Transactions in GIS

Research Article

Transactions in GIS, 2013, 17(3): 463–481

Discovering Spatial Interaction Communities from Mobile Phone Data

Song Gao,* Yu Liu,† Yaoli Wang‡ and Xiujun Ma§

**Department of Geography, University of California, Santa Barbara*
†Institute of Remote Sensing and Geographical Information Systems, Peking University
‡Department of Geography, University of Georgia
§Key Laboratory of Machine Perception, Peking University

Abstract

In the age of Big Data, the widespread use of location-awareness technologies has made it possible to collect spatio-temporal interaction data for analyzing flow patterns in both physical space and cyberspace. This research attempts to explore and interpret patterns embedded in the network of phone-call interaction and the network of phone-users' movements, by considering the geographical context of mobile phone cells. We adopt an agglomerative clustering algorithm based on a Newman-Girvan modularity metric and propose an alternative modularity function incorporating a gravity model to discover the clustering structures of spatial-interaction communities using a mobile phone dataset from one week in a city in China. The results verify the distance decay effect and spatial continuity that control the process of partitioning phone-call interaction, which indicates that people tend to communicate within a spatial-proximity community. Furthermore, we discover that a high correlation exists between phone-users' movements in physical space and phone-call interaction in cyberspace. Our approach presents a combined qualitative-quantitative framework to identify clusters and interaction patterns, and explains how geographical context influences communities of callers and receivers. The findings of this empirical study are valuable for urban structure studies as well as for the detection of communities in spatial networks.

1 Introduction

In geographic information science and regional geography there is a tradition of research on spatial interaction processes between different sub-regions, based on population immigration, commuting travel activities, and disease diffusion (Clarke 1996, Giuliano and Small 1993, Goodchild and Janelle 1984, Guo 2007, Jang and Yao 2011, Roy and Thill 2003, Tobler 1976, 1988). For regional studies, the functional region is defined by regional geographers based on interaction between its distinctive land-use zones (Johnston et al. 1981). Representative forms of interaction between different zones include human movement, commodity flow, resource allocation, and information communication. For the past several decades, studies of spatial interaction processes have mainly been based on census datasets (Jang and Yao 2011, Rae 2009).

In the age of Big Data, the widespread use of location-awareness devices (LAD), such as mobile phones and GPS-enabled devices, has made it possible to collect large-scale individual trajectories for analyzing human activity patterns and spatial interactions in both physical

Address for correspondence: Xiujun Ma, Key Laboratory of Machine Perception (MOE), Peking University, Beijing, China. E-mail: maxj@cis.pku.edu.cn
Acknowledgements: This research is supported by National Natural Science Foundation of China (Grant No. 41271386). We would like to thank Professor Michael F. Goodchild for discussion, advice and critical comments on the revision. We also appreciate insightful comments and suggestions from the reviewers.

 doi: 10.1111/tgis.12042

space and cyberspace. Mobile communications have shifted social ties from "linking people-in-places to linking people wherever they are and from place-to-place to person-to-person" (Kwan 2007). It is valuable for researchers and policy makers to understand how regional functions interact with human movements and telecommunication flows derived from these new information sources, and to understand in turn how these information and communication technologies (ICT) influence urban space (Batty and Miller 2000, Foss and Couclelis 2009, Ren and Kwan 2007, Shaw and Yu 2009, Torrens 2008, Yu and Shaw 2008, Yuan et al. 2012). The effect of ICT on human movements includes three types: substitution, stimulation, and modification (Mokhtarian and Meenakshisundaram 1999). Townsend (2000) has argued that the use of mobile phones leads the intensification of urban activity and productivity. Some researchers have employed mobile phone data for regionalization (Ratti et al. 2010), for discovering various types of social divides (Walsh and Pozdnoukhov 2011), and for finding linguistic separation (Expert et al. 2011). Other studies have demonstrated that mobile subscribers' movements are determined by geographical and socio-economic factors (Kang et al. 2010, 2012b, Leskovec and Lang 2008, Onnela and Arbesman 2011). Recently, more attention has been placed on deriving temporal characteristics of feature types from location-based services (Ye et al. 2011) and on extracting land-use from spatio-temporal trajectories (Liu et al. 2012b). Nevertheless there are still some important questions that deserve attention: for example, what is the relationship between interaction in the physical world and that in cyberspace, how to exploit and analyze patterns of spatial interaction, and how to design efficient economical and administrative boundaries based on spatial interaction?

Many analyses of spatial interaction datasets have been based on predefined areal units. Tobler (1976), for example, has analyzed state-to-state college attendance, and Rae (2009) has analyzed flows of immigrants. Other researchers have developed various techniques and models to find patterns in spatial interaction data (Adrienko and Adrienko 2011, Fischer et al. 1993, Roy and Thill 2003, Young 2002). In addition, researchers have taken the spatial interaction network as a graph in which areal zones are transformed into nodes while interaction flows embedded in space are represented by weighted edges. A number of graph partitioning methods found in the literature of complex networks and computer science have been applied to research on community detection or pattern discovery. Tobler's First Law (TFL) of geography says that near things are more related than distant things (Tobler 1970). Based on this proposition, it can be inferred that the strength of spatial interaction between communities is subject to geographic constraints which can be discovered through the use of the community-detection algorithms of social network analysis. However, popular methods for social network analysis rarely consider the relationship between physical space and cyberspace, although some research has taken spatial effects into account (Eagle et al. 2009, Expert et al. 2011, Guo 2009, Guo and Wang 2011, Guo et al. 2012, Onnela and Arbesman 2011). Goodchild et al. (2000) mention that we should demand an explicit accounting for spatial interaction and spatial dependence in empirical models of social networks. Spatialized social network analysis (SSNA) facilitates the understanding of social behaviors that relate to the structure of the network as well as to relative location-context in physical space. For example, Radil et al. (2012) use SSNA to investigate the network structure and geographical context of gang violence. Thiemann et al. (2010) analyze a human travel network generated by the circulation of banknotes, and find the effective boundaries partially overlap with existing administrative borders and also physical barriers like rivers and mountains. Barthélemy (2011) thoroughly explains the related work of understanding how the spatial constraints affect the structure and properties of these spatial networks.

Table 1 Data format of mobile-phone call-detail records

User	Receiver	Date	Start Time	End Time	Duration (seconds)	Base Station	Long	Lat
serveNub1	oppNub1	2007-07-23	09:25:10	09:28:20	190	A	127.495	50.243
serveNub1	oppNub2	2007-07-23	12:15:32	12:15:52	20	B	127.502	50.241

In general, the spatial effects on networks include: (1) spatial constraints on the connectivity patterns of nodes embedded in geographical locations; (2) physical networks like roads and railways, which are affected by spatial topology; and (3) restrictions on long-distance links due to economic costs.

In this study, we attempt to explore the community patterns hidden in the geographically embedded telecommunication network of phone-call interaction and the network of phone-users' movements. We adopt a hierarchical agglomerative clustering algorithm based on a Newman-Girvan modularity metric and an alternative modularity function incorporating a gravity model to study the dynamics of spatial interaction communities. Thus, the goal of this article is to discover communities from the networks of spatial interactions revealed by phone-call records and to use geographical context to explain these patterns. To this end, we geo-reference the detected spatial communities in Google Earth, further analyze some spatial communities of interest, and discover previously unknown spatially coherent regions and isolated enclaves. The empirical studies show how geographical context influences patterns of spatial interaction. To the best of our knowledge, this is the first effort to connect and compare the community patterns embedded in a phone-call dataset in cyberspace with phone-users' movements in the physical world.

The remainder of the article is organized as follows. In Section 2, we introduce the mobile phone datasets of the study area and preprocessing procedures for extracting networks of spatial interaction flow, as well as dynamic spatial structures of phone-call activities. In Section 3, we describe the method of community detection, and specifically the graph-partitioning algorithms. In Section 4, we analyze the resulting urban communities and compare patterns of call interaction and movement interaction. In Section 5, we summarize the main conclusions and discuss further work.

2 Data Description and Processing

2.1 Mobile Phone Data

There are mainly two types of mobile-phone datasets: individual call detail records (CDR) and aggregated call-volume data. On the one hand, a set of anonymized CDR which contains call information of phone user, receiver, base-station locations, date, time, duration, etc. (Table 1) is maintained by mobile network operators. Every time a call is made, each phone is geo-referenced to a nearby mobile base station that has a unique *longitude/latitude position*, which can be used to estimate the location of the phone user and further quantify human mobility (González et al. 2008, Kang et al. 2010, 2012b, Song et al. 2010). On the other hand, the aggregated data represents the total call volume of each mobile base station in different time intervals, such as hourly, daily, or weekly. This type of call volume data has been collected to

Table 2 Volumes of records on different days of a week

Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
11,190,182	10,895,508	10,922,212	10,706,807	11,019,933	9,829,428	9,443,295

understand the spatio-temporal structures of urban activities, population distribution and traffic flow (Gao et al. 2013, Kang et al. 2012a, Ratti et al. 2006).

In this research, the dataset is the first type listed above, and includes a week of over 74,000,000 phone call records (as shown in Table 2) of nearly 1,000,000 mobile subscribers in Harbin, a large city in northeastern China. The coverage area of each mobile base station can be approximated as a Thiessen polygon for call activity analysis and termed a "cell". In this Thiessen partition, all phone calls within a given polygon are closer to the corresponding mobile base station than to any other station. Generally, urban core areas have a higher density of mobile cells where the average distance between mobile base stations is approximately one kilometer (the value of average separation depends on the size of the sampling area). The average distance between base stations in the whole study area is 7.95 km. Each geo-referenced phone call which connects the locations of two mobile phone users could be represented as an origin-destination (OD) pair in geographical space or as a link in a network of telecommunication interaction (*TeleFlow)*. Also, we can get corresponding positioning movements (*MoveFlow*) of each mobile subscriber by connecting his or her series of geo-referenced call records. Therefore, based on the phone-call activities, two types of spatial interaction processes could be extracted (see Figure 1).

2.2 Interaction Data Representation

Let *G_TeleFlow (V, E)* be a weighted-undirected network graph of phone call flows where Thiessen polygons of mobile base stations are transformed into nodes *(V)* while interactions among stations are represented by weighted edges *(E)*. Let W_{ijt} represent the total call flow between cell i and cell j during time interval t (hourly, daily, weekly). Similarly, let *G_MoveFlow (V, E)* be a weighted undirected network graph of human movements and let M_{ijt} represent the total movement flow between cell i and cell j during time interval t, including movement flows both from i to j and from j to i. Note that we can also get the weighted directed network graph of call activity interactions by adding the direction of flow. But in this research we only focus on the strength of spatial interaction between two zones without direction. Figure 2 shows the two networks of interaction embedded in geographical space. It is difficult to clearly interpret the patterns directly from these networks since edges intersect and overlap. Thus grouping or graph-partitioning tools need to be employed to discover spatial-interaction communities.

2.3 Distance Decay of Spatial Interactions

Inspired by previous literature about human mobility patterns following scaling laws (Brockmann et al. 2006), we examine the distance distributions of network of call interactions *G_TeleFlow (V, E)* and network of movements *G_MobileFlow (V, E)* to determine if both distributions showed the characteristics of distance decay. We find that nearly 90% of all interac-

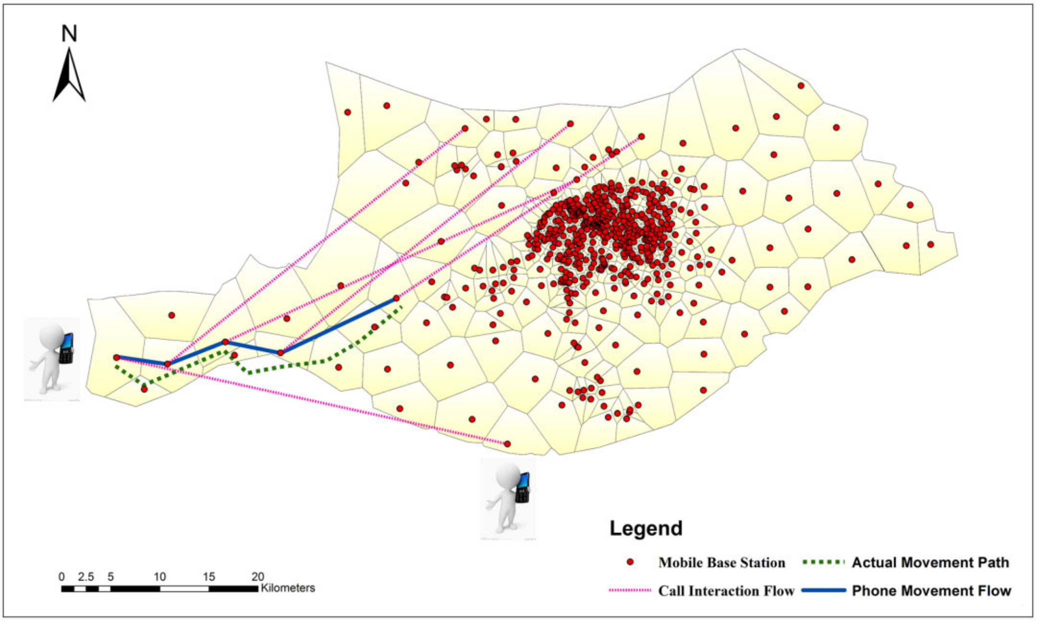
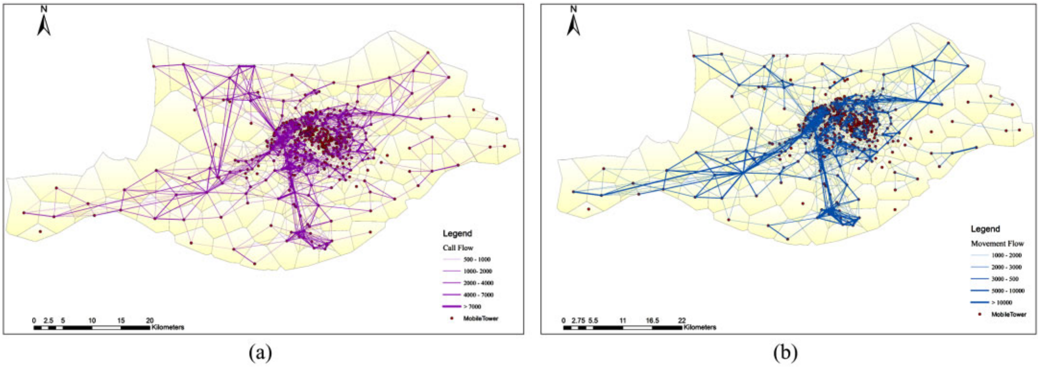


Figure 1 Spatial distribution of mobile base stations and an illustration of the two types of interaction flows extracted from mobile phone CDR



(a) (b)

Figure 2 (a) Visualizing the graph of phone-call interaction; and (b) Visualizing the movements of phone users in physical space. Each dot represents a node of a Thiessen polygon while a link implies an interaction between two polygons representing the coverage of mobile base stations. Note that small volumes of flow are not visualized

tions occur across distances less than 20 km (Figure 3a). We fit a power law which has the scale-free characteristics that can be better interpreted for different spatially embedded interaction networks with varying sizes:

$$P \propto d^{-\beta} \tag{1}$$

where P denotes the probability of having an occurrence of spatial interaction at the distance d between two localities, and β is the decay parameter (Taylor 1971). The results show that the

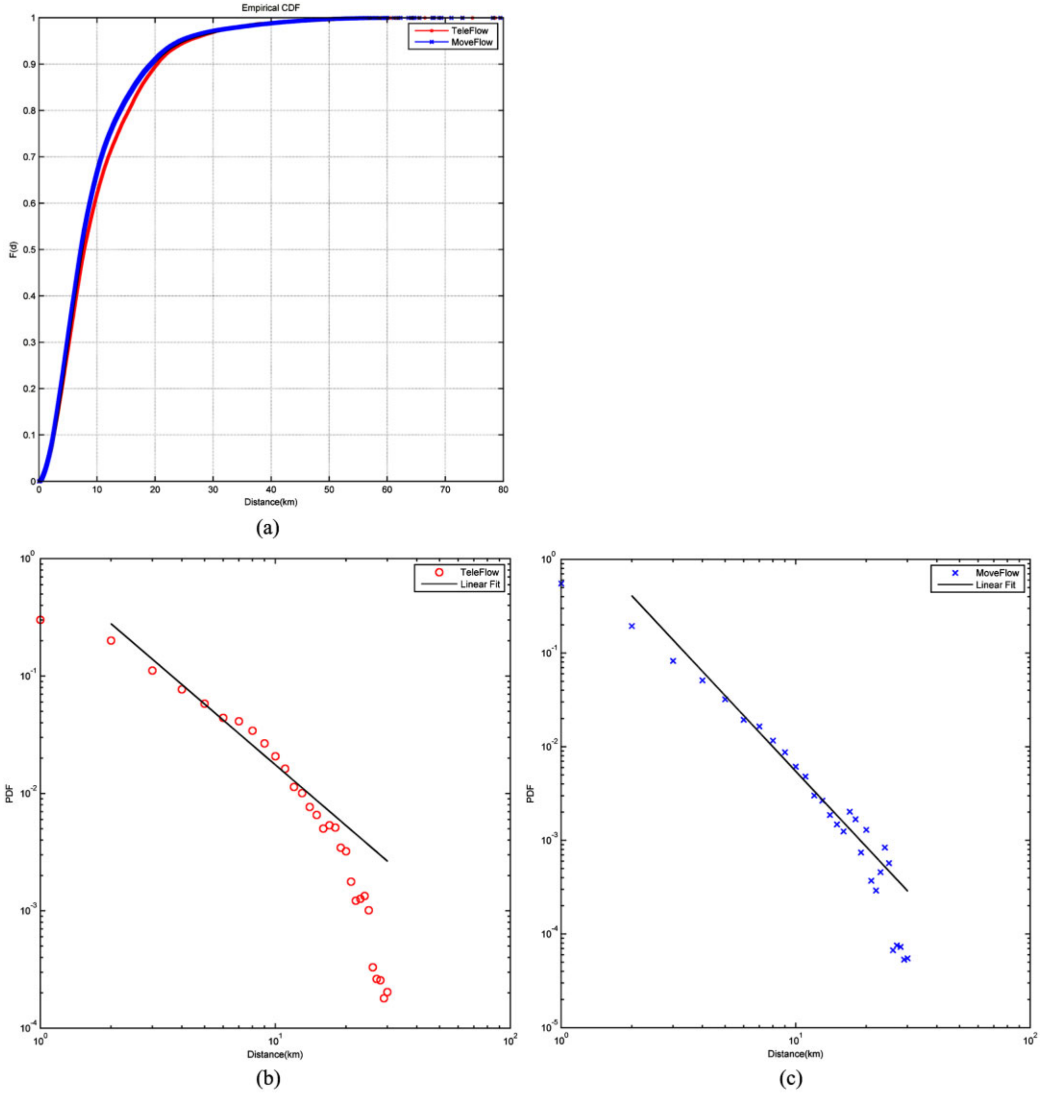


Figure 3 Statistical analysis of spatial interactions showing distance-decay effects: (a) Cumulative probability function (CDF) of distance distributions in two interaction networks: 89.47% of phone-call interactions and 90.98% of movements occur across distances less than 20 km; (b) Network of calls *G_TeleFlow*: the power-law fit features a decay parameter $\beta = 1.45$ on log-log plot; and (c) Network of movements *G_MoveFlow:* the power-law fit features a decay parameter $\beta = 1.60$ on a log-log plot. Note that raw data were aggregated and binned in 1 km intervals for linear least squares fitting, and we only fitted the interactions occurring across distances less than 30 km

spatial interaction of call flows has a smaller decay parameter ($\beta = 1.45$) than movement flow ($\beta = 1.60$), which means the friction effect of distance is relatively weaker in cyberspace than in physical space. But the goodness of fit clearly deteriorates beyond distances of approximately 15 km (Figure 3). We speculate that this may result from an edge effect which constrains long-distance interactions beyond the urban boundary of Harbin.

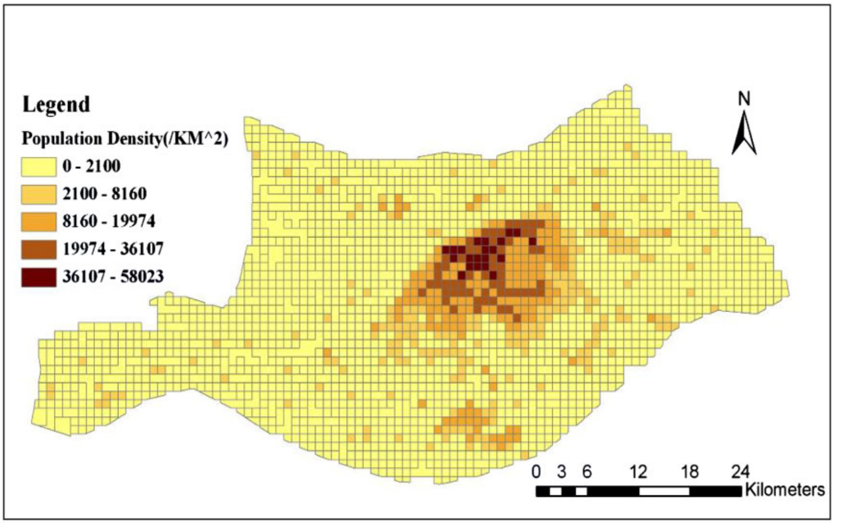


Figure 4 Spatial distribution of LandScan population density in the study area

This result merits comparison to previous research which found different friction effects of distance. For example, González et al. (2008) fit an exponent of 1.75 for human mobility, Liu et al. (2012a) fit an exponent of 1.2 for taxi trips, Xiao et al. (2013) fit an exponent of 0.4 for air-passenger flow, and Liu et al. (2014) fit an exponent of 0.2 for the frequency of co-occurrences on web pages.

2.4 Population Distribution Data

In order to better understand the characteristics of intra-urban human mobility and spatial interactions, we examine the population distribution in the study area extracted from the LandScan™ 2008 global population dataset developed by Oak Ridge National Laboratory. Note that the data quality depends on the estimation uncertainty of LandScan in the study area, because so far it is the finest spatial resolution (30 arc-seconds or approximately 1 km) data available and represents an estimation of average population over 24 hours. It uses a novel algorithm integrating GIS and remote sensing imagery techniques with socio-economic census data including population, places of work, journey to work, and other mobility factors (Dobson et al. 2000). The population distribution data of the study area was extracted into Esri Grid format (Figure 4). We converted the population data into the same unit of mobile cells using the procedure mentioned by Kang et al. (2012a).

3 Methods

3.1 Community Detection Algorithm

In the study of complex networks, a community is defined as a subset (group) of a network; the nodes of the network can be grouped into sets of nodes so that each community is densely connected internally. The identification of such densely connected groups is called community detection. Finding communities within an arbitrary network can be a difficult task. The number of communities, if any, within the network is typically unknown and the communities are often of unequal size or density. Despite these difficulties, however, several methods for community finding have been developed and employed with varying levels of success. Popular community detection methods can be classified into two groups: graph partitioning and hierarchical clustering. Graph partitioning divides a network graph into a set of non-overlapping groups, while hierarchical clustering seeks to build a hierarchy of clusters of nodes, such that for each cluster there are more internal than external connections. The hierarchical clustering method includes two types: hierarchical agglomerative and hierarchical divisive algorithms, according to whether a provisional cluster's edges are added or removed.

Newman and Girvan (2004) propose a *modularity* metric to evaluate the quality of a particular division of a network into communities. Modularity compares a proposed division to a null model in which connections between nodes are random. It is defined as the sum of differences between the fraction of edges falling within communities and the expected value of the same quantity under the random null model.

$$Q = \sum_{k} \sum_{ij \in C} (realflow_{ijk} - estflow_{ijk}) \tag{2}$$

where k is the number of partition communities, $realflow_{ijk}$ gives the actual fraction of interactions between nodes i and j within the same community C, and $estflow_{ijk}$ represents the expected values under the random null model or other theoretical models. If the fraction of edges within communities is no better than the null model, the modularity Q=0, while Q=1 indicates the most robust community structure. In practice, modularity values of different real world networks with varying sizes fall into the range 0.3 to 0.7 (Newman 2006).

One of the most widely used techniques for community detection is the Newman modularity maximization method (Newman 2004). In practice, in view of the time complexity of this algorithm and its feasibility in real-world applications, several alternative algorithms have been developed to support large-scale real networks and to achieve good modularity of community detection. One such algorithm is a fast greedy hierarchical agglomeration (Clauset et al. 2004) whose time complexity on a graph with n vertices and m edges is $O(md\log n)$, where d is the depth of the dendrogram describing the community structure. As introduced in Section 2.2, *G_TeleFlow (V, E)* is the weighted undirected graph of phone-call interaction and *G_MoveFlow (V, E)* is the weighted undirected graph of human movements. We apply this fast hierarchical agglomerative clustering algorithm based on the modularity metric to these two graphs to find whether natural boundaries exist within mobile-cell regions (spatial communities). Note that the algorithm does not consider geographic locations in building clusters. For more detailed steps and the data structures used for computation, please refer to Clauset et al. (2004).

3.2 Incorporating Gravity Model

As noted earlier, spatial networks are affected by spatial constraints. In order to detect hidden socio-economic, structural or cultural connections, therefore, it would be desirable to first take out the effects of spatial distance, by comparing observed interactions to a null model that includes expected spatial effects. Gravity models are an obvious and natural choice for such a null model, since they represent the patterns of spatial interaction that urban structures usually follow (Expert et al. 2011, Fotheringham 1981, 2011).

Several previous studies have employed the method of spatial partitioning to detect community structure and visualize such cluster-to-cluster spatial interaction graphs (Guo 2009, Guo and Liu 2010). Here we propose a community detection algorithm based on a modified modularity function that incorporates a gravity model. Going back to Equation (2), we calculate $estflow_{ijk}$ values as estimations of interaction flow in a community C_k according to the classic gravity model (Liu et al. 2013, Lukermann and Porter 1960, Tobler 1976):

$$estflow_{ij} = \frac{P_i P_j}{d_{ij}^{\beta}} (i, j \in C_k) \tag{3}$$

where P_i represents the population in the cell i and comes from the same community with cell j, and can be calculated based on LandScan population; d_{ij} is the spatial distance between the cells; and the distance-decay parameter β can be obtained by fitting a gravity model (Fotheringham 1981, Kang et al. 2013, O'Kelly et al. 1995).

Note that the absolute value of difference between the actual flow and estimated flow is tiny for small-area cells in urban core regions, but the change rate of difference is higher than that for large-area cells. We consider the alternative fraction format of gravity-modularity for detecting communities:

$$Q_g = \sum_{k} \sum_{ij \in C} \frac{realflow_{ijk}}{estflow_{ijk}} \tag{4}$$

A bottom-up fast greedy approach is adopted and the algorithm is searching for a partition that maximizes Q_g as follows: Step 1 – Each cell starts in its own independent cluster of community and the value of modularity among all pairs of cells for all communities can be calculated; Step 2 – Find a pair of cells which have the maximum value of $realflow_{ijk}$ $/estflow_{ijk}$ and merge them into a community; and Step 3 – Recalculate the population and estimated flow for the new merged community of cells and then repeat Step1 to calculate Q_g until all cells come into one big community so multilevel-community structures can be found. By definition, Q_g favors spatial communities made of cells which are more connected than expected for that distance. Compared to the random null model, Q_g is expected to uncover communities driven by non-spatial factors. We conduct various controlled experiments based on Q and Qg to discover spatial structures of interaction communities over time.

4 Experiment Results

4.1 Discovering Spatial Interaction Communities of Call Flow

At the beginning, we aggregate the total volume of phone-call interaction for each pair of cells on the network *G_TeleFlow* over different hours of a day and then apply the algorithm of maximizing modularity Q to partition these graphs of a week-long dataset. Table 3 shows the statistical results of the community detection and the relatively stable partitions on both weekdays and weekends, considering the number of communities, size and modularity values. A peak of call interaction among cells happened on Thursday (56,070 edges). From the sequence of maps of spatial community structures of *G_TeleFlow* in Figure 5 we can see the spatially coherent communities clearly in a whole-week cycle. Most cells of outer suburbs have been aggregated into the same community in both weekdays and weekends (filled color with yellow, green, blue and violet); whereas some urban core cells have been merged into different communities on varying days. Keep in mind that such natural-boundary characteristics of spatially coherent regions can be found even without taking into account the geographical

Table 3 Community detection results of phone-call interaction (*G_TeleFlow*)

Day	Nodes	Edges	Community Number	Size Min	Size Max	Size Avg	Modularity
Monday	609	41,960	10	27	120	61	0.528
Tuesday	608	40,902	10	26	129	61	0.533
Wednesday	609	40,649	10	29	121	61	0.538
Thursday	609	56,070	8	27	134	76	0.405
Friday	608	54,091	8	27	132	76	0.422
Saturday	605	48,673	8	28	134	75	0.438
Sunday	607	46,506	8	30	131	75	0.446

adjacency in the algorithm. In other words, it demonstrates that the physical space has distance decay effect on call activities in cyberspace. Previous studies also show that natural boundaries can be shaped in multiple spatial scales because of physical barriers or social divides (Ratti et al. 2010, Walsh and Pozdnoukhov 2011). What inspired us most are how these divisions shape and what geographical context is behind them. More detailed explanations are provided below.

The emergence of these isolated enclave polygons reflect strong call interactions between cells that come from the same community. In order to further understand the geographical context of cells, we geo-reference these polygons of spatial communities in Google Earth, and visually interpret some examples of interest based on some labels, landmarks and other semantic descriptions about these locations. Figure 6 shows more detailed information about four enclave regions of spatial communities. Firstly, cell *A* locates in the overpass intersection of the ring and airport expressways which is near a large residential suburb area of this city, and a high volume of call interaction make it merged to the northern spatial community (yellow) of official cells. Secondly, cell *B* has been merged into the same distant community on Monday, Thursday and Friday although there is a physical division of a river flowing between them, whereas it aggregates into a nearby spatial adjacent community on weekends. Cell *B* corresponds to the governmental buildings which has strong connections with the eastern cells (green) of the Central Business District (CBD) on weekdays. Official or business call-contacts strengthen two spatially separated cells, and thus partition them into the same community. However, the call flow between them becomes weak on weekends, so that cell *B* becomes isolated. In addition, we have checked the call direction on raw datasets and find that most of the cells in the CBD have larger numbers of outflow than inflow calls from the governmental cluster. Thirdly, cell *C* has a strong link to the southern cells (red) during the week and they are assigned to the same community. With visual exploratory analysis and looking up these place labels on websites, we find that it locates near the railway station which covers a wood processing plant, food brewery wholesale market and a residential village. There may be business/industry communications that make these cells aggregated into the same community. Fourthly, cell-group *D* has been merged into the same community (orange) with the spatially separated western region. *D* covers a famous local farm and implies a business connection with the city community. In order to identify whether physical movements also exist between these spatially separated cells, we will refer to the partition results of the network of movements.

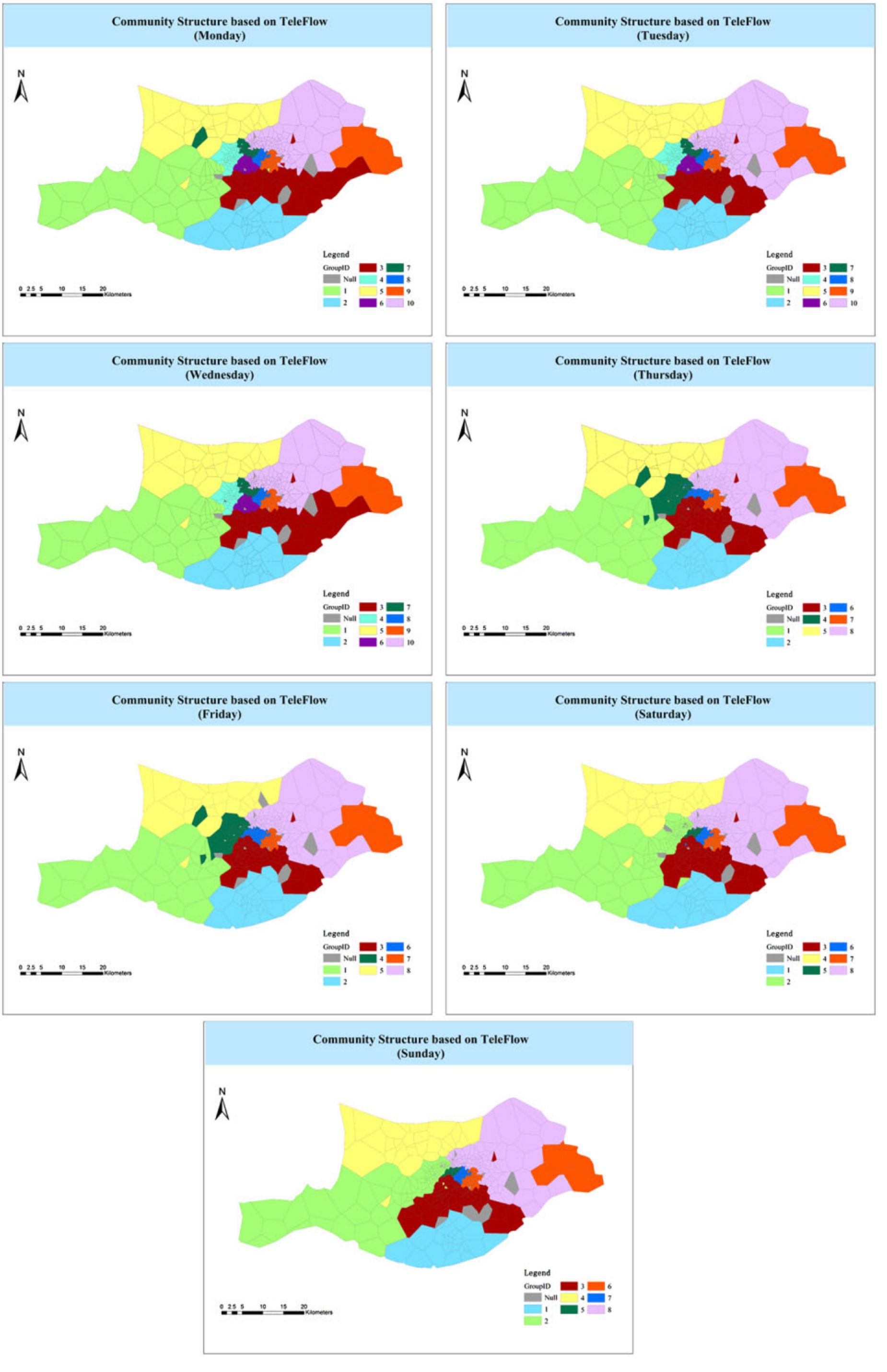


Figure 5 Community structures of phone-call networks throughout a week. Different colors represent different community groups on each day, and the gray cells (Null) show that they are not assigned to any community

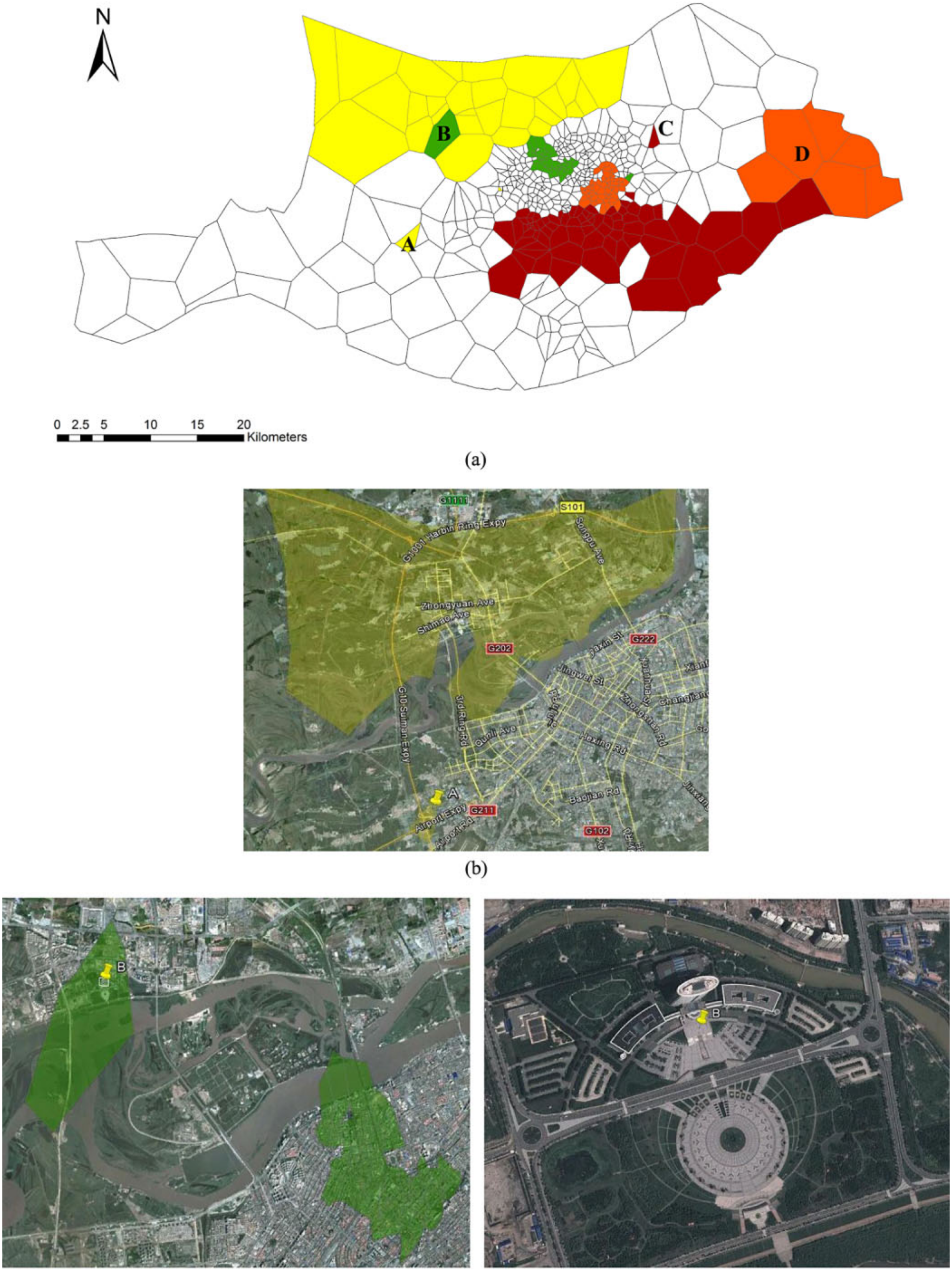


Figure 6 Examples of the differentiated geographical context of isolated regions in spatial communities which have strong call interactions with cells in the same community. Note that the cell-polygons are distorted in Google Earth because of different map projections: (a) geographic locations of these cells; (b) cell A; (c) cell B; (d) cell C; and (e) cell-group D

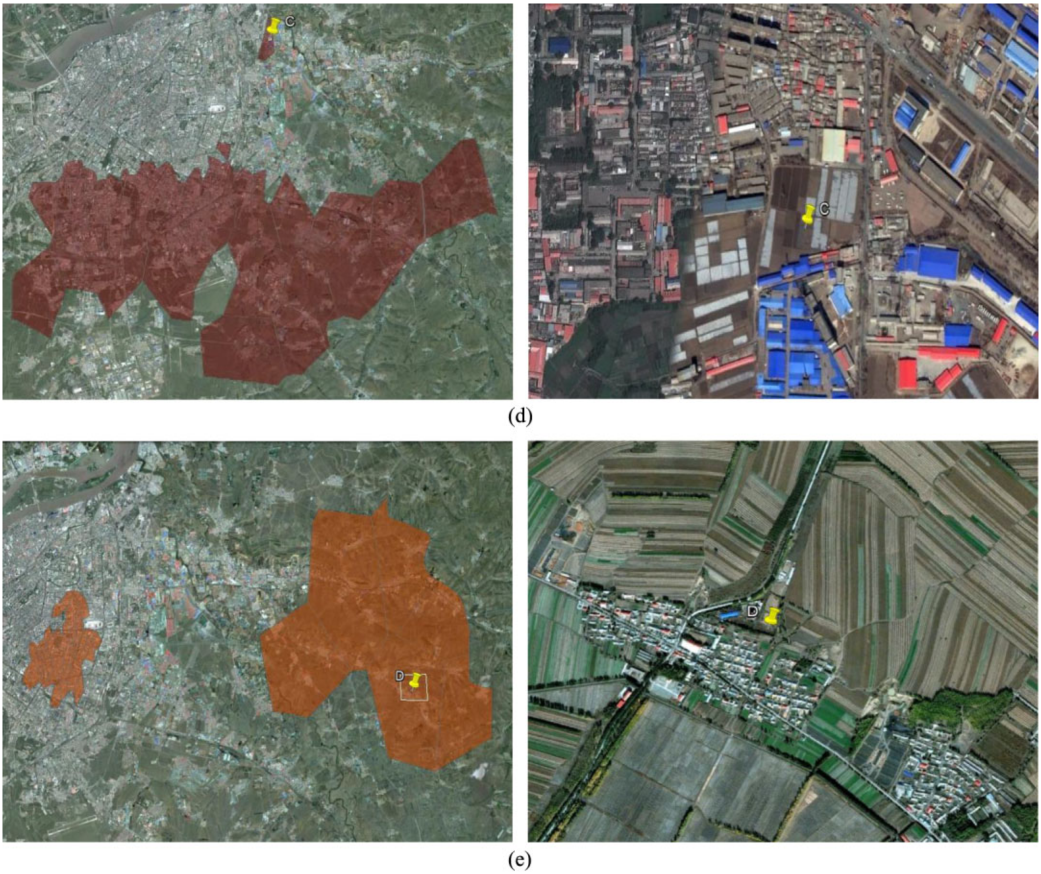


Figure 6 Continued

4.2 Discovering Spatial Interaction Communities of Movement Flow

Similarly, we partition the network of phone-users' movements *G_MoveFlow*. The results show similar community structures on weekdays considering the number of communities, size and modularity values (Table 4). However, the number of movement flows on Sunday increases greatly when residents have higher probabilities of travel. Figure 7 presents the structures of spatial interaction communities based on movements over the whole week. We find that spatial-adjacent cells have higher probabilities to be merged into the same community. Nevertheless, these isolated communities may uncover differentiated geographical contexts which affect patterns of human movements.

In general, the partition structures of movement interaction are similar to that of call interaction. We attempt to examine the statistical correlation coefficients between them (Table 5). There are high correlations existing between movements in physical space and their phone call activities in cyberspace (maximum $R^2 = 0.865$). However, a high correlation does not mean a causal relationship between them (Yuan et al. 2012). We need further exploratory analysis of individual interactions to find potential causal relations, for example, Calabrese et al. (2011) have found that close to 70% of mobile users who call each other frequently

Table 4 Community detection results of movement flow (*G_MoveFlow*)

Day	Nodes	Edges	Community Number	Size Min	Size Max	Size Avg	Modularity
Monday	610	66,915	8	27	143	76	0.538
Tuesday	610	66,219	8	29	143	76	0.538
Wednesday	610	66,258	10	20	144	61	0.539
Thursday	610	66,276	9	27	143	68	0.534
Friday	610	66,980	9	9	143	68	0.536
Saturday	608	65,530	9	22	115	68	0.537
Sunday	610	93,272	8	20	118	76	0.413

shared the same location at least once during the period examined and made phone calls before the face-to-face meetings.

The following part will mainly compare those cells (labeled in Figure 6) mentioned in the previous discussion on the network of call interaction. Firstly, cell *A* is merged into the same spatially separated community both on the network of call interaction *(G_TeleFlow)* and on the network of movements *(G_MoveFlow)*. It locates near the airport expressway and its community is shaped by strong phone call connections and also movements. Secondly, cell B with the governmental office buildings has strong call connections with the CBD but lacks physical movement flow. Thus, it has not been merged into the same community on the *G_MoveFlow*. Thirdly, cell *C*, associated with southern cells, is merged into the same community on both networks; this may result from the spatial allocations of markets and stores (Church 2002). Finally, cell-group D has not been merged into the same community again on *G_MoveFlow*. The accessibility of transportation or land-use types may explain such division but this needs further detailed data analysis (Liu and Zhu 2004, Miller and Wu 2000).

4.3 Community Detection Results Based on Gravity-modularity

According to the gravity-model, the intensity of interaction between two cells depends on two factors: (1) the attraction between them, usually proportional to their masses; and (2) the distance effect. The detection of communities only based on the strength of call interaction or movements (Section 4.1 and 4.2) reflects the combined effects of the two factors, and therefore the partition results appear as spatially coherent regions. By introducing the gravity-model for estimating flow on the network of *G_TeleFlow*, we can take out the distance effect in partitioning the network. Figure 8 shows a controlled community-detection experiment based on the gravity-modularity function (Equation 4), illustrating that the more aggregated communities found in the urban core region are not spatially coherent compared with the results in Section 4.1. This finding offers insight into the reaction of non-space factors, but we need more detailed social and geographical backgrounds to explain the result at length.

5 Conclusions

Mobile phone datasets serve as a novel and high spatio-temporal resolution source for understanding the dynamic spatial interaction processes in urban space. By partitioning both the

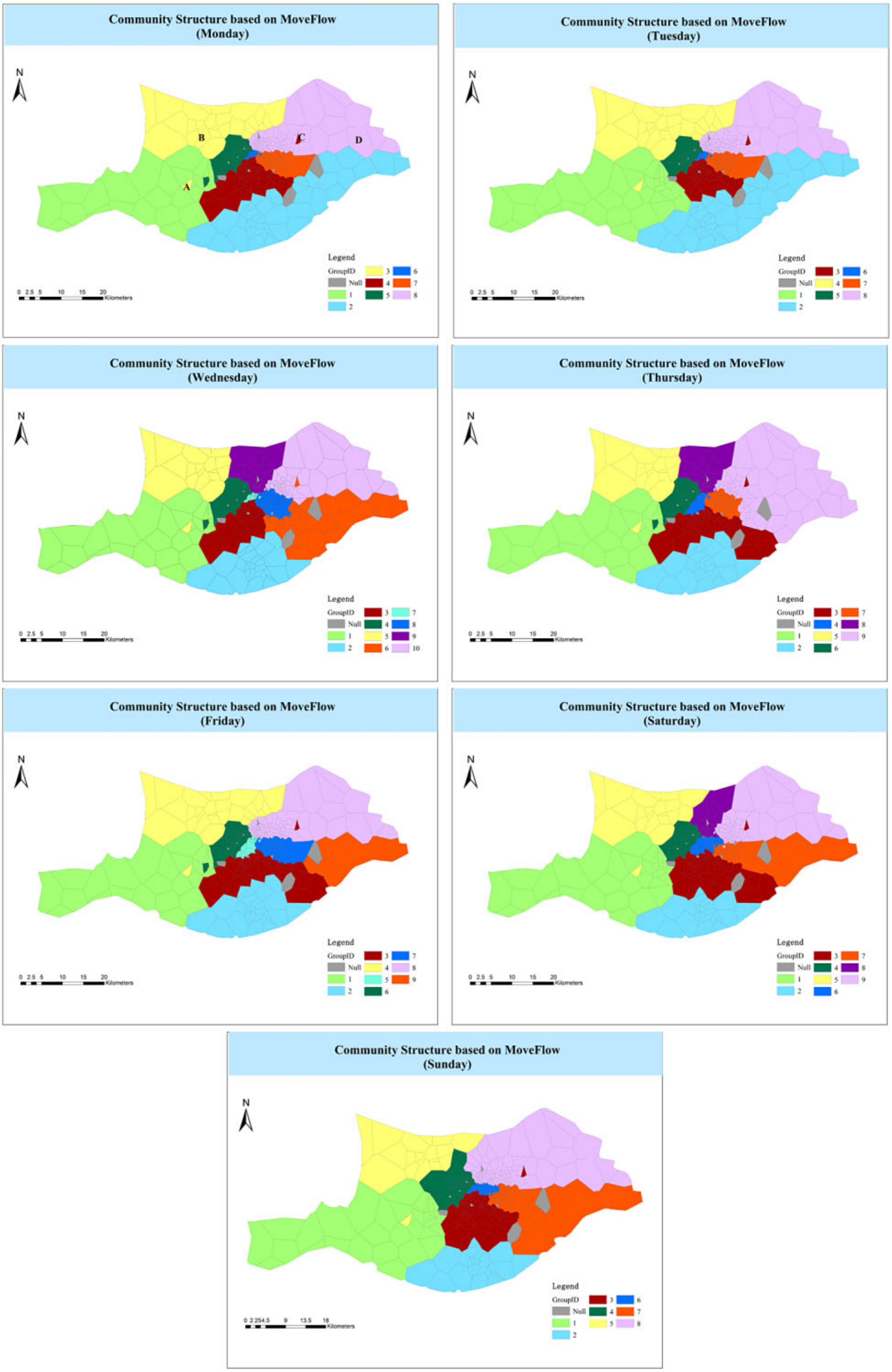


Figure 7 Community structures of movement-flow networks deriving from mobile phone users throughout a week

Table 5 Correlation coefficients between call interaction and movement flow in all pairs of cells

	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
R^2	0.857	0.852	0.852	0.848	0.852	0.857	0.865

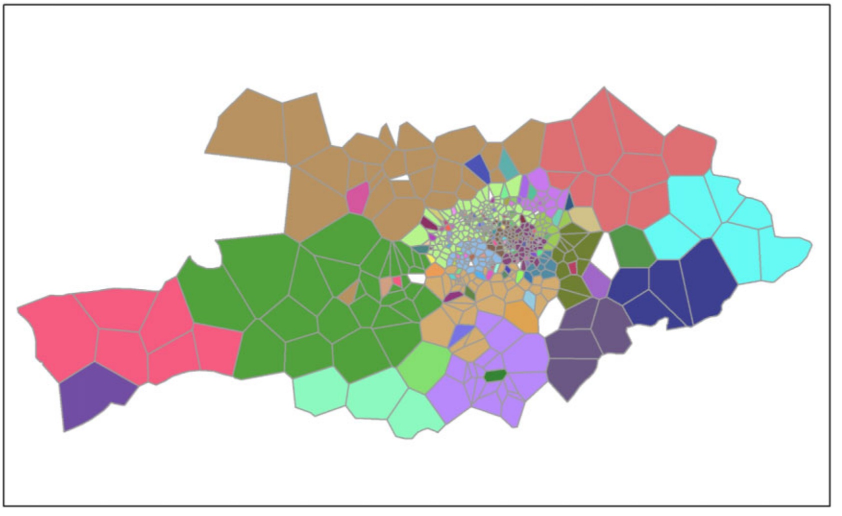


Figure 8 One example of community detection based on gravity-modularity

network of call interaction and the network of movements, we verify the distance decay effect on the detection of geographically embedded communities. The results indicate that people tend to communicate within a spatial-proximity community. This yields insight into how the structure of social interactions might be interpreted in differentiated spaces and places. Our approach presents a combined qualitative-quantitative framework to identify clusters and interaction patterns in spatially embedded networks, and explains how geographical context influences such patterns as revealed by distinct land use features of mobile cells, as well as temporal characteristics of interactions (e.g. office cells-central business district). The findings of this empirical study can be utilized by researchers and urban managers to explore the dynamic spatial interaction patterns between different regions of the city over time and may help to guide transportation planning and other potential applications, e.g. infrastructure construction projects.

Furthermore, we would like to regard this research as a beginning of detecting the spatial interaction communities based on mobile phone datasets. Further research is required to understand the causal relations and semantics of these communities if more detailed land-use or socio-economic data are available. For example, high resolution land-use types could help towards an understanding of the overall geography of these spatial communities; detailed travel-to-work survey and social background information of phone users may better explain human behaviors with ICT in urban areas. In addition, we would like to conduct experiments based on directed weighted-networks with finer temporal resolution (e.g. hourly) for future work. Last but not least, we still need to evaluate the findings in various regions and different spatio-temporal scales.

References

Adrienko N and Adrienko G 2011 Spatial generalisation and aggregation of massive movement data. *IEEE Transactions on Visualization and Computer Graphics* 17: 205–19

Batty M and Miller H J 2000 Representing and visualizing physical, virtual and hybrid information spaces. In Janelle D and Hodge D (eds) *Information, Places, and Cyberspace: Issues in Accessibility*. Berlin, Springer-Verlag: 133–46

Barthélemy M 2011 Spatial networks. *Physics Reports* 499: 1–101

Brockmann D, Hufnagel L, and Geisel T 2006 The scaling laws of human travel. *Nature* 439: 462–65

Calabrese F, Smoreda Z, Blondel V D, and Ratti C 2011 Interplay between telecommunications and face-to-face interactions: A study using mobile phone data. *PLoS One* 6: e20814

Church R L 2002 Geographical information systems and location science. *Computers and Operations Research* 29: 541–62

Clarke K 1996 On epidemiology and geographic information systems: A review and discussion of future directions. *Emerging Infectious Diseases* 2: 85–92

Clauset A, Newman M, and Moore C 2004 Finding community structure in very large networks. *Physical Review E* 70: 1–6

Dobson J, Bright E, and Coleman P 2000 LandScan: A global population database for estimating populations at risk. *Photogrammetric Engineering and Remote Sensing* 66: 849–57

Eagle N, Pentland A S, and Lazer D 2009 Inferring friendship network structure by using mobile phone data. *Proceedings of the National Academy of Sciences, USA* 106: 15274–78

Expert P, Evans T S, Blondel V D, and Lambiotte R 2011 Uncovering space-independent communities in spatial networks. *Proceedings of the National Academy of Sciences, USA* 108: 7663–68

Fischer M, Essletzbichler J, Gassler H, and Trichtl G 1993 Telephone communication patterns in Austria: A comparison of the IPFP-based graph-theoretic and the intramax approaches. *Geographical Analysis* 25: 224–33

Foss S and Couclelis H 2009 Throwing space back in: A tale of Indian fishermen, ICT and travel behavior. *Journal of Transport Geography* 17: 134–40

Fotheringham A S 1981 Spatial structure and distance-decay parameters. *Annals of the Association of American Geographers* 71: 425–36

Fotheringham A S 2011 A new set of spatial-interaction models: The theory of competing destinations. *Environment and Planning A* 15: 15–36

Gao S, Wang Y, Gao Y, and Liu Y 2013 Understanding urban traffic-flow characteristics: A rethinking of betweenness centrality. *Environment and Planning B* 40: 135–53

Giuliano G and Small K A 1993 Is the journey to work explained by urban structure? *Urban Studies* 30: 1485–1500

González M C, Hidalgo C A, and Barabasi A L 2008 Understanding individual human mobility patterns. *Nature* 453: 779–82

Goodchild M F and Janelle D G 1984 The city around the clock: Space-time patterns of urban ecological structure. *Environment and Planning A* 16: 807–20

Goodchild M F, Anselin L, Appelbaum R P, and Harthorn B H 2000 Toward spatially integrated social science. *International Regional Science Review* 23: 139–59

Guo D 2007 Visual analytics of spatial interaction patterns for pandemic decision support. *International Journal of Geographical Information Science* 21: 859–77

Guo D 2009 Flow mapping and multivariate visualization of large spatial interaction data. *IEEE Transactions on Visualization and Computer Graphics* 15: 1041–48

Guo D and Liu S 2010 A graph-based approach to vehicle trajectory analysis. *Journal of Location Based Services* 4: 183–99

Guo D and Wang H 2011 Automatic region building for spatial analysis. *Transactions in GIS* 15: 29–45

Guo D, Liao K, and Jin H 2012 Partitioning geographically embedded complex networks. In *Proceedings of the Seventh International conference on Geographic Information Science,* Columbus, Ohio: 1–4

Jang W and Yao X 2011 Interpolating spatial interaction data. *Transactions in GIS* 15: 541–55

Johnston R, Gregory D, and Smith D 1981 *The Dictionary of Human Geography*. Oxford, Blackwell

Kang C, Gao S, Lin X, Xiao Y, Yuan Y, Liu Y, and Ma X 2010 Analyzing and geo-visualizing individual human mobility patterns using mobile call records. In *Proceedings of the Eighteenth International Conference on Geoinformatics,* Beijing, China: 1–7

Kang C, Liu Y, Ma X, and Wu L 2012a Towards estimating urban population distributions from mobile call data. *Journal of Urban Technology* 19: 3–21

Kang C, Ma X, Tong D, and Liu Y 2012b Intra-urban human mobility patterns: An urban morphology perspective. *Physica A* 391: 1702–17
Kang C, Zhang Y, Ma X, and Liu Y 2013 Inferring properties and revealing geographical impacts of intercity mobile communication network of China using a subnet data set. *International Journal of Geographical Information Science* 27: 431–48
Kwan M P 2007 Mobile communications, social networks, and urban travel: Hypertext as a new metaphor for conceptualizing spatial interaction. *Professional Geographer* 59: 434–46
Leskovec J and Lang K 2008 Statistical properties of community structure in large social and information networks. In *Proceedings of the Seventeenth International Conference on the World Wide Web,* Beijing, China: 695–704
Liu S and Zhu X 2004 An integrated GIS approach to accessibility analysis. *Transactions in GIS* 8: 45–62
Liu Y, Kang C, Gao S, Xiao Y, and Tian Y 2012a Understanding intra-urban trip patterns from taxi trajectory data. *Journal of Geographical Systems* 14: 463–83
Liu Y, Wang F, Xiao Y, and Gao S 2012b Urban land uses and traffic "source-sink areas": Evidence from GPS-enabled taxi data in Shanghai. *Landscape and Urban Planning* 106: 73–87
Liu Y, Wang F, Kang C, Gao Y, and Lu Y 2014 Analyzing relatedness by toponym co-occurrences on web pages. *Transactions in GIS* 18: in press
Lukermann F and Porter P 1960 Gravity and potential models in economic geography. *Annals of the Association of American Geographers* 50: 493–504
Miller H and Wu Y 2000 GIS software for measuring space-time accessibility in transportation planning and analysis. *Geoinformatica* 4: 141–59
Mokhtarian P and Meenakshisundaram R 1999 Beyond tele-substitution: Disaggregate longitudinal structural equations modeling of communication impacts. *Transportation Research Part C: Emerging Technologies* 7: 33–52
Newman M E J 2004 Fast algorithm for detecting community structure in networks. *Physical Review E* 69: 066133
Newman M E J and Girvan M 2004 Finding and evaluating community structure in networks. *Physical Review E* 69: 026113
Newman M E J 2006 Modularity and community structure in networks. *Proceedings of the National Academy of Sciences, USA* 103: 8577–82
O'Kelly M E, Song W, and Shen G 1995 New estimates of gravitational attraction by linear programming. *Geographical Analysis* 27: 271–85
Onnela J and Arbesman S 2011 Geographic constraints on social network groups. *PLoS One* 6: e16939
Radil S M, Flint C, and Tita G E 2012 Using social network analysis to investigate geographies of gang rivalry, territoriality, and violence in Los Angeles. *Annals of the Association of American Geographers* 100: 307–426
Rae A 2009 From spatial interaction data to spatial interaction information. *Computers, Environment and Urban Systems* 33: 161–78
Ratti C, Williams S, and Frenchman D 2006 Mobile landscapes: Using location data from cell phones for urban analysis. *Environment and Planning B* 33: 727–49
Ratti C, Sobolevsky S, Calabrese F, Andris C, Reades J, Martino M, Claxton R, and Strogatz S H 2010 Redrawing the map of Great Britain from a network of human interactions. *PLoS One* 5: e14248
Ren F and Kwan M P 2007 Geovisualization of human hybrid activity-travel patterns. *Transactions in GIS* 11: 721–44
Roy J R and Thill J C 2003 Spatial interaction modelling. *Papers in Regional Science* 83: 339–61
Shaw S and Yu H 2009 A GIS-based time-geographic approach of studying individual activities and interactions in a hybrid physical–virtual space. *Journal of Transport Geography* 17: 141–49
Song C, Qu Z, Blumm N, and Barabási A L 2010 Limits of predictability in human mobility. *Science* 327: 1018–21
Taylor P J 1971 Distance transformation and distance decay functions. *Geographical Analysis* 3: 221–38
Thiemann C, Theis F, Grady D, Brune R, and Brockmann D 2010 The structure of borders in a small world. *PLoS One* 5: e15422
Torrens P M 2008 Wi-Fi geographies. *Annals of the Association of American Geographers* 98: 59–84
Tobler W 1970 A computer movie simulating urban growth in the Detroit region. *Economic Geography* 46: 234–40
Tobler W 1976 Spatial interaction patterns. *Journal of Environmental Systems* 6: 271–301
Tobler W 1988 The quadratic transportation problem as a model of spatial interaction patterns. In Coffey W J (ed) *Geographical Systems and Systems of Geography: Essays in Honor of William Warntz*. London, Methuen: 75–88

Townsend A M 2000 Life in the real-time city: Mobile telephones and urban metabolism. *Journal of Urban Technology* 7: 85–104

Walsh F and Pozdnoukhov A 2011 Spatial structure and dynamics of urban communities. In *Proceedings of the 2011 Workshop on Pervasive Urban Applications (PURBA)*, San Francisco, California: 1–8

Xiao Y, Wang F, Liu Y, and Wang J 2013 Reconstructing gravitational attraction of major cities in China from air passenger flow data 2001–2008: A particle swarm optimization approach. *Professional Geographer* 65: 265–82

Ye M, Janowicz K, and Mülligann C 2011 What you are is when you are: The temporal dimension of feature types in location-based social networks. In *Proceedings of the Nineteenth ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, Chicago, Illinois: 102–11

Young D 2002 A new space-time computer simulation method for human migration. *American Anthropologist* 104: 138–58

Yu H and Shaw S 2008 Exploring potential human activities in physical and virtual spaces: A spatio-temporal GIS approach. *International Journal of Geographical Information Science* 22: 409–30

Yuan Y, Raubal M, and Liu Y 2012 Correlating mobile phone usage and travel behavior: A case study of Harbin, China. *Computers, Environment and Urban Systems* 36: 118–30

