and Vira, B., 2010. Biodiversity Conservation: Challenges Beyond 2010. *Science*, 329 (5997), 1298–1303.

Ratti, C., Sobolevsky, S., Calabrese, F., Andris, C., Reades, J., Martino, M., Claxton, R., and Strogatz, S.H., 2010. Redrawing the map of Great Britain from a network of human interactions. *PLoS ONE*, 5 (12).

Rauch, J.E., 2001. Business and Social Networks in International Trade. *Journal of Economic Literature*, 39 (4), 1177–1203.

Resnick, M.D., Bearman, P.S., Blum, R.W., Bauman, K.E., Harris, K.M., Jones, J., Tabor, J., Beuhring, T., Sieving, R.E., Shew, M., Ireland, M., Bearinger, L.H., and Udry, J.R., 1997. Protecting Adolescents From Harm Longitudinal Study. *Jama*, 278 (10), 823–832.

Rindfuss, R.R., Jampaklay, A., Entwisle, B., Sawangdee, Y., Faust, K., and Prasartkul, P., 2004. The Collection and Analysis of Social Network Data in Nang Rong, Thailand. *In*: M. Morris, ed. *Network Epidemiology: A Handbook for Survey Design and Data Collection*. Oxford University Press, 175–200.

Rinzivillo, S., Mainardi, S., Pezzoni, F., Coscia, M., Pedreschi, D., and Giannotti, F., 2012. Discovering the Geographical Borders of Human Mobility. *KI Künstliche Intelligenz*, 26 (3), 253–260.

Robinson, J.G., 1993. The Limits to Caring: Sustainable Living and the Loss of Biodiversity. *Conservation Biology*, 7 (1), 20–28.

Ross, E., 2013. The Role of Small NGOs: Building Quality International Education. *Harvard International Review*, 35 (1), 40–44.

Sailer, K. and McCulloh, I., 2012. Social networks and spatial configuration—How office layouts drive social interaction. *Social Networks*, 34 (1), 47–58.

Salonen, M., Toivonen, T., Cohalan, J.-M., and Coomes, O.T., 2012. Critical distances: Comparing measures of spatial accessibility in the riverine landscapes of Peruvian Amazonia. *Applied Geography*, 32 (2), 501–513.

Sandbrook, C. and Adams, W.M., 2012. Accessing the Impenetrable: The Nature and Distribution of Tourism Benefits at a Ugandan National Park. *Society & Natural Resources*, 25 (9), 915–932.

Sarkar, D., Chapman, C.A., Angom, S.C., Valenta, K., Kagoro, W., and Sengupta, R., n.d. Research stations as conservation instruments provide long term community benefits through social connections. *Under review*.

Sarkar, D., Chapman, C.A., Kagoro, W., and Sengupta, R., 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.

Sarkar, D., Sieber, R., and Sengupta, R., 2016. GIScience Considerations in Spatial Social Networks. *In*: J.A. Miller, D.O. Sullivan, and N. Wiegand, eds. *Lecture Notes in Computer Science*. Springer International Publishing, 85–98.

Scellato, S., Mascolo, C., Musolesi, M., and Latora, V., 2010. Distance Matters: Geo-social Metrics for Online Social Networks. *In*: *Proceedings of the 3rd Wonference on Online Social Networks*. 8–1811111.

Scellato, S., Noulas, A., Lambiotte, R., and Mascolo, C., 2011a. Socio-spatial Properties of Online Location-based Social Networks. *In*: L.A. Adamic, R.A. Baeza-Yates, and S. Counts, eds. *ICWSM*. The AAAI Press.

Scellato, S., Noulas, A., Lambiotte, R., and Mascolo, C., 2011b. Socio-Spatial Properties of Online Location-Based Social Networks. *ICWSM*, 11, 329–336.

Schipper, D. and Spekkink, W., 2015. Balancing the Quantitative and Qualitative Aspects of Social Network Analysis to Study Complex Social Systems. *Complexity, Governance & Networks*, 1 (1), 5–22.

Scott, J., 2017. *Social Network Analysis*. SAGE Publications Ltd.

Sempsey, J.J., 1998. The death of distance: How the communications revolution will change our lives. *Journal of the American Society for Information Science*, 49 (11), 1041–1042.

Sen, P., Dasgupta, S., Chatterjee, A., Sreeram, P.A., Mukherjee, G., and Manna, S.S., 2003. Small-world properties of the Indian railway network. *Physical Review E*, 67 (3), 036106.

Serrat, O., 2017. Social Network Analysis. *In*: *Knowledge Solutions*. Singapore: Springer

Singapore, 39–43.

Sevtsuk, A., Huang, S., Calabrese, F., and Ratti, C., 2009. Sevtsuk, A., Huang, S., Calabrese, F., & Ratti, C. (2009). Mapping the MIT campus in real time using WiFi. *In*: *Handbook of Research on Urban Informatics: The Practice and Promise of the Real-Time City*. 326–337.

Skupin, A. and Agarwal, P., 2008. Introduction: What is a Self-Organizing Map? *In*: *Self-Organising Maps*. Chichester, UK: John Wiley & Sons, Ltd, 1–20.

Skupin, A. and Fabrikant, S., 2003. Spatialization methods: a cartographic research agenda for non-geographic information visualization. *Cartography and Geographic Information Science*, 30, 95–119.

Smith, D.M., 1999. Geography, community, and morality. *Environment and Planning A*, 31 (1), 19–35.

Snijders, T. a. B., 1996. Stochastic actor-oriented models for network change. *The Journal of Mathematical Sociology*, 21 (1–2), 149–172.

Sobolevsky, S., Szell, M., Campari, R., Couronné, T., Smoreda, Z., and Ratti, C., 2013. Delineating geographical regions with networks of human interactions in an extensive set of countries. *PLoS ONE*, 8 (12).

Stommel, H., 1963. Varieties of Oceanographic Experience: The ocean can be investigated as a hydrodynamical phenomenon as well as explored geographically. *Science*, 139 (3555), 572–576.

Strandburg-Peshkin, A., Twomey, C.R., Bode, N.W.F., Kao, A.B., Katz, Y., Ioannou, C.C., Rosenthal, S.B., Torney, C.J., Wu, H.S., Levin, S.A., and Couzin, I.D., 2013. Visual sensory networks and effective information transfer in animal groups. *Current Biology*, 23 (17), R709–R711.

Struhsaker, T., 1999. *Ecology of an African Rain Forest: Logging in Kibale and the Conflict Between Conservation and Exploitation*. 1st ed. University Press of Florida.

Struhsaker, T.T., Struhsaker, P.J., and Siex, K.S., 2005. Conserving Africa’s rain forests: Problems in protected areas and possible solutions. *Biological Conservation*, 123 (1), 45–

54.

Sullivan, D.O., 2014. *Handbook of Regional Science*. Berlin, Heidelberg: Springer Berlin Heidelberg.

Thiemann, C., Theis, F., Grady, D., Brune, R., and Brockmann, D., 2010. The Structure of Borders in a Small World. *PLoS ONE*, 5 (11), e15422.

Tijssen, R.J.W. and Winnink, J.J., 2017. Capturing ‘R&D excellence’: indicators, international statistics, and innovative universities. *Scientometrics*, 1–13.

Times Higher Education, 2017. World University Rankings 2018 methodology [online]. *timeshighereducation.com*. Available from: https://www.timeshighereducation.com/world-university-rankings/methodology-world-university-rankings-2018 [Accessed 19 Jan 2018].

Tita, G.E. and Radil, S.M., 2011. Spatializing the Social Networks of Gangs to Explore Patterns of Violence. *Journal of Quantitative Criminology*, 27 (4), 521–545.

Tobler, W.R., 1970. A Computer Movie Simulating Urban Growth in the Detroit Region. *Economic Geography*, 46, 234.

Tranos, E. and Nijkamp, P., 2012. *The Death of Distance Revisited: Cyberplace, Physical and Relational Proximitie*. Tinbergen Institute Discussion Paper.

USGS, 2006. Shuttle Radar Topography Mission.

Valente, T.W., 2010. *Social Networks and Health: Models, Methods, and Applications*. Oxford University Press.

Valente, T.W., Coronges, K., Lakon, C., and Costenbader, E., 2008. How Correlated Are Network Centrality Measures? *Connections*, 28 (1), 16–26.

Verdery, A.M., Entwisle, B., Faust, K., and Rindfuss, R.R., 2012. Social and spatial networks: Kinship distance and dwelling unit proximity in rural Thailand. *Social Networks*, 34 (1), 112–127.

Vertesi, J., 2008. Mind the Gap: The London Underground Map and Users’ Representations of Urban Space. *Social Studies of Science*, 38 (1), 7–33.

Wagner, C.S. and Leydesdorff, L., 2005. Network structure, self-organization, and the growth of international collaboration in science. *Research Policy*, 34 (10), 1608–1618.

Wagner, C.S., Park, H.W., and Leydesdorff, L., 2015. The Continuing Growth of Global Cooperation Networks in Research: A Conundrum for National Governments. *PLOS ONE*, 10 (7), e0131816.

Ward, M.D. and Gleditsch, K.S., 2008. *Spatial Regression Models*. SAGE Publications Ltd.

Wasserman, S., 1994. *Social network analysis: Methods and applications*. Cambridge University Pres.

Wasserman, S. and Faust, K., 1999. *Social network analysis: Methods and applications*. Cambridge, UK.: Cambridge University Pres.

Wasserman, S. and Iacobucci, D., 1998. Sequential social network data. *Psychometrika*, 53 (2), 261–282.

Watts, D.J., Dodds, P.S., and Newman, M.E.J., 2002. Identity and search in social networks. *Science (New York, N.Y.)*, 296 (5571), 1302–5.

Watts, D.J. and Strogatz, S.H., 1998. Collective dynamics of 'small-world' networks. *Nature*, 393 (6684), 440–442.

Wen, T.-H., Chen, D.-R., and Tsai, M.-J., 2010. Identifying geographical variations in poverty-obesity relationships: empirical evidence from Taiwan. *Geospatial health*, 4 (2), 257.

WESP, 2014. *Country Classification System*. World Economic Situation and Prospects.

Wilshusen, P.R., Brechin, S.R., Fortwangler, C.L., and West, P.C., 2002. Reinventing a Square Wheel: Critique of a Resurgent 'Protection Paradigm' in International Biodiversity Conservation. *Society & Natural Resources*, 15 (1), 17–40.

Wineman, J., Kabo, F., and Davis, G., 2009. Spatial and social networks in organizational innovation. *Environment and Behavior*, 41 (3), 427–442.

Wong, L.H., Pattison, P., and Robins, G., 2006a. A spatial model for social networks. *Physica A: Statistical Mechanics and its Applications*, 360 (1), 99–120.

Wong, L.H., Pattison, P., and Robins, G., 2006b. A spatial model for social networks. *Physica A: Statistical Mechanics and its Applications*, 360 (1), 99–120.

Wu, H. and Li, Z.-L., 2009. Scale Issues in Remote Sensing: A Review on Analysis, Processing and Modeling. *Sensors*, 9 (3), 1768–1793.

Wuestman, M., Hoekman, J., and Frenken, K., 2018. The Complex Geography of Scientific Knowledge Accumulatio. *In*: *DRUID Academy Conference*.

Wylie, J.L., Cabral, T., and Jolly, A.M., 2005. Identification of networks of sexually transmitted infection: a molecular, geographic, and social network analysis. *The Journal of infectious diseases*, 191 (6), 899–906.

Ye, M., Yin, P., and Lee, W.-C., 2010. Location recommendation for location-based social networks. *In*: *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems - GIS '10*. New York, New York, USA: ACM Press, 458.

Yuan, M., Nara, A., and Bothwell, J., 2014. Space-time representation and analytics. *Annals of GIS*, 20 (1), 1–9.

Zaggl, M.A., 2017. Manipulation of explicit reputation in innovation and knowledge exchange communities: The example of referencing in science. *Research Policy*, 46 (5), 970–983.

Zhong, C., Arisona, S.M., Huang, X., Batty, M., and Schmitt, G., 2014. Detecting the dynamics of urban structure through spatial network analysis. *International Journal of Geographical Information Science*, 28 (11), 2178–2199.

Ziersch, A.M., Baum, F.E., Macdougall, C., and Putland, C., 2005. Neighbourhood life and social capital: the implications for health. *Soc.Sci.Med.*, 60 (1), 71–86.

Zimmerer, K.S., 1994. Human Geography and the "New Ecology": The Prospect and Promise of Integration. *Annals of the Association of American Geographers*, 84 (1), 108–125.

Zimmerer, K.S., 2003. Geographies of Seed Networks for Food Plants (Potato, Ulluco) and Approaches to Agrobiodiversity Conservation in the Andean Countries. *Society & Natural Resources*, 16 (7), 583–601.

Zimmerer, K.S., 2006. Cultural ecology: at the interface with political ecology - the new geographies of environmental conservation and globalization. *Progress in Human Geography*, 30 (1), 63–78.

Zimmerer, K.S., 2010. Biological Diversity in Agriculture and Global Change. *Annual Review of Environment and Resources*, 35 (1), 137–166.

Zuckerman, E.W., 2008. Why social networks are overrated (a 3+ when they are at best a 2) [online]. *OrgTheory.net*. Available from: http://orgtheory.wordpress.com/2008/11/14/why-social-networks-are-overrated-a-3-when-they-are-at-best-a-2/.