DETECTING ASSOCIATED COMMUNITIES IN SOCIAL NETWORK AND URBAN ACTIVITY SPACES

by

YAOLI WANG

(Under the Direction of Xiaobai Yao)

ABSTRACT

Clusters in social network tell the heterogeneity of people's connections, and clusters in geographical movement network display the difference in movement pattern. I test whether the two clusters show similar pattern to understand the complexity between social network and movement behaviors for implications on future urban structure that helps maintaining face-to-face social connections. I do community detection simultaneously and independently in both networks drawn from a mobile phone call dataset in Jiamusi, China. I involve distance decay to detect clusters due to long-distance geographical movements. I also do community detection in social network and project the social communities into geographical space by anchor points to examine whether long-distance movement communities are spatially associated with social communities. The result testifies my argument that people still require physical interaction in social life, even in the era of information.

INDEX WORDS: Social network, Urban activity space, Community detection, Distance decay, Cell phone call dataset

DETECTING ASSOCIATED COMMUNITIES IN SOCIAL NETWORK AND URBAN ACTIVITY SPACES

by

YAOLI WANG

B.S., Peking University, China, 2012

A Thesis Submitted to the Graduate Faculty of the University of Georgia in Partial Fulfillment of the Requirements for the Degree

MASTER OF SCIENCE

ATHENS, GEORGIA

2014

DETECTING ASSOCIATED COMMUNITIES IN SOCIAL NETWORK AND URBAN ACTIVITY SPACES

by

YAOLI WANG

Major Professor: Xiaobai Yao
Committee: Lan Mu
Steven Holloway
Clio Andris

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
May 2014

ACKNOWLEDGEMENTS

I sincerely appreciate the aegis of Dr. Xiaobai Yao in my two years' study in the Department of Geography at the University of Georgia. Without her support, I would not have had my research fruits or the manuscript attached in my thesis. I gratefully thank my committee members, Dr. Lan Mu, Dr. Steven Holloway, and Dr. Clio Andris for their guidance in my study and research. I benefited a lot from Dr. Mu's programming class and from Dr. Holloway's instruction of statistical analysis. Dr. Clio Andris infused me with great ideas and helped me with some data pre-processing during my summer internship with her at the Santa Fe Institute. I also appreciate Dr. Yu Liu at the Peking University, China, for releasing the dataset to me, and Chaogui Kang, a doctoral student with Dr. Liu, for running code for me remotely to draw sample data. I thankfully appreciate the funding from the *International Geographic Informational Fund* to grant me the student travel grant for my research presentation at the *2014 Annual Meeting of the Association of American Geographers*, Tampa, FL. I acknowledge the funding from the *Special Fund of Key Laboratory of Eco Planning & Green Building, Ministry of Education (Tsinghua University), China* to support my manuscript publication. Last but not least, as an international student, the support from my parents and friends are so valuable! I love them!

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

Reasons for This Research

Urban sprawl has been a predominant trend in contemporary society. Landscape Urbanism argues that urban design should be a flexible, open-ended, strategic and non-hierarchical process instead of a static plan as is traditionally done (Chaeles Waldheim 2002; Charles Waldheim 2006), which should make advantage of the existing resources on the ground (e.g., the Sustainable Park proposed by Cranz and Boland (Cranz and Boland 2004)), implement ecological infrastructure design that prevents urban encroachment of wilderness (e.g., (Yu, Wang, and Li 2011)), to name a few. The doctrine of Landscape Urbanism hence more or less encourages, or at least accepts urban sprawl (Koolhaas and Mau 1998).

However, distance decay, as elaborated by Tobler's First Law of Geography (Tobler 1970), has been widely observed in concrete geographical space (Eldridge and Jones 1991; Gao, Wang, et al. 2013; Fotheringham 1981; Kang et al. 2010). Thus, urban sprawl will definitely generate higher impedance of distance. On the other hand, as argued by Cairncross (2001), the development of telecommunication has diminished the importance of distance, especially in terms of organizational businesses, given the prosperities of, for example, numerous international companies. In this sense, the telecommunication network seems able to disperse everywhere free of geographic constraints. However, by evidence from the geography of internet usage, Kolko (Kolko 2000) proved that internet, as a good substitute simply for long-distance communication which benefits remote cities, is actually a complement, instead of a substitute, for face-to-face communications. Not only have observation results confirmed that distance decay does exist in virtual space, as shown by evidences from telecommunication (Ratti et al. 2010; Gao, Liu, et al. 2013; Walsh and Pozdnoukhov 2011) and social media check-in data (Liu et al. 2014), but also some studies argue that telecommunication is not functionally identical to face-to-face communication (Flaherty, Pearce, and Rubin 1998; Duke 2001).

Zooming in to an urban scale, the *linked activity spaces* (Wang et al.) suggests that friends (in this research, defined as a pair of persons with any possible social connection, e.g., business, family, classmate, friends, etc.) tend to use the same urban infrastructure more often than by random chance (Figure 1.1). By saying random chance, we mean the possibility that two random selected persons utilize the same urban infrastructure approximated by Point of Interest (POI), like parks, bars, restaurants, etc.). Such finding indicates that friends are spatially clustered in urban activity space. Inspired by that, we hypothesize that agents who are clustered by the relation of friendship in social network are more likely to have their physical activity spaces belonging to the same cluster as well. Moreover, while the *linked activity spaces* demonstrate more about local pattern (of short-distance movements), we aim to find out the mutual interplay between social network and long urban trips. By saying "long", we particularly refer to the trips longer than the expectation from the average distance decay.

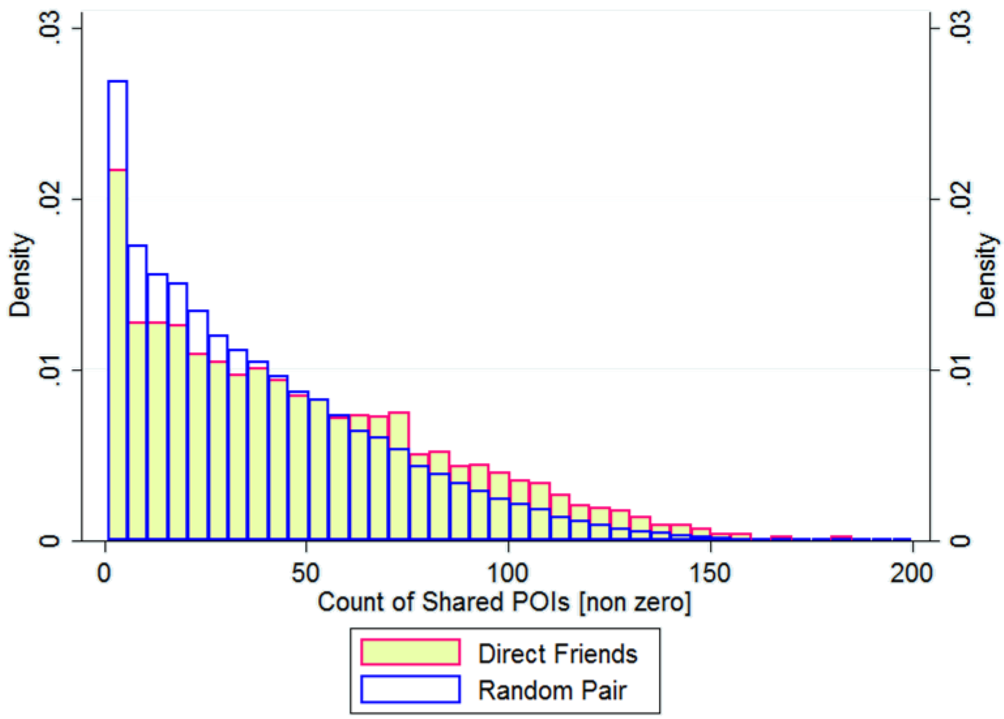


Figure 1.1 The count of shared POIs by friends against that by randomness (Wang et al.). This is calculated by counting POIs in the overlapped activity spaces of each possible pair of users or each pair of friends. The horizontal axis depicts the number of POIs in the overlapped activity spaces, POIs within in which are shared by both users. Density is the frequency distribution: pink frame bars shows what percentage of all pairs of friends share the corresponded count of POIs, while blue frame bars displays the same thing but of random pairs of users, i.e., users either friends or non-friends. It demonstrates that friends tend to share more POIs than randomness, indicating potential spatial clusters in their activity spaces.

Objective of This Research

While many works discuss the issue of smart growth and urban sprawl from diverse angles (Geller 2003; Ewing et al. 2003; Greca et al. 2011), this research particularly adopts community detection in network analysis to examine the association between the clustering patterns from the social network built via telecommunication and the long-distance urban movements. If there is significant similarity between the social and the movement clustering patterns, I would argue that, in terms of social connections, urban sprawl should not be extensively encouraged, despite its advantages as argued by Landscape Urbanism (Livesey 2009; Chaeles Waldheim 2002; Koolhaas and Mau 1998).

This work hence will leverage a mobile phone dataset, doing community detections independently in social network and in urban movement network. The social network community detection is initially conducted in a completely virtual network without spatial information, the result from which will be projected into urban space via each user's mostly visited cell phone towers (i.e., anchor points). For unban movement network, distance decay is involved to detect communities caused by long-distance trips. The similarity of patterns from social and movement communities is measured by an index called *relative cardinality*, and hypothesis test is utilized to ensure the significance of the similarity.

A Review of the Major Methods

Despite that "community" may have varied specific definitions according to its different application situations, the basic idea is the intra-community interaction should be stronger or closer than inter-community ones. Community is an important mesoscopic structure, which is caused by the heterogeneity in network, and which is about the presence of high densities of connection in some regions while of low densities in other regions (M. Newman and Girvan

2003). In spatial network, community structure is identical to the concept of spatial heterogeneity, for which the “spatial cluster” is defined as high densities of connections, say, communication strength, traffic flow intensity, not necessarily short distance between two nodes.

Community detection is the set of methods and process to partition a network into smaller scale networks while meeting the aforementioned requirements. Among those different methods (Barthélemy 2011) for community detection, we adopt the modularity-based method, initially proposed by M. E. J. Newman & Girvan (2004). Modularity is a measure quantifying the difference between really observed connection strength and the expected connection strength. This method sets a null model to quantify the theoretical network structure, i.e., a network with the intact topological connection structure but with different connection strength (link weight).

The reasons to choose it are due to two partially overlapped reasons. Firstly, compared with other methods, community detection considers the significance of detected communities against randomness (by null model). The phenomenon called “small world”(Watts and Strogatz 1998; Watts, Dodds, and Newman 2002; Amaral et al. 2000) supports that, instead of growing evenly, network innately have clusters; high degree nodes are inclined to attract new nodes more rapidly through evolution (Liben-Nowell and Novak 2005; Watts, Dodds, and Newman 2002; Guimerà et al. 2005; Thiemann et al. 2010; Ratti et al. 2010; Grady et al. 2012). Hence, if we consider the random clustering mechanism to show a robust community structure beyond randomness, modularity-based method is a convenient way.

Secondly, it allows the flexibility to either include or exclude distance into community detection (Expert et al. 2011). By plugging the distance decay model into modularity as the approximation to the expected connection strength, I am able to detect the communities formed by long-distance urban movements, because the null model predicts that the cell phone towers with short distance *should* have higher cell phone call volume than the towers with long distance, and thus short-distance towers are not assigned to the same community simply due to the absolute high volume between them.

Innovations of This Research

Transportation network and mobile phone network are two areas that intensively use network analysis to understand our complex social systems with regard to spatial relationship (Kang et al. 2012; Lambiotte and Blondel 2008; Liben-Nowell and Novak 2005; Grady et al. 2012; Onnela et al. 2011; Guo 2009; Gao, Liu, et al. 2013). Other works exploring social network also start referring to the underlying geographic space to explain the intensity of social interaction or the structure of social network (Steinhaeuser and Chawla 2008; Watts and Strogatz 1998; Correa et al. 2008; Vespignani 2009; Onnela et al. 2011; Eagle, Pentland, and Lazer 2009; Girvan and Newman 2002; Expert et al. 2011). Although some of the previous works referred above try to link social and spatial components together to explain their results, their analyses are still anchored in one field, either social or spatial. None of them has conducted analysis in both social and spatial networks *in parallel* and compared them with each other, nor has any of them concentrated particularly on long-distance movements on an urban scale. For example, Expert et al. (2011) introduced gravity model into social network community detection, arguing that such method tells the story *beyond* space (Expert et al. 2011, Figure 1.2). They realized that even virtual connection is constraint by space, but if removing spatial factor, they can uncover a real pattern depicting the underlying nature of social network. However, despite their realization of the spatial influence on social connection, they didn’t perform any analysis with physical movements.

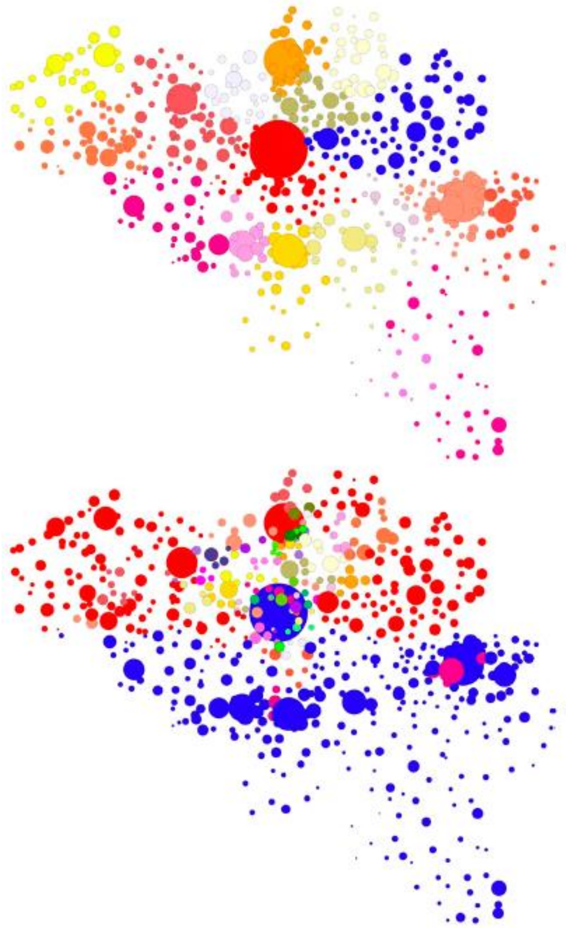


Figure 1.2 Community detection results without involving distance into modularity (upper) and with distance in modularity (lower). The former demonstrate good spatial continuity while the latter displays some scattered parts belong to the same community. Adopted from Expert et al. (2011).

This study, in contrast, proposes to perform social network analysis and spatial movement analysis independently initially, and then to juxtapose them to see if social network and physical movement yield similar clustering pattern. The research design is aimed to decipher the effect of social network on the generation of long urban trips. If the results demonstrate similar patterns, there will be good argument for the assumption that social connection is associated with long urban movements, which protests against urban sprawl.

Additionally, there are methodological innovations. Firstly, I revise the form of modularity from subtraction to division. As I will elaborate in the methodology section, the previous form of modularity is problematic because it is biased by the size of the community. I will show how the new version of modularity I propose can adjust for it. Secondly, I generate a new method to make the social communities and the spatial communities comparable. Since there are few people doing community detection in social and spatial networks separately, no approach has been proposed about projecting social communities into urban space or about comparing the similarity between two community patterns.

CHAPTER 2

DETECTING ASSOCIATED COMMUNITIES IN SOCIAL NETWORK AND URBAN MOVEMENT SPACES[1]

[1] Wang, Y. and Yao, X. To be submitted to *PLOSONE*.

Abstract

The development of Information and Communication Technology makes the relationship between virtual social network and face-to-face interaction more complex. In this study, we base our points from the perspective of network analysis to discuss the association between social network and physical movements. As communities (i.e., clusters) emerge in both social network and geographical movement network, it is important to find whether the clusters show similar spatial patterns. This helps to understand the influence of telecommunication-based virtual social links on urban trips, whether replacement or supplement. We propose to do community detection simultaneously and independently in social and movement networks drawn from a mobile phone call dataset in Jiamusi, China. The social network community detection is originally performed without considering any geographic information, after which the social communities are projected into urban space. We involve distance decay into the modularity-based community detection of movement network to detect clusters particularly due to long-distance geographical movements. With random permutation and hypothesis test, we find significant association between the spatial patterns of long-distance movement communities and social network communities. The statistical result, followed by specific discussions contextualized in the city, testifies our argument that social connections yield meaningful indications on long urban trips. Information technology in social life, therefore, is not a substitute, but instead, a supplement for urban-scale physical interaction, which disputes intense urban sprawl.

Introduction

The complex relationship between the “virtual” space built up via telecommunication and the geographic impedance for physical interaction has attracted much research interests. Distance decay, as elaborated by *Tobler’s First Law of Geography* that “everything is related to everything else, but near things are more related than distant things" (Tobler 1970), has been widely observed in concrete geographical space (Eldridge and Jones 1991; Gao, Wang, et al. 2013; Fotheringham 1981). Admittedly, as argued by Cairncross (Cairncross 2001), the development of telecommunication has diminished the importance of distance, especially for organizations and companies that rely on the remote transmission of information to do global businesses. However, by evidence from the geography of internet usage, Kolko (Kolko 2000) proved that internet, which is presumably a good substitute simply for long-distance communications, is actually a complement, instead of a substitute, for face-to-face communications. Real world observations confirm that distance decay *does* exist in virtual space, as shown by evidences from telecommunication (Ratti et al. 2010; Gao, Liu, et al. 2013) and social media check-in data (Liu et al. 2014). Wang et al. (Wang et al.) further quantitatively proved that people with social links via telecommunication are significantly more inclined to have geographic proximity in their daily activity spaces. All these previous studies testify that interactions in the virtual space are indeed influenced by geographic distance. Expert et al. (Expert et al. 2011) explained the distance decay in virtual space by "homophily" (McPherson, Smith-Lovin, and Cook 2001) and "focus constraint" (Feld 1981); they tried to remove the bias of geographic proximity from social network and detected social communities (i.e., clusters in social network) *beyond* space to uncover hidden structural or cultural similarity.

However, the influence of distance decay may play out differently in the virtual space than that in the physical space. Previous studies argue that telecommunication is not functionally identical to face-to-face communication (Flaherty, Pearce, and Rubin 1998; Duke 2001). From a perspective of psychology, Salomon (Salomon 1985) proposed that technology is not a substitute for personal urban trips. He brought up the ambiguity of the potential influence of communication technology on urban trip generation, arguing that people cannot bear being homebound for long. Thus, when studying the effect of telecommunication on individual level and within cities, it is even more difficult to predict the influence of distance. Studying telecommunication interactions *within* a city raises the need to consider the physical movements of callers, unlike the *inter*-city interactions for which researchers can generally georeference individual callers to respective cities (as in (Expert et al. 2011)).

In this work, we aim to uncover, by network analysis, the uncertainty of the relationship between telecommunication-based social links and *urban-scale* face-to-face interactions. The aforementioned works observed that geographical distance decay may *constrain* social network, resulting in the "virtual" distance decay. This study, however, tends to investigate the relationship the other way around: what is the attraction between social network and geographical movements? Is the generation of *long-distance* urban trips associated with the existence of social links located physically far apart? Do they reinforce or replace each other? Simply, we want to answer the question: are social links in the virtual space associated with *long-distance* physical urban movements? If yes, we would argue that telecommunication is complementary to physical contact, because, instead of being widely dispersed in space, telecommunication is partially implemented, and thus implicated by urban movements. Note that the "long-distance" is a relative concept compared with the average movements in a particular city. We are primarily interested in long-distance urban trips because telecommunication usually plays a more important role in long-distance events, and we argue that the decrease in the amount of social links (with longer distance) does not mean the decay of their importance, as is suggested in the literature (Apicella et al. 2012).

We examine the association between social connections and long-distance urban movements by leveraging cell phone call records, one type of the most popular telecommunications. The "social connections" here is used in a broad sense, including both business and personal social contacts. We treat both social connections and urban movements as networks, by which the interactions between different participants or places are best represented. We apply community detection method to measure the pattern of relations among nodes. *Community* in network science means "cluster", i.e., a subgroup of nodes among which connections are stronger while inter-community connections are much weaker (M. Newman and Girvan 2003). *Community detection* is a method to partition the whole network into closely related clusters as communities (Steinhaeuser and Chawla 2008; Guimera, Sales-Pardo, and Amaral 2004; Wang 2012; Onnela et al. 2011; Girvan and Newman 2002; Palla, Barabási, and Vicsek 2007; Expert et al. 2011). We adopt modularity-based community detection (M E J Newman and Girvan 2004) to ensure the robustness of the detected community structure contingent on the expected connection strength. Modularity is a measure alike residual, which quantifies the gap between the observed and the expected values.

Our basic methodological idea is to perform community detection in the social network and the caller's physical movements in the urban space independently. Both social network and movements are extracted from cell phone call records. For social network, we start community detection with an entirely space-irrelevant network (i.e., a network containing no geographic information at all), and then map the resulted communities into urban space based on the most frequently utilized cell phone towers for each user. For physical movements, we apply a modified community detection method that rule out the influence of distance decay so that the identified communities result from other underlying factors than distance. Finally, we create a similarity measure for the two community patterns and conduct hypothesis test to verify the associations between social connections and physical urban movements from the perspective of network analysis.

The innovations of our work are twofold. Firstly, although there are many studies doing social or spatial network community detection (Steinhaeuser and Chawla 2008; Comber, Brunsdon, and Farmer 2012; Onnela et al. 2011; Expert et al. 2011; Gao, Liu, et al. 2013; Ratti et al. 2010), few of them has examined both social and spatial networks simultaneously and compared them. This study is one of the first to reflect the association between the two networks and give hint to urban future. Secondly, we are one of the earliest to involve distance decay in *movement* network to discover long trips. Although Expert et al. (Expert et al. 2011) have considered distance decay in community detection, their study subject is *social* network.

The findings of the study suggest that interactions in the physical space and in the social network space in a city are mutually following each other and may reinforce each other. It informs future urban development. Our findings suggest that face-to-face interactions are natural results of social links, which supports the importance of social connections in urban life as elaborated by Gehl's book *Life between buildings* (Gehl 1987). This provides important implications to the issue of urban sprawl. According to the scaling law in cities (Bettencourt 2013), a more sprawled city definitely demands more urban infrastructures, and thus requires higher costs to maintain social networks.

Dataset and Methods

Study Area and Data Preprocessing

A case study is performed to illustrate the idea and to test the hypothesized association. The study area is the metropolitan part of the city of Jiamusi (marked by the red circle), a prefectural level city located in Heilongjiang province bordering Russia in northeastern corner of China. The red circle marks the location of Jiamusi (Figure 2.1). The city sits along the Songhua River, and is well connected via railway, waterway, and highway with Harbin, the capital city of Heilongjiang Province. Our study area covers the metropolitan area, including Xiangyang District, Yonghong District (merged to Suburban in 2006), Qianjin District, and Dongfeng District. The city has 2,552,097 people in total (2010), and 850,750 in the built-up area of the metropolitan Jiamusi (2010). The city serves as a producer of wood pulp and newsprint as well as an international trade harbor near the China-Russia border. Because of its advantageous location, Jiamusi is the economic center of eastern Heilongjiang Province.

Figure 2.1 Location of the city of Jiamusi in Heilongjiang. Adopted from (Ctrip). The map of China does not show the islands in the South China Sea.

On the map of the metropolitan Jiamusi (Figure 2.2) shows the city's resource distribution and land use by overlapping the satellite image of the city with the boundary of the built-up area, the mark of downtown, the locations of POIs, population density distribution (LandScan^{TM} data (©Oak Ridge National Laboratory)), and the major transportation infrastructures, e.g., a railway station, and two airports. The downtown area, marked by the light-blue star, is located at about the geometric center of the built-up area delineated by the white built-up area boundary. The population density (displayed by the grid data) in the city center can be as high as about 27 thousands of people per square kilometers, which is much higher (red color on Figure 2.2) than the surrounding suburban and rural areas (blue color on Figure 2.2). Such high density is possible in China, especially on university campus. To approximate the distribution of city infrastructure and resources, we use the same set of points of interest (POIs, yellow dots on the map) as (Wang et al.), including services, transport, and recreational areas. POIs are much denser near the city center, signifying the high attraction of this area, while become more dispersed towards periphery. Such pattern is highly likely to influence people's activity spaces and social network distribution. The city demonstrates a homocentric pattern in terms of the distributions of both population and POIs. Regarding transportation infrastructure, there is a railway station on the southeast side of the downtown area, and two airports on east and west side of the city, respectively. The airport on the west is named Jiaxi Airport, which is an airport specifically targeted to farmland usage (for example, spreading seeds), while the one on the eastern side is a normal airport for the public.

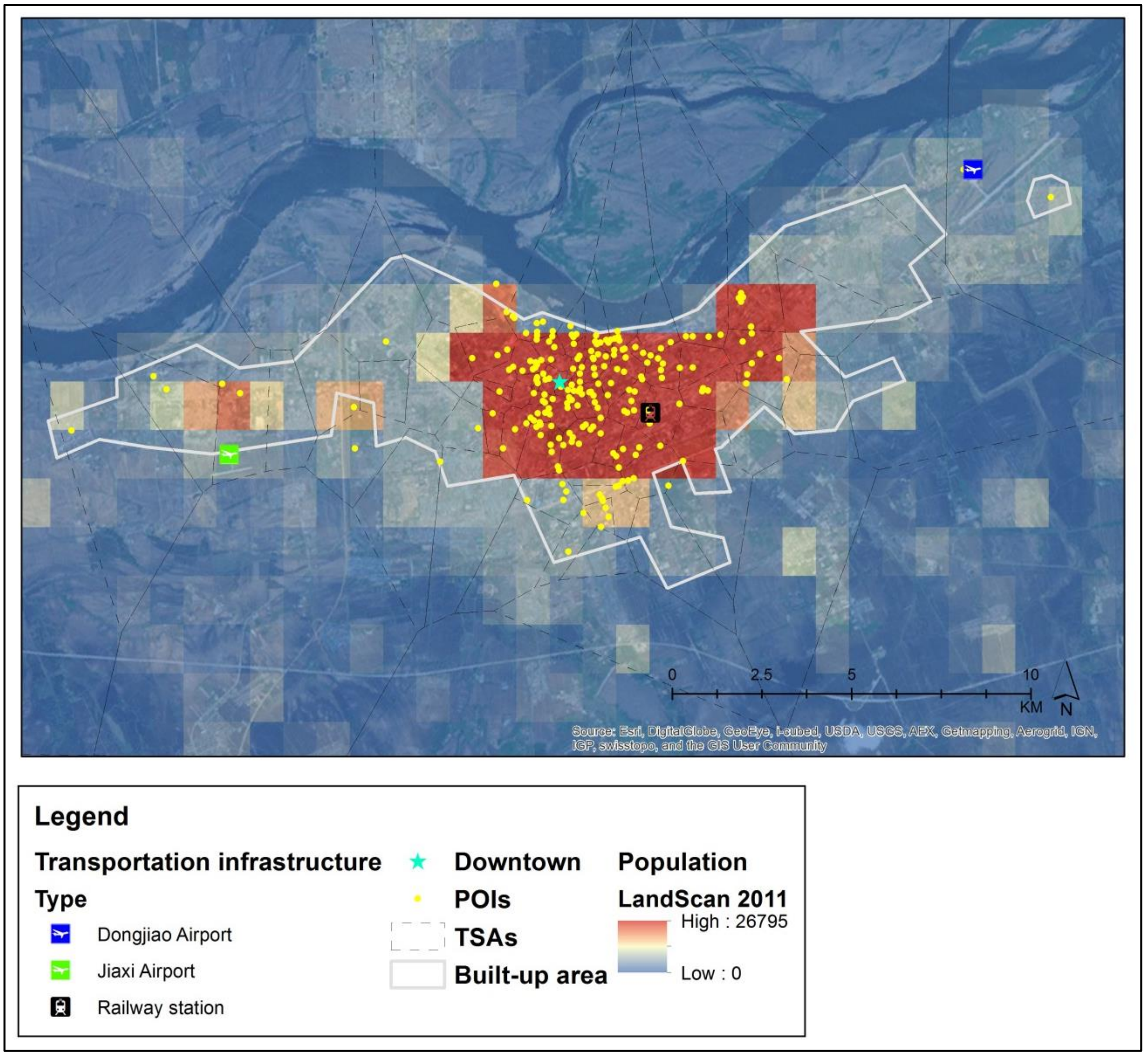


Figure 2.2 Map of metropolitan Jiamusi with population distribution and POIs.

The research dataset is a set of anonymized cell phone call records from one telecommunication company. It contains roughly 1.6 million call records per day times 31 days during an undisclosed month. The total number of cell phone users during the month is 324,752, about 38.2% of the built-up area population. There are 96 cell phone towers (represented by red dots in Figure 2.3) in the study area, all of which are located on the south bank of Songhua River without physical barrier between them. The towers are denser near the city center while more distributed in suburban area. Since the range served by a particular tower may vary because of the limit of call volume load and other technical reasons, we utilize Voronoi diagrams to approximate the tower service area (TSA) covered by each tower (Figure 2.3). But there is possibility that a person in the TSA of one tower is drawn to another tower.

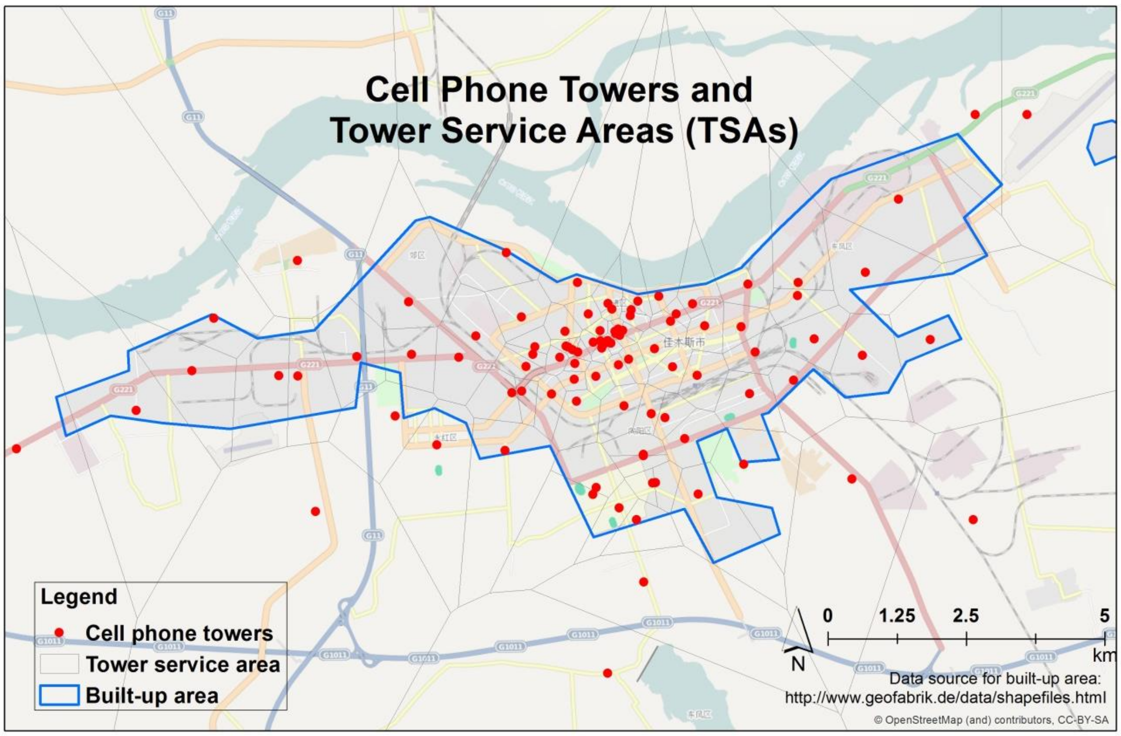


Figure 2.3 The cell phone towers with corresponding TSAs in Jiamusi, China.

In this study, we concentrate on two types of networks, social network and movement network, both of which are weighted network. Social network treats each cell phone user as a node, and the existence of calls as link. The weight of each link is the number of calls during the time period under study. In the caller's movement network, each node is a cell phone tower which represents the corresponding TSA. A link between two nodes is weighted by the movement flows between the pair of TSAs. The complete raw data are pre-processed and prepared in two forms, *Aggregation 1* as social network and *Aggregation 2* as human movement trajectory (Table 2.1). *Aggregation 1* is extracted by ranking *Caller* and *Receiver IDs*, aggregating by the same *Caller* and *Receiver*, counting the number of records of the same *Caller* and *Receiver* as *Count of Calls*, and summing up the corresponding call durations as *Total Duration*. *Aggregation 2* is drawn out by copying each call record to both *Caller* and *Receiver* (collectively termed users hereafter) and ranking records by *UserID* while keeping all the other information as the original.

Table 2.1 Data structure of the raw data and processed data

<table>
<tr><th>Original Table</th><th>Caller</th><th>Receiver</th><th>Caller Location (x, y)</th><th>Receiver Location (x, y)</th><th>Start time</th><th>Duration</th></tr>
<tr><td>Aggregation 1</td><td>User1</td><td>User2</td><td>-</td><td>-</td><td>Count of Calls</td><td>Total Duration</td></tr>
<tr><td>Aggregation 2</td><td colspan="2">UserID</td><td colspan="2">Location (x, y)</td><td>Start time</td><td>Duration</td></tr>
</table>

Because of the impossibility to process the huge size of the entire dataset and the issues of data copyright as well as confidentiality, we are allowed to draw out a randomly selected sample dataset by a hierarchical method. We initially select a random sample of 150 users, and then draw out their friends (1^{st} degree ties), their friends' friends (2^{nd} degree ties), and their friends' friends' friends (3^{rd} degree ties), in total four layers. The friendship in this case study is defined as the following: between any two users, if the count of calls between any two users is no less than 10 or the duration of call is no less than 10 minutes, the two users are connected by friendship. Those "friends" who do not meet such requirement will be left out. We also set a criterion to ensure users are active so that we can extract the database of physical movements. This criterion requires that each user must show up at least 3 cell phone towers during the whole month. The data selection yields a well-connected network with 8,231 users and 47,297 friendship links. The first 150 users produced 460 first degree ties, 2,242 second degree ties, and 11,316 third degree ties. On average each sampled user has 44 friends, 328 calls and 6.18 hours' call during the month.

The original dataset is organized day by day, which allows us to analyze the temporal changes of the detected communities by day of the week and find contiguous patterns throughout the week. After sampling the 8231 cell phone users, we extract social network and spatial movement flow data day by day within a full week (from Sunday to Saturday). Further analyses are performed on each of the seven sets of data as well as the comparisons among them. We need to clarify three things. Firstly, not all the users sampled from the entire dataset during the whole month make phone calls on each day of the selected week. The final dataset of the seven days is a subset of the whole month, with 7,906 users (about 2.4% of the total cell phone users during the whole month) and 730,620 calls (i.e., sum of link weights). We name the ones who do make calls "active users" (Table 2.2). The temporal dynamics varying from weekends to weekdays are noticeable. On *D1* and *D7*, less people make calls than on weekdays. On Sunday (*D1*), each person makes fewer calls on average. People generally make 3.36 calls per user per day on weekdays, but slightly more calls on Saturday. It is possible that Saturday is usually a time to hang out with friends when people rely on calls to negotiate time and find others, while Sunday is a day for rest and religion ritual when calls are less necessary. Secondly, for social network, although the link weights are computed by day, the degrees (i.e., the total number of calls from or to a user) of the sampled users are calculated based upon the records within the whole week. We think this method makes more sense to represent a robust social network structure that is relatively stable within a week. This is also to solve the problem of data scarcity on some days (many of the sampled users have no calls on one specific day) and to smooth extreme cases (e.g., on a special day one person may have significantly more calls than usual). Finally, for similar reasons, the distance decay function in the urban movement community detection (UMCD) is approximated by fitting the inter-TSA flows of all the seven days.

Table 2.2 Number of active users and average number of calls per user on each of the selected seven days

Day	***D1***	***D2***	***D3***	***D4***	***D5***	***D6***	***D7***
#ActiveUsers	6448	6789	6733	6732	6807	6846	6703
Avg. #calls/user	2.32	3.37	3.35	3.40	3.39	3.37	3.40

Before generating the urban movement network, we simplify the trajectories (i.e., a series of visited TSAs by each sampled user in time sequence) of trips by the following way. Based upon the size of the study area and the average travel situation, we choose the time threshold of thirty (30) minutes to filter data. Any two successive trips with a time gap shorter than 30 minutes should be merged, and thus the intermediate stops are removed. Given that we care about the inter-TSA movement, we exclude intra-TSA movements (i.e., two calls happened consecutively in the same TSA). This is done for each of the seven days.

Modified Modularity and the Greedy Technique for Community Detection

Previous studies (Liben-Nowell and Novak 2005; Watts, Dodds, and Newman 2002; Guimerà et al. 2005; Thiemann et al. 2010; Ratti et al. 2010; Grady et al. 2012; Arenas and Diaz-Guilera 2007) have found that network innately have clusters, i.e., high degree nodes attract new nodes more rapidly through evolution. Consequently, even a random network is not evenly distributed. Modularity (M E J Newman and Girvan 2004) Q was initially defined as a measure of the quality or strength of a detected community structure (eq. (1)), but later on has been directly used to detect significant community structure to show clusters caused by meaningful reasons other than random clusters in networks (M E J Newman 2006). The measure can be explained as the residual weights of edges after taking the expected weights from the observed.

$$Q = \sum_i \left(e_{ii} - a_i^2\right) = Tr\mathbf{e} - \| \mathbf{e}^2 \| \text{, where } a_i = \sum_j e_{ij} \qquad (1)$$

In equation (1), the total fraction of (weighted) links with both ends in the same community i is denoted as e_{ii}, and the square of the total quantity of (weighted) links with at least one end in that community is a_i^2. In a broad sense, modularity indicates the deviation of the observed fraction of (weighted) edges within each community (denoted by e_{ii}) from the expected fraction of (weighted) edges in the same community (represented by a_i^2). The formula that quantifies the expected fraction is called null model. By comparing against the null model, modularity discloses a significant community structure caused by other reasons instead of by the intrinsic random property of a network (Danon et al. 2005).

Our modularity calculation is modified from Newman and Girvan's model. The general form of our modified formula of modularity is defined by equation (2), where $e_{m,n}$ is the link weight between node *m* and *n* that both belong to the same community *i*, and *R*, *E* represent *real* connections and *expected* connection between the two nodes, respectively. This function is basically a multiplication-division alternative to Newman and Girvan's. We propose to use ratio, i.e., division, between *real* and *expected* edge weights to remove the potential influence of the community size. When considering two communities, *A* and *B* (degrees and connections shown in Table 2.3), the larger community with more nodes (i.e., *A* in this case) is more likely to yield bigger difference (e_{ii} - a_i^2), which is very likely to be caused by having more nodes in community *A* than *B* instead of stronger interplay between *A* and *B*. This change is mathematically denoted as each individual term in equation (2), i.e., the part within the bracket. Secondly, community detection is aimed not only 1) to group nodes with strong interactions together, but also 2) to partition weak connections into different groups. We hence use multiplication upon the terms to get the final modularity, instead of adding the ratios up, because addition is not sensitive to small ratios and thus leads to the failure of aim 2) and may contribute to the resolution limit in community detection which results in the misleading merging of smaller communities (Fortunato and Barthélemy 2007). In this study, we also release the constraint that the total of $R(e_{m,n})$ should be equal to that of $E(e_{m,n})$ as suggested by (Reichardt and Bornholdt 2006) and (Expert et al. 2011) to adjust for the resolution limit. Therefore, we do not require to use fraction by normalizing $R(e_{m,n})$ and $E(e_{m,n})$ with their total. We will elaborate later on how to decide the number of communities based upon such revision.

$$Q = \prod_{i} \prod_{m,n \in i} (R(e_{m,n}) / E(e_{m,n})) \qquad (2)$$

Table 2.3 Example of the bias of community size to detection result		
	A	***B***
$\boldsymbol{e_{ii}}$	0.50	0.075
$\boldsymbol{a_i^2}$	0.41	0.042
$\boldsymbol{e_{ii} - a_i^2}$	0.09	0.033
$\boldsymbol{e_{ii} / a_i^2}$	1.22	1.8

In social network, for each community i, $R(e_{m,n})$ stands for the observed social link weights between each particular pair of nodes m and n, and $E(e_{m,n})$ is the quantity for expected social links within that community. The specific formula of social network modularity (Q_{SN}) is shown by equation (3), where $l_{m,n}$ is the number of connections between two egos m and n of the same community i, M is the total weights of links over the whole network, and d_m, d_n are the weighted degree centrality for m and n, respectively.

$$Q_{SN} = \prod_i \prod_{m,n \in i} \left(\frac{l_{m,n}}{d_m d_n \frac{1}{M}} \right) \tag{3}$$

Regarding movement network, $R(e_{m,n})$ is the observed movement flows (from the origin to the destination in a trip) between intra-community TSAs, and $E(e_{m,n})$ is the estimated flow within community i by our theoretical model. The detailed formulas for the movement network will be demonstrated below.

Community detection targets to maximize the value of *Q* for the optimal community partition. Since we revise the modularity into a division-multiplication form, many *Q*-optimization algorithms (Pons and Latapy 2005; M E J Newman 2006; Barthélemy 2011; Mark E. J. Newman 2004), such as spectral method, cannot be used because some properties of the adjacency matrix are no longer kept. We therefore employ the very basic greedy technique that generates a hierarchical clustering dendrogram for detection result and stops when all the connected nodes have been merged to one community. This method may not be safe due to the possibility of local optima (Barthélemy 2011; Guo 2009), but we argue that it works for our study because we utilize hypothesis test to compare results later on.

For each iteration, the two communities (initially isolated nodes) that yield the biggest increase of *Q* will be merged as a new community. Using *Q(n)* to represent the modularity at the *n*-th iteration step, we give the formula of the increase of modularity (λQ) by equation (4), where *Int(n-1)* stands for the set of (weighted) inter-community edges at step *n-1* between the two communities i_1 and i_2 that are merged at step *n*.

$$\lambda Q = \frac{Q(n)}{Q(n-1)} = \prod_{e_{m,n} \in Int(n-1)} (R(e_{m,n}) / E(e_{m,n})) \tag{4}$$

Distance Decay in Urban Movement Community Detection (UMCD)

While the original modularity by Newman and Girvan (M E J Newman and Girvan 2004) makes sense in an abstract topological graph in which the expected flow between two nodes are proportional to the product of their degrees, it overlooks the spatial properties of the system that flows between near places are usually stronger than between remote ones. As a result, the identified community structure is under the mixed influences of distance decay and all other factors, which makes no surprise to find spatial contiguity in the detected communities even without superimposed requirement of adjacency. Expert et al (Expert et al. 2011) tried to deal with the shortcoming by *explicitly* introducing distance decay function in the calculation of expected values so that the resulting community structure shows the influence from other non-spatial factors. Accordingly, we argue that the traditional modularity of the popular Newman and Girvan model is not sufficient to depict the urban movement network. It not only lacks the geographical meaning of the weight of node (i.e., geographic entities such as a TSA, a city, or a state), but also absences the consideration of distance decay. Node weight in a geographical context usually represents the area or population of a place. For example, Guo (Guo 2009) did community detection in US immigration flow network by involving population of each state into his null model. In this work, we assume that the movement flow between two TSAs is proportional to the product of their populations, inspired by the gravity model. The population is approximated by the LandScan™ data (©Oak Ridge National Laboratory) because we do not have access to the accurate census data at present. Regarding distance decay, Expert et al. propose a method that involves decay model into modularity (eq. (5)), where the distance decay function $f(d_{ij})$ is given by equation (6).

$$Q_{Spa} = \frac{1}{2m} \sum_{C \in P} \sum_{i,j \in C} \left(A_{ij} - N_i N_j f\left(d_{ij}\right) \right) \tag{5}$$

$$f\left(d_{ij}\right) = \frac{\sum_{i,j|d_{ij}=d} A_{ij}}{\sum_{i,j|d_{ij}=d} N_i N_j} \tag{6}$$

Here N_i, N_j denote the population of place/community i and j, respectively, and d is the center value of the distance bin $[d - \Delta, d + \Delta]$ (Supporting Information I). Similar to our social network modularity, we transform the addition-subtraction form to a multiplication-division form and finally get equation (7), where $f(d_{ij})$ is also denoted by equation (6). Q_{UM} means the modularity for urban movements.

$$Q_{UM} = \prod_{C \in P} \prod_{i,j \in C} \frac{A_{ij}}{N_i N_j f\left(d_{ij}\right)} \qquad (7)$$

In this study, we also discard the constraint of spatial adjacency following some previous studies (e.g., (Liu et al. 2014; Ratti et al. 2010; Gao, Liu, et al. 2013)), because enforcing such a requirement on detection process may hide intrinsic patterns of places that are really closely related.

It is worth attention that our aim is to remove spatial clustering from the *movement* network, which is different from Expert et al. who include distance decay in *social* network. Because what is put into the null model is excluded from the observed connections, we remove the relatively frequent short urban trips to find the hidden community structure formed by relatively weak *long*-distance trips. We introduce the decay only in the network of urban movements, and claim that, in our social network, no geographic locations are related to the cell phone users at the stage of community detection.

The distance decay function, different from the traditional method of fitting a gravity model with parameters, is approximated by inversion from empirical data. We adopt such method for three reasons (see Chapter 3 Supporting Information I). Firstly, we have proved that equation (6) is distance decay in nature. Such proof guarantees that the function reserves the gist of distance decay. Secondly, equation (6) makes sense from a technical perspective. In fact, equation (6) can be treated as a distance decay model by aggregating individual plots into bins before data fitting. Even if we adopted the traditional method to fit a gravity model in a normal formula with parameters, we still need to aggregate data. Thirdly, equation (6) is more robust for empirical data, because we do not really know what function the real data follows. The function of the empirical data might not be a perfect gravity model such as power law or exponential curve as is usually utilized; it might be an irregular step-function, which cannot yield a good distance decay coefficient for the whole range of data. Therefore, the formula of distance decay should be a deterrence function $f(d_{ij})$ to be fitted with the empirical data. The fitted distance decay function using equation (6) is shown in Figure 2.4. Interestingly, there is a minor peak at about 14 kilometers. We speculate that it is caused by the commutation between the city center and the airport, since the distances between the centroids of the TSAs in the city center and of the TSA where the eastern side public airport is located are roughly 14 kilometers. This minor peak substantiates that the distance decay function is not a regular one that can be well depicted by a perfect mathematical formula, and thus supports our selection of equation (6).

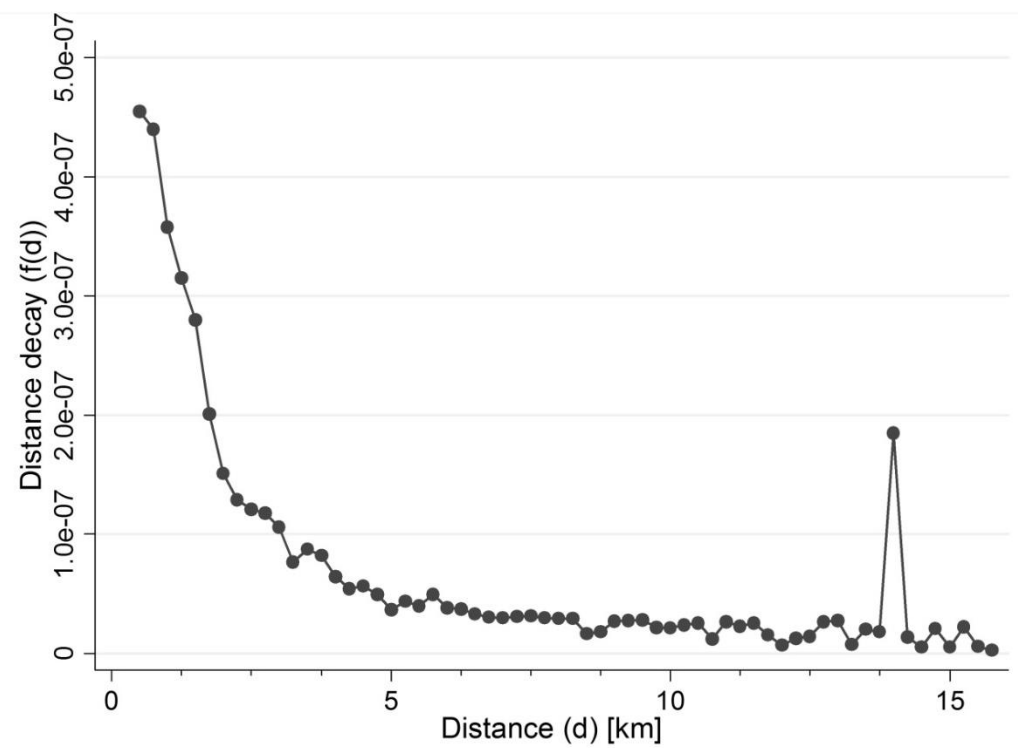


Figure 2.4 Distance decay of the spatial movement (without intra-TSA flow).

Mapping Social Network Communities into Urban Space

In our research design, we carry out social network community detection (SNCD) in parallel with UMCD, but in a virtual social space without any geographical location information. This method reverses from Expert et al.'s method to Newman and Girvan's traditional way for SNCD, despite the changed form from subtraction to division. It may be perceived initially that our goal is similar as Expert et al's (Expert et al. 2011). However, we do not intend to remove the role of distance from social network communities for two reasons. For one thing, we want to understand whether the social network matches urban movement pattern. If we eliminate spatial factor from social network, we may not be able to observe the underlying association. For another, Expert et al's method (Expert et al. 2011) is not adoptable in our work because we conduct SNCD on mobile individuals (i.e., each node in the social network represents a user), while social network in (Expert et al. 2011) is on an aggregated level (i.e., a node stands for a city made up of all users in the city). In this study, each user in a city can have multiple frequently visited places, so it is difficult to tie a user to only one geographic location.

However, as we intend to compare the detected patterns from SNCD and UMCD, we still need to map SNCD results into urban space. We therefore rely on anchor point (AP), defined as the mostly visited places in a person's daily activity space (Huang et al. 2010), to solve this problem. Cell phone users expose themselves in urban space by making calls that are transmitted via the geo-located cell phone towers, so we use the top three mostly visited cell phone towers by a particular user as his/her APs. We set the threshold relative to each of the users because some of them are active who make many calls while others have much less locations exposed by cell phone.

The challenge is that we cannot simply assign the towers to the community each user belongs to, because a tower can be frequently visited by many users who belong to different communities. To handle this problem, we create a weighted co-community matrix (***WCCM***) for the AP towers, which is an AP similarity matrix. Then a hierarchical clustering method utilizes this matrix to complete its process. The basic idea is to cluster AP towers according to how closely they are related with each other in the social network. The ***WCCM*** is built by two steps: 1) to measure how closely an AP AP_x is associated with a social network community i (denoted by $w(AP_x \rightarrow i)$ in eq. (8)), and 2) to calculate the probability that a pair of APs, AP_x and AP_y, belong to the same community (eq. (9)). $WCCM_{x,y}$ is the cell value of *(x,y)* in the ***WCCM***.

$$w(AP_x \rightarrow i) = \sum_{u \in i} \overline{\lambda Q}_u * p(AP_{ux}) \tag{8}$$

$$WCCM_{x,y} = \sum_{i} w(AP_x \rightarrow i) * w(AP_y \rightarrow i) \tag{9}$$

To calculate equation (8), we adopt two criteria that form a community-user-AP chain: 1) the AP-user link is the visiting frequency of a specific user (u) to one of his/her APs; and 2) the user-community link is the extent to which a user belongs to the assigned social network community. We represent the user-community link by the average contribution a link makes to the increase of modularity, denoted as $\overline{\lambda Q}_u$, where u stands for one of the users in the community i who was an isolated node before joining this community. It is calculated by dividing λQ (eq. (4)) by the count of links (i.e., the cardinality of set *Int(n-1)* denoted by |*Int(n-1)*|) between the two merged communities (including one-node community). Although in its calculation, all the inter-community edges are involved, $\overline{\lambda Q}_u$ is only assigned to a node u when the node is merged into a community for the first time, even if the community is merged to another community later on. This is because only newly added edges will affect the increase of modularity. The second term in equation (8), $p(AP_{ux})$, quantifies a user (u)'s visiting frequency to a particular AP AP_x normalized by the frequency to all the towers visited by this user. While equation (8) illustrates how a single AP is related to a community, equation (9) quantifies the co-occurrence of two APs in the same community. As a result, the hierarchical clustering assigns those cell phone towers into the same number of communities per day as that of the urban movement communities.

Decision on the Number of Detected Communities

Since the resulted community is a hierarchical clustering dendrogram, the number of detected communities is flexible depending on which level to cut the dendrogram (M E J Newman and Girvan 2004). Although sometimes we can decide the number of communities subjectively, we adopt an objective method to approach an optimized community structure. Modularity-based community detection is aimed to maximize the value of modularity, so the greedy technique each time merges two nodes/communities that yield the biggest λQ (eq. (4)). Ideally, the best community structure is reached when modularity reaches its peak value. Mathematically, it can be derived that modularity reaches its top value exactly when the increasing rate λQ is larger than or equal to one for the last time in the iterative process. After this point, all the rest increasing rates of modularity become smaller than one (Figure 3.1 in Supporting Information II).

This method works well for the decision of the number of urban movement communities during which we find a peak value of Q (Figure 3.1), but fails to do so for the social communities for which we do not know where the peak value of Q *really* is. The final step of merging in the social network yields a single community while the λQ is still greater than one (Figure 3.2). If we followed our original decision principle, the social community result would be a meaningless single community losing all its inner structure. Nevertheless, we find that many merging steps before the final step only yield a small value of λQ, which do not increase the Q significantly (Figure 3.2). We therefore propose to use the number of communities at the merging step from which on all the λQs are smaller than the lower bound of the 95% confidence interval of λQs. Despite the outlier in the list of λQs, this approach cuts off those merging steps that cannot yield significant increase of modularity (see Supporting Information II). Listed in Table 2.4 is the final result of the numbers of detected communities for each day in social network and urban movement network. The number of communities is derived by subtracting the number of merging steps from the total number of nodes (users or towers) in each network. Note that the total number of nodes in the social network varies for each day.

Table 2.4 The numbers of detected communities for each day

1) # social communities							
Day	***D1***	***D2***	***D3***	***D4***	***D5***	***D6***	***D7***
#Communities	438	1815	2521	2858	2527	2059	29
2) # movement communities							
Day	***D1***	***D2***	***D3***	***D4***	***D5***	***D6***	***D7***
#Communities	38	35	35	37	39	40	41

Similarity between SNCD and UMCD Structures and Hypothesis Test

To judge whether the SNCD pattern is similar to the UMCD pattern, we define a measure called similarity coefficient (*SC*) as equation (10), which shares a similar idea to the coefficient of areal correspondence (Taylor 1977). Here, $SNCD_i$ is the set of the elements in the *i*-th SNCD community and *j* for UMCD, and *N* is the total number of communities for a specific day. *SC* is the sum of the relative quantities of shared elements by two communities. The relative quantity is the normalized ratio of the number of shared elements (i.e., individual towers) in each pair of the SNCD and UMCD communities to the number of the elements in the union of the two sets. As the *SC* denotes, the more similar the two patterns, the higher the ratio.

$$SC = \frac{1}{N(N-1)} \sum_{i}^{N} \sum_{j,j \neq i}^{N} \frac{\left| SNCD_i \cap UMCD_j \right|}{\left| SNCD_i \cup UMCD_j \right|} \qquad (10)$$

To guarantee that the similarity between SNCD and UMCD is significant against randomness, we conduct a t-test on *SC*s based upon 100 times' permutations of social network. Each permutation is performed by randomly assigning social network connection (link) as well as connection strength (link weight) between cell phone users while keeping the users' original APs intact. We then map the social network communities into space via the APs, and calculate the *RC*. We repeat such process for each of the seven days under study. By keeping the original anchor points for each user in the random permutations, we avoid the bias caused by tower usage frequency (i.e., some towers are more intensely used than the others), and ensure that social network structure is the only factor that makes a difference. The random permutations form a sample distribution, to which the *SC* between the observed social network communities and urban movement communities will be compared. The null hypothesis is that there is no significant similarity between the spatial pattern of the observed social network communities and that of the urban movement communities. But if the *SC* from the observed case is significantly larger than the average *SC* from the random permutations, it signifies a significant similarity in the community pattern of social network and urban movements.

Results and Discussions

The Association between Social Network Communities and Urban Movement Communities

The detected results from SNCD and UMCD are shown in Figure 2.5. For social network (column B is the result from the observed social network and C is that from one of the randomly permuted social networks), people are more likely to call others whose calls are covered by the TSAs of the same color, though not necessarily the same TSA. Regarding urban movements (column [A]), people generally have higher probability to travel to the TSAs of the same color than the chance quantified by distance decay model. We call the occurrence of such movements "long trips", meaning the trips longer than the constraint by distance.

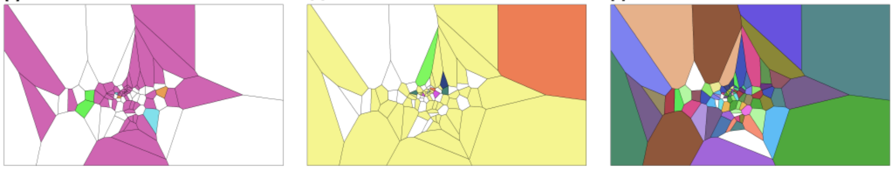
Why is there stronger movement attraction among certain TSAs? Does it have anything to do with social network? We confirm the association not only by the visually apparent similarity between the spatial pattern of long trips and that of the real social network against the result from the random social network (Figure 2.5), but also by the quantitative result from the hypothesis test. The two-tail t-test result is displayed in Table 2.5 to measure the similarity. For each of the seven days, *SC* from the observed SNCD and UMCD (*oSC*) is significantly larger than the randomly generated *SC* sample distribution (*rSC*).

Since *SC* plays a similar function as correlation coefficient, our result confirms that social network is associated with (i.e., shows a similar community pattern as) *long-distance* urban movement network, which indicates that social and geographical behaviors are mutually affected. When somebody develops a social link, no matter by a physical or virtual means, for example, an on-site interview or an online admission to a school, the link then becomes an "attraction" for the person's behaviors if finally she is accepted by the employer or school. On the other hand, a visit to some place exposes a person to more potential opportunities for building new social connections. Although it is hard to tell which is the cause and which is the result, social network or urban movement, we think the former plays a more active role in our study. The long-distance trips that cost higher time and monetary budget are less likely to take place for no certain reasons, which is different from a local-scale random wandering.

The similarity between social communities and movement communities confirms our assumption that the social network contributes to, and gives implications on the generation of long trips. If telecommunication constructed a virtual social space sufficient to substitute for physical interactions, we would possibly see people develop social connections everywhere though may have never personally been there. However, the social network under our observation is not randomly diffused in space. People visit where they know somebody, particularly when they spend more time and money to make a trip longer than the usual case quantified by the distance decay model.

Day 1 Pattern

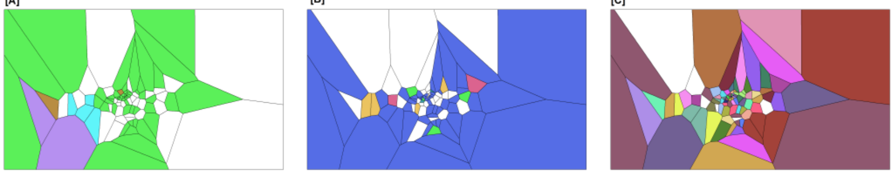
[A] [B] [C]



Colors are randomly assigned;
the same color stands for the same community.

Day 2 Pattern

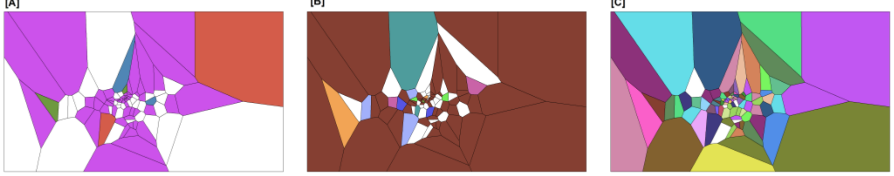
[A] [B] [C]



Colors are randomly assigned;
the same color stands for the same community.

Day 3 Pattern

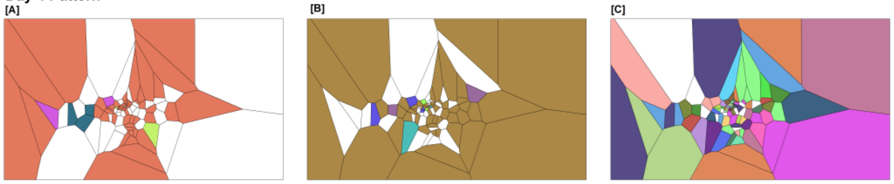
[A] [B] [C]



Colors are randomly assigned;
the same color stands for the same community.

Day 4 Pattern

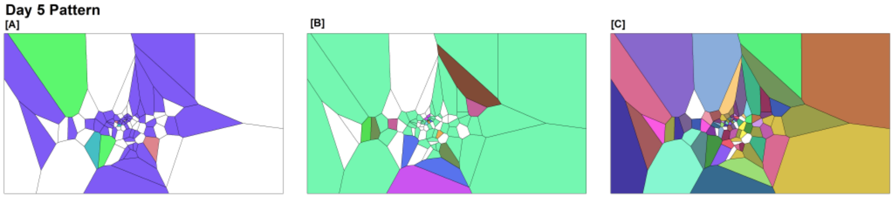
[A] [B] [C]



Colors are randomly assigned;
the same color stands for the same community.

Day 5 Pattern

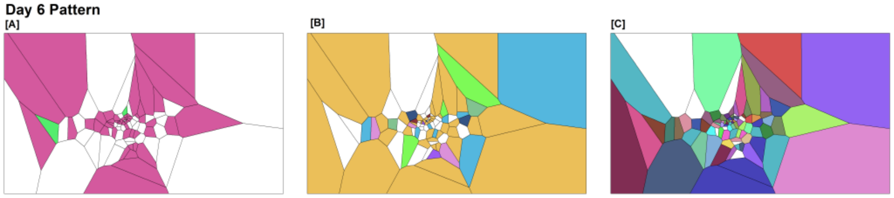
[A] [B] [C]



Colors are randomly assigned;
the same color stands for the same community.

Day 6 Pattern

[A] [B] [C]



Colors are randomly assigned;
the same color stands for the same community.

Day 7 Pattern

[A] [B] [C]

Colors are randomly assigned;
the same color stands for the same community.

Figure 2.5 Detected social network and urban movement communities for each of the seven days. [A] Detected urban movement communities: the pattern illustrates communities caused especially by long-distance travel; [B] Detected social network communities based upon the observed social network; [C] Detected social network communities based upon one of the simulated social networks. All the TSAs with white color are assigned to an isolated community that contains only the TSA itself. All the TSAs with the same color other than white are assigned to the same community with multiple TSAs.

Table 2.5 t-test (two-tail) result for the *similarity coefficient* (*SC*)

(Sample size = 100, degree of freedom = 99)

Day	**Mean of the *rSC*s**	**[95% Conf. Interval of *rSC*]**		***oSC***
D1	0.0154	0.0152	0.0156	0.0335*
D2	0.0168	0.0166	0.0170	0.0356*
D3	0.0166	0.0164	0.0168	0.0351*
D4	0.0157	0.0155	0.0159	0.0306*
D5	0.0150	0.0149	0.0152	0.0262*
D6	0.0143	0.0141	0.0144	0.0252*

D7	0.0139	0.0138	0.0141	0.0259*
*p-value < 0.0001 on a 0.95 confidence level				

The Implicit Nature of Jiamusi and the Linked Activity Spaces (LAS)

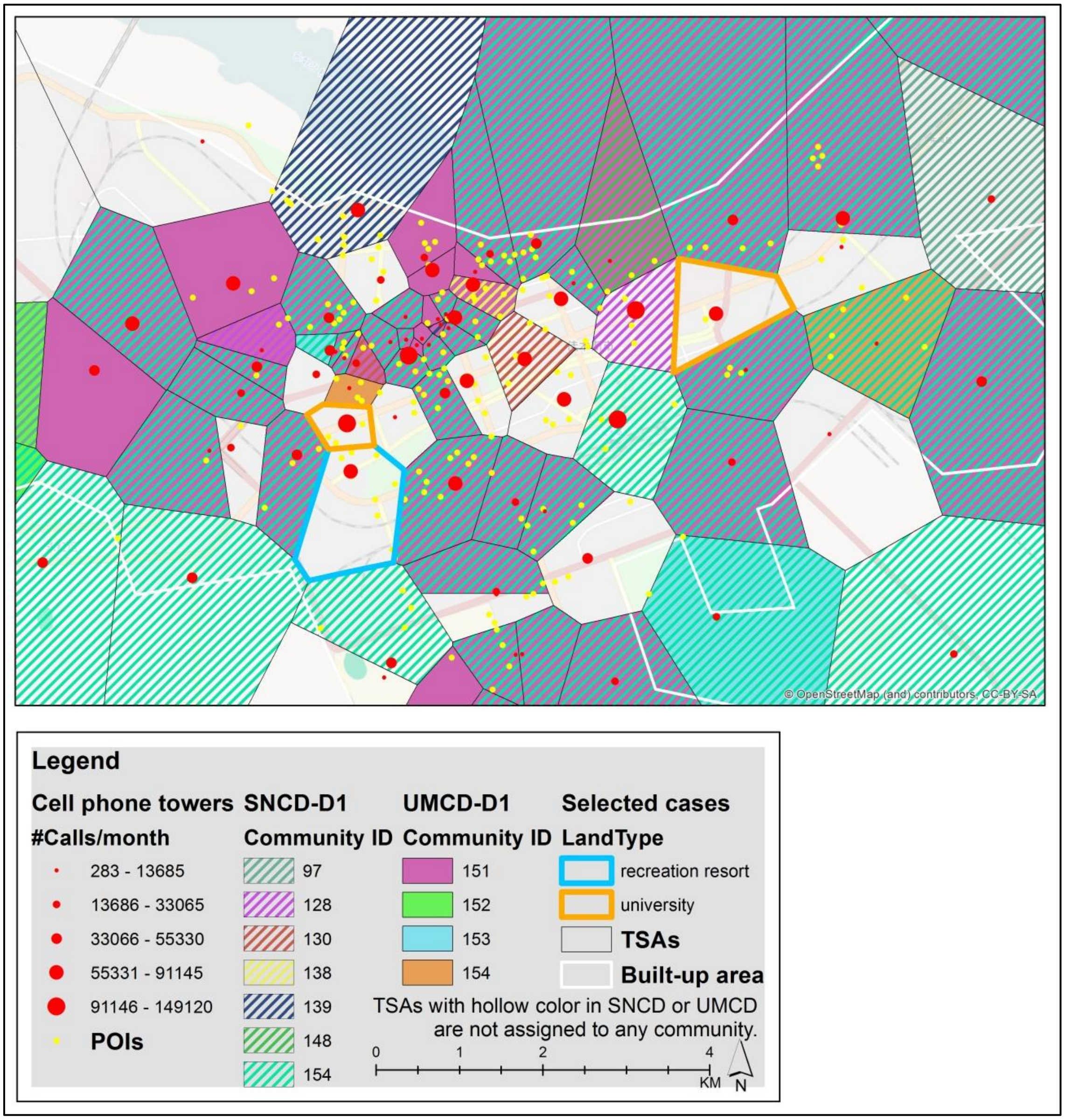
We then look into some details to understand what the association between SNCD and UMCD results can jointly tell the nature of the city. Due to a lack of access to accurate land use and census data in the city, we utilize Google Map (©2014 AutoNavi, Google), ©OpenStreetMap, and the POIs digitized from Google Map by (Wang et al.) to speculate the ground truth. Figure 2.6 overlaps the ©OpenStreetMap, the POIs, the SNCD result, and the UMCD result together. In order to display SNCD and UMCD simultaneously, we render SNCD results with hash-line colors (transparent between lines) and UMCD with solid colors. Different hash line colors represent different SNCD communities, while different solid colors represent different UMCD communities. When the results are overlaid (for each day separately, here we just show *Day 1*), readers can see the color of the underneath layer (i.e., UMCD) through the slots of SNCD results. Stand-alone TSAs in both SNCD and UMCD results are transparent. Though we only illustrate the result on *Day 1*, our analyses are based upon the temporal patterns. We find some interesting results.

Firstly, self-contained TSAs usually suggest some typical land use types where people can either avoid calls or just call people within the same region. By saying “self-contained”, we mean TSAs have stronger inner-region interplay rather than with outside. In Figure 2.6 (a), the SNCD results are colored by hash lines with slots in between so that we can see the colors of UMCD underneath. The visualization method is the same in Figure 2.6 (b) and (c). A TSA with only solid color means it is assigned to a multi-member community (i.e., not stand-alone) in UMCD, but is stand-alone in SNCD; similarly, a TSA with only hash line color (for which the basemap can be seen) is assigned to a multi-member community in SNCD, but is stand-alone in UMCD. Therefore, if a TSA is stand-alone in both communities, its color will be transparent on the map. We just show the data on *Day 1,* but the selected cases (i.e., TSAs) are stand-alone TSAs throughout the seven days, which ensures that they are typical stand-alone ones. Some TSAs are consistently not assigned to other communities (transparent on the map) but instead, stand by themselves in both social and movement communities. We select not only self-contained TSAs but also the ones with high call volume (having larger red dot at the centroid) because we want to understand why the users in those TSAs are active but isolated from the outside. The *selected cases* on Figure 2.6 (a) are some typical examples. The land cover images suggest that the TSA with blue-frame mainly covers a recreation resort, and the TSAs with orange-frame serve universities. People go to the recreation resort together for fun, so in that meantime, people may call to find each other, but have no need to contact the outside. On university campus, students contact their classmates and professors who are mostly on campus as well, so there are no strong connections with outside. The social ties and activity spaces are concentrated and overlapped. The calls to people off-campus are mostly for families in another city, the signals of which are not involved in our dataset. Calls do not happen with places a person never goes and knows nobody, so the border of a university separates people inside and outside.

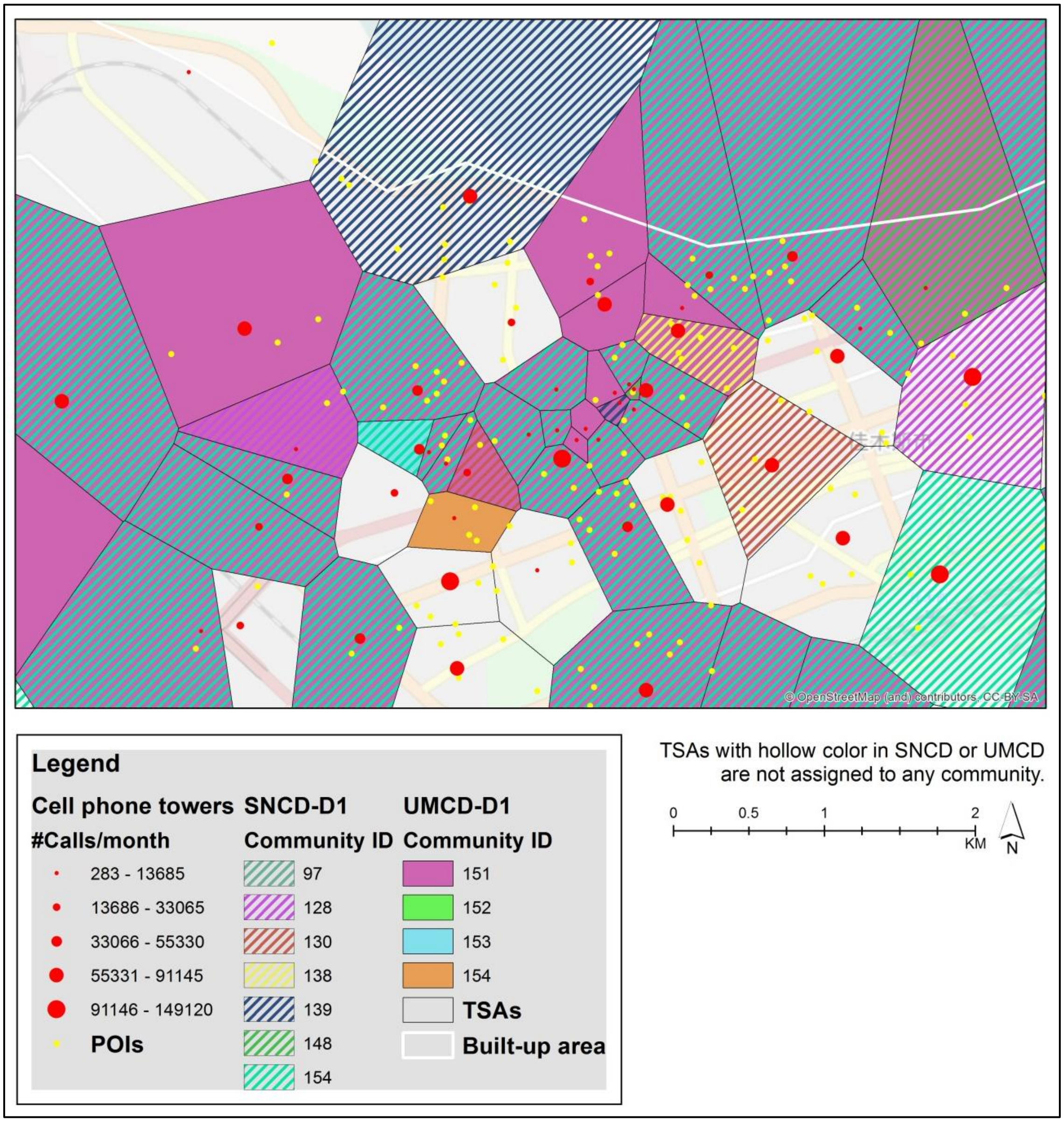
Additionally, when zooming in to the downtown area (Figure 2.6 (b)), urban movements seem cover a wider range than social network. Especially in the dense small TSAs near the center, the call volumes are even quite low (with small red dot tower), which shows that cell phones may not be intensely used at the city center. The continuous magenta color representing the dominant community in UMCD implies people's strong cross-boundary movements in downtown. Referring to *linked activity spaces* (LAS, (Wang et al.)), we would infer that downtown is a place full of resources and infrastructures that make the budget of face-to-face interactions low, which therefore makes virtual social connections less indispensable. Moreover, downtown attracts such a high amount of people that a lower percentage of them are in friendship (Figure 10 and 11 in (Wang et al.), i.e., both friends and strangers enjoy the resource here). Different cliques of social connections will share one TSA, which makes adjacent TSAs be more easily assigned to different communities and thus demonstrates a trivial social community pattern. When resources are so sufficient and convenient for physical connections, not only people in the same social circle, but also people from different groups of social ties will be tied together in geographical space.

On the other hand, if people are located out of reach to each other, social network via telecommunication is a good supplementary. Comparing column [A] and [B] in Figure 2.5, social network communities are more dispersed than urban movements. A TSA on the peripheral area, for example, the one with white-frame shown by Figure 2.6 (c), relies more on virtual telecommunication to be connected with the city. The white-frame TSA is consistently stand-alone in UMCD throughout the seven days, which means its movement interaction with other areas is not very strong, or at least no stronger than expected by the distance decay model. However, this TSA on six out of the seven days is assigned to the dominant social network community containing the majority of TSAs. From Google Map, we find this TSA mainly covers part of a farm and its affiliated hospitals in southeastern Jiamusi metropolitan area. A farmland affiliated hospital located at a relatively remote rural place is apparently not a popular POI to visit for the average public. But the patients at the hospital, mostly the local residents, need cell phones to contact their family members or friends. Telecommunication in this case can be regarded as an extension for face-to-face communication.

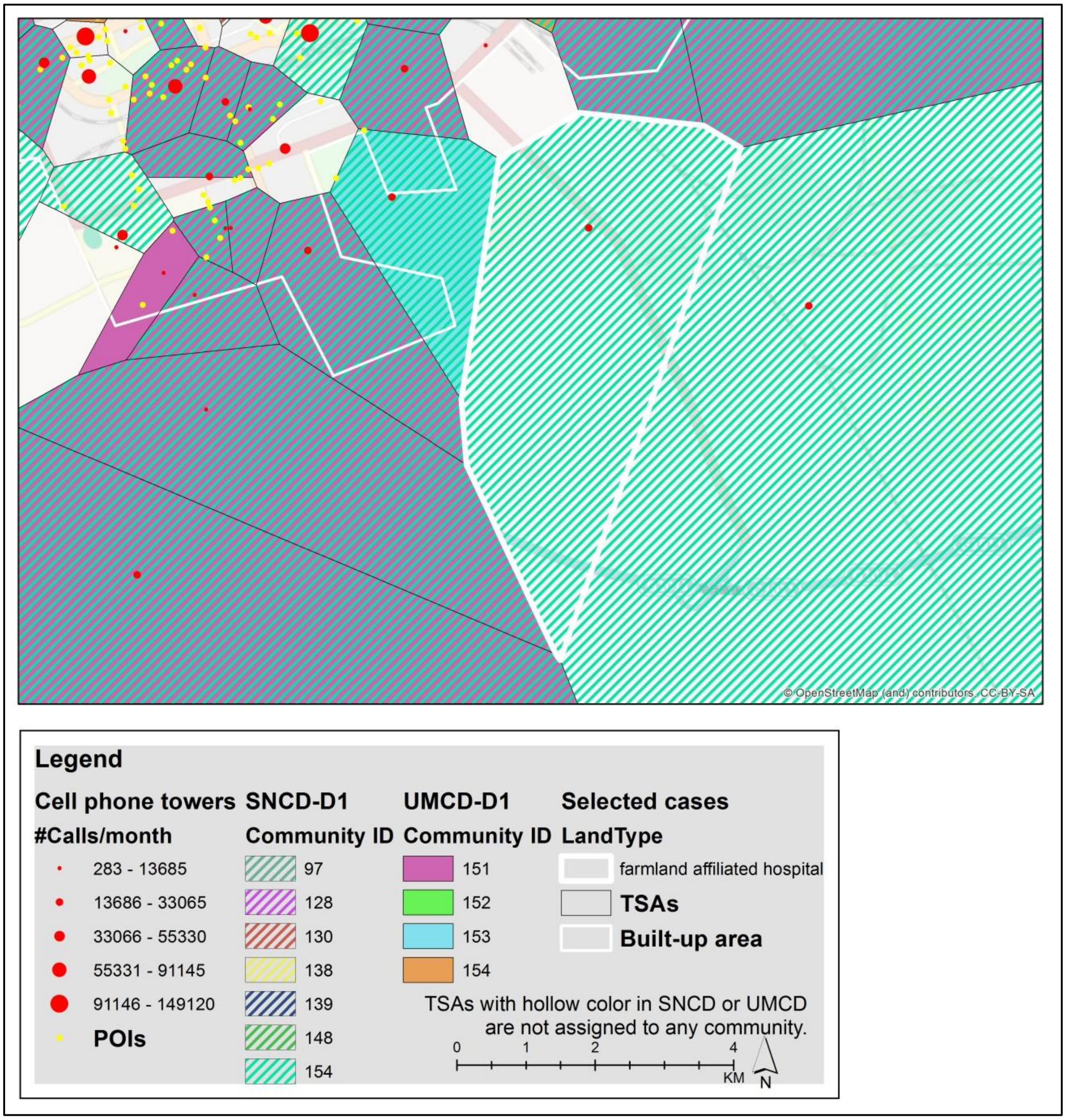
Indicated by the above specific case analysis, physical interactions usually happen when urban resources and infrastructures are well served, i.e., the cost is low and the attraction is high, like in the city center. People won't go to the places where they do not know anybody or have to spend a long time and distance. LAS (Wang et al.) specifies that people's daily activity spaces are linked together via their social connections; friends' movements are more or less overlapped with each other to achieve physical communications, otherwise a purely virtual social tie is unlikely to be preserved.



a. An overlapped map showing the stand-alone TSAs in both SNCD and UMCD results in the built-up area of Jiamusi.



b. An overlapped map showing the SNCD and UMCD results in downtown area on *Day 1*.



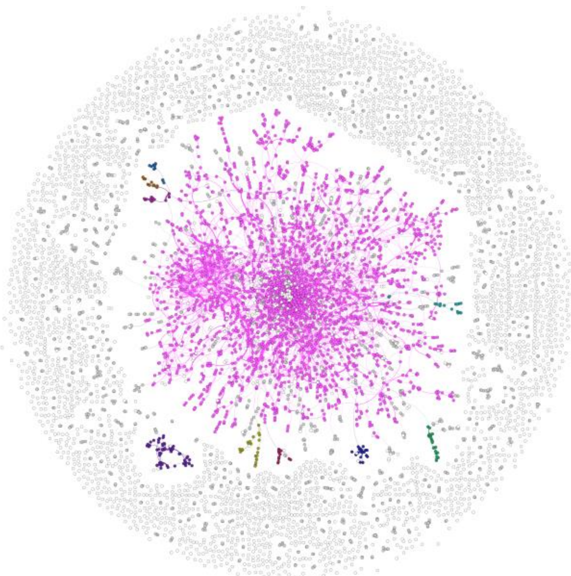
c. The overlapped map of SNCD and UMCD on *Day 1* showing a TSA in suburban area that has strong social network connections with other areas but no strong enough movement flows with the outside.

Figure 2.6 The overlapped map of *Day 1*'s SNCD and UMCD results on the basemap of ©OpenStreetMap with digitized POIs.

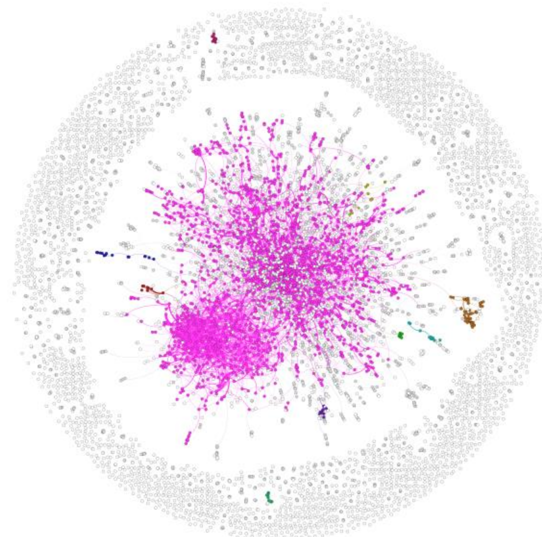
Hints from the Spatial Distribution of Social Network in Jiamusi

To further understand what social network means in urban space, we look at the specific spatial distribution of social clusters detected by SNCD. We visualize the social network structure with nodes (i.e., cell phone user) colored by the assigned SNCD community (Figure 2.7). In Figure 2.7, each different color stands for a different community a user is assigned to. The white ones are stand-alone communities with only one member. The light grey dots represent the trivial communities with a very small number of users (more than one but mostly less than ten). The colorful communities are (top 10) largest communities with multiple members. The pink community in the center represents the dominant community, which is the one with the majority of nodes. We do this for each of the seven days and find some common regulations. First of all, there is a dominant community for each day (e.g., the pink one on each day). Secondly, there are a few minor communities that are loosely connected with or even isolated from the major body, but form a tight clique within themselves (e.g., the dark purple community on the bottom left corner of *Day 1* in Figure 2.7).

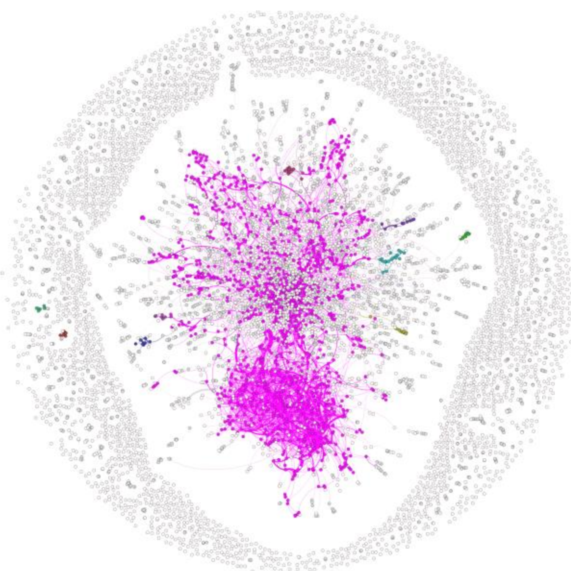
Besides, we also see variations across different days. *Day 1* and *Day 7* (i.e., weekends) both yield a well-connected major community (the pink one in the center) with much larger size than the other trivial ones. But on *Day 2* to *Day 6* (i.e., weekdays), the size differences between the major community in pink and the other minor ones are smaller. This finding corresponds with the fact that the numbers of detected communities on *Day 2* to *Day 6* are much more than on *Day 1* and *Day 7* (Table 2.4 (1)). It literally means people are more connected on weekends. However, given that there are more active users and slightly higher average number of calls per person than Sunday (Table 2.2), the lower connectedness is not a result of fewer calls. It is possible that people make calls to more diverse groups of people, including business and personal calls, which makes the connections within one particular group less intense and the communities more distributed.



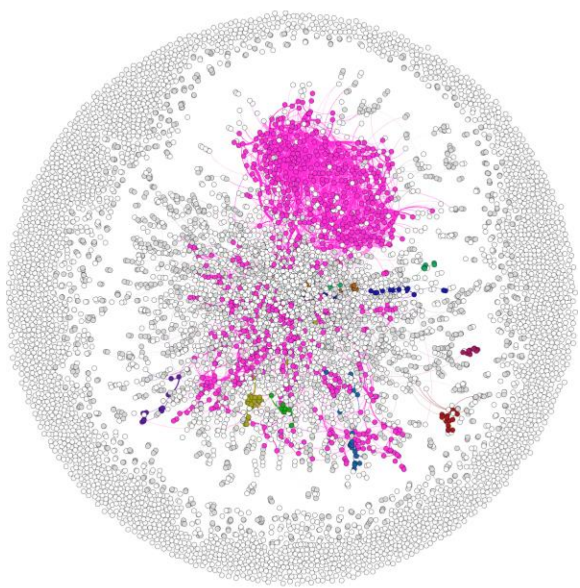
Day 1 (64.0% of active users, i.e.,33.4% of total users in pink)



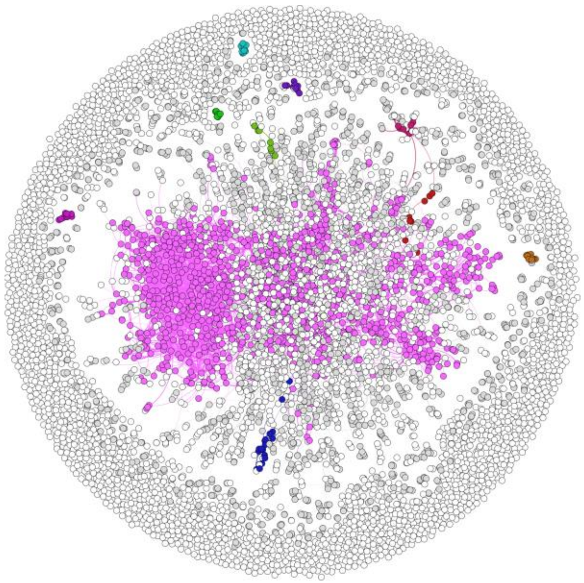
Day 2 (40.9% of active users, i.e., 24.2% of total users in pink)



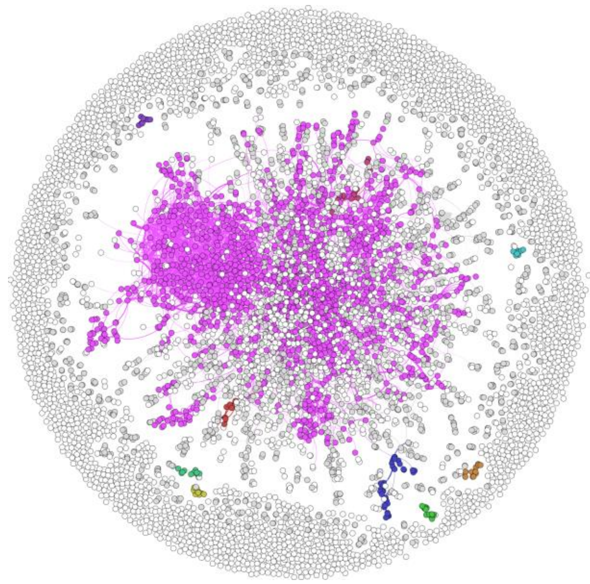
Day 3 (25.8% of active users, i.e., 15.1% of total users in pink)



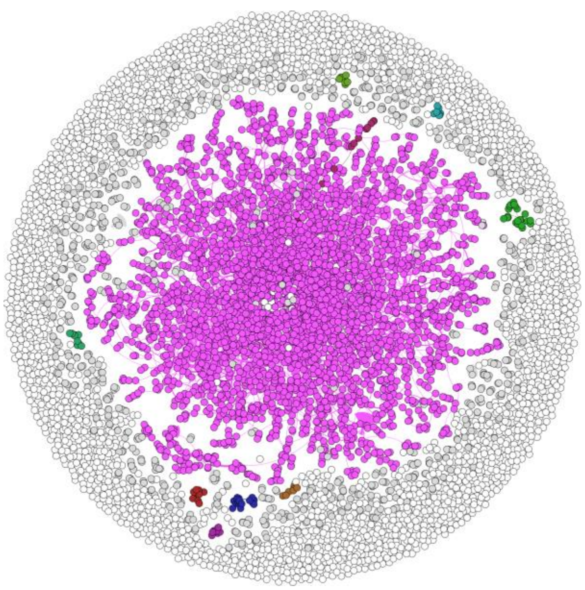
Day 4 (16.2% of active users, i.e., 9.4% of total users in pink)



Day 5 (25.1% of active users, i.e., 14.7% of total users in pink)



Day 6 (34.5% of active users, i.e., 20.8% of total users in pink)



Day 7 (84.0% of active users, i.e., 47.3% of total users in pink)

Figure 2.7 Visualizations for social network on each of the seven days. Made by Gephi 0.8.2.

It is not the social network alone, but its spatial distribution that raises our interest. The major pink community initially draws our attention. By mapping the frequency of each cell phone tower being the AP of the people in the major pink community (*Day 1* in Figure 2.7), we find the APs are mainly clustered in the city center, especially the areas where both POIs and TSAs are very dense (Figure 2.8). It corresponds with the finding by (Wang et al.) that people with higher degree (i.e., number of unique friends) usually take advantage of the city center at which resources are rich and to which accessibility is high. The urban center is hence the most efficient place for people to meet and talk. The pattern is also supposed to correlate with population density, which is also associated with urban resource distribution. In China, a majority of people live in the city center to get better access to resources, which is different from the US.

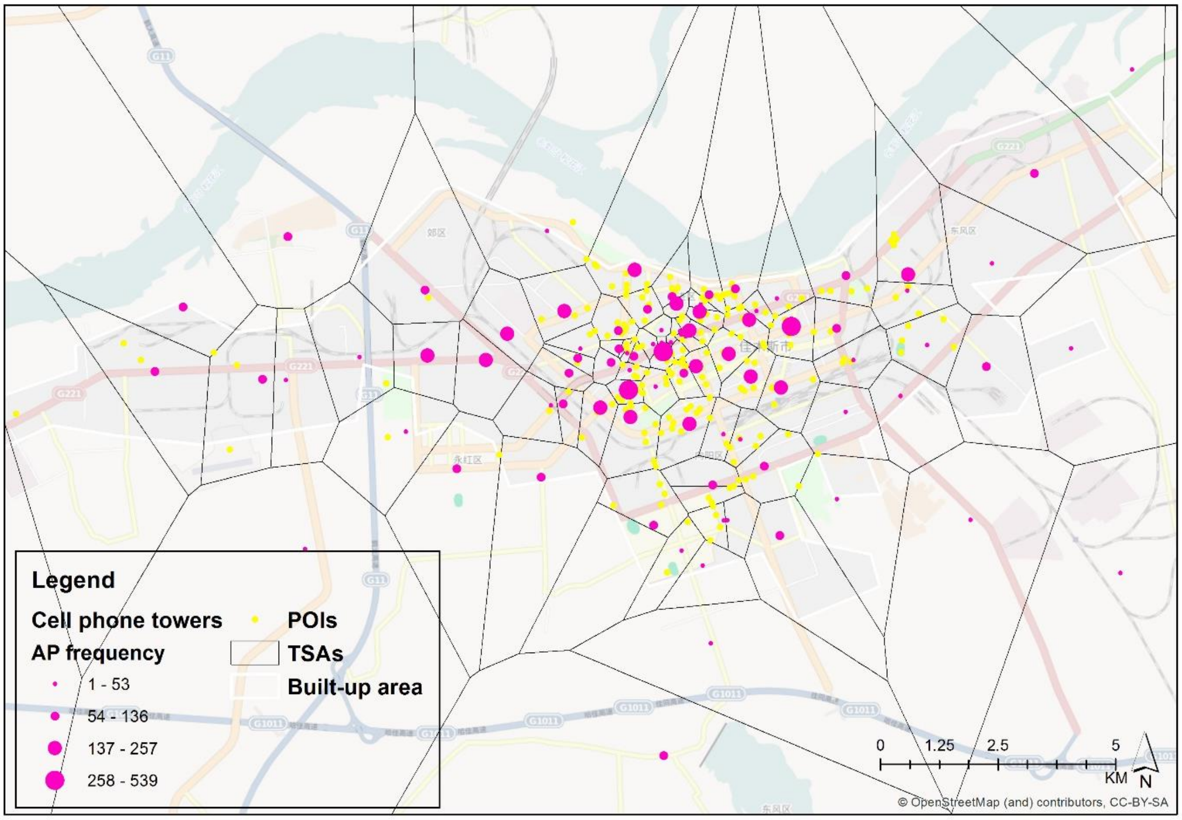


Figure 2.8 The frequency of each cell phone tower being the AP of members in the dominant community on *Day 1*.

We have shown above (Figure 2.6 (a)) that the self-contained TSAs indicate some particular land use types, so we are involved to know if the "self-contained" cell phone users, namely, the isolated cliques, are spatially meaningful. We select *Day 1* as a representative because there is an obvious isolated clique (i.e., the purple in the bottom left). After mapping the four mostly used APs by the members in the isolated purple community on Day 1 (Figure 2.7), we find their locations are mainly in suburban area (Figure 2.9). Surprisingly, the TSAs corresponding to the top four APs are not of no relation. The TSA in green-frame is a farmland, and the one in pink-frame is a botanical park exhibiting agricultural products, new types of plants, as well as selling food like farmers' market. So it is reasonable that people on the farm have connections with who are in the park, either to negotiate business or to contact their partners. The TSA to the west of the park covers a village, where the farmers may live. The most unexpected but very reasonable thing is the TSA in blue frame. We know there is an airport in that area, and assume that farmers need to contact the airport for product transportation. But we wonder why they have to choose the farther Jiaxi Airport instead of the closer Dongjiao Airport. We then find that Jiaxi Airport is an airport specifically for farmland usage, not open to the public.

The findings above thus reinforce the association between social network and urban movements. It also substantiates our argument that social network is not randomly dispersed in space. Instead of "reaching" anywhere by cell phone, people contact where they can or need to visit to make the physical connections happen. Meanwhile, telecommunication, as a supplement, makes such connections a lot easier.

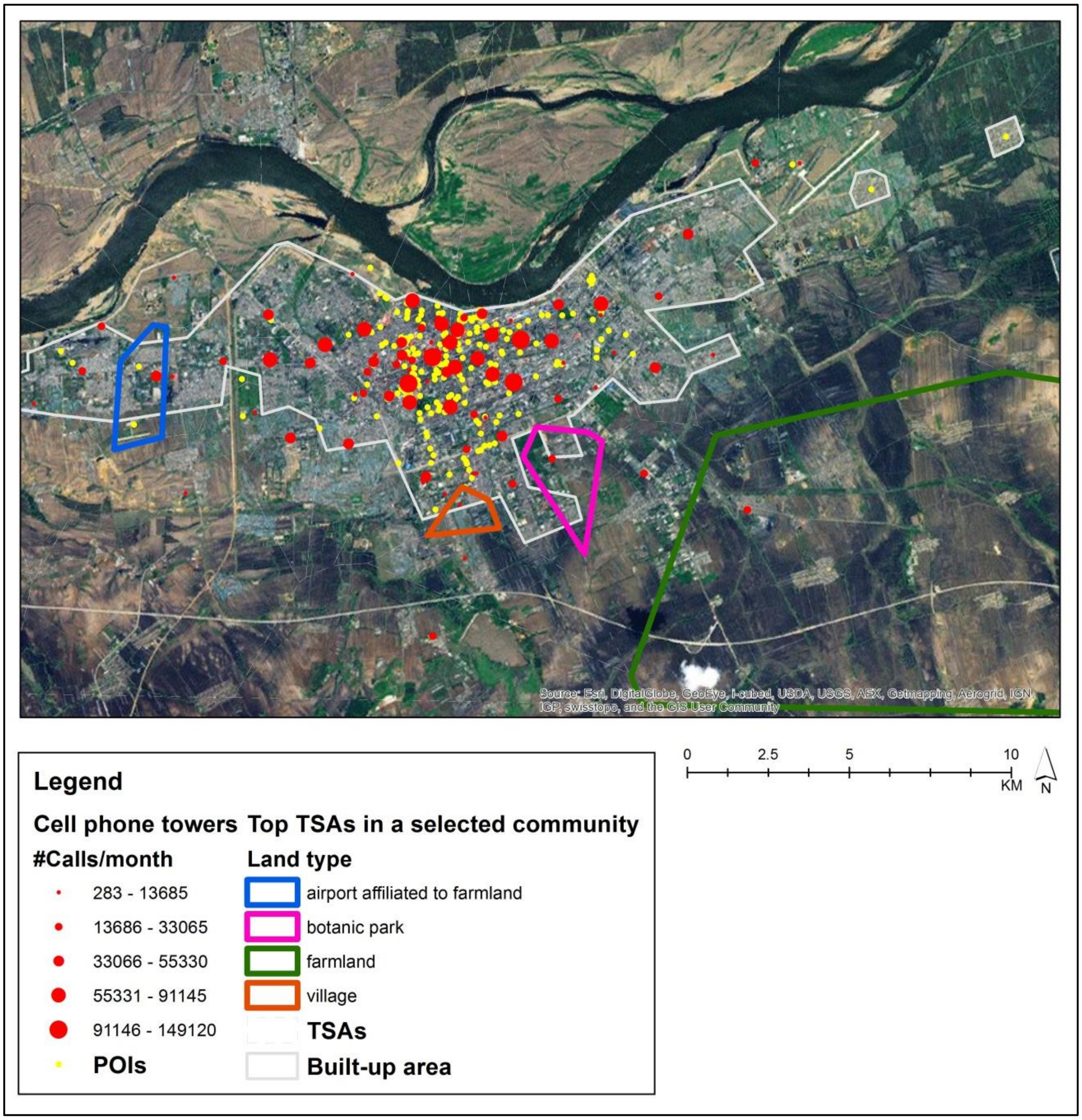


Figure 2.9 The top 4 mostly used TSAs utilized by the users of the selected isolated purple community on *Day 1*.

Implications for Urban Future

The implications of our findings are further than simply the justification for the association between social network and urban movement communities. As cities are growing super rapidly, it is time to ponder on what the urban future should be like. We answer this question in terms of the influence of urban form on social life. One topic of the heated discussions between New Urbanism and Landscape Urbanism concentrates on future urban form. While New Urbanism focuses on neighborhood design, architectural style, and transit-oriented development that will create a high-density walkable city center and a compact urban shape (Al-hindi and Till 2001; Vanderbeek and Irazabal 2007), Landscape Urbanism argues that urban design should be a flexible, open-ended, strategic and non-hierarchical *process* instead of a static plan as is traditionally done (Chaeles Waldheim 2002; Charles Waldheim 2006), including making advantage of the existing resources on the ground (e.g., the Sustainable Park proposed by Cranz and Boland (Cranz and Boland 2004)), implementing ecological infrastructure design that prevents urban encroachment of wilderness (e.g., (Yu, Wang, and Li 2011)), to name a few. The doctrine of Landscape Urbanism thus more or less encourages, or at least accepts urban sprawl (Koolhaas and Mau 1998).

However, although we embrace the environmentally friendly principles proposed by Landscape Urbanism, we protest against urban sprawl based on our research findings from the perspective of social network instead of environment. Given that social network plays a significant role on the generation of long urban trips, urban sprawl as well as its consequent higher travel cost may demolish social connections whose importance has been discussed in details by Gehl (Gehl 1987). Indeed, we see from Figure 2.5 that the dominant social network community appears covering a wider range in urban space than the dominant urban movement community, which probably indicates that virtual connection via information technology can to some extent conquers the impedance of space, and thus is less affected by urban sprawl. However, we still question the potential of information technology in substituting physical interactions.

Our work develops a method to detect the associated communities in social network and urban movement network as an empirical justification for not only the integrity of physical communications in maintaining social ties, but also its priority over virtual links in terms of people's preference. The significant similarity between SNCD and UMCD patterns controverts the randomness in the spatial distribution of social network. Social network is not random because people develop it when face-to-face communication must be achievable, which explains its similarity to UMCD patterns. We have demonstrated that in the downtown area (Figure 2.6 (b)), physical movements are so frequent and convenient that the virtual connections are much less essential. Telecommunication is just an extension of our touchable life when budgets do not allow concrete interactions, as is the suburban farm case (Figure 2.6 (c)).

Compared with the finding of significant overlap in people's daily activity spaces from LAS (Wang et al.), we would say the association between social connection and spatial movement is even stronger since we emphasize on trips longer than quantified by distance decay. Although long trips happen less frequently than short ones due to higher cost, they are likely to be driven by more crucial reasons that make people willing to pay for the higher budget. The significant similarity between the patterns in SNCD and UMCD therefore proves the importance of social network as a motivation of long urban trips. The growth of urban size, consequently, may facilitate the generation of more long trips in order to maintain social ties in a physical way, which exacerbates traffic congestion; the worse traffic situation in turn will cost higher budget to meet social partners and potentially harm the stability of social connections. If averagely people have to spend one hour or more to meet somebody, except that this is a must, the meeting frequency will be highly possible to fall drastically.

We think our methodology is transplantable to other cities, but the results are specific to this study area. Case studies in other city may yield similar findings, given alike urban size and structure. For instance, Jiamusi metropolitan area is generally homocentric, so another homocentric city with similar geometric shape and scale is possible to get indistinguishable results. Limited by the access to data, currently our discussions are only based upon monocentric cities. But we are interested in comparing different community patterns detected in compact monocentric and mega multi-nuclear cities. When a city grows very big, it usually “split” into multiple functional zones; what are the consequences in that case? The localness of social network may be enhanced because people cannot afford higher space-time budget on physical communication, and thus lose long-distance social links. On the other hand, there may emerge more long-distance social ties if people are willing to pay higher cost to travel longer, which, nevertheless, definitely cause more transportation and environment trouble as encountered by many mega cities such as Beijing, Tokyo, and London.

CHAPTER 3

SUPPORTING INFORMATION[2]

[2] Wang, Y. and Yao, X. To be submitted to *PLOSONE* as Supporting Information.

I. Proof for Expert el al.'s Deterrence Function as an Alternative to Distance Decay Function

Expert el al.'s modularity model is (adopted from eq. [3] in (Expert et al. 2011)):

$$Q = \frac{1}{2m}\sum_{C\in P}\sum_{i,j\in C} | A_{ij} - P_{ij} | \qquad \text{(S1)}$$

where P_{ij} is the expected flow between *community* i and j (eq. (S2)):

$$P_{ij} = N_i N_j f\left(d_{ij}\right) \qquad \text{(S2)}$$

The traditional gravity model is:

$$P_{ij} = k\frac{N_i N_j}{{d_{ij}}^b} \qquad \text{(S3).}$$

By plugging equation (S3) into the left side of equation (S2), we get

$$f(d_{ij}) = \frac{k}{{d_{ij}}^b} = \frac{P_{ij}}{N_i N_j} \qquad \text{(S4).}$$

If we have observed flow A_{ij} between each pair of places i and j, we can get the estimated value of b (namely, $\hat{b}$) by fitting data with equation (S4), i.e., by substituting P_{ij} with A_{ij} to get equation (S5), which is an <u>individual</u> level function for distance decay, since it fits the curve by <u>each</u> particular data plot.

$$f(d_{ij}) = \frac{k}{{d_{ij}}^{\hat{b}}} = \frac{A_{ij}}{N_i N_j} \qquad \text{(S5)}$$

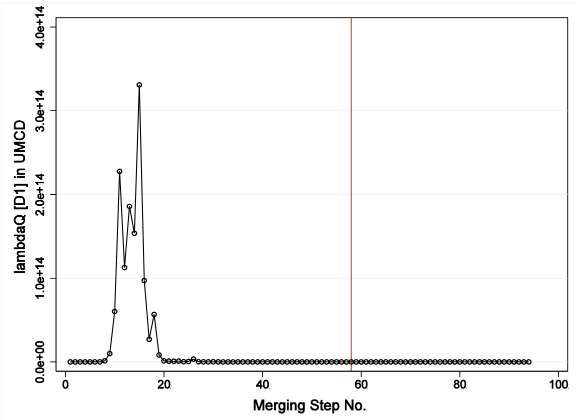
However, in Expert et al.'s method, distance decay is approximated by

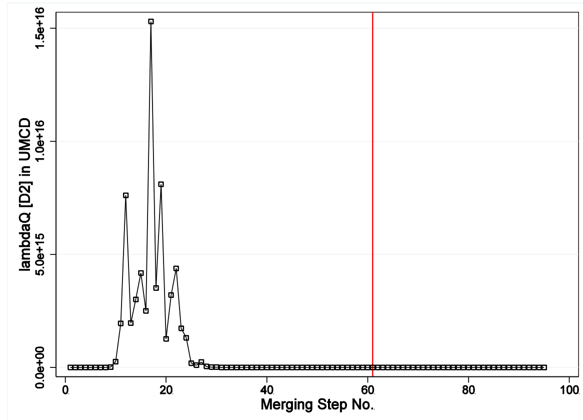
$$f\left(d_{ij}\right) = \frac{\sum_{i,j|d_{ij}=d} A_{ij}}{\sum_{i,j|d_{ij}=d} N_i N_j} \qquad \text{(S6),}$$

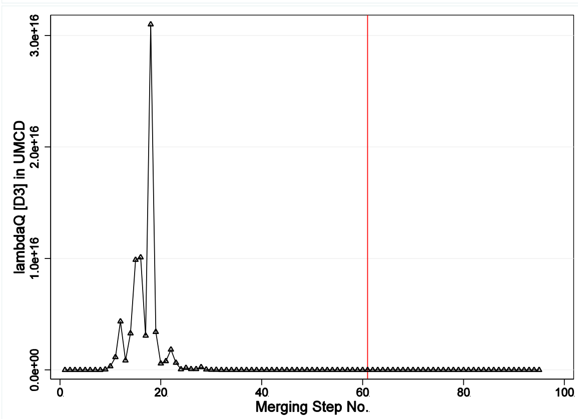
where the numerator is an aggregation of flows at a distance bin centered at d (namely, a distance band $[d - \Delta, d + \Delta]$), and the denominator is the aggregation of the products of node sizes at the two ends of each of the flows, i.e., A_{ij}. Up to here, it is obvious to see the correspondence between equation (S5) derived from traditional gravity model and equation (S6) proposed by Expert et al. (Expert et al. 2011).

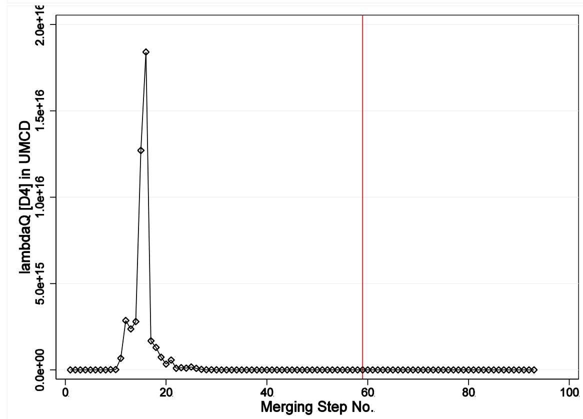
II. Decision on the Number of Social Network Communities

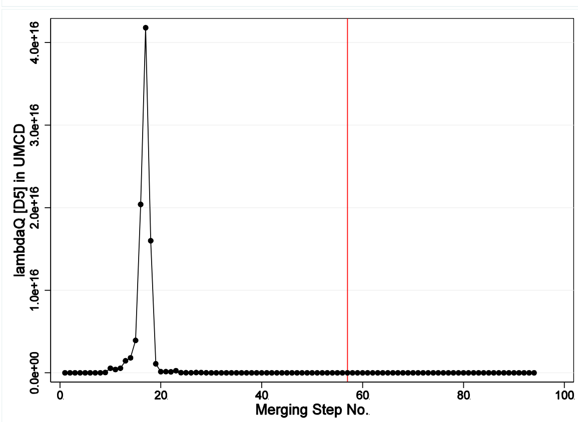
In urban movement community detection (UMCD), we decide the number of communities by watching the trend of the increase rate of modularity (λQ) against the ordinal number of merging step (n). Figure 3.1 demonstrates the change trend, where the x-axis is the ordinal number of the merging step, and y-axis is the λQ for each step. We find that the change of modularity generally yields a single-peak pattern for each of the seven days, with one global maximum of λQ at which point Q grows fastest and multiple local maxima (minor peaks). The peak value appears roughly when n is between 15 and 20, which signifies the merging of some existing smaller communities from the previous merging steps. The final number of communities in UMCD is decided by the n^* after which step no $\lambda Q>1$, i.e., the modularity will decrease. There is no λQ before n^* that satisfies $\lambda Q<1$. The cut-off step number, which demonstrates how many merging steps are needed to get the optimal community pattern, are: 58, 61, 61, 59, 57, 56, and 55 (shown by the red reference line in Figure 3.1).

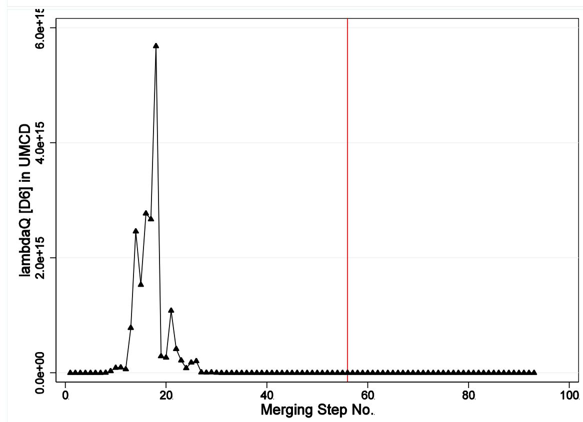

lambdaQ [D1] in UMCD
0.0e+00
1.0e+14
2.0e+14
3.0e+14
4.0e+14
0
20
40
60
80
100
Merging Step No.


lambdaQ [D2] in UMCD
0.0e+00
5.0e+15
1.0e+16
1.5e+16
0
20
40
60
80
100
Merging Step No.


lambdaQ [D3] in UMCD
0.0e+00
1.0e+16
2.0e+16
3.0e+16
0
20
40
60
80
100
Merging Step No.


lambdaQ [D4] in UMCD
0.0e+00
5.0e+15
1.0e+16
1.5e+16
0
20
40
60
80
100
Merging Step No.


lambdaQ [D5] in UMCD
0.0e+00
1.0e+16
2.0e+16
3.0e+16
4.0e+16
0
20
40
60
80
100
Merging Step No.


lambdaQ [D6] in UMCD
0.0e+00
2.0e+15
4.0e+15
0
20
40
60
80
100
Merging Step No.

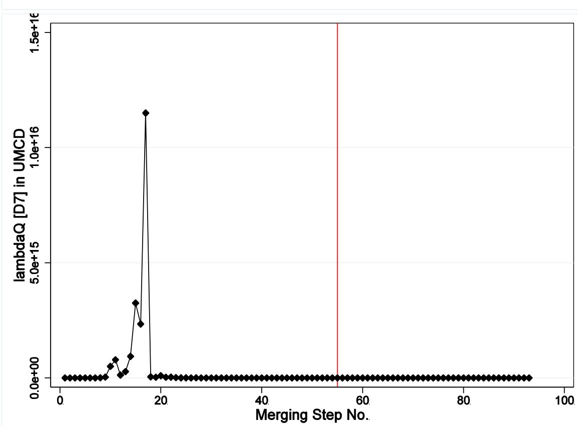

lambdaQ [D7] in UMCD
0.0e+00
5.0e+15
1.0e+16
0
20
40
60
80
100
Merging Step No.

Figure 3.1 The λQ of each merging step in the UMCDs for each of the seven days.

Regarding social network community detection (SNCD), Figure 3.2 shows the λQ–n plot, where the red reference line is the merging step n^* that yields an optimized social network community structure. We see a similar single-peak pattern as UMCD, but it is very noticeable that each of the seven days yields a long tail, which indicates that there may be many merging steps that do not contribute meaningfully to the increase of modularity, compared with the peak value. We find that at the final merging step when all the nodes are merged into one big community, λQ is still greater than one, meaning that the modularity is increasing all the way.

Therefore, we cannot use the threshold 1 for λQ to decide the number of the optimized communities. We propose to test the distribution of the values of λQ, which is different from the λQ-n plot, and claim that the optimized community structure occurs at the merging step n^* when λQ is significantly lower than the average of λQ (i.e., λQ is smaller than the lower bound of the 95% confidence interval) hereafter. The rationale is to get rid of the merging steps that yield no meaningful contribution to the general increase of modularity. Table 3.1 demonstrates the mean value and 95% confidence interval of λQ for each day, where the lower bound is in bold.

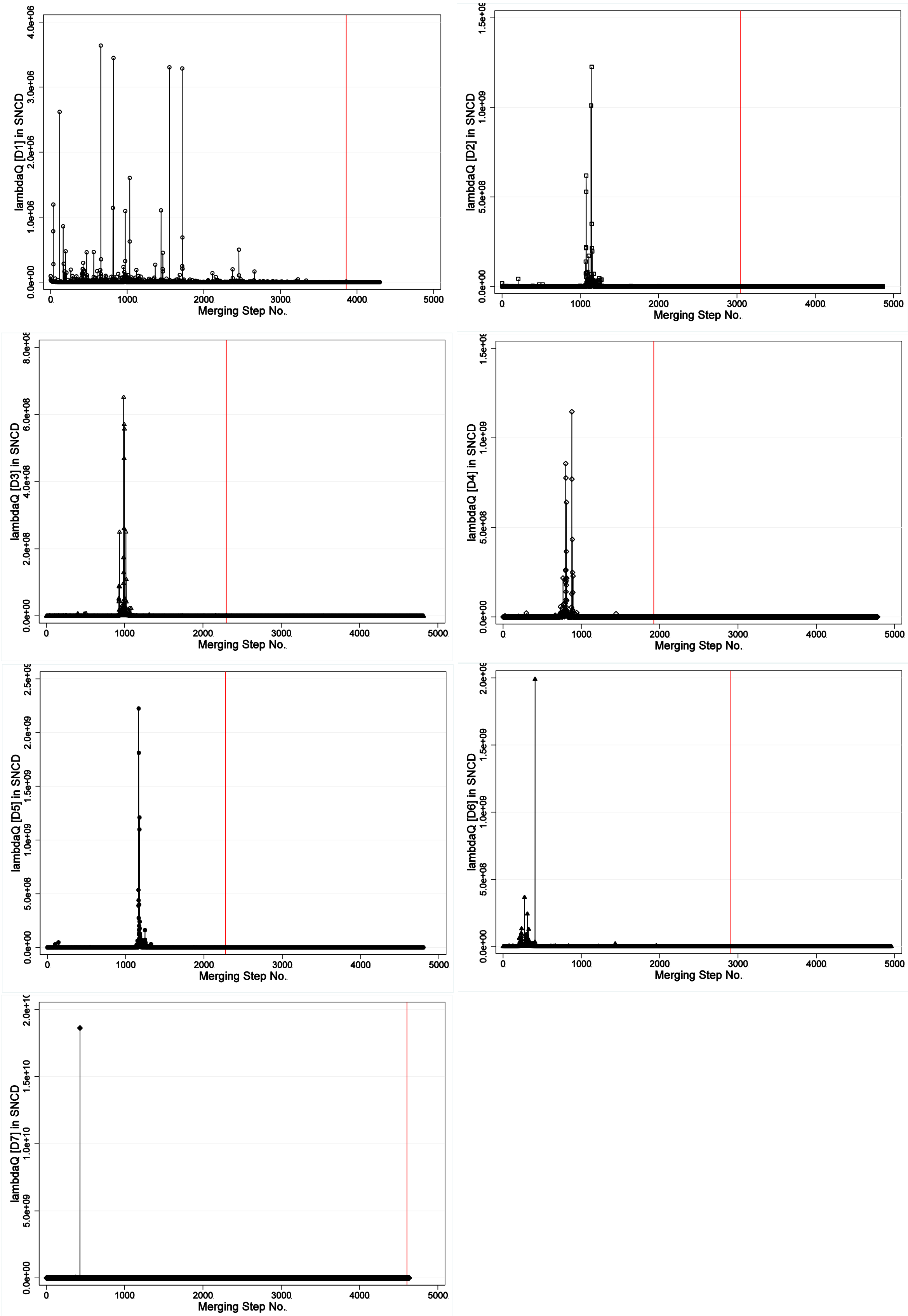

lambdaQ [D1] in SNCD
Merging Step No.
lambdaQ [D2] in SNCD
Merging Step No.
lambdaQ [D3] in SNCD
Merging Step No.
lambdaQ [D4] in SNCD
Merging Step No.
lambdaQ [D5] in SNCD
Merging Step No.
lambdaQ [D6] in SNCD
Merging Step No.
lambdaQ [D7] in SNCD
Merging Step No.

Figure 3.2 The λQ of each merging step in the SNCDs for each of the seven days.

Table 3.1 95% Confidence Interval for λQ in SNCD

Day	**Mean of λQ**	**[95% Conf. Interval]**	
D1	10576.12	**6869.044**	14283.2
D2	1423356	**658737.4**	2187975
D3	1022708	**511192.4**	1534224
D4	1997455	**1120147**	2874762
D5	2473024	**1052318**	3893731
D6	1110521	**291238.6**	1929803
D7	4052008	**-3823432**	1.19e+07

CHAPTER 4

CONCLUSIONS

In this work, I leverage the cell phone call dataset for community detection in social network and urban movement network independently and comparably. By comparing the community pattern from the social network and from the urban movement network, I find a noticeable association between them. My argument is that, while, on a local scale, social connection and geographical movements are mutually influenced, the association in my study justifies the attraction of social network to long-distance urban trips.

The indications of my findings are further than simply the justification for the association between social network and urban movement communities. As cities are growing super rapidly, it is time to ponder on what the urban future should be like. I answer this question in terms of urban form. One topic of the heated discussions between New Urbanism and Landscape Urbanism concentrates on future urban form. While New Urbanism focuses on neighborhood design, architectural style, and transit-oriented development that will create a high-density walkable city center and a compact urban shape (Al-hindi and Till 2001; Vanderbeek and Irazabal 2007), Landscape Urbanism argues that urban design should be a flexible, open-ended, strategic and non-hierarchical *process* instead of a static plan as is traditionally done (Chaeles Waldheim 2002; Charles Waldheim 2006), including making advantage of the existing resources on the ground (e.g., the Sustainable Park proposed by Cranz and Boland (Cranz and Boland 2004)), implementing ecological infrastructure design that prevents urban encroachment of wilderness (e.g., (Yu, Wang, and Li 2011)), to name a few. The doctrine of Landscape Urbanism thus more or less encourages, or at least accepts urban sprawl (Koolhaas and Mau 1998).

However, although I appreciate the environmentally friendly principles proposed by Landscape Urbanism, I protest against urban sprawl based on my research findings from the perspective of social network instead of environment. Given my results that social network demands play a significant role on the generation of long urban trips, urban sprawl as well as its consequent higher travel cost may demolish social connections whose importance has been discussed in details by Gehl (Gehl 1987). Indeed, I see from Figure 3 that the dominant social network community appears covering a wider range in urban space than the dominant urban movement community, which probably indicates that virtual connection via information technology can to some extent conquers the impedance of space. However, we still question the potential of information technology in substituting physical interactions. The observations confirm the existence of geographical proximity in social network space. By *linked activity spaces*, Wang et al. (Wang et al.) have found that people with social ties are more likely to form spatial clusters in their daily activity spaces. We would say the association between social connection and spatial movement is even stronger since we emphasize on long-distance trips. Although long trips happen less frequently than short ones because of higher budget, they are more likely to be driven by more crucial reasons that make people willing to spend for the high budget. The significant similarity between the patterns from SNCD and UMCD therefore proves the importance of social network as a motivation of long urban trips.

Limited by the access to data, currently our discussions are only based upon monocentric cities. But we are interested in comparing different community patterns detected in compact monocentric and mega multi-nuclear cities. When a city grows very big, it usually “split” into multiple functional zones; what are the consequences in that case? Our argument on social network's property of being local may be enhanced because people cannot afford higher space-time budget, or may be diminished if people are willing to pay higher cost, the latter of which, nevertheless, definitely cause more trouble as many mega cities such as Beijing, Tokyo and London.

REFERENCES

Al-hindi, Karen Falconer, and Karen E Till. 2001. "(Re)placing the New Urbanism Debates: Toward an Interdisciplinary Research Agenda." *Urban Geography* 22 (3): 189–201.

Amaral, L A N, A Scala, M Barthélémy, and H E Stanley. 2000. "Classes of Small-World Networks." *Proceedings of the National Academy of Sciences of the United States of America* 97 (21): 11149–11152. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=17168&tool=pmcentrez&rendertype=abstract.

Apicella, Coren L, Frank W Marlowe, James H Fowler, and Nicholas a Christakis. 2012. "Social Networks and Cooperation in Hunter-Gatherers." *Nature* 481 (7382) (January 26): 497–501.

Arenas, Alex, and A Diaz-Guilera. 2007. "Synchronization and Modularity in Complex Networks." *The European Physical Journal-Special Topics* 1: 1–9.

Barthélemy, Marc. 2011. "Spatial Networks." *Physics Reports* 499: 1–101.

Bettencourt, Luís M A. 2013. "The Origins of Scaling in Cities." *Science* 340 (6139) (June 21): 1438–41.

Cairncross, Frances. 2001. *The Death of Distance: How the Communications Revolution Will Change Our Lives*. *Journal of the American Society for Information Science*. Vol. 49. Boston: Harvard Business Press.

Comber, a.J., C.F. Brunsdon, and C.J.Q. Farmer. 2012. "Community Detection in Spatial Networks: Inferring Land Use from a Planar Graph of Land Cover Objects." *International Journal of Applied Earth Observation and Geoinformation* 18 (August): 274–282.

doi:10.1016/j.jag.2012.01.020. http://linkinghub.elsevier.com/retrieve/pii/S0303243412000220.

Correa, C D, T Crnovrsanin, C Muelder, Zeqian Shen Zeqian Shen, R Armstrong, J Shearer, and Kwan-Liu Ma Kwan-Liu Ma. 2008. *Cell Phone Mini Challenge Award: Intuitive Social Network Graphs Visual Analytics of Cell Phone Data Using Mobivis and Ontovis*. *2008 IEEE Symposium on Visual Analytics Science and Technology*. Ieee. doi:10.1109/VAST.2008.4677391. http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4677391.

Cranz, Galen, and Michael Boland. 2004. "Defining the Sustainable Park: A Fifth Model for Urban Parks." *Landscape Journal* 23 (2-04): 102 – 120.

Ctrip. "Heilongjiang Transportation." *Http://english.ctrip.com/travel/region/heilongjiang/transportation/*. http://images4.c-ctrip.com/target/english/754/763/579/4583BE54B2F341F9AC472B8E8B97F1D7.Jpg.

Danon, Leon, Albert Díaz-Guilera, Jordi Duch, and Alex Arenas. 2005. "Comparing Community Structure Identification." *Journal of Statistical Mechanics: Theory and Experiment* 2005 (09) (September 26): P09008.

Duke, Shearlean. 2001. "E-Mail: Essential in Media Relations, But No Replacement for Face-To-Face Communication." *Public Relations Quarterly* 46 (4): 19–22.

Eagle, Nathan, Alex Sandy Pentland, and David Lazer. 2009. "Inferring Friendship Network Structure by Using Mobile Phone Data." *Proceedings of the National Academy of Sciences of the United States of America* 106 (36) (September 8): 15274–8. doi:10.1073/pnas.0900282106.

http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2741241&tool=pmcentrez&rendertype=abstract.

Eldridge, J D, and J P Iii Jones. 1991. “Warped Space: A Geography of Distance Decay.” *The Professional Geographyer* 43 (4): 500–511. doi:10.1111/j.0033-0124.1991.00500.x. http://www3.interscience.wiley.com/cgi-bin/abstract/119347580/ABSTRACT.

Ewing, Reid, Rolf Pendall, Don Chen, and Smart Growth America. 2003. “Measuring Sprawl and Its Transportation Impacts.” *Transportation Research Record* 1831: 175–183.

Expert, Paul, Tim S Evans, Vincent D Blondel, and Renaud Lambiotte. 2011. “Uncovering Space-Independent Communities in Spatial Networks.” *Proceedings of the National Academy of Sciences of the United States of America* 108 (19) (May 10): 7663–7668.

Feld, Scott L. 1981. “The Focused Organization of Social Ties.” *American Journal of Sociology* 86 (5) (March): 1015.

Flaherty, Lisa M., Kevin J. Pearce, and Rebecca B. Rubin. 1998. “Internet and Face-to-face Communication: Not Functional Alternatives.” *Communication Quarterly* 46 (3) (June): 250–268.

Fortunato, Santo, and Marc Barthélemy. 2007. “Resolution Limit in Community Detection.” *Proceedings of the National Academy of Sciences of the United States of America* 104 (1) (January): 36–41.

Fotheringham, AS. 1981. “Spatial Structure and Distance-Decay Parameters.” *Annals of the Association of American Geographers* 71 (3): 425–436. http://www.tandfonline.com/doi/abs/10.1111/j.1467-8306.1981.tb01367.x.

Gao, Song, Yu Liu, Yaoli Wang, and Xiujun Ma. 2013. “Discovering Spatial Interaction Communities from Mobile Phone Data.” *Transactions in GIS* 17 (3) (June 17): 463–481.

Gao, Song, Yaoli Wang, Yong Gao, and Yu Liu. 2013. “Understanding Urban Traffic-Flow Characteristics: A Rethinking of Betweenness Centrality.” *Environment and Planning B: Planning and Design* 40 (1): 135–153.

Gehl, Jan. 1987. *Life Between Buildings: Using Public Space*. New York: Van Nostrand Reinhold.

Geller, AL. 2003. “Smart Growth: A Prescription for Livable Cities.” *American Journal of Public Health* 93 (9): 1410–1416.

Girvan, M, and MEJ Newman. 2002. “Community Structure in Social and Biological Networks.” *Proceedings of the National Academy of Sciences of the United States of America* 99 (12): 7821 – 7826.

Grady, D, R Brune, C Thiemann, and Dirk Brockmann. 2012. “Modularity Maximization and Tree Clustering: Novel Ways to Determine Effective Geographic Borders.” *Handbook of Optimization in Complex Networks Theory and Applications*.

Greca, Paolo La, Luca Barbarossa, Matteo Ignaccolo, Giuseppe Inturri, and Francesco Martinico. 2011. “The Density Dilemma. A Proposal for Introducing Smart Growth Principles in a Sprawling Settlement within Catania Metropolitan Area.” *Cities* 28 (6) (December): 527–535. doi:10.1016/j.cities.2011.06.009. http://linkinghub.elsevier.com/retrieve/pii/S0264275111000758.

Guimerà, R, S Mossa, A Turtschi, and L A N Amaral. 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.

Guimera, R, M Sales-Pardo, and LAN Amaral. 2004. “Modularity from Fluctuations in Random Graphs and Complex Networks.” *Physical Review E* 70: 025101.

Guo, Diansheng. 2009. “Flow Mapping and Multivariate Visualization of Large Spatial Interaction Data.” *IEEE Transactions on Visualization and Computer Graphics* 15 (6): 1041–1048.

Huang, Wenhao, Zhengbin Dong, Nan Zhao, Hao Tian, Guojie Song, Guanhua Chen, Yun Jiang, and Kunqing Xie. 2010. “Anchor Points Seeking of Large Urban Crowd Based on the Mobile Billing Data.” In *Advanced Data Mining and Applications*, 346–357.

Kang, Chaogui, Song Gao, Xing Lin, Yu Xiao, Yihong Yuan, Yu Liu, and Xiujun Ma. 2010. “Analyzing and Geo-Visualizing Individual Human Mobility Patterns Using Mobile Call Records.” In *Geoinformatics, 2010 18th International Conference on*, edited by Liu Y And Chen A, 1–7. Beijing: IEEE.

Kang, Chaogui, Xiujun Ma, Daoqin Tong, and Yu Liu. 2012. “Intra-Urban Human Mobility Patterns: An Urban Morphology Perspective.” *Physica A: Statistical Mechanics and Its Applications* 391 (4) (February): 1702–1717.

Kolko, Jed. 2000. “The Death of Cities? The Death of Distance? Evidence from the Geography of Commercial Internet Usage.” *The Internet Upheaval: Raising Questions, Seeking Answers in Communications Policy*: 73–98.

Koolhaas, Rem, and Bruce Mau. 1998. *S, M, L, XL*. 2nd ed. Monacelli Press.

Lambiotte, R, and VD Blondel. 2008. “Geographical Dispersal of Mobile Communication Networks.” *Physica A: Statistical …* 387: 5317 – 5325. http://www.sciencedirect.com/science/article/pii/S0378437108004342.

Liben-Nowell, D, and Jasmine Novak. 2005. "Geographic Routing in Social Networks." *Proceedings of the National Academy of Sciences of the United States of America* 102 (33): 11623 – 11628.

Liu, Yu, Zhengwei Sui, Chaogui Kang, and Yong Gao. 2014. "Uncovering Patterns of Inter-Urban Trip and Spatial Interaction from Social Media Check-In Data." Edited by Peter Csermely. *PLoS ONE* 9 (1) (January 17): e86026.

Livesey, Graham. 2009. "A Look at Landscape Urbanism." *Canadian Architect*: 45 – 47.

McPherson, M, L Smith-Lovin, and JM Cook. 2001. "Birds of a Feather: Homophily in Social Networks." *Annual Review of Sociology* 27: 415–44.

Newman, M E J. 2006. "Modularity and Community Structure in Networks." *Proceedings of the National Academy of Sciences* 103 (23) (June 6): 8577–8582.

Newman, M E J, and M Girvan. 2004. "Finding and Evaluating Community Structure in Networks." *Physical Review. E, Statistical, Nonlinear, and Soft Matter Physics*.

Newman, Mark E. J. 2004. "Fast Algorithm for Detecting Community Structure in Networks." *Physical Review E* 69 (6): 066133.

Newman, MEJ, and M Girvan. 2003. "Mixing Patterns and Community Structure in Networks." *Statistical Mechanics of Complex Networks* 67 (2): 21.

Onnela, Jukka-Pekka, Samuel Arbesman, Marta C González, Albert-László Barabási, and Nicholas A Christakis. 2011. "Geographic Constraints on Social Network Groups." Edited by Matjaz Perc. *PLoS ONE* 6 (4): 7.

Palla, G, A L Barabási, and T Vicsek. 2007. "Quantifying Social Group Evolution." *Nature* 446: 664–667.

Pons, Pascal, and Matthieu Latapy. 2005. "Computing Communities in Large Networks Using Random Walks." *Journal of Graph Algorithms and Applications* 10 (2): 191 – 218.

Ratti, Carlo, Stanislav Sobolevsky, Francesco Calabrese, Clio Andris, Jonathan Reades, Mauro Martino, Rob Claxton, and Steven H Strogatz. 2010. "Redrawing the Map of Great Britain from a Network of Human Interactions." Edited by Olaf Sporns. *PLoS ONE* 5 (12): 6.

Reichardt, J, and Stefan Bornholdt. 2006. "Statistical Mechanics of Community Detection." *Physical Review E*: 1–16. http://pre.aps.org/abstract/PRE/v74/i1/e016110.

Salomon, Ilan. 1985. "Telecommunications and Travel: Substitution or Modified Mobility?" *Journal of Transport Economics and Policy* 19 (3): 219–235.

Steinhaeuser, K, and NV Chawla. 2008. "Community Detection in a Large Real-World Social Network." *Social Computing, Behavioral Modeling, and Prediction*: 168 – 175.

Taylor, PJ. 1977. "Chapter 5: Areal Association." In *Quantitative Methods in Geography: An Introduction to Spatial Analysis*, 396. Boston: Houghton Mifflin.

Thiemann, Christian, Fabian Theis, Daniel Grady, Rafael Brune, and Dirk Brockmann. 2010. "The Structure of Borders in a Small World." Edited by Gonzalo De Polavieja. *PLoS ONE* 5 (11): 7.

Tobler, WR. 1970. "A Computer Movie Simulating Urban Growth in the Detroit Region." *Economic Geography* 46 (2): 234–240.

Vanderbeek, Michael, and Clara Irazabal. 2007. "New Urbanism as a New Modernist Movement: A Comparative Look at Modernism and New Urbanism." *Traditional Dwellings and Settlements Review* 19 (1): 41–57.

Vespignani, Alessandro. 2009. “Predicting the Behavior of Techno-Social Systems.” *Science* 325 (5939): 425–8. doi:10.1126/science.1171990. http://www.ncbi.nlm.nih.gov/pubmed/19628859.

Waldheim, Chaeles. 2002. “Landscape Urbanism: A Genealogy.” *Praxis* 4: 10 – 17.

Waldheim, Charles. 2006. *The Landscape Urbanism Reader*. *New York*. Princeton Architectural Press.

Walsh, F, and A Pozdnoukhov. 2011. “Spatial Structure and Dynamics of Urban Communities.” In *First Workshop on Pervasive Urban Applications (PURBA) at Ninth International Conference on Pervasive Computing*. San Francisco.

Wang, Yaoli. 2012. “Modularity Measure for Community Detection in Urban Space: A Case Study with Mobile Phone Network”. Beijing: Peking University.

Wang, Yaoli, Chaogui Kang, Luís M A Bettencourt, Yu Liu, and Clio Andris. “Linked Activity Spaces: Embedding Social Networks in Urban Space.” In *Computational Approaches for Urban Environments*, edited by Marco Helbich, Jamal Jokar Arsanjani, and Michael Leitner. Springer. (in press)

Watts, D J, P S Dodds, and M E J Newman. 2002. “Identity and Search in Social Networks.” *Science* 296 (5571): 4.

Watts, D J, and S H Strogatz. 1998. “Collective Dynamics of ‘Small-World’ Networks.” Edited by Mark Newman, Albert-László Barabási, and Duncan J Watts. *Nature* 393 (6684): 440–2. doi:10.1038/30918. http://www.ncbi.nlm.nih.gov/pubmed/9623998.

Yu, Kongjian, Sisi Wang, and Dihua Li. 2011. “The Negative Approach to Urban Growth Planning of Beijing, China.” *Journal of Environmental Planning and Management* 54 (9) (November): 1209–1236.

APPENDICES

A. A List of Abbreviations

AP – Anchor point

LAS – Linked activity spaces

POI – Point of interest

SNCD – Social network community detection

TSA – Tower service area

UMCD – Urban movement community detection