Uncovering Patterns of Inter-Urban Trip and Spatial Interaction from Social Media Check-In Data

Yu Liu*, Zhengwei Sui, Chaogui Kang, Yong Gao

Institute of Remote Sensing and Geographical Information Systems, Peking University, Beijing, China

Abstract

The article revisits spatial interaction and distance decay from the perspective of human mobility patterns and spatially-embedded networks based on an empirical data set. We extract nationwide inter-urban movements in China from a check-in data set that covers half a million individuals within 370 cities to analyze the underlying patterns of trips and spatial interactions. By fitting the gravity model, we find that the observed spatial interactions are governed by a power law distance decay effect. The obtained gravity model also closely reproduces the exponential trip displacement distribution. The movement of an individual, however, may not obey the same distance decay effect, leading to an ecological fallacy. We also construct a spatial network where the edge weights denote the interaction strengths. The communities detected from the network are spatially cohesive and roughly consistent with province boundaries. We attribute this pattern to different distance decay parameters between intra-province and inter-province trips.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: liuyu@urban.pku.edu.cn

Introduction

A number of social media websites that support geo-tagged information submission and sharing have been recently introduced and achieved great commercial success. Various functions have been provided by these websites, such as social networking (Facebook), micro-blogging (Twitter), photo sharing (Flickr), and location based check-in (Gowalla and Foursquare). Each website has millions of registered members and their submissions form an important type of big data. Since much information is user-generated and associated with particular locations, Goodchild coined the term volunteered geographical information (VGI) for it [1]. In this paper, we use "check-in record" to denote a piece of geo-tagged content posted by a user. A check-in record generally includes a short textual message, a photo, and the time and location indicating when and where the message was posted. With a check-in data set, we can extract the footprints of large volumes of individuals. Although the trajectory of one particular person is rather stochastic, we can find underlying patterns when the number of trajectories increases. An interesting example is a map depicting the last 500 million check-in points on Foursquare that clearly demonstrate the human activity distribution across the world (https://foursquare.com/infographics/500million). Much research has been conducted using check-in data, sometimes with additional data such as social ties between users, collected from various sources. Several strands of status quo work can be identified. At the individual level, human mobility patterns [2,3] and geographical impacts on social networks [4,5] are investigated. At the aggregate level, these data enables us to study spatial activity distributions and spatial interactions between regions [6].

Recently, human mobility patterns have drawn much attention in the areas of physics [7], geography [8,9], and computer science [10], with the availability of multi-sourced trajectory data [11]. However, these studies either do not distinguish motion patterns at different spatial scales or focus on intra-urban trip patterns. It is natural that inter-urban trips have different mechanisms from those of intra-urban trips. For example, one in general has two frequently revisited anchor points (i.e. home and workplace) and commutes occupy a large proportion in intra-urban trips. On the contrary, we can only find one anchor point, corresponding to his (or her) home town, from an individual's trajectory at the inter-urban scale. However, whether there exists different mechanisms account for different human mobility patterns at and across different scales remains a research question. Little comparison research on this point has been done due to the lack of individuals' inter-urban trajectories. Clearly, a check-in data set makes an investigation of inter-urban mobility possible for its large spatio-temporal coverage.

In this research, we use a social media check-in data set submitted by about half millions users to study the inter-urban trip patterns. At the collective level, these trips represent spatial interaction strengths between cities. Our research serves three purposes. First, we intend to reveal the underlying distance effect in the trips extracted from check-in records. Second, we try to link patterns at the collective level of spatial interactions versus the individual level of human movements, and to make a comparison with intra-urban patterns revealed from mobile phone or taxi data sets. Last, we investigate the implications of distance decay effect in regionalizing the study area based on spatial interactions between cities.

Background

This section summarizes research in three areas: spatial interaction, human mobility pattern, and spatially-embedded network. The first is a fundamental topic in geographical applications, and the last two have recently drawn much attention in both geographical and physical studies, with the availability of spatio-temporally-tagged big data. This research reveals the underlying connections among them using empirical data set.

1 Distance Decay Effect in Spatial Interactions

Spatial interactions between geographical entities such as cities and regions help us to understand spatial structure of a region and plan an efficient spatial configuration. In practice, interaction strength can be measured by volumes of passengers [12], migration flows [13], trade flows, currency flows, telecommunications [14–16], or even toponym co-occurrences [17]. Due to the complexity of spatial interaction involving pairs of multiple spatial nodes, much research has also been conducted on effectively visualizing spatial interactions and delineating meaningful sub-regions [18,19].

Most spatial interaction systems are governed by the distance decay effect [20], which is in general expressed in the gravity model [21]. Derived from Newton's law of gravity, the gravity model in geographical applications is formulated as $I_{ij} = \frac{kP_iP_j}{f(d_{ij})}$, where I_{ij} and d_{ij} denote the interaction from i to j and distance between two places, and P_i and P_j are repulsion of place i and attraction of place j, respectively. If we do not distinguish the two directions, I_{ij} denotes the sum of flows from i to j and j to i. Meanwhile, P_i and P_j are often approximated by the sizes of the places. The gravity model has been widely used for estimating traffic and migration flows. In the model, distance decay function can vary with applications of interest (e.g., traffic flow versus migration) and technological renovation [22], and one fixed decay function might not fit all problems. Population size might not be able to accurately describe the ability of repulsion or attractiveness of places. A number of studies have then used the observed interaction strengths and distances between geographical entities to fit the gravity model, resulting in the theoretical size (or nodal attraction) of each entity and the distance friction function $f(d)$. Wang [23] summarized several forms of $f(d)$, among which the power law function $d^{-\beta}$ is widely used. The distance decay parameter β reveals the distance impacts on interaction behavior due to the scale free property of $d^{-\beta}$. We can compare different interaction behaviors using their β values. A greater β implies faster decay effect and the interactions are more influenced by distance.

A number of practical methods have been developed for fitting the gravity model, including linear programming [24] and the simplified algebraic method [25,26]. Recently, the particle swarm optimization (PSO) method was introduced to fit the gravity model [11]. The merit of this method is twofold. First, it works well for interaction networks with low density, that is, the interactions of certain pairs of nodes are absent. Second, we can use different distance friction functions beyond the power law when optimizing the model to estimate the nodal attractions.

2 Human Mobility Patterns

Understanding human mobility patterns can help us in many fields including epidemic control and traffic management [27–29]. A number of data sources are introduced to study human mobility patterns. They include mobile phone call records [7,30], GPS (Global Positioning System) enabled taxi trajectories [8,9,31], smart card records in public transportation systems [32], and check-in data [2,3].

A number of measurements can be used to quantify human mobility patterns [33,34]. Among them, the distribution of displacements is extensively investigated. Existing studies reveal that the probability of a movement with distance Δd, denoted by $P(\Delta d)$, decreases with an increase of Δd, indicating the distance decay effect. Different studies suggest that $P(\Delta d)$ can be fitted by different statistical distributions such as power law $P(\Delta d) \sim \Delta d^{-\beta}$ [2,10], exponential law $P(\Delta d) \sim exp(-k\Delta d)$ [30,31], or exponentially truncated power law $P(\Delta d) \sim exp(-k\Delta d)\Delta d^{-\beta}$ [7,9]. The parameters in the above distributions are critical in applications such as epidemic or virus diffusion [28,35]. Particularly, when $P(\Delta d)$ follows a power law distribution in which $1 < \beta < 3$, and the direction distribution is uniform, the trajectory can be modeled by a Lévy flight.

Various models have been proposed to interpret the observed human mobility patterns. They takes into account different influencing aspects such as population characteristics [7], individuals' activities (e.g. returning to particular points, [36]), geographical environments [30,37–39], and distance effects [3,9,40]. These aspects are central to human geography such that the big-data-based human mobility research can shed light on understanding human environment interactions from a new perspective.

3 Spatially-embedded Network

Given a set of geographical entities with known interaction strengths between them, we can construct a spatially-embedded network (or spatial network), in which each node is located in space so that the distance between each two nodes can be measured [41]. A spatial network may be tangible (e.g. street networks) or intangible (e.g. flight networks or networks constructed from social media). With the advances in complex network research, many geographical studies introduce complex network methods into geographical analyses [42–43].

In complex network analyses, detecting communities is an important task. Given a network, a community is a subset with relatively dense node-to-node connections. Many algorithms have been proposed for detecting communities, including the Girvan-Newman method [45], multilevel method [46], fastgreedy method [47], infomap method [48], walktrap method [49], and others. In a community detection procedure, the modularity of a graph is widely used for measuring how good a division is. For a weighted graph, the modularity is computed as.

$$Q = \frac{1}{2m} \sum_{i,j} \left(A_{ij} - \frac{k_i k_j}{2m} \Delta(c_i, c_j) \right) \qquad (1)$$

where m is the number of edges, A_{ij} is the edge weight between nodes i and j, k_i and k_j are the sum weights of edges linked to the two nodes. c_i and c_j denote the community of i and j and $\Delta(x,y)$ equals 1 when $x = y$ and 0 when $x \neq y$.

For a spatial network, a community corresponds to a region, which may be spatially connected or disconnected (i.e. with enclaves). Community detection methods are therefore extended to take into account specific spatial characteristics, such as adjacency constraint [17] and distance effect [50], for regionalization. However, some research directly uses conventional community detection methods for spatial networks, including global flight networks [51], telephone communication networks [52,53], and networks constructed from movements [54,55]. It is interesting that such networks yield spatially connected regions, and some regions coincide with administrative units rather well.

For instance, De Montis et al. reported that the communities obtained from commuters' flows of Sardinia, Italy, in many cases match administrative configurations [54]. In this research, we try to interpret the spatial connectedness using the distance decay effect.

Materials

1 Data Description

This research uses a check-in data set collected from a major Chinese LBSNS (location-based social network service) provider, which can be viewed as a counterpart of Foursquare in the western world. We obtained the data set due to the collaboration between our laboratory (Geosoft@PKU) and the LBSNS provider. The data set contains check-in records posted by approximately 521,000 registered users in one year, from September 2011 to September 2012. Note that fake check-ins exist in the data set. A fake check-in record means that the distance between its real location, denoted by geographical coordinates, and its declared venue is greater than a threshold. After filtering out fake check-ins, we obtain about 23,500,000 records. The heat map of all check-in points clearly highlights the urbanized areas in China (Figure 1A). For the sake of qualitative analyses, the data set also records place names to describe the footprints. All place names are pre-defined and correspond to different levels of administrative units. From the data set, we identified 370 places in total, including the 4 Direct-Controlled Municipalities (Beijing, Shanghai, Tianjin, and Chongqing), Macau, Hong Kong, 332 prefecture-level units, 13 county-level units, and 19 cities in Taiwan. There are actually 333 prefecture level units in China up to 2013. In the data set, we do not find any check-in record inside Shigatse, Tibet. The administrative division system of China is rather complicated and the readers may refer to a Wikipedia entry (http://en.wikipedia.org/wiki/Administrative_divisions_of_the_People's_Republic_of_China) and a webpage in Chinese (http://www.gov.cn/test/2005-06/15/content_18253.htm) for a better understanding. In this research, a place is abstracted to a point that is the capital city's (or town's, in very rare cases) location of the units. All check-in locations inside the place are captured to the point so that we can investigate the aggregate level of spatial interaction (see Tables S1 and S2 in the Supporting Information for spatial interactions and distances between cities, and geographical locations of the 370 cities). For simplicity, we use the term "city" for a cluster of check-in points. It is natural that the total check-ins within a city would be positively correlated with its size. This is confirmed by the distribution of check-ins (Figure 1B), which is consistent with the rank size distribution of Chinese cities [56].

Given a user, his or her trajectory can be formalized as $\{<City_1, T_1>, <City_2, T_2>, ..., <City_n, T_n>\}$, where n is the check-in number of the user, and the $City_i$ was visited at time T_i ($1 \leq i \leq n$). Figure 2A plots the distribution of all users' check-in numbers, which follow a power law distribution well. From the footprints of each user, we can extract the cities that he or she visited. The distribution of visited cities is shown in Figure 2B also illustrates a heavy tailed distribution. Among all users, 237,000 (45.6%) individuals have visited at least two cities, and we can thus construct inter-urban scale trajectories for these users (Figure 2C).

2 Data Evaluation: a Comparative Approach

The inter-urban movements extracted from check-in records are associated with representativeness issues. In other words, not all individuals are registered users of a LBSNS. According to the statistics of Foursquare (http://www.factbrowser.com/tags/foursquare/), a large proportion of its registered users are young and the users are likely to check in at particular places such as airports. The same is true for the Jiepang data set. To evaluate data, we introduce the flight passenger data of year 2011 as a comparison. The flight data set includes 79 cities and 541 pairs of flows, denoted by T_{fij} for cities i and j. T_{fij} is the number of passengers between cities i and j in 2011. From the check-in data, we also compute the trip numbers T_{cij} for the 541 city pairs. T_{cij} and T_{fij} are roughly positively correlated with a low $R^2 = 0.533$ (Figure 3A). The low R^2 indicates that the check-in records either capture movements beyond the flight data or underestimate a number of flight trips. To investigate the first case, we calculate T_{cij}/T_{fij} and select 50 pairs with the highest T_{cij}/T_{fij} (Figure 3B). Three trends are found in the 50 city pairs. First, the distances between 32 pairs of cities (64%), depicted in blue color in Figure 3B, are less than 1,000 km. Within this distance interval, flight trips are not dominant and railway travels is a major competitor in China. Hence, we can obtain relatively more movements from check-in records than from flight data. Additionally, a direct flight line does not exist for very short distance city pairs, such as Beijing-Tianjin and Guangzhou-Shenzhen. The trips between these city pairs can be estimated by the check-in data. Second, as a new mobile application, a LBSNS has different acceptance rates across the country, depending on the regional ICT (information and communications technology) development level. Among the 50 city pairs, 44 pairs (88%) include Shanghai or Beijing. This can be attributed to the high ICT development level of the two cities and more registered users relative to the other cities. Last, it is interesting that the city list covers some China's top tourism destination places (e.g. Jiuzhaigou in Sichuan, Lijiang in Yunnan, Zhangjiajie in Hunan, Sanya in Hainan, and Guilin in Guangxi), indicating a person is more likely to check-in when he (or she) is on a tour. It is natural since one may be excited during a tour and want to share some new contents with his (or her) friends via a social media application. This leads to a high check-in probability and we can thus extract more trips. On the contrary, the low T_{cij}/T_{fij} values can always be interpreted by either long trip distances or low ICT development levels. Note that we only introduce the flight data as a comparison due to the data limitation. If we have the flows of other transportation modes such as railway, similar results can still be obtained, in other words, flows derived from different datasets are not well correlated and their ratios are influenced by the features of cities.

Results

At present, most human mobility research is conducted based on a large population instead of a single person [7–9,30]. Hence, the movement frequency between two places can be used to measure the interaction strength between them. Distance has been widely accepted as an important factor in both individual movements [3,9,40] and collective spatial interactions [20,21]. There is little research on linking these two aspects. Hence, we construct a spatially-embedded interaction network and introduce the gravity model to quantify the distance impact behind the network and to examine whether the distance decay can reproduce the observed displacement distribution, which is critical in human mobility studies. Network science provides a new perspective to understand spatial interactions. Recently, much literature has introduced community detection methods to regionalize a study area [52,57]. The distance decay effect, however, has not been considered in such studies, despite its
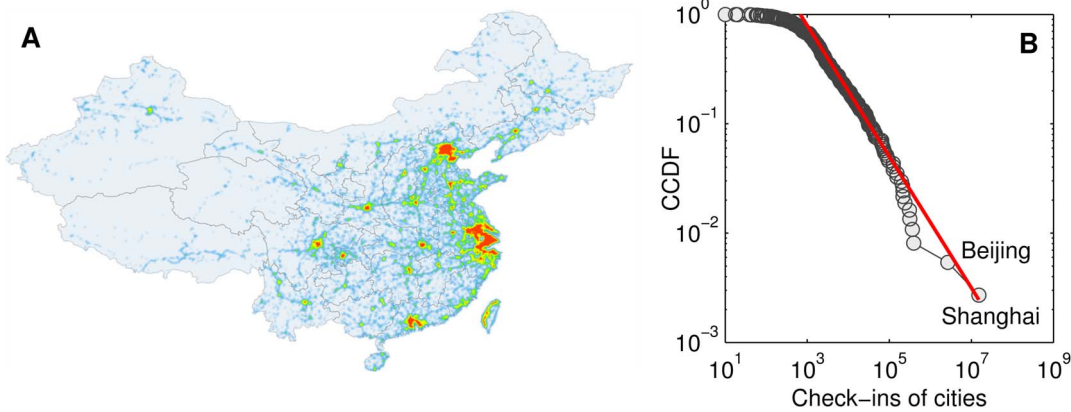
Figure 1. Heat map of all check-in points and frequency distribution of check-ins in the 370 cities. (A) The map, created using density estimation, clearly depicts the distributions of cities and transportation networks in China. Note that The South China Sea Islands are not shown for simplicity. (B) As shown by the CCDF (complementary cumulative distribution function), the frequency distribution exhibits a heavy tail characteristic. Shanghai and Beijing, the two biggest cities in China, have the most check-in records.
doi:10.1371/journal.pone.0086026.g001

importance in modeling spatial interactions. In the following sections, we focus on displacement distribution and community detection, two important topics in human mobility patterns and spatially-embedded networks, using two fundamental concepts in geographical analyses: spatial interaction and the distance decay effect.
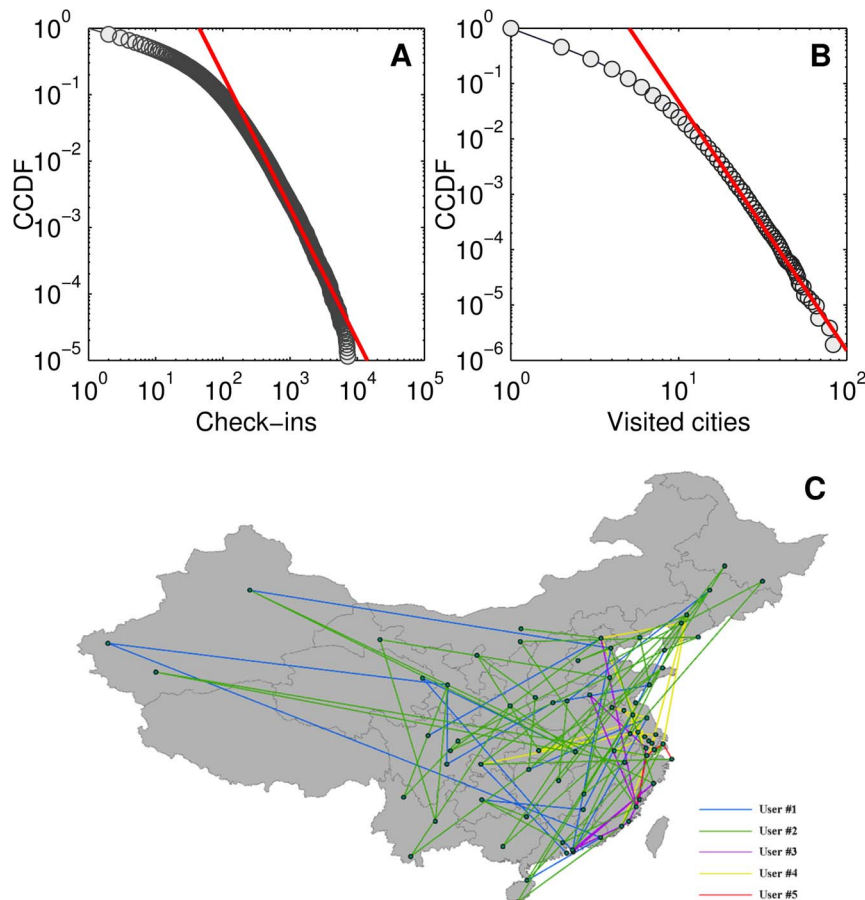

Figure 2. Characteristics of check-ins from the perspective of users. For each user, we compute the number check-ins, N_h, and the number of visited cities, N_c, so that the inter-urban movements can be extracted. Note that N_h and N_c are not well correlated, since a user may check in many times in the same city. (A) Complementary cumulative distribution of N_h. (B) complementary cumulative distribution of N_c. One user visited 83 cities, which is the maximum of all users. (C) Five anonymous example individuals' trajectories.
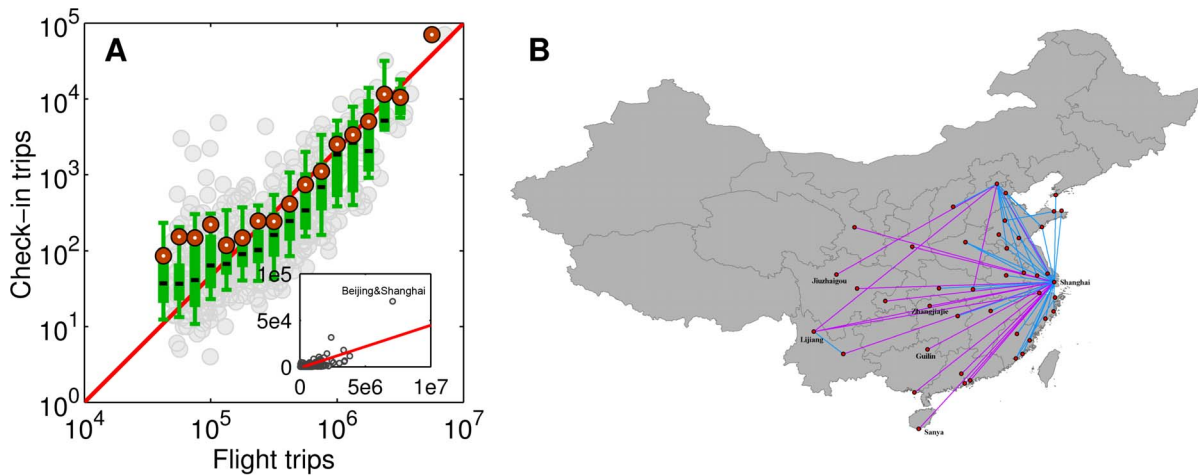doi:10.1371/journal.pone.0086026.g002

Figure 3. Comparison between trips extracted from check-in records, denoted by T_{cij}, and flight trips T_{fij}. (A) Scatter plot of T_{cij} versus T_{fij}, indicating a weak positive correlation. (B) 50 city pairs with the top highest T_{cij}/T_{fij}.
doi:10.1371/journal.pone.0086026.g003

1 Fitting the Gravity Model

From the extracted trajectories, we can compute both the check-in number for each city and the movement between each two cities. An undirected weighted network, denoted by G, is constructed from the interaction strengths (Figure 4A). Note that the movements between cities are actually directed, and we sum the flows in two directions to represent the interaction strengths. G has 370 nodes and 15101 edges (graph density = 0.351). In terms of other statistics of G, the graph diameter is 3, the average degree $\langle k \rangle = 81.6$, the average shortest path $\langle l \rangle = 1.781$, and the average clustering coefficient $\langle C \rangle = 0.657$. Compared with a random network, the relatively low $\langle l \rangle$ and high $\langle C \rangle$ suggest that G has properties of a small world network.

The edge weights follow a power law distribution (Figure 4B). It is similar to the spatial interaction distributions identified from different data sets [15,16,32]. Kang et al. have argued that such a power law distribution mainly derives from the city size distribution, given that its distance decay effect is weak [15].

In this research, we quantitatively estimate the distance decay effect by fitting the gravity model. Because of the low graph density, we adopt the PSO method to find the best fit. According to the PSO method, we try different β values, from 0.1 to 2.0 with a step of 0.1, in the gravity model. The goodness of fit (GOF) is measure using the correlation coefficient between the observed and estimated interactions. For each fixed β value, say 1.0, the PSO method is used to search the best GOF, where each particle is a 370-dimensional vector denoting the theoretical sizes of all cities.

The maximum GOF = 0.985 is achieved when β = 0.8. The exponent is close to the value observed from air passenger flows in China [11] but lower than the distance parameters, which vary between 1.0 and 2.0, estimated using intra-urban movement data [9,30]. Figure 5 plots the relationship between the estimated interactions and real interactions between cities. The high GOF indicates that the inter-urban interactions are governed by the gravity model with a power law distance decay effect.
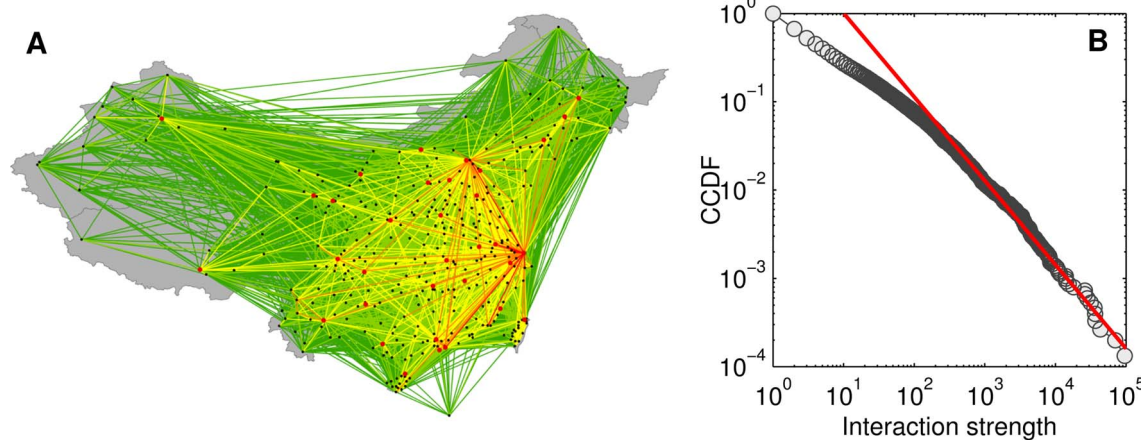

Figure 4. Characteristics of interaction strengths between the 370 cities. (A) Interaction map of the 370 cities. The red lines indicate stronger interactions. The maximum value is 137,847, which is the number of trips between Shanghai and Suzhou, extracted from the check-in data set. The red dots represent capital cities of provinces in China. (B) Complementary cumulative distribution of edge weights (or interaction strengths) between cities.
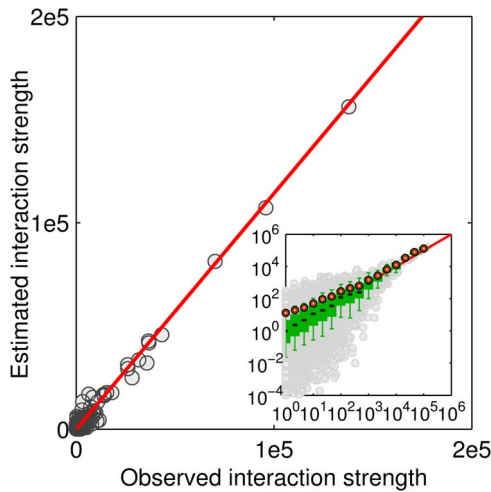doi:10.1371/journal.pone.0086026.g004

Figure 5. Plot of estimated versus observed interaction strengths when β = 0.8, indicating the observed inter-urban interactions can be well fitted using the gravity model. The inset depicts the correlation in a log-log scale. Note that the estimated interaction strengths for some city pairs are less than 1 and thus negative values exist in the log-log plot.
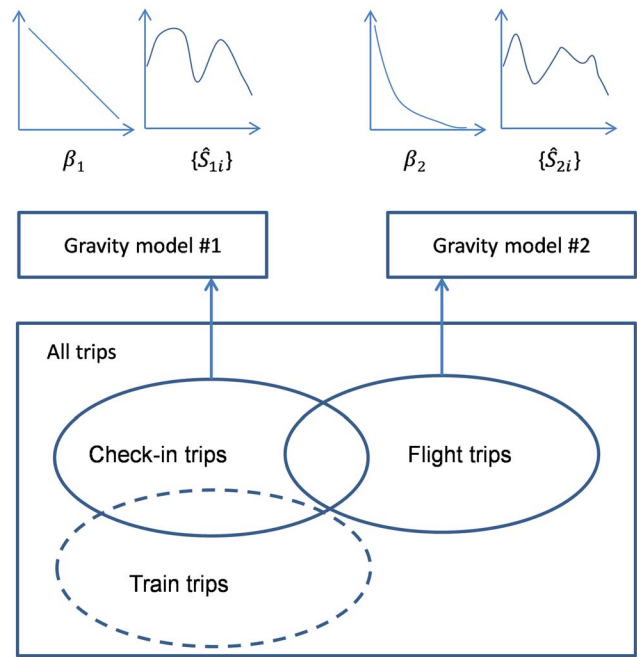doi:10.1371/journal.pone.0086026.g005

Figure 6. Different data sets represent different aspects of "the ground truth" of human movements and thus can be used for revealing different roles of the same city. Given two known data sets for the same group of places, we can obtain two gravity models, denoted by $\hat{I}_{1ij} = \hat{S}_{1i}\hat{S}_{1j}/d_{ij}^{\beta_1}$ and $\hat{I}_{2ij} = \hat{S}_{2i}\hat{S}_{2j}/d_{ij}^{\beta_2}$. β_1 and β_2 represent the distance effect in the two interaction systems. Additionally, $\hat{S}_{1i}$ and $\hat{S}_{2i}$ indicate the importance of city i according to different data sets. For example, in the flight network, the attraction of Beijing is a bit greater than that of Shanghai, according to Ref. [11]. With regard to the check-in data, on the contrary, Shanghai is much more important than Beijing. Such a difference is caused by the fact that Beijing is China's political center with more flight lines but Shanghai exceeds Beijing in both economy and ICT development. If we have a third inter-urban interaction data set for example collected from railway passenger flows, similar comparative investigation can also be conducted.
doi:10.1371/journal.pone.0086026.g006

Some research has pointed out shortcomings of the gravity model [40,57]. They have argued that the gravity model cannot well predict human movements at both the individual level and the collective level. Recently, Masucci et al. compared the gravity model versus the radiation model developed in Ref. [40] using several empirical data sets but did not reach a clear conclusion [58]. In these studies, populations are directly used in the gravity model as the nodes' sizes. However, the "mass" of a place that leads to its observed interaction strengths does not necessarily have a positive linear correlation to that place's population. As shown in Figure 1b, the check-in number in Shanghai is much greater than that in Beijing, although the populations of the two cities are close. Hence, we suggest do not negate the gravity model easily. The appropriate way to adopt the gravity model is fitting the model according to the real interaction strengths and distances between places instead of predicting interactions directly based on populations or other similar attributes.

As pointed in Section 3.2, check-in data only partially capture inter-urban movements and there exist sampling biases. Sampling biases also exist in the air passenger data or even a data set collected based upon other transport modes (e.g. railway) as users of different modes are often correlated with their socio-economic attributes. A single data set represents one aspect of human trips and thus they might not be consistent with each other. It is interesting that the check-in data and the air travel data can be well fitted by different gravity models with different distance parameters and theoretical size sets. Figure 6 demonstrates a framework for integrating different interaction systems. Suppose we have N cities and K interaction systems constructed from different data sources. Let I_{kij} denote the flow strength between i and j in the kth system and d_{ij} is the distance ($i,j=1,...,N$ and $k=1,...,K$). Since $\{I_{kij}\}$ and $\{d_{ij}\}$ are known, ideally, we can get K gravity models. For the kth model, the distance parameter β_k and the theoretical sizes $\{\hat{S}_{k1},...,\hat{S}_{kN}\}$ can be estimated. Since a city plays different roles in various interaction systems, a comparative investigation on the derived theoretical size set $\{\hat{S}_{1i},...,\hat{S}_{Ki}\}, 1 \leq i \leq N$ helps us better understand the ith city. In conclusion, although the check-in data are biased, they still obey

geographical laws and we can obtain particular findings from the data that might be hidden in other data.

In terms of the human mobility pattern, the displacement distribution can be well fitted by an exponential distribution $P(\Delta d) \sim \exp(-\alpha \Delta d)$, where $\alpha = 0.003$ and Δd is measured in kilometers (Figure 7). It is interesting that the inter-urban displacement distribution do not have a heavy tail. Similar distributions have been observed using other data sets such as taxi trajectories [9,31] and mobile phone records [30]. Liu et al. attributed the observed distribution to the impact of geographical environments such as population distribution [9]. With regard to the inter-urban trips, the cities' locations and their populations will influence the displacement distribution.

The exponential displacement distribution is seemingly inconsistent with the power law distance decay, which implies a slower distance decay effect. Liu et al. [9] and Liang et al. [39] suggested that the observed displacement distribution can be well interpreted by integrating the inherent distance decay effect with geographical heterogeneity, and thus proposed a probability form of the gravity model:

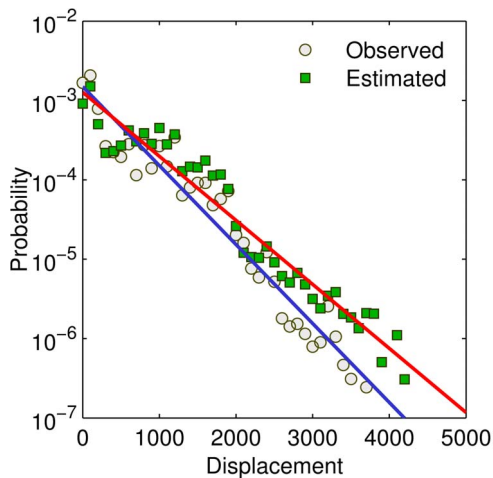$$Pr(T_{ij}) = \frac{kP_iP_j}{d_{ij}^{-\beta}} \qquad (2)$$

Figure 7. Displacement distributions of observed and estimated trips. The observed displacements follow an exponential distribution. We can find a small peak when $\Delta d \approx 1200$ km, since the distances between a number of big cities in China, such as Beijing-Shanghai, Beijing-Wuhan, Shanghai-Guangzhou, and Shanghai-Shenzhen, are all approximately 1200 km. There are a great number of trips for these city pairs (see Figure 4A). The closeness of two best fit lines indicates that the gravity model provides a reasonable explanation of the observed mobility pattern.
doi:10.1371/journal.pone.0086026.g007

where T_{ij} denotes the event that there is a movement between i and j. We adopt the estimated city sizes and distance decay parameter $\beta = 0.8$ to randomly generate the same number of synthetic trips with the number of observed trips using the Monte Carlo simulation approach. The distributions of both observed and synthetic displacements are shown in Figure 7. We can see that the two distributions match well, further confirming the underlying gravity model.

As mentioned earlier, inter-urban (or region) interaction is a traditional topic in geographical analyses, while human mobility patterns have recently drawn much attention thanks to the availability of big trajectory data. This research indicates that the aggregate level of spatial interactions and individual level of movements can be viewed as two sides of the same coin. If the collective spatial interactions can be interpreted by the gravity model (Figure 5), then it is possible that the individual level movements are governed by the gravity model with an identical distance decay parameter (Figure 7). Note that there are some efforts to introduce the gravity model or similar models for mobility patterns. For example, Bazzani et al. proposed a chronotopic model that takes into account attractivity to simulate intra-urban movements [59]. Recently, Liang et al. proved that the exponential displacement distribution can be obtained from the gravity model [39]. Besides the gravity model, some models are built based on benefits (or opportunities) and thus distance plays an indirect role [3,38,40]. These models take into account the decision when individuals plan a trip to a random destination, such as finding a restaurant. However, many inter-urban movements, such as returning to hometown during holidays, are nonrandom, implying that the benefit based models do not apply.

It should be pointed out that an ecological fallacy exists in extending collective level statistics to the individual level. Although various existing models, including the gravity model in this research, closely reproduce the observed displacement distribution, it is still questionable that each individual's movements follow the same gravity model. Figure 8 demonstrates two extremely

contrary cases with the same collective statistics. The data set contains four individuals' (denoted by #1, #2, #3, and #4) trajectories, the displacement distributions of which are represented using different colors. The first plot (Figure 8A) actually represents the model described by Equation 2, that is, each individual's movements exhibit a clear distance decay effect. However, we cannot deny the situation depicted in Figure 8B, where each individual moves with a roughly fixed distance, just like the commute trips inside a city. Most real individual level movements are the mixture of the two cases. We need further studies to decouple them using more detailed trajectory data sets.

2 Identifying Network Communities

For a spatially-embedded network, the community detection method can help us to reveal its structure. In this research, we create a Voronoi diagram based on the 370 cities and merge Voronoi polygons containing cities in the same community so that all communities can be spatialized and visualized. The multilevel algorithm developed by Blondel et al. [46] is adopted to optimize the modularity measure. Additionally, considering that most community detection algorithms are associated with randomness, and thus different iterations of the same algorithm will yield slightly different results, the method proposed in Ref. [57] is adopted. We perform the algorithm 20 times and the result is depicted in Figure 9, where the thicker borders indicate that they are boundaries in more resulting maps of community detection.

From Figure 9, we can draw two conclusions about the partition result. First, all communities are spatially connected, although we do not impose the adjacency constraint during the procedure. Second, a number of communities roughly coincide with the administrative units, that is, the provincial units in China. Provinces such as Jilin, Henan, Guizhou, and Guangdong are clearly delineated in the resulting map. Note that the slight inconsistency between community boundaries and province boundaries is partially due to that the Voronoi polygon instead of the actual administrative area of a city is used to visualize the communities.

The partition pattern has been observed from various spatial networks [51–54]. The first feature, spatial connectedness, can be attributed to the distance decay effect in spatial interactions. Because of the distance decay, closer places generally have stronger interactions and thus are likely to be classified in to the same community. However, the second observation, i.e., coincidence of the identified communities with administrative units, has not been well interpreted yet. We suggest that the distance decay effect is different for intra-province trips and inter-province trips. Due to the political characteristics of China, city pairs within the same administrative unit are typically more socioeconomically integrated than cities located in disparate administrative units, indicating a high frequency of intra-province movements. In other words, the distance decay effect in intra-province trips is weaker than that in inter-province trips. Unfortunately, the number of intra-province city pairs (2053, about 1/7 of the total city pairs) extracted from the check-in data is small and cannot fit the gravity model very well. We simply redraw Figure 5 in a log-log plot and use different symbols to distinguish intra-province and inter-province interactions (Figure 10). It is clear that intra-province interactions are in general greater than inter-province interactions when compared with the estimated interactions computed from the gravity model, suggesting that the gravity model with $\beta = 0.8$ underestimates the intra-province interactions and we should use a smaller exponent instead. In other words, administrative boundaries play a role of obstacle for human inter-urban movements and communications. This provides a reasonable explanation to the
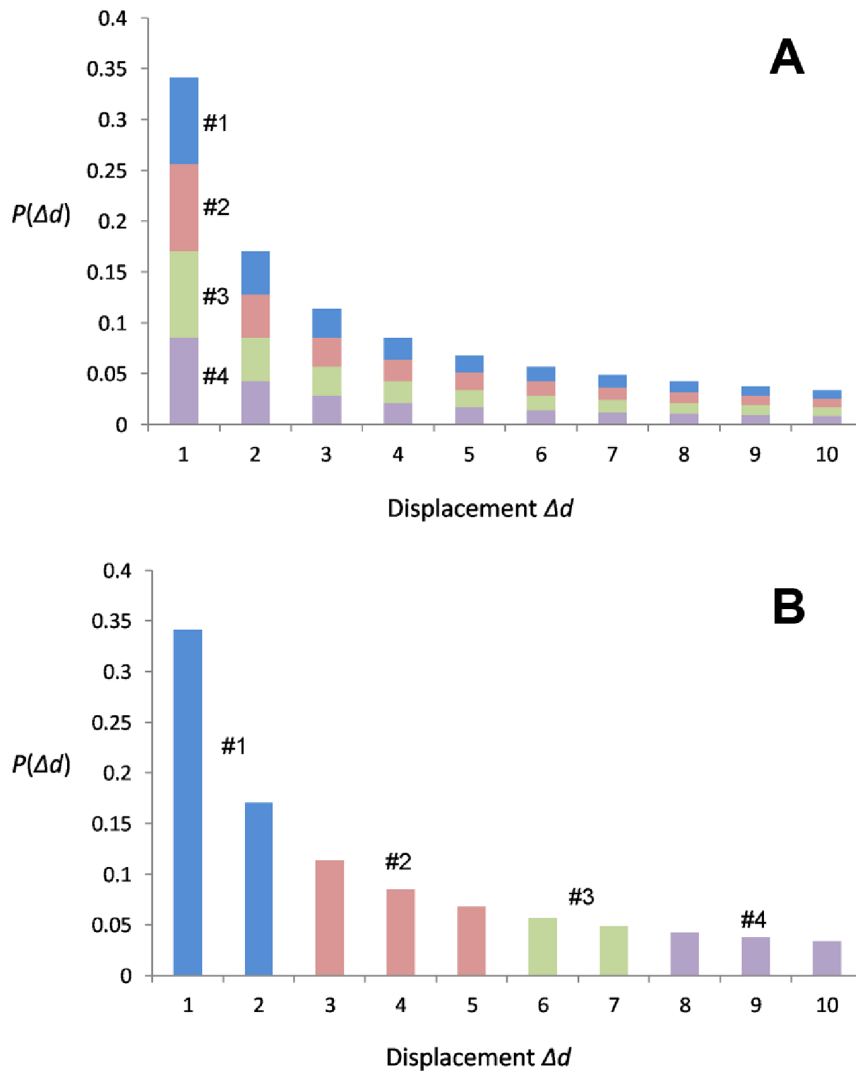
Figure 8. Different individual level movement patterns may lead to the same collective statistics. (A) The four persons' movements are all influenced by the distance decay effect. (B) Distance decay effect is not clear for each person. However, the four persons' movements collectively exhibit the distance decay effect.
doi:10.1371/journal.pone.0086026.g008

community detection result, as well as the findings reported based on different data sets [52,54,57].

Discussion

Human mobility patterns have been a hot research topic in many areas. However, existing studies do not differentiate movements at different spatial scales. Particularly, due to the data limitation, little literature has investigated nationwide inter-urban trips. For the first time, this research adopts the check-in data to analyze inter-urban movements. Our findings include the following four aspects. First, the inter-urban displacements follow an exponential distribution and do not have a heavy-tail property. This distribution is similar to that observed in intra-urban movements. Liu et al. suggested that the geographical environment is a reason for the thin tail in intra-urban displacement distributions. For inter-urban trips [9], this impact still exists. If all cities in this research are identical and have the same mass value in Equation 2, the movements will obviously follow a power law distribution. It is the size and location characteristics of all cities

that lead to the difference between power law distance decay and exponential displacement distribution.

Second, the spatial interactions reflected by the check-in data can be well fitted by the gravity model. This confirms again the power law distance decay effect in spatial interactions, which has been observed from many different data sets. Some existing research has argued that the gravity model cannot well predict spatial interactions if the place populations are directly used as the masses in the model. This research, on the contrary, illustrates that fitting the gravity model to estimate both the places' theoretical sizes and the distance decay function is an appropriate approach.

Third, this research points out the connection between spatial interactions and human mobility patterns. The distance decay function $d^{-\beta}$ can also be used to interpret individuals' movements. The distance parameter $\beta = 0.8$ is less than those estimated from intra-urban movements, indicating a weaker distance decay effect. We also clarify the ecological fallacy issue in modeling human mobility patterns. Hence, a safe statement is that we "cannot reject" an individual level model if the statistics of the synthetic trajectories generated based on the model match the observed

Figure 9. Communities detected from the interaction network *G*. We run the multilevel algorithm 20 times, each of which yields a partition. By merging the Voronoi polygons of cities in the same community, a partition can be visualized. Regions with thicker borders indicate that they occur in more partitions.
doi:10.1371/journal.pone.0086026.g009

statistics. To construct a precise individual level model requires long-term and detailed trajectory data.

Last, by constructing a spatially-embedded network from the check-in data, we regionalize China's territory using a community detection method. The result exhibits a similar pattern to previous studies, in which most communities are spatially consecutive and coincide with geographical units (provinces in the case of this research). Such patterns can also be attributed to the distance decay effect that generally influences closer cities to form stronger connections and thus be clustered together. We also find a difference between the distance decay effects in intra-province and inter-province trips. It is this difference that makes interactions between cities in the same province relatively stronger and therefore classified into the same community.
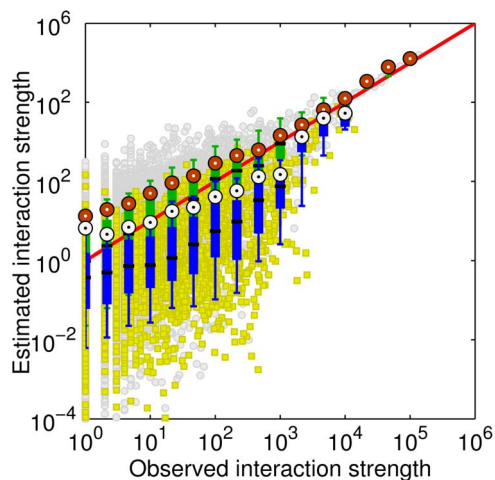
Human mobility patterns and spatially-embedded networks have drawn much attention in recent complexity science studies, where much literature focuses on finding the underlying geographical impacts. Meanwhile, spatial interactions in different spatial scales are widely investigated in geographical analyses. Distance obviously plays an important role in human mobility patterns, spatial interactions, and spatially-embedded networks. The distance decay effect decreases the probabilities of long-distance movements as well as the interaction strengths between faraway places, and consequently shapes the topological structures of spatial networks. Based on an empirical data set, this research makes an initial effort to bridge the three concepts using the distance decay effect. Inversely, with the rapid development of complexity science, human mobility patterns and spatially-embedded networks provide a new perspective and new tools to revisit conventional geographical analyses. This is especially valuable in the era of big data since it is becoming easier for us to collect various data for representing movements, measuring interactions, and constructing spatial networks.

Supporting Information

Table S1 Number of trips and distances between the 370 cities. Values in the upper triangular matrix are trip numbers and the distances are in the lower triangular matrix. "na" indicates no trip observed from the check-in data.
(XLSX)

Table S2 Geographical coordinates of the 370 cities.
(XLSX)

Acknowledgments

Figure 10. Log-log plot of estimated versus observed interaction strengths when β = 0.8. The yellow rectangles and gray circles represent interactions between cities in one province and two different provinces, respectively. It is clear that the gravity model underestimates intra-province trips.
doi:10.1371/journal.pone.0086026.g010

Author Contributions

Conceived and designed the experiments: YL ZS. Performed the experiments: ZS CK. Analyzed the data: YL ZS CK. Contributed reagents/materials/analysis tools: SZ. Wrote the paper: YL YG.

References

1. Goodchild MF (2007) Citizens as sensors: The world of volunteered geography. GeoJournal 69: 211–221.
2. Cheng ZY, Caverlee J, Lee K, Sui DZ (2011) Exploring millions of footprints in location sharing services. Proceeding of ICWSM 2011 81–88.
3. Noulas A, Scellato S, Lambiotte R, Pontil M, Mascolo C (2012) A tale of many cities: Universal patterns in human urban mobility. PLoS ONE 7: e37027.
4. Crandall D, Backstrom D, Cosley D, Suri S, Huttenlocher D, et al. (2010) Inferring Social Ties from Geographic Coincidences. Proceedings of the National Academy of Sciences of the USA 107: 22436–22441.
5. Cho E, Myers SA, Leskovec J (2011) Friendship and mobility: User movement in location-based social networks. Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining 55–62.
6. Cranshaw J, Schwartz R, Hong J, Sadeh N (2012) The Livehoods project: Utilizing social media to understand the dynamics of a city. Proceedings of the 6th International AAAI Conference on Weblogs and Social Media (ICWSM-12), June, Dublin, Ireland.
7. González MC, Hidalgo CA, Barabási A-L (2008) Understanding individual human mobility patterns. Nature 453: 779–782.
8. Jiang B, Yin J, Zhao S (2009) Characterizing the human mobility pattern in a large street network. Physical Review E 80: 021136.
9. Liu Y, Kang C, Gao S, Xiao Y, Tian Y (2012) Understanding intra-urban trip patterns from taxi trajectory data. Journal of Geographical Systems 14: 463–483.
10. Rhee I, Shin M, Hong S, Lee K, Chong S (2008) On the Levy-walk nature of human mobility. Proceedings of IEEE INFOCOM 2008 924–932.
11. Lu Y, Liu Y (2012) Pervasive location acquisition technologies: Opportunities and challenges for geospatial studies. Computers, Environment and Urban Systems 36: 105–108.
12. Xiao Y, Wang F, Liu Y, Wang J (2013) Reconstructing gravitational attractions of major cities in China from air passenger flow data 2001–2008: A particle swarm optimization approach. The Professional Geographer 65: 265–282.
13. Flowerdew R, Lovett A (1988) Fitting constrained Poisson regression models to interurban migration flows. Geographical Analysis 20: 297–307.
14. Guldmann J-M (1999) Competing destinations and intervening opportunities interaction models of inter-city telecommunication flows. Papers in Regional Science 78: 179–194.
15. Mu L, Liu R (2011) A heuristic alpha-shape based clustering method for ranked radial pattern data. Applied Geography 31: 621–630.
16. Kang C, Zhang Y, Ma X, Liu Y (2013) Inferring properties and revealing geographical impacts of inter-city mobile communication network of China using a subnet data set. International Journal of Geographical Information Science 27: 431–448.
17. Liu Y, Wang F, Kang C, Gao Y, Lu Y (2013) Analyzing relatedness by toponym co-occurrences on web pages. Transactions in GIS (In press). doi: 10.1111/tgis.12023.
18. Guo D (2009) Flow mapping and multivariate visualization of large spatial interaction data. IEEE Transactions on Visualization and Computer Graphics 15: 1041–1048.
19. Yan J, Thill J-C (2009) Visual data mining in spatial interaction analysis with self-organizing maps. Environment and Planning B 2009 36: 466–486.
20. Miller HJ (2004) Tobler's First Law and spatial analysis. Annals of the Association of American Geographers 94: 284–289.
21. Fotheringham AS, O'Kelly ME (1988) Spatial interaction models: Formulations and applications. Boston: Kluwer Academic Publishers. 244 p.
22. Taaffe EJ, Gauthier HL, O'Kelly ME (1996). Geography of Transportation. New Jersey: Prentice Hall. 422 p.
23. Wang F (2012) Measurement, optimization and impact of health care accessibility: a methodological review. Annals of the Association of American Geographers 102: 1104–1112.
24. O'Kelly ME, Song W, Shen G (1995) New estimates of gravitational attraction by linear programming. Geographical Analysis 27: 271–285.
25. Shen G (1999) Estimating nodal attractions with exogenous spatial interaction and impedance data using the gravity model. Papers in Regional Science 78: 213–220.
26. Shen G (2004) Reverse-fitting the gravity model to inter-city airline passenger flows by an algebraic simplification. Journal of Transport Geography 12: 219–234.
27. Belik VV, Geisel T, Brockmann D (2009) The impact of human mobility on spatial disease dynamics. Proceedings of International Conference on Computational Science and Engineering (CSE'09) 932–935.
28. Wesolowski A, Eagle N, Tatem AJ, Smith DL, Noor AM, et al. (2012) Quantifying the impact of human mobility on Malaria. Science 338: 267–270.
29. Gao S, Wang Y, Gao Y, Liu Y (2013) Understanding urban traffic flow characteristics: A rethinking of betweenness centrality. Environment and Planning B: Planning and Design 40: 135–153.
30. Kang C, Ma X, Tong D, Liu Y (2012) Intra-urban human mobility patterns: An urban morphology perspective. Physica A: Statistical Mechanics and its Applications 391: 1702–1717.
31. Liang X, Zheng X, Lü W, Zhu T, Xu K (2012) The scaling of human mobility by taxis is exponential. Physica A: Statistical Mechanics and its Applications, 391: 2135–2144.
32. Roth C, Kang SM, Batty M, Barthélemy M (2011) Structure of urban movements: Polycentric activity and entangled hierarchical flows. PLoS ONE 6: e15923.
33. Yuan Y, Raubal M, Liu Y (2012) Correlating mobile phone usage and travel behavior - A case study of Harbin, China. Computers, Environment and Urban Systems 36: 118–130.
34. Csájia BC, Browet A, Traag VA, Delvenne J-C, Huens E, et al. (2013) Exploring the mobility of mobile phone users. Physica A: Statistical Mechanics and its Applications, 392: 1459–1473.
35. Balcan D, Colizza V, Gonçalvesa B, Hu H, Ramasco JJ, et al. (2009) Multiscale mobility networks and the spatial spreading of infectious diseases. Proceedings of the National Academy of Sciences of the USA 106: 21484–21489.
36. Song C, Koren T, Wang P, Barabási A-L (2010) Modelling the scaling properties of human mobility. Nature Physics 6: 818–823.
37. Han X, Hao Q, Wang B, Zhou T (2011) Origin of the scaling law in human mobility: Hierarchy of traffic systems. Physical Review E 83: 036117.
38. Simini F, Maritan A, Néda Z (2013) Human mobility in a continuum approach. PLoS ONE 8: e60069.
39. Liang X, Zhao J, Dong L, Xu K (2013) Unraveling the origin of exponential law in intra-urban human mobility. Scientific Reports. 3: 2983.
40. Simini F,González MC, Maritan A, Barabási A-L (2012) A universal model for mobility and migration patterns. Nature 484: 96–100.
41. Barthélemy M (2011) Spatial networks. Physics Reports 499: 1–101.
42. De Montis A, Barthelemy M, Chessa A, Vespignani A (2007) The structure of interurban traffic: A weighted network analysis. Environment and Planning B: Planning and Design 34: 905–924.
43. Wang F, Antipova A, Porta S (2011) Street centrality and land use intensity in Baton Rouge, Louisiana. Journal of Transport Geography 19: 285–293.
44. Wang J, Mo H, Wang F, Jin F (2011). Exploring the network structure and nodal centrality of China's air transport network: A complex network approach. Journal of Transport Geography 19: 712–721.
45. Girvan M, Newman MEJ (2002) Community structure in social and biological networks. Proceedings of the National Academy of Sciences of the USA 99: 7821–7826.
46. Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E (2008) Fast unfolding of communities in large networks. Journal of Statistical Mechanics: Theory and Experiment 2008: P10008.
47. Clauset A, Newman MEJ, Moore C (2004) Finding community structure in very large networks. Physical Review E 70: 066111.
48. Rosvall M, Bergstrom CT (2008) Maps of random walks on complex networks reveal community structure. Proceedings of the National Academy of Sciences of the USA 105: 1118–1123.
49. Pons P, Latapy M (2006) Computing communities in large networks using random walks. Journal of Graph Algorithms and Applications 10: 191–218.
50. Expert P, Evans TS, Blondel VD, Lambiotte R (2011) Uncovering space-independent communities in spatial networks. Proceedings of the National Academy of Sciences of the USA 108: 7663–7668.
51. Guimerà R, Mossa S, Turtschi A, Amaral LAN (2005) The worldwide air transportation network: Anomalous centrality, community structure, and cities' global roles. Proceedings of the National Academy of Sciences of the USA 102: 7794–7799.
52. Ratti C, Sobolevsky S, Calabrese F, Andris C, Reades J, et al. (2010) Redrawing the map of Great Britain from a network of human interactions. PLoS ONE 5: e14248.
53. Gao S, Liu Y, Wang Y, Ma X (2013) Discovering spatial interaction communities from mobile phone data. Transactions in GIS 17: 463–481.
54. De Montis A, Caschili S, Chessa A (2013) Commuters networks and community detection: A method for planning sub regional areas. The European Physical Journal Special Topics 215: 75–91.
55. Kang C, Sobolevsky S, Liu Y, Ratti C (2013) Exploring human movements in Singapore: A comparative analysis based on mobile phone and taxicab usages. Proceedings of the 2nd International Workshop on Urban Computing (UrbComp'13), Chicago, USA.
56. Anderson G, Ge Y (2005) The size distribution of Chinese cities. Regional Science and Urban Economics 35(6), 756–776.
57. Thiemann C, Theis F, Grady D, Brune R, Brockmann D (2010) The structure of borders in a small world. PLoS ONE 5: e15422.

58. Masucci AP, Serras J, Johansson A, Batty M (2013) Gravity versus radiation models: On the importance of scale and heterogeneity in commuting flows. Physical Review E 88: 022812.

59. Bazzani A, Giorgini B, Servizi G, Turchetti G (2003) A chronotopic model of mobility in urban spaces. Physica A: Statistical Mechanics and its Applications 325: 517–530.