Chapter 13
Linked Activity Spaces: Embedding Social Networks in Urban Space

Yaoli Wang, Chaogui Kang, Luís M.A. Bettencourt, Yu Liu, and Clio Andris

Abstract We examine the likelihood that a pair of sustained telephone contacts (e.g. friends, family, professional contacts, called "friends") uses the city similarly. Using call data records from Jiamusi, China, we estimate a proxy for the daily activity spaces of each individual subscriber by interpolating the points of geo-located cell towers he or she uses most frequently. We then calculate the overlap of the polygonal activity spaces of two established telephone contacts, what we call *linked* activity spaces.

Our results show that friends and second-degree friends (e.g. friends of friends) are more likely to geographically overlap than random pairs of users. Additionally, individuals with more friends and with many network triangles (connected groups of three friends) tend to congregate in the city's downtown at a rate that surpasses randomness. We also find that the downtown is used by many social groups but that each suburb only hosts one or two groups. We discuss our findings in terms of the need for a better understanding of spatialised social capital in urban planning.

Y. Wang
Department of Geography, University of Georgia, 120K, Geog-Geol Building, Athens, GA 30602, USA

Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, USA

C. Kang
Institute of Remote Sensing and Geographical Information Systems, Peking University, Beijing 100871, China

MIT Senseable City Laboratory, 9-209, 77 Massachusetts Avenue, Cambridge, MA 02139, USA

L.M.A. Bettencourt
Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, USA

Y. Liu
Institute of Remote Sensing and Geographical Information Systems, Peking University, Beijing 100871, China

C. Andris (✉)
Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, USA

Department of Geography, The Pennsylvania State University, 302 Walker Building, University Park, PA 16802, USA
e-mail: clio@psu.edu

M. Helbich et al. (eds.), *Computational Approaches for Urban Environments*, Geotechnologies and the Environment 13, DOI 10.1007/978-3-319-11469-9_13

Keywords Activity space • Daily movement • Call data records • Mobile phones • Social networks • Friendship • Relationships • Cities • Built environment

Abbreviations

CDRs	Call data records
GIS	Geographic information system(s)
KML	Keyhole markup language
LAS	Linked activity spaces
POIs	Points of interest

13.1 Introduction

In this chapter, we present a methodology that can help elucidate how groups of friends, family and professional contacts use the city. We know that cities are comprised of two interacting components, social networks and physical infrastructure, and that the social dynamics of encounters in urban space form the backbone of city life (Bettencourt 2013). Yet our ability to model social networks and social capital in urban spaces is very limited. This presents a problem because often our behavior results from the influence of others (Salganik and Watts 2008). The establishment, discovery and maintenance of our social ties are guided by the city. These ties will also affect how we use the city: where we choose to meet, live and work.

Within a city, it remains an open question as to whether a citizen benefits most from having his or her social contacts nearby or dispersed. At one extreme, dispersed contacts can expose the ego to new neighborhoods and a variety of urban knowledge (such as finding the quickest post office, the best doctor or an exciting new restaurant), due to their variety of experiences in diverse parts of a city. Yet, it may be more difficult and more expensive to meet spatially dispersed contacts. Having friends in disparate parts of the city is also more likely to lead to a social network where one has few friends "in common" with other friends, which can be a key strength of social networks.

At the other extreme, a socially tight neighborhood forms trusted bonds through multiple channels of social validation (Centola and Macy 2007) and through increased exposure to one another in the outdoors and through neighborhood institutions such as local schools. Proximal social contacts can meet conveniently, benefiting elderly and the mobility challenged and, in some cases, poorer or immigrant communities who likely rely on friends and family for help with amenities such as child care. Yet, in these enclosed neighborhoods, information and social capital from other parts of the city may be less accessible (Granovetter 1983) resulting in missing or unsupportive social systems across the city (Granovetter 1973).

How do urbanites organize their lives to balance their need for information and accessibility with its costs? In order to answer this question, we must measure and model the social network of egos within the urban built environment.

13.1.1 The State of Social/Spatial Modeling

Everything happens somewhere: examining social life as extricated from the influence of the built environment results in an unrealistic view. Yet the methods available for understanding the clustering and dispersion of a set of individual social networks in geography are limited, as social network and urban spatial models have matured in separate domains, and are analyzed in separate spheres, through social network analysis and geographic information systems (GIS), respectively (Andris 2011). Network methods are also rarely used by those who study city form (Sevtsuk and Mekonnen 2012). Social networks represent influences and social capital as graph configurations of nodes (agents) and links (e.g., edges) between nodes where primary metrics are connectivity and embeddedness; alternatively, spatial (e.g., GIS) models are represented in a contiguous topological plane, where adjacency and proximity are primary metrics (Andris 2011).

As a result, social/spatial phenomena are often explained separately by those inclined toward computational sociology or geography, respectively. One example is the study of obesity, where social networks (Christakis and Fowler 2007) and city form (Papas et al. 2007) are examined as causal factors, but not in the same study. To obtain a clearer picture of the mechanisms surrounding obesity, one should consider social ties and the built environment as coincidental factors – as these influences can compound. Similarly, research showing how students use a college campus in space and time via WiFi usage describes the flexibility of meeting places due to mobile computing (Sevtsuk et al. 2009), but could be extended to assess social gatherings in time and place, as do Eagle et al. (2009) on the same college campus, during a similar time period. Eagle et al. (2009) show the temporal social patterns of dyadic (pairwise) relationships in terms of calls, SMS messages and colocation, and alludes to the role of the campus in providing the backdrop for social groups and pairs. When combined with Sevtsuk et al. (2009), this study could provide the social ties within a spatial setting to uncover where friends meet, where they travel on the campus and how these factors can be leveraged to create a better campus environment.

This is not to say that datasets on interpersonal communication and movement have not been embedded into geography; analyses of interplace networks of social flows such as commodities, telecommunications, migration, and commuting are common in computational urban research (examples abound). Yet, these represent place-to-place aggregate flows instead of person-to-person flows and thus do not directly express the decisions of individuals. Small-scale examples of spatially embedded social networks describe gang membership (Radil et al. 2010; Papachristos et al. 2013), transportation (Frei and Axhausen 2011; Arentze et al. 2012), and epidemiology (Emch et al. 2012). We take these initiatives a step further by creating a general method that can respond to patterns of human socialization in a built environment. These studies can elucidate where and when (different types of) relationships form and could be used to advise architects, urban, and transportation planners in creating places that support and create social connectivity.

In working toward this goal, those looking to examine social/spatial problems are aided by the recent proliferation of large datasets evidencing human social contact and movement (such as GPS or cell tower usage records) in the city (Reades et al. 2007). The integration of human movement and activity data, such as information from GPS traces (Gao et al. 2013), check-in data (Cho et al. 2011), online social networks (Scellato et al. 2011), and photo-sharing sites (Crandall et al. 2010; Girardin et al. 2008; Sun et al. 2013), into urban models are providing new windows on how humans use the built environment. Specifically, the use of mobile telephone calls to understand city usage patterns are becoming a cornerstone of modern urban informatics, planning, and transportation (Ratti et al. 2006). We take advantage of mobile telephone call data to test our research questions about the locality or dispersion of social ties in the city.

Further, the relative convenience of colocation for friends can be evaluated. Calabrese et al. (2011) find that in 94 % of telephone calling partners, one partner constantly travels further to meet. On average, the partner traveling further travels 3 times further to meet. This method uses travel time and distance, which is important for logistics. However, we extend this concept by incorporating the built environment into these compromises, to show where in the city friends are likely to meet. By spatially-linking the respective activity spaces of two friends in the GIS, we can better understand how the city is able to provide places for friends to meet, and assess the travel needs to do so—i.e. it is relatively easy for friends with spatially-overlapping activity spaces to meet face-to-face.

13.1.2 Linked Activity Spaces

We use cell phone call data records (CDRs) to model "friendships" (i.e., interpersonal relationships) as a social network, inferred by the frequency calls between two agents, and the sets of locations visited by each member of the social network within the city (i.e., activity spaces). A pair of activity spaces of an ego and alter are called *linked activity spaces* (LAS) if the ego and alter are friends (i.e., contacts) in the dataset. The two activity spaces of friends are modeled within the GIS and spatially analyzed for similarity, via the number of "third places" shared among the pair (following Rosenbaum 2006). Moreover, we analyze the social network as a whole to find whether high-degree *egos* (a.k.a. those with many friends), *triangles* (groups of three agents) and *communities* use the city in significantly similar ways.

We have four main hypotheses for the analysis of LAS. (1) We expect that friends' activity spaces will overlap more often than a random pair of activity spaces, indicating that friends use the city more similarly than a random pair of people. (2) We also hypothesize that egos with high degrees or high clustering coefficients (see Jackson 2010) will be more associated with the city center, as this denser environment tends to have more meeting places, diverse services, commercial areas, and nightlife. (3) In terms of city form and groups, we believe that central areas will play an enhanced role in supporting "clique-like" and modular groups instead of being a mixing pot for many groups. We expect the downtown area to host tight-knit social groups who do not venture to the suburbs often. (4) Finally, we expect that suburban POIs will accommodate individuals from diverse social groups, as these agents are likely visit different parts of the city using automobiles.

This chapter proceeds as follows. We first describe the study area and the setting of the CDR dataset. We then describe, in the methods section, how we delineate each user's activity spaces. We analyze how *linked activity spaces* (LAS) are spatially correlated in an urban environment by shared points of interest (POIs). We conclude with a discussion of the usefulness of this method, its drawbacks, potential applications, and future work.

13.2 Study Area and Dataset

Our study area is the city of Jiamusi, located in northeastern China, with a population 2.5 million (est. 2010). This industrial city serves as a producer of wood pulp and newsprint and participates in the global economy via a thriving international trade harbor. The urban core of Jiamusi is nearly 18 by 10 km in spatial extent, and its residents travel on average 1 km a day (Kang et al. 2012).

13.2.1 Dataset and Sampling

We focus on calls made within the city area and exclude long-distance calls. We use a CDR (call data record) dataset of mobile cell phone calls from an undisclosed mobile phone provider in China.

The original CDR dataset contains nearly 424,000 users over 31 days. Users are anonymized in the dataset. Combined, users make an average of 1,600,000 calls daily. In the 31-day time span of our dataset, each user participates in an average of 328 calls for a total duration of 6.15 hours. Each record of a mobile phone call contains the start time, call duration, and locations of the caller and receiver. The locations are geo-referenced to one of 96 cell towers closest to the mobile phone's location (Table 13.1). The dataset does not include text messages (e.g., SMS).

We process the dataset into two parts: a social network of agents (social network in Table 13.1) and the activity spaces of each agent (spatial patterns in Table 13.1). We filter the network by including only those who use at least three cell towers during the study period in order to eliminate users who may be confined to their home and thus interact with the city differently than a typical mobile user. Also, an individual may have multiple mobile phones, and a phone with fewer than three cell towers used may represent a "secondary" or less frequently used device. In the social network, the *number of calls* is determined between a unique pair of users, and *duration* is the sum of call time between the two users. The network

Table 13.1 Call data record (CDR) variables with original data fields (*top*)

Original table	Caller	Receiver	Caller location (x, y)	Receiver location (x, y)	Start time	Duration
Social network	User 1	User 2	–	–	Number of calls	Total duration between users 1 and 2
Spatial patterns	User 1		Location (x, y)	Location (x, y)		Duration

A social network and spatial data summary table are listed in the *middle* and *bottom* rows, respectively

is undirected in order to reflect each member's inclination to participate in the conversation regardless of the initiator (Calabrese et al. 2011). In other words, records showing that A calls B, or B calls A, are summed to represent a connection between unique, undirected pair A, B. Each pair must have either 10 + mutual phone calls or 10+ min of total call duration in the given month to be considered friends. This process eliminates non-friend calls such as sales calls, as these do not represent persistent relationships. Our resultant dataset has an average of 11.55 calls per friendship connection (with a 95 % confidence interval (c.i.) of [11.35, 11.76]) and an average of 12.52 min for each link (95 % c.i. is [12.22, 12.81]).

The spatial patterns table contains the locations of each user, which are combined to geo-locate a pair of callers in the social network. The coordinates of the cell phone tower where a user places or receives a call are summed and weighted by the number of calls the user places or receives at that cell tower location. We use the resulting set of weighted locations to represent the user's geographic activity pattern (such as Carrasco et al. 2006), which are known to capture "anchor points" (Golledge 1999) such as home and workplace (or school), as they are the most visited locations for the average traveler and, thus, frequent calling points (Schönfelder and Axhausen 2003).

13.2.2 Sampling

We sample the large CDR dataset by selecting a random sample of 150 "seed" users and retrieve their contacts (first-degree ties), second-degree and third-degree ties, in a method similar to Kurant et al. (2011). The number of seed users is calibrated based on our ability to visualize and computationally analyze the resultant dataset. We also choose this method over a random sample of all users (e.g., choosing 20,000 random users and the possible network that might form between them) because the seed method ensures that retrieved nodes have connections (since we select friends, then friends of friends). This method also is able to find groups, whereas in a random sample of the network, nodes may not be connected. This configuration yields a network that is focused on the social interactions of a small sample of users. As a result, this “core” social network does not resemble a complete social network’s typical degree distribution (such as Albert and Barabási 2002), traversability, or density (Newman and Park 2003).

13.2.3 Social Network and Geographic Characteristics

The degree values for the core network range from 1 to 344, with a mean degree of 43.8 (Fig. 13.1). The diameter of the network is 12. Clustering coefficient values range from 0 to 1, with a mean coefficient of 0.11. The network is visualized in Fig. 13.2. Using Spearman’s correlation statistic, we find that users with more contacts make shorter calls, while those with fewer friends speak for longer.

Our dataset includes 96 cell phone towers. The caller’ or receiver’s location is approximated to the site of the cell tower (which is offset slightly by the telephone provider), although generally the caller or receiver could be found anywhere in the signal radius around the cell tower. We note that it is also possible that a caller’s call may be routed to a cell tower that is not the closest to him or her, as the closest cell tower may be saturated with calls or out of service, though we cannot account for such situations. Some towers are used by many subscribers, while others are used by few: the 10 most popular cell phone towers are used by at least 20 % of the population whereas the 40th–91st most popular towers are each used by less than 10 % of the population.

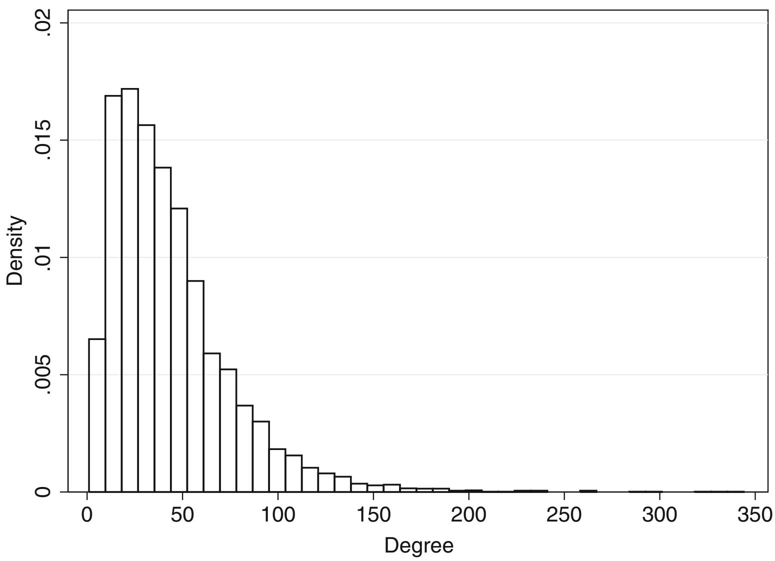

Fig. 13.1 The degree distribution of the calling network, comprised of 8,231 sampled users, is right-skewed with mean 44 and is well described by a log-normal distribution

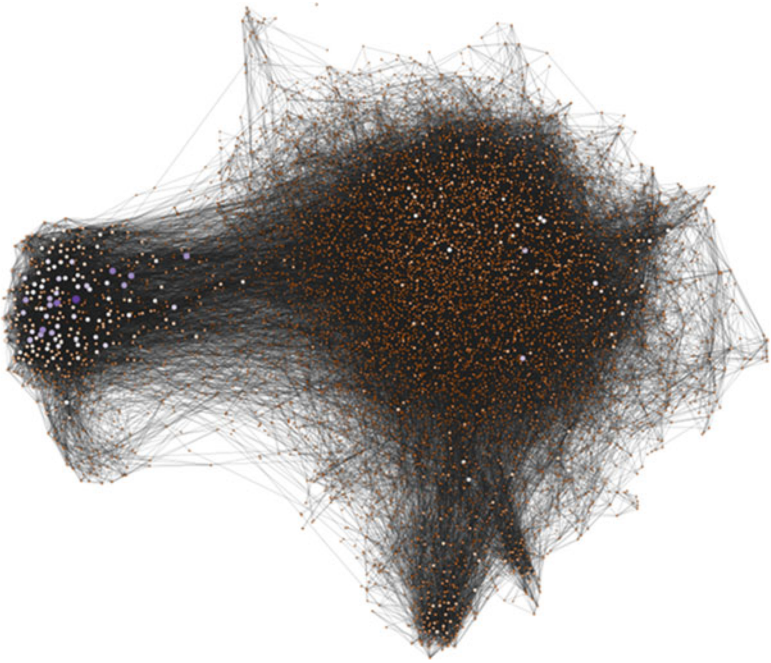

Fig. 13.2 The aspatial social network is visualized with a force-directed method that places nodes in feature space based on their density of linkages in the Gephi computing environment (Bastian et al. 2009). *Larger* nodes denote higher degree and *smaller* nodes indicate smaller degree

13.3 Methods

13.3.1 Creating Activity Spaces

To represent how the user moves in a city such as Jiamusi, we build activity spaces (e.g. Axhausen et al. 2002; Axhausen 2007) that likely encompass a user's home, work, and "third places" (Ahas et al. 2009; Schneider et al. 2013). We choose a polygon method in order to represent the area surrounding the cell towers where the user is likely to be found, since he or she uses the nearby towers. This polygon will also likely encapsulate the areas that are convenient for a user to travel between work and home.

These activity spaces summarize a user's set of frequently visited points (e.g., cell towers) by an ellipse that encapsulates 68 % (i.e., one standard deviation) of points visited by capturing points that are concentrated in the center and neglecting sparse points in the periphery (as in Carrasco et al. 2006). Ellipses are first centered on the mean geometric center of a user's tower locations (mean of x coordinates and y coordinates; repeating values are allowed if a user visits towers more than once).

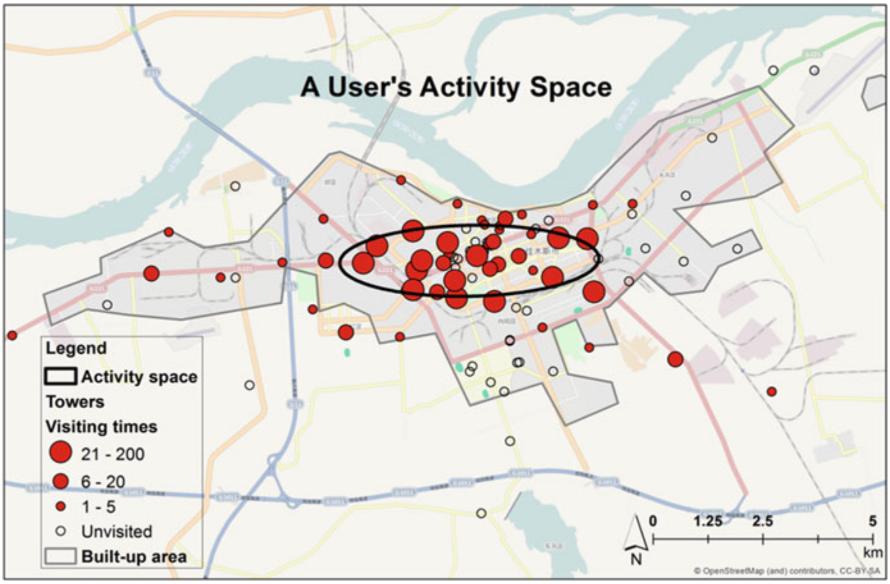

Fig. 13.3 An individual activity space is represented as an ellipse (in *black*). The ellipse captures a user's most frequently used towers, shown as *red circles*, where larger *red circles* indicate more frequent visits

The value of the standard deviation is calculated for all x coordinates to obtain an axis, and y coordinates to obtain a second, perpendicular axis. The ellipse is tilted in a direction that captures the major axis (long edge) of the distribution (see Mitchell 2005).

As mentioned, the ellipse does not typically encompass all visited towers and excludes those not frequently visited, to represent daily activity space (see Fig. 13.3). It captures the essence of the central tendency, dispersion, and direction of the user's travel patterns without including infrequent cell tower usage (such as a traveler's phone call from the airport).

13.3.2 Assessing Overlap of Activity Spaces

After each individual is assigned an activity space, we quantify the similarity between an ego's and an alter's activity spaces. A method for finding whether two activity spaces are similar is not straightforward. The percent overlap between two activity spaces will not account for how much physical area two friends' spaces share. Additionally, using the area that two activity spaces share does not tell us how big their spaces are, i.e., whether this shared area is actually "convenient" relative to their whole activity spaces. Also, with these two methods, we will not

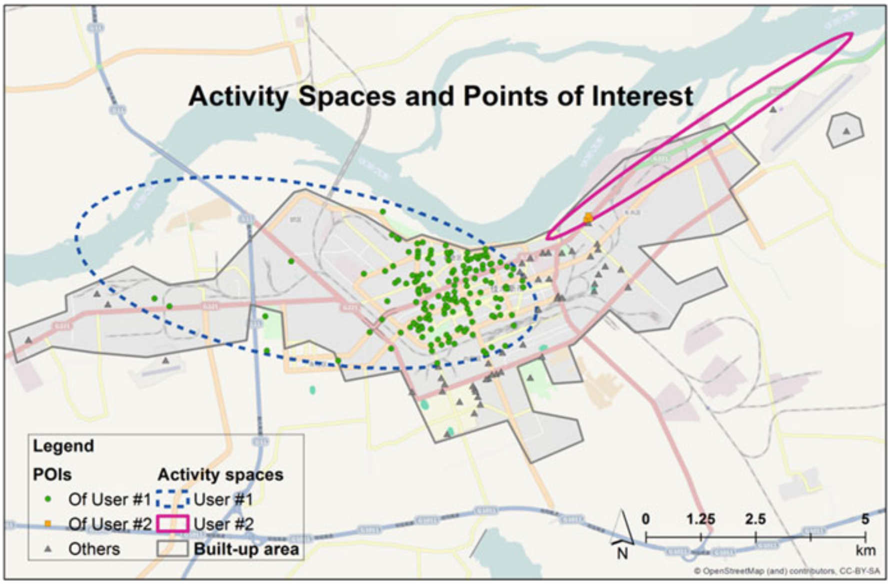

Fig. 13.4 Two single activity space ellipses in *dashed blue* and *solid pink* intersect underlying Points of Interest (POIs). The POIs intersected by the dashed ellipse are *green circles*. Those intersected by the *solid pink* ellipse are *orange squares*. POIs that do not intersect either ellipse are *grey triangles*

be able to understand what amenities and places for meeting each individual has in his or her activity space. We overcome these drawbacks by creating a third layer of relevant points, as suggested by geometric probability theory (Santaló 2004).

Agents' activity spaces are qualified by the points of interest (POIs) they spatially intersect (Fig. 13.4). These POIs are where "optional" activities are likely to occur (Gehl 1987), including services, transport, and recreational areas as a proxy for how agents use the city. POIs are landmarks for social interactions, as third places and the activities performed in third places have been shown to be essential for relationships, social health, and quality of life (Rosenbaum 2006). Friends may visit POIs to dine or to do business.

POIs are selected and digitized by the authors from *Google Maps* (2013) with data provided by *AutoNavi,* which operates under an open-use license. Although it is default to gauge how well this set of geographic information (VGI) reflects actual POIs in the city, it is used as a proxy for all POIs in the city (Coleman 2010; Neis and Zipf 2012). The digitized set of POIs are retrieved as a keyhole markup language (KML) file and analyzed in the Esri *ArcMap 10.1* environment. POIs include the city's recreation spots (including parks, internet cafés, personal wellness centers), commercial centers (restaurants and bars, stores, markets, and shopping centers), public services and institutions (hospitals, post offices, police stations), transportation centers (airports, train stations), and named villages in the suburban area. POIs of similar types (e.g., restaurants) in a 50 m radius are grouped into a single POI to eliminate redundant information.

As mentioned, two activity spaces are considered *linked* if their corresponding nodes are connected via the social network. Thus, friendships are embedded in geographic space through the two activity spaces. Each unique pair of linked activity spaces (LASs) is compared by the number of common POIs shared by both (in absolute number and percent of the user's total POIs). In Fig. 13.5, two pairs of linked activity space ellipses share POIs, where the pair demarcated with a solid blue line shares more POIs. In Fig. 13.6, an ego ellipse is in focus (in solid pink) and shares POIs to various extents with each of his or her alters (in dashed lines). We use a statistical t-test to compare friends' typical number of shared POIs versus that of a random pair of users. Our hypothesis is that friends share more POI points than random pairs.

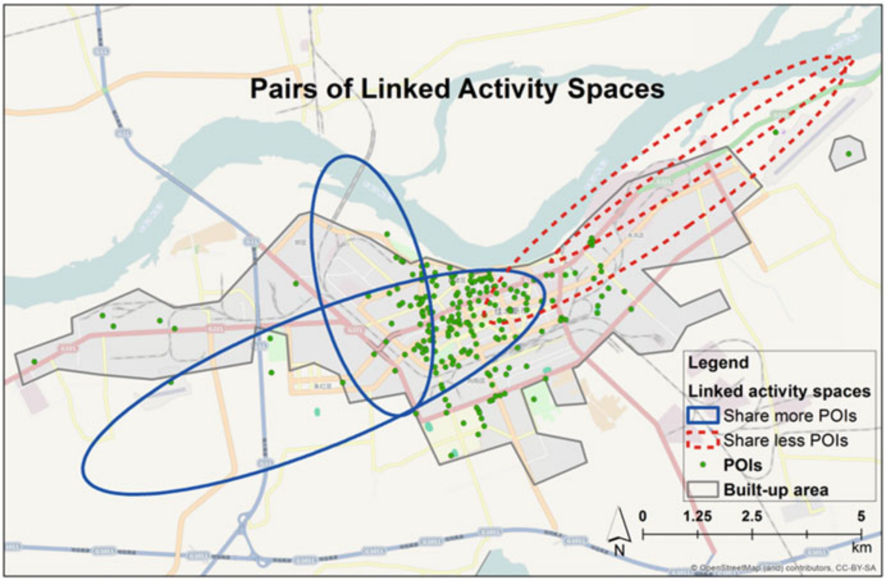

Fig. 13.5 This image shows how linked activity spaces overlap and can be quantified using the number of POIs found in the intersection of activity spaces. These POIs can be a proxy for convenience of shared meeting points. More specifically, this figure shows two pairs of linked activity spaces, in *solid blue* and *dashed red lines*. The pair of *solid blue* ellipses intersects more common POIs than the two *dashed red* ellipses. However, this set of ellipses has a particular place of intersection that is west of the most general POI cluster trend

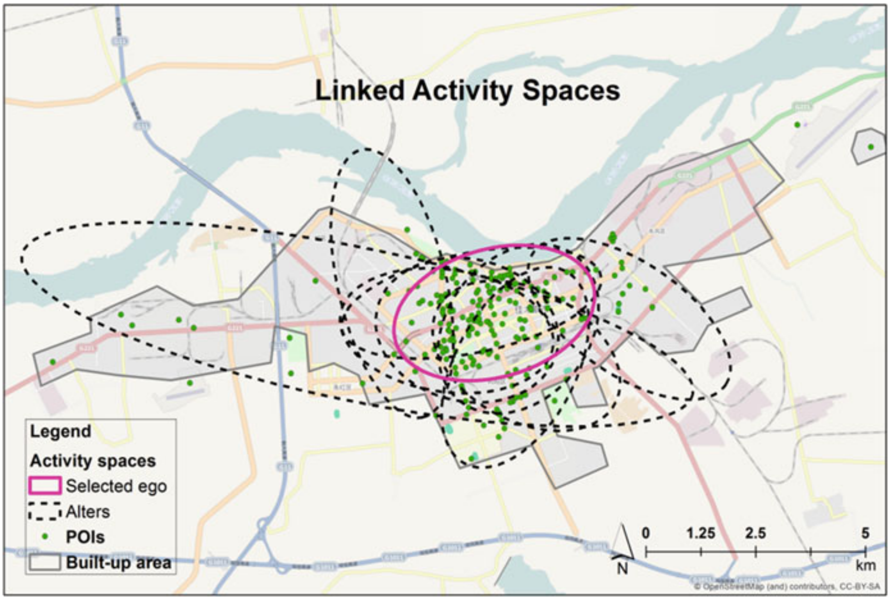

Fig. 13.6 An ego in focus (*solid pink*) shares POIs to various extents with each of his or her alters (*dashed lines*). The pattern shows that the focal ego and friends use the downtown area. One friend seems to use the Western area of the city more than the other friends

13.4 Results

13.4.1 Dyadic Relationships

We find that a pair of friends is more likely than a random pair to use the same places in Jiamusi. 11 % of friends and 50 % of random pairs share no POIs. Of pairs who share POIs, friend pairs share an average of 55.8 POIs, and a random pair shares 45.77 out of 212 total possible POIs (Fig. 13.7a). A t-test using the ellipse activity-space method yields a t-value of 56.03 (degree of freedom (df)) = 44,706, p-value < 0.001) and allows us to reject the null hypothesis that the mean difference between these two groups is insignificant, indicating that friends share more POIs than random pairs.

Second-degree friends also utilize urban infrastructure significantly more similarly, in terms of shared POIs, than random users (Fig. 13.7b). The t-value is 94.82 (df = 152,070, p-value < 0.001). The count of POIs shared by second-degree friends on average is 55.23 compared to 45.77 for random pairs.

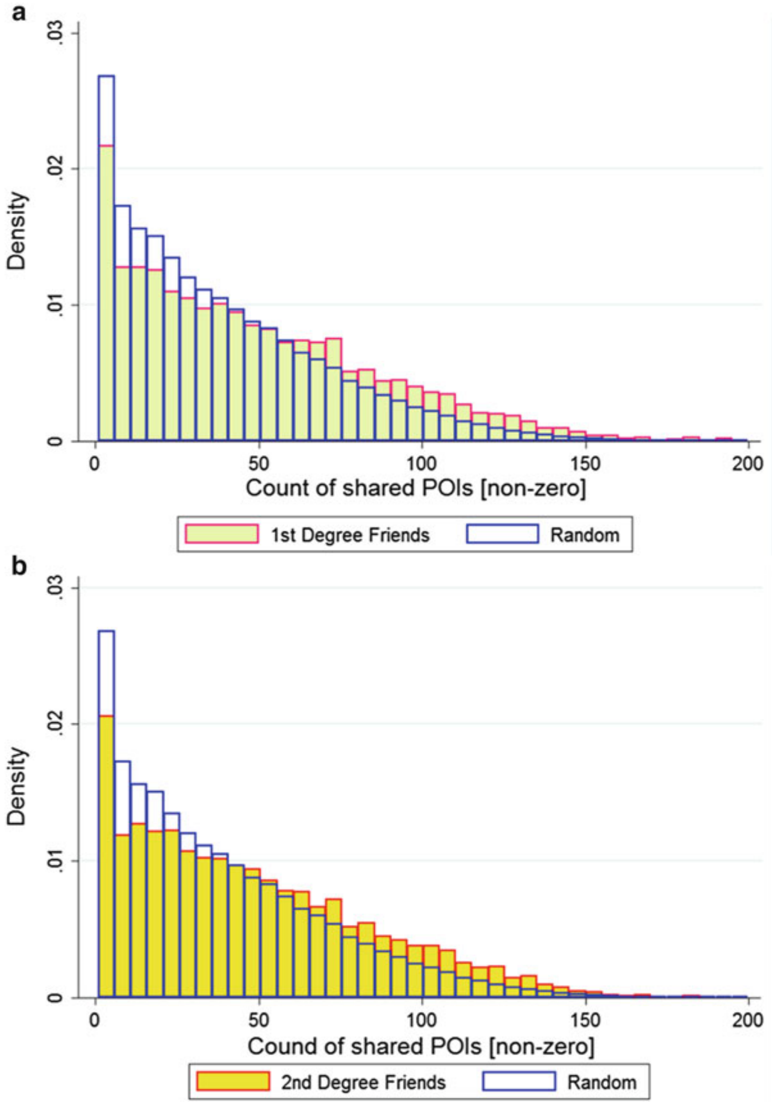

Fig. 13.7 (**a**) The probability density distribution of shared POIs of a random pair (in *blue*) and a linked pair (i.e. friends) (in *red*) show that friends have a greater probability of sharing POIs (as depicted by taller *red bars* towards the tail). We exclude pairs (random or linked) who share no common POIs. (**b**) The distribution of second-degree linked activity spaces shows that second-degree friends share more POIs on average than random pairs

Additionally, it is rare that an ego's alters will visit a POI that the ego does not visit. The *average* proportion of total egos who have visited a specific POI (e.g., "flower park") is proportional to the average percentage of their friends who have also visited the POI. For example, consider a group of 100 egos, each of whom has unique 50 alters (summing to a total of 5,000 friends). If an average of 50 % of one ego's alters (25 alters) have visited a certain POI, then there is a ½ chance the egos has visited there as well. If the average at another park is 20 % of all alters (10 alters), there is about a 1/5th chance that the egos has visited this park as well. The chance that an ego has visited a POI is *1.036 times* the number of total alters who have visited the POI, with an r^2 correlation of 0.968. This means there are few, if any, POIs where one frequents and his or her friends do not frequent. Conversely, there are also few POIs where one does not frequent yet a high percentage of his or her friends visit often.

13.4.2 Social Personas

In traditional social network analysis, a user's role in the network can reflect his or her importance and prominence in various facets of social life, such as providing information about new job opportunities to ones alters. For instance, a figure with a special role in a social network (i.e., a figure with many friends, or who is a "common friend" between poorly connected groups) can be identified through social network metrics such as betweenness centrality, degree centrality, or brokerage statistics (Jackson 2010). In one case, it has been shown that those with higher network centrality live in more central places on the Euclidean grid of longitude and latitude for the network (Onnela et al. 2011).

Confirming our second hypothesis, we find that high-degree users use the city center more often than expected, given a random set of users. The top 1 % of high-degree users (equating to users with 150 or more friends) concentrate at the urban center. A Fisher-Snedecor test (F-test) of ANOVA yields a p-value of 0.006 (95 % c.i.) signifying that the spatial variance between the high-degree users' activity spaces is significantly lower than the spatial variance between activity spaces of the universal population. This result illustrates high-degree agents' proclivity toward high-density areas that are shown to be more innovative, dynamic, and energetic environments (Bettencourt 2013).

We do not find significant spatial patterning with ego characteristics such as clustering coefficient (Jackson 2010), which measures whether one's friends are also friends themselves.

Using Spearman's correlation statistic, we find no significant relationship between user degree or total talking time (call duration) and the size of the user's activity space.

13.4.3 Community and City Form

We hypothesized that central areas play an enhanced role in supporting social communities of friends. The results do not confirm this hypothesis in the sense that the central area of Jiamusi does not favor tight-knit social circles but instead hosts heterogeneous groups of friends.

We define "groups" of friends (i.e., communities) in the Gephi environment (Bastian et al. 2009) with Blondel et al.'s (2008) community detection (modularity) algorithm that assigns each node (friend) to a cluster. This process produces 58 clusters with a modularity value of 0.732. Note that a modularity statistic of 1.0 indicates that communities are partitioned "perfectly," so that a node *i* does not connect with other nodes *j*, if *j* are not in *i*'s modularity group. Smaller modularity values indicate that connections across groups occur more frequently. Roughly, this indicates that, on average, those assigned to a cluster call within the same cluster 73.2 % of the time.

Each network agent (node) is assigned one modularity group. These agents are denoted by their ellipse centroid (geographic centers) in Fig. 13.8. Agents denoted by yellow squares or teal triangles (Fig. 13.8) are examples from two social network clusters with significant spatial clusters that differ from the overall distribution,

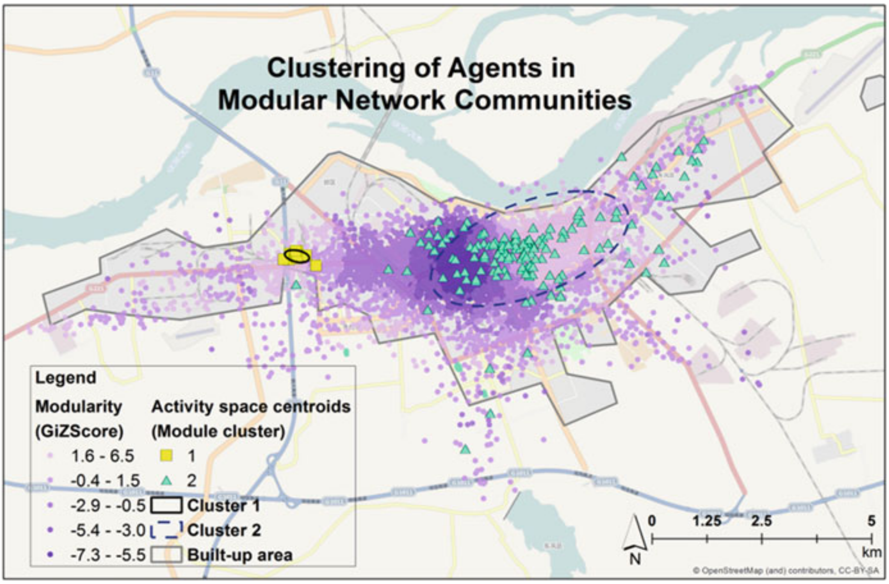

Fig. 13.8 A modularity algorithm is applied to the "aspatial" social network. Then, members of two separate modularity clusters are now mapped in *yellow squares* (denoting one group) or *teal triangles* (denoting a separate group) to show how social network groups use the city. In the downtown core, covered in *dark purple*, clusters are significantly mixed, meaning that many social groups use the downtown but are not confined to the region

deviating westward and north-eastward, respectively. The yellow squares and teal triangle clusters are examples of spatially embedded social groups that are *also* significantly clustered in a way that deviates from the expected spatial distribution of agents the city.

Including these two examples, the spatial distribution of 58 social clusters, in total, does form statistically significant "hot spots" (significantly dense clusters) and "cold spots" (a mixture of modular groups) as shown with the Getis-Ord Gi* statistic (Getis and Ord 1992). Hot spots (light purple areas in Fig. 13.8) contain agents of the same modular group in two major regions. Cold spots (dark purple areas in Fig. 13.8) cover the downtown, signifying that the groups that frequent the downtown are not clustered in the downtown, but have other group members around the city.

Another prominent pattern of community configuration, at a more local scale, is the prevalence of social triangles in the network. A social triangle can be defined as a group three nodes who connect to one another (Latapy 2008) and, pragmatically, will have meeting needs that are different and more complex than those of a dyad but perhaps not as complex as a modular group, which can contain many nodes. In our dataset, agents with the most social triangles cluster in the downtown area. However, this may be an artifact of the high-degree users' downtown, as they are likely to have more social triangles.

The distribution of high-triangle nodes (denoted in green circles and black stars in Fig. 13.9) follows a series of parallel roads downtown. Those with the highest

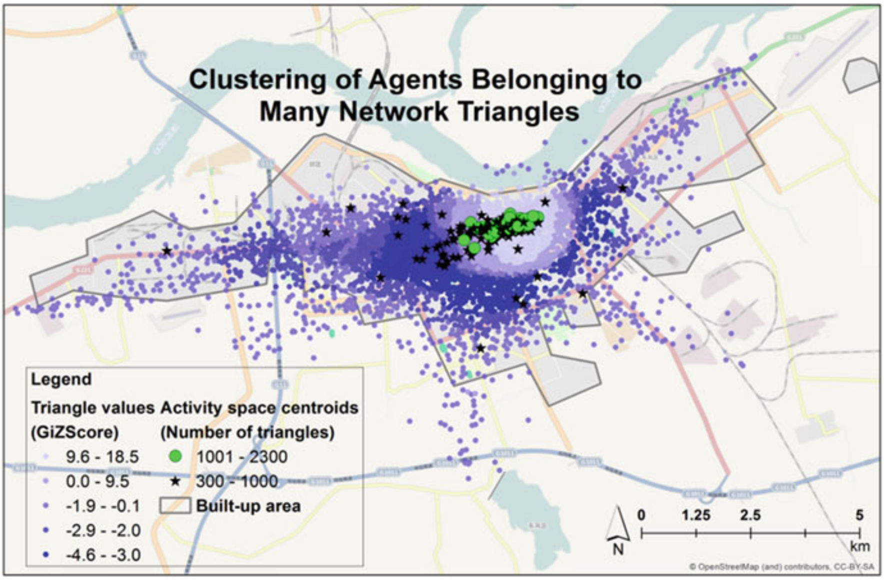

Fig. 13.9 A clustering of agents who are a part of many social triangles (i.e. groups of three friends) congregate toward the urban core (*green circles* and *black stars*). This clustering is statistically significant in one area east of downtown (denoted by *light to dark blue*), via statistical Gi* Z scores

number of social triangles (green circles as activity space centroids) form a tight linear cluster in the core area. This pattern is statistically significant, showing high Getis-Ord GI* Z scores with p-values > .0001 in the area slightly east of the downtown (in light color, Fig. 13.9). P-values are not significant in other parts of the city. Interestingly, centroids covering this neighborhood also saw the most significant modularity clusters (Fig. 13.8).

We also hypothesized that peripheral areas may be more mixed. We reject this hypothesis as the peripheral areas seem to host more cohesive communities than the downtown area. The POIs shared by linked activity spaces are more frequently on the periphery of the city. A POI can have from .01 % to 2.8 % of its pairs shown to be friends. A POI with a high value indicates that it is located in a convenient area for linked activity spaces (i.e. friends) to meet, where POIs with a lower value is more likely to host non-friend individuals. The POI hosting 2.8 % linked activity spaces is located near many popular hotels and the city's largest park – a notable tourist site in the city. This site might be a popular meeting spot for friends, but it may also be the artifact of many local business calls to one another in nearby buildings.

More generally, this ratio increases further from the city center (Figs. 13.10 and 13.11), so that POIs on the outskirts of the city are more likely to host pairs of contacts. This may be because many people have activity spaces that stretch into the downtown for work, but have their social contacts closer to their residences in another part of the city. Though agents with more contacts tend to frequent the dense downtown, given a POI with 100 random people, a user is more likely to find a friend within this 100 people if he or she is at a suburban POI, than an urban

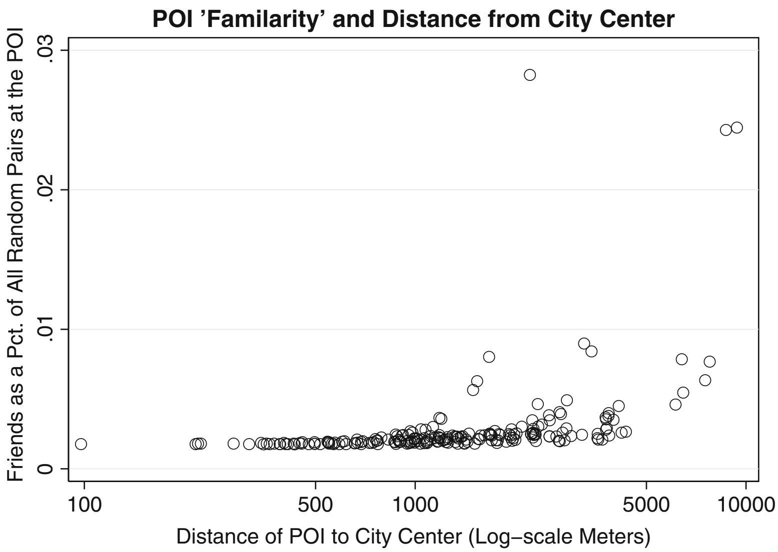

Fig. 13.10 As the POIs are located farther away from the city center there is a higher ratio of friend to non-friend pairs using these points

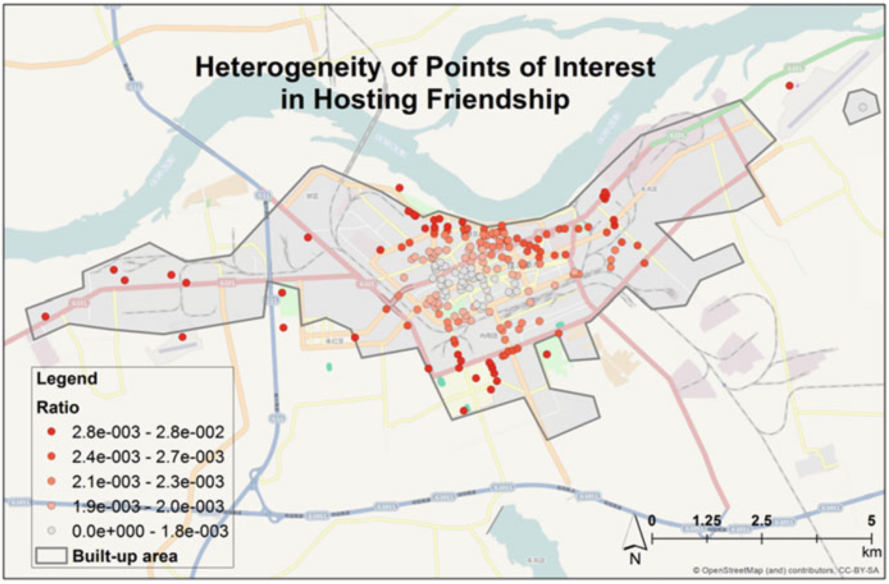

Fig. 13.11 The ratio of POI usage for friends vs. random pairs shows that friends are more likely to use the same POIs if they live on the periphery of the city. These POIs are denoted in *red*, where up to 2.8 % of their usage is linked with friendship

POI. In other words, although the downtown core attracts more people, the majority seem to be mutual strangers, while in the suburban area POIs serve as intentional meeting points. We interpret this finding with care, as there are fewer POI points on the periphery of the city, thus reducing the granularity and precision to capture activity spaces found in the city's outskirts. For instance, an activity space in the shape of a narrow line can be captured via the dense, granular points in downtown, but such a detailed structure could not be defined in the periphery, since there are so few cell towers and POIs to delineate a more precise activity space.

In summary, friends and even second-degree friends tend to use common points of interest in the city. Additionally, high-degree users (i.e., those with many friends) tend to be associated with downtown locations (central business district), but those with many social triangle friendships center in a neighborhood east of the central business district. The downtown core hosts many heterogeneous social groups instead of small tightly knit social clusters.

13.5 Discussion

We leveraged social/spatial data from call data records in a new way that emphasizes social relationships embedded in urban physical space. In this section, we respond to our initial hypotheses regarding how dyads (pairs), social network personas, triads (groups of three), and communities use the city.

First, the LAS method allows us to find the extent to which a pair of friends is more likely to use the same places in the city more than a random pair, so that a user i is more likely to have frequented the same POIs as a friend j than a random user r.

Second, we find that egos with high degrees are inclined toward the city center, while egos with high clustering coefficients show no significant spatial correlation. These results are not necessarily intuitive, as community members in suburban areas might also have high degrees, but do not seem to. Also, users who cling to the urban core (or tight suburban neighborhoods) might also be expected to be part of a number of "cliques" or friend groups; however, this also does not seem to be the case.

Third, we expected that central areas would play an enhanced role in supporting "clique-like" and modular groups, but we found, counter to our expectation, that the downtown was indeed a mixing pot for many groups. We do, however, find two specific neighborhoods that tend to harbor enclosed ("clique-like") social groups. Moreover, triads of friends are likely to use the downtown area.

Also, we had expected that peripheral areas were more mixed as suburbanites often have more access to automobiles and, thus, may not choose to live next to their contacts if they can drive to other parts of the city to visit. Counter to our initial hypothesis, peripheral areas show less frequent mixing of social groups and friends than any other part of the city.

13.5.1 Utility

We find these results useful for theoretical and practical issues in planning. First, in Jiamusi, we find that the ratio of friends' shared POIs to random pairs' shared POIs is 59:46. This ratio can be considered an indicator of clustered socialization. When high, this ratio shows that friends tend to group in certain parts of the city, or use the same amenities in a city. This ratio represents an urban feature that can be compared across cities, over time, and as it correlates with urban features such as population, crime rates or traffic.

Next, a planner can use these findings, for example, to discuss the merits of different models of urbanism. For example, while New Urbanism focuses on neighborhood design, architectural style, and transit-oriented development for high-density walkable cities (Al-hindi and Till 2001; Vanderbeek and Irazabal 2007), and Landscape Urbanism argues that urban design should be flexible and open-ended, (Waldheim 2002), by leveraging existing resources (Cranz and Boland 2004) and preserving wilderness (Yu et al. 2011). The LAS method results can be used to probe the adverse consequences of urban sprawl, such as its challenges for social life (Gehl 1987).

This analysis can be used to plan the location of third places (Rosenbaum 2006), such as restaurants, parks, coffee shops, theaters, and other facilities. Locations could be found by determining places that are convenient for pairs or groups of friends to meet, and combining this with other criteria, such as low traffic or areas known to be safe for pedestrians.

This method can also be used to understand the size, temporal persistence, and location (thus, level of accessibility to other places, environmental quality of the land) of certain places, such as ethnic or working neighborhoods (such as homes near a factory), where members form a dense group of ties – e.g., neighbors are likely to know and depend on one another. The method can show where these neighborhoods are and how they expand and contract over time. This can be useful for investigations of urban social capital (Granovetter 1983), cultural assimilation (De Blij and Murphy 1986), or models of epidemiology or idea spreading.

It is clear that we are only at the beginning of understanding how interpersonal relationships manifest themselves in the built environment. Yet it is a phenomenon; we experience daily as we meet colleagues at work, family at home, and perhaps friends in third places. The tension between the costs of movement in cities and the need for access to the possibilities of the city also guide our decisions about raising families through the choice of neighborhoods and school districts, as well as migration, through the choice of leaving established social circles for new circles (or vice versa).

13.6 Conclusion

A healthy city is built on strong social networks (Gilchrist 2009), but we still do not know what kinds of ties exist in cities and neighborhoods nor the detailed social dynamics that creates and changes them. Because of these limitations, we cannot currently use social network structures (clustered, decentralized, hub-spoke, etc.) as cause or effect variables in assessing planning choices for new or existing neighborhoods and cities. However, as this type of data becomes richer, such studies will become increasingly possible,

Our ability to socialize with others is affected by urban planning and government decisions regarding low-income housing, immigration reform, and health codes, such as the number of people to an urban residence or the choice of building subdivisions versus condominiums (Farber and Li 2013) or narrow versus wide roads (Montgomery 2013). The spatial and social clustering of ties changes with the creation and dissolution of institutions, such as firms, universities, military bases, sports franchises, and religious institutions. Less socialization may also lead to stagnated mobility, not due to a lack of accessibility, but to a reduced need for third places to socialize (Rosenbaum 2006), and visit others' homes.

Urban planners, geographers, government officials, civil engineers, and transportation planners that focus on improving social life in their city may be able to more directly improve residents' quality of life (Cacioppo and Patrick 2008), more so than traditional national level economic stimuli (Montgomery 2013). Instead, planners and geographers have worked toward better urban environments investigating residents' accessibility to amenities (such as hospitals), travel time to work, and social justice issues such as susceptibility to industrial and environmental hazards (Cutter et al. 2003). In addition to these variables, we should emphasize the importance of social capital within the city.

We cannot infer these social patterns from city form, or social networks alone as the connection between these variables is statistical and likely scale dependent. Thus, our task with this chapter was to illustrate how the linked activity space method allows for the integration of information from a social network into geographic space, a combination that is rarely investigated in detail (Andris 2011).

Although we use a call dataset record (CDR) for our analysis, this method can be employed to any dataset that has both evidence of social ties between agents and the geo-location of the agents. Other mechanisms for telecommunications (such as Skype, Google Video/Chat, Viber, and Whats-App) can be substituted for mobile phone calls. If these data are available, it may be worth considering the combination of the CDR dataset for interesting results on which modes of communication are popular in general, or in certain parts of the city, or during certain time frames. We may be able to capture the growth of one mode over another, over a longer time period. One exciting prospect is to see which neighborhoods make more international calls, or calls to other cities.

We do find a number of methodological and pragmatic challenges to this type of research. Many of these challenges stem from the nascent state of big data analysis that will perhaps become more reliable and complete in the future. Nevertheless, there are issues with these data that can be addressed: CDR datasets do not capture an ego with alters who do not appear in the social network. Multiple cell phones per person and multiple people per cell phones do not ensure that the telephone number is a proxy for an individual's communication patterns. Without figures on the provider's market penetration rate is difficult to understand friendships via calls to users who use a different provider. We also note a number of subjective decisions in creating a meaningful sample, such as the minimum number of towers frequented in order to be included in the dataset, the number of seed users, and number of friend "levels" to draw from the networks. None of these issues is a fundamental limitation, so we look forward to future datasets that can overcome some of these difficulties. We hope to see more research on the integration of social networks and urban spaces in the future as a unique window into how urban form and social function shape each other.

Acknowledgments This research was partially supported by the Army Research Office Minerva Program (grant no. W911NF-121A −0097), the John Templeton Foundation (grant no. 15705), the Bill and Melinda Gates Foundation (grant no. OPP1076282), the Rockefeller Foundation, the James S. McDonnell Foundation (grant no. 220020195), the National Science Foundation (grant no. 103522), the Bryan J. and June B. Zwan Foundation, and the University of Georgia.

References

Ahas R, Silm S, Saluveer E, Järv O (2009) Modelling home and work locations of populations using passive mobile positioning data. In: Gartner G, Rehrl K (eds) Location-based services and telecartography II, Lecture notes in geoinformation and cartography. Springer, Heidelberg/Berlin, pp 301–315

Albert R, Barabási A-L (2002) Statistical mechanics of complex networks. Rev Mod Phys 74:7

Al-hindi KF, Till KE (2001) (Re)placing the new urbanism debates: toward an interdisciplinary research agenda. Urban Geogr 22:189–201

Andris C (2011) Methods and metrics for social distance. Dissertation, Massachusetts Institute of Technology

Arentze T, van den Berg P, Timmermans H (2012) Modeling social networks in geographic space: approach and empirical application. Environ Plan A 44:1101–1120

Axhausen KW (2007) Activity spaces, biographies, social networks and their welfare gains and externalities: some hypotheses and empirical results. Mobilities 2:15–36

Axhausen KW, Zimmermann A, Schönfelder S, Rindsfüser G, Haupt T (2002) Observing the rhythms of daily life: a six-week travel diary. Transportation 29:95–124

Bastian M, Heymann S, Jacomy M (2009) Gephi: an open source software for exploring and manipulating networks. In: Proceedings of the international conference on weblogs and social media (ICWSM), San Jose, CA, 17–20 May 2009. AAAI Press, Menlo Park

Bettencourt LMA (2013) The origins of scaling in cities. Science 340:1438–1441

Blondel V, Guillaume J-L, Lambiotte R, Lefebvre E (2008) Fast unfolding of communities in large networks. J Stat Mech Theor E 10:P10008

Cacioppo JT, Patrick W (2008) Loneliness: human nature and the need for social connection. WW Norton & Company, New York

Calabrese F, Smoreda Z, Blondel V, Ratti C (2011) Interplay between telecommunications and face-to-face interactions: a study using mobile phone data. PLOS ONE 7:e20814

Carrasco J A, Miller E, Wellman B (2006) Spatial and social networks: the case of travel for social activities. Paper presented at the 11th international conference on travel behavior research, Kyoto, Japan, 16–20 August 2006

Centola D, Macy M (2007) Complex contagions and the weakness of long ties. Am J Sociol 113:702–734

Cho E, Myers SA, Leskovec J (2011) Friendship and mobility: user movement in location-based social networks. In: Proceedings of the 17th ACM SIGKDD international conference on knowledge discovery and data mining (KDD), San Diego, CA, 21–24 August 2011. ACM, New York

Christakis N, Fowler J (2007) The spread of obesity in a large social network over 32 years. New Engl J Med 357:370–379

Coleman DJ (2010) Volunteered geographic information in spatial data infrastructure: an early look at opportunities and constraints. Paper presented at the GSDI 12 world conference, Singapore, 19–22 October 2010

Crandall DJ, Backstrom L, Cosley D, Suri S, Huttenlocher D, Kleinberg J (2010) Inferring social ties from geographic coincidences. Proc Natl Acad Sci U S A 107:22436–22441

Cranz G, Boland M (2004) Defining the sustainable park: a fifth model for urban parks. Landsc J 23:102–120

Cutter SL, Boruff BJ, Shirley WL (2003) Social vulnerability to environmental hazards. Soc Sci Q 84:242–261

De Blij HJ, Murphy AB (1986) Human geography: culture, society and space. Wiley, New York

Eagle N, Pentland AS, Lazer D (2009) Inferring friendship network structure by using mobile phone data. Proc Natl Acad Sci U S A 106:15274–15278

Emch M, Root ED, Giebultowicz S, Ali M, Perez-Heydrich C, Yunus M (2012) Integration of spatial and social network analysis in disease transmission studies. Ann Assoc Am Geogr 102:1004–1015

Farber S, Li X (2013) Urban sprawl and social interaction potential: an empirical analysis of large metropolitan regions in the United States. J Trans Geogr 31:267–277
Frei A, Axhausen KW (2011) Modeling spatial embedded social networks. Working paper 685: transport and spatial planning, IVT, ETH Zurich, Zurich, Switzerland
Gao S, Wang Y, Gao Y, Liu Y (2013) Understanding urban traffic-flow characteristics: a rethinking of betweenness centrality. Environ Plan B 40:135–153
Gehl J (1987) Life between buildings: using public space (trans: Koch J). Van Nostrand Reinhold, New York
Getis A, Ord JK (1992) The analysis of spatial association by use of distance statistics. Geogr Anal 27:286–306
Gilchrist A (2009) The well-connected community: a networking approach to community development, 2nd edn. Policy Press, London
Girardin F, Calabrese F, Dal Fiore F, Ratti C, Blat J (2008) Digital footprinting: uncovering tourists with user-generated content. IEEE Pervas Comput 7:36–43
Golledge RG (1999) Human wayfinding and cognitive maps. In: Golledge R (ed) Wayfinding behavior: cognitive mapping and other spatial processes. Johns Hopkins University Press, Baltimore, pp 5–45
Granovetter M (1973) The strength of weak ties. Am J Sociol 78:1360–1380
Granovetter M (1983) The strength of weak ties: a network theory revisited. Soc Theor 1:201–233
Jackson MO (2010) Social and economic networks. Princeton University Press, Princeton
Kang C, Ma X, Tong D, Liu Y (2012) Intra-urban human mobility patterns: an urban morphology perspective. Physica A 391:1702–1717
Kurant M, Markopoulou A, Thiran P (2011) Towards unbiased BFS sampling. IEEE J Sel Area Commun 29:1799–1809
Latapy M (2008) Main-memory triangle computations for very large (sparse (power-law)) graphs. Theor Comput Sci 407:458–473
Mitchell A (2005) The ESRI guide to GIS analysis, vol 2. ESRI Press, Redlands
Montgomery C (2013) Happy city: transforming our lives through urban design. Farrar, Straus and Giroux, New York
Neis P, Zipf A (2012) Analyzing the contributor activity of a volunteered geographic information project – the case of OpenStreetMap. Int J Geo-Inf 1:146–165
Newman ME, Park J (2003) Why social networks are different from other types of networks. Phys Rev E 68:036122
Onnela JP, Arbesman S, González M, Barabási AL, Christakis N (2011) Geographic constraints on social network groups. PLoS ONE 6:e16939
Papachristos A, Hureau D, Braga A (2013) The corner and the crew: the influence of geography and social networks on gang violence. Am Sociol Rev 78:417–447
Papas MA, Alberg AJ, Ewing R, Helzlsouer KJ, Gary TL, Klassen AC (2007) The built environment and obesity. Epidemiol Rev 29:129–143
Radil SM, Flint C, Tita G (2010) Spatializing social networks: using social network analysis to investigate geographies of gang rivalry, territoriality, and violence in Los Angeles. Ann Assoc Am Geogr 100:307–326
Ratti C, Williams S, Frenchman D, Pulselli RM (2006) Mobile landscapes: using location data from cell phones for urban analysis. Environ Plan B 33:727–748
Reades J, Calabrese F, Sevtsuk A, Ratti C (2007) Cellular census: explorations in urban data collection. IEEE Pervas Comput 6:30–38
Rosenbaum MS (2006) Exploring the social supportive role of third places in consumers' lives. J Serv Res 9:59–72
Salganik MJ, Watts DJ (2008) Leading the herd astray: an experimental study of self-fulfilling prophecies in an artificial cultural market. Soc Psychol Q 71:338–355
Santaló LA (2004) Integral geometry and geometric probability. Cambridge University Press, Cambridge

Scellato S, Noulas A, Lambiotte R, Mascolo C (2011) Socio-spatial properties of online location-based social networks. In: Proceedings of the international conference on weblogs and social media, Barcelona, 17–21 July 2011, pp 329–336
Schneider CM, Belik V, Couronné T, Smoreda Z, González MC (2013) Unravelling daily human mobility motifs. J R Soc Interface 10:20130246
Schönfelder S, Axhausen KW (2003) Activity spaces: measures of social exclusion? Arbeitsbericht Verkehrs- und Raumplanung, 140, Institut für Verkehrsplanung und Transportsysteme (IVT), ETH Zürich, Zürich
Sevtsuk A, Mekonnen M (2012) Urban network analysis: a new toolbox for ArcGIS. Rev Int Géomatique 22:287–305
Sevtsuk A, Huang S, Calabrese F, Ratti C (2009) Mapping the MIT campus in real time using WiFi. In: Foth M (ed) Handbook of research on urban informatics: the practice and promise of the real-time city. IGI Global, Hershey, pp 326–337
Sun Y, Fan H, Helbich M, Zipf A (2013) Analyzing human activities through volunteered geographic information: using Flickr to analyze spatial and temporal pattern of tourist accommodation. In: Krisp JM (ed) Progress in location-based services. Springer, Heidelberg/Berlin, pp 57–69
Vanderbeek M, Irazabal C (2007) New Urbanism as a New Modernist movement: a comparative look at Modernism and New Urbanism. Tradit Dwell Settl Rev 19:41–57
Waldheim C (2002) Landscape urbanism: a genealogy. Praxis 4:10–17
Yu K, Wang S, Li D (2011) The negative approach to urban growth planning of Beijing, China. J Environ Plan Man 54:1209–1236