Analysis of urban agglomeration structure through spatial network and mobile phone data

Abstract

Urban agglomeration is an important strategy used to promote economic development and urbanization in China. Understanding the structure of urban agglomeration is therefore essential for policy makers and planners. In this study, the Beijing-Tianjin-Hebei urban agglomeration (BTHUG) is explored through a proposed spatial network analytical framework and a large amount of mobile phone dataset (over 20 million users). We first construct a weight-directed spatial interaction network based on an origin-destination matrix derived from the dataset. Several network metrics (i.e., degree, strength, the rich-club coefficient, and the assortativity coefficient) and three selected community detection algorithms (i.e., Infomap, Louvain and Regionalization) are applied and compared to reveal the structure of the BTHUG. A four-level hierarchical structure is defined and observed: One global center, two local centers, major cities that have low mobility flow but strong linkages with the three centers, and peripheral cities that have low mobility flow and weak linkages with the three centers. Especially, the results imply that the spatial structure of BTHUG is over-dependent on the Global center (i.e., Beijing and northern Langfang). Further, ignoring spatial interaction patterns in top-down administrative planning for urban agglomeration may lead to ineffective integrated development. The implications for BTHUG planning are discussed.

Key words: Spatial structure; network analysis; urban agglomeration; mobile phone data.

1. Introduction

Urban agglomerations have rapidly expanded in many developing countries. It is now an important strategy for accelerating economic development and urbanization in China. The development of urban agglomerations leads to huge flows of people, vehicles, goods, and capital among cities, thus increasing openness and connectivity across city borders. This has led to an economic revolution in mega-city regions and has attracted increasing research attention in recent years (Taylor et al. 2002, Taylor and Derudder 2015, Wu et al. 2016, Zhang et al. 2018, Liu et al. 2016, Li and Phelps 2017, Li and Phelps 2019). Top-down administrative planning has become the main mechanism for urban agglomeration integrated development since the "opening-up" of China in the 1970s, and it is controlled by the central government (Wu 2016). Hence, since the emergence of urban agglomeration integrated development emerged, it has witnessed increasing flows of people, vehicles, goods, and capital along with rapid economic development. However, a fundamental question is whether the "top-down" administrative planning for urban agglomeration integrated development reflects the spatial interaction of human activities, and what are the potential impacts? It is thus important to understand the spatial structure of urban agglomeration by developing methods that should be founded upon the spatial interaction patterns.

The administrative structure of a city or a city-cluster is to some extent static. Nevertheless, Batty (2013) proposed that an urban system should be regarded as "systems of networks and flows”. Many researchers have conducted empirical studies to explore the “dynamics” of the spatial structure of an urban region, by taking the spatial interaction network approach. A spatial network is typically generated by representing a spatial unit (e.g., neighborhood, traffic analysis zone, or grid) as the node of a network, and the intensity of flow (people, vehicle, etc.) between nodes as weighted links. The “spatial structure” of an urban area can be detected from such a spatial network (e.g., Liu et al. 2015, Kempinska et al. 2018). The spatial structure of an urban region shapes and is shaped by spatial interaction patterns. Despite potential issues such as data uncertainty (Xu et al. 2020; Steiger et al. 2015; Huang and Wong 2016), spatial network analysis can be used to effectively explore the spatial structure with good performance (Wang et al. 2017, Wang et al. 2019), particularly with increasingly available human mobility data (e.g., taxi GPS trajectories, social media, smart card data, etc.).

The considerable geographic extent and a “top-down” administrative system are two of the main characteristics of China’s urban agglomeration development (Wu 2002). The Beijing-Tianjin-Hebei urban agglomeration (BTHUG) is the biggest urban agglomeration located in northern China, which includes the national capital, Beijing, the municipality of Tianjin, and the province of Hebei. Many studies in the past tend to explore the spatial structure of the BTHUG through its local social economical factors (Kuang et al., 2014), impervious surface coverage (Cao et al. 2018), and spatial connection intensity (e.g., transport connection and economic connection). However, the main mechanism for urban agglomeration integrated development in China is top-down administrative planning which controlled by central government. Thus, it should be paid attention to how human mobility and spatial interaction can facilitate the spatial structure analysis of BTHUG using the mobile phone dataset under the constraint of top-down administrative planning. The aim of this study is to explore the spatial structure of the BTHUG. The research questions can be summarized as: (1) What are the statistical characteristics of spatial interactions in BTHUG? (2) How these statistical characteristics reflect the spatial structure of BTHUG? (3) Are there significant differences among "top-down" administrative cities and the structures identify by different community detection algorithms (e.g., Infomap, Louvain and Regionalization)? If so, what is the difference? And (4) what is the observed spatial structure of BTHUG shaped by "top-down" administrative planning and the "bottom-up" spatial patterns?

To answer the questions, this study proposed a spatial network analysis framework based on a large-scale mobile phone dataset. The dataset covers 20 million users (nearly 20% of the BTHUG population) and includes various types of demographic information. It can therefore reflect real-world human mobility patterns. We first extract an origin-destination spatial interaction matrix from the mobile phone dataset to build a spatial interaction network *G*. A set of network metrics (i.e., degree, strength, a rich-club coefficient, and an assortativity coefficient) are applied to explore statistical characteristics of spatial interaction in BTHUG. The result reveals a ***global (core) center*** area and the linkages between this and other counties in the BTHUG. Then, the top-down administrative cities and three different outcomes from selected community detection algorithms (i.e., Infomap, Louvain and Regionalization) are compared to further discuss the inconsistency in detecting communities by using different community detection algorithms. Last, the detected communities via Infomap, which has the best match with top-down administrative cities, is applied to identify the ***local centers***, ***major cities***, and ***peripheral cities***. The results provide deeper insights into the urban agglomeration structures and spatial network collaboration.

2. Related work on urban agglomeration structures

Traditionally, urban structure refers to the spatial arrangement or layout of land uses, such as zonal (Burgess 2008), sectoral (Hoyt 1939), and multiple nuclei models (Harris and Ullman 1945). Early studies of urban structure mainly focused on static land use at the city level. With the rapid development of information technology and urbanization, big data of the mobility of people, transport vehicles, and freight has become increasingly available. Researchers have begun using big data to detect the dynamics of urban structures through spatial network analysis (e.g., Zhong et al. 2014; Shaw et al. 2016). Hence, the spatial structure of urban regions turns to the arrangement of spatial units (e.g., grid squares, neighborhoods, blocks, etc.) and a set of relationships arising out of the distribution of these units and the underlying interactions. These consist of people, freight, and capital (Rodrigue 2016, Wu 2020). From this perspective, the spatial structure of urban areas both shapes and is shaped by spatial interaction patterns. In terms of the methodology, by extracting the origin-destination (OD) matrices from mobility big data (e.g., taxi GPS trajectories, social media, and smart card data), a spatial interaction network can be constructed that reveals the spatial structure (Louail et al. 2015), in which the ODs are regarded as the nodes of the network.

The basic properties of spatial networks (i.e., their degree and strength) can provide an overview of travel demand and interactions across urban space (Zhao et al. 2018). Zhong et al. (2014) identified the connectivity and centrality of an urban space using a centrality quantity measurement of the spatial network (i.e., betweenness and PageRank). Wei et al. (2018) found a typical oligarchic spatial structure characteristic in China's population flow network by using the rich-club coefficient and the assortativity coefficient. Unlike the centrality measurement, the rich-club and assortativity coefficients emphasize that a few powerful nodes dominate the structure of a network, and will thus not only form a cohesive cluster among themselves, but will also maintain their connections with peripheral nodes. These coefficients are therefore widely used to explore various types of network structure, such as two-level (e.g., "rich-poor") or hierarchical structures (e.g., Xing et al. 2016, Ducruet et al. 2016).

The spatial structure in a spatial network can be detected using "community detection" algorithms. This organization can then be projected onto an urban region and depict the borders that subdivide the space into different clusters according to the spatial interaction patterns, which are called the "bottom-up" borders, in contrast to the "top-down" administrative borders (Yin et al. 2017). The hypothesis behind the bottom-up borders is that nodes (i.e., spatial units) with strong interactions will form a module (i.e., community), and the divisions between the modules are the borders. In the past decade, the Infomap, Louvain and Regionalization methods are widely applied to detect the community structure for a spatial network (Guo 2008, Lengyel et al. 2015, Guo et al. 2018).

Many studies suggest that although the structures of network communities in the geographic space generally correspond well with top-down administrative borders, the presence of inconsistent borders indicates that human movements do not necessarily follow them. First, it has been suggested that closely connected spaces, in the form of network communities, follow the effects of spatial proximity. That is, the interaction strength between two spaces decreases as the geographical distance between them increases (Fotheringham 1981). Second, research demonstrated how the intrinsic demands of human movement and resources (e.g., trade, urban freight) can reshape the structure of space (Ratti et al. 2010, Gao et al. 2013, De Montis et al. 2013).

Studies using mobility big data analytics and new technological methods have laid the foundations for a better understanding of human mobility behavior, and its relationship with the spatial structure of city regions in particular. However, most have focused on the city level. Some urban agglomerations or city-clusters have emerged, such as the New York bay area, Tokyo bay area, and Yangtze River Delta urban agglomeration. Most countries, including China, break down the urban region space into a system of multi-level hierarchical administrative units (e.g., provinces, prefecture-level cities, counties, and towns). The Chinese government recently initiated a national strategy to establish urban agglomerations, such as the BTHUG, which was first proposed by the central government in 2015 (Zhang 2016). Some researchers have assessed urban agglomeration structures using census or surveying data, but these are costly and time-consuming in the early stage (Castells 2011). Flow data (e.g., taxi GPS trajectories, social media, and smart card data), both within a city and among cities, are becoming increasingly available due to advances in information technology, and thus enable further insights into the structures of cities and city-clusters, and offer unprecedented levels of resolution.

Current spatial network analysis techniques are convenient for exploring the structure of urban agglomerations because they are often tied to mobility-related big data (e.g., taxi GPS trajectories, social media, smart card data) that can be used to examine spatial interaction patterns (Liu et al. 2012, Liu et al. 2014). However, the use of spatial networks via mobility big data presents a level of uncertainty, as the revealed spatial structure of an urban agglomeration may be affected by the source of the data. For example, social media data (e.g., geo-tagged photos, geo-located tweets) are sparse and irregular in time and space (Xu et al. 2020). In addition, the data sample cannot cover the entire population (i.e., social media data, taxi GPS trajectories, and smart card data only cover specific users). Thus, the identified spatial structure of urban agglomerations based on spatial networks constructed via these datasets may not reflect the spatial interaction patterns in the real world (Steiger et al. 2015, Huang and Wong 2016). This issue has prevented the establishment of a systematic understanding of the spatial structure of urban agglomeration based on spatial interaction patterns. Mobility big data should thus be considered along with various types of demographic information to accurately establish spatial interaction patterns and spatial structures.

We address these issues by combining mobile phone data with a spatial network approach. We analyze a human mobility dataset that captures the spatial interaction patterns in the BTHUG, and apply methods that can reveal the spatial structure of the BTHUG. We aim to provide an understanding of this structure that is based on real-world spatial interaction patterns.

3. Dataset and spatial network construction

3.1 Study area

Our study area is the national capital region of China, or the Beijing-Tianjin-Hebei urban agglomeration (BTHUG), which is the biggest urbanized megalopolis in northern China and has an area of 217,156 km^2 (see figure 1). The BTHUG includes 204 counties that belong to 13 cities. Beijing is the capital and the center of politics, economics, and culture in China. Tianjin is one of the country's four directly governed municipalities (the others are Beijing, Shanghai, and Chongqing), and Hebei is a province with 11 prefecture-level cities. The study area had a population of 110 million and a GDP of CNY66,474 hundred million in 2014.

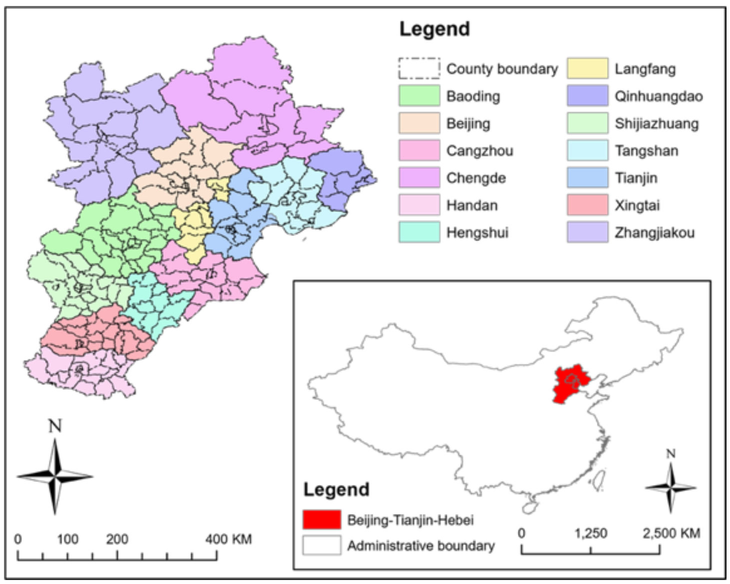

Figure 1. Study area of Beijing-Tianjin-Hebei (BTH) with 13 cities and 204 counties.

3.2 Dataset and BTHUG spatial network construction

Figure 2 presents the details of the dataset and data preprocessing for spatial network construction. The mobile phone data collected by China Unicom Co., Ltd. from November 2 to November 7, 2015, constituted the main dataset used in this work. This includes 20 million users (nearly 20% of the BTH region's population) and 266,214 cell towers. Note that November 2 to November 7, 2015 are normal weekdays in China. We randomly selected this period after excluding some special holiday periods (e.g., Spring Festival, National Day holidays, etc.). Further, county services as the basic unit in the urban agglomeration integrated development in China (Ma 2005, Yeh and Chen 2020). Hence, the administrative county units of the BTH provide another dataset. Each mobile data record tracks various attributes of a user, such as a unique ID, the date, the time, and the connected cell tower, every 30 minutes or when phone communication starts (i.e., a call/SMS). Note that the users' IDs and cell phone numbers were re-assigned as sequence numbers (e.g., 1, 2, 3...) before data preprocessing to protect personal privacy.

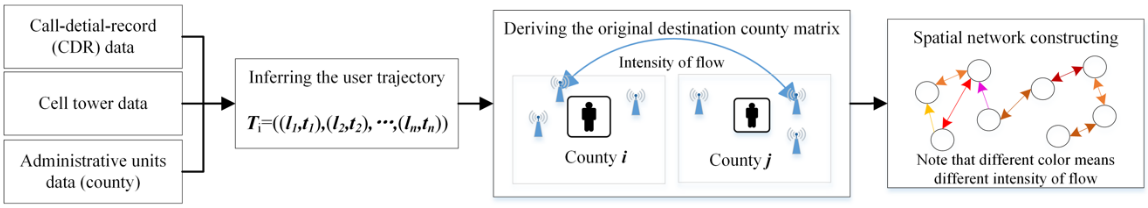

Figure 2. Details of dataset and preprocessing: (1) Inferring the user trajectory; (2) Deriving the origin-destination county matrix *M*; (3) Constructing the weighted-directed spatial network.

To infer each user's movements among counties, we first perform a spatial join analysis in ArcMap of the association between cell tower data and administrative county units. The cell tower data provide information pertaining to the county that the cell tower is located in. Second, each user's trajectory among counties is denoted as a tuple list of $T = \{(l_1, t_1), (l_2, t_2), \ldots (l_n, t_n)\}$, where l_i is the user's location (i.e., county) at time t_i. Third, we infer the stay county chain of each user by setting the stay time threshold to be not less than six hours, and the stay county chain of each user is presented as $L = (c_1, c_2, \ldots c_n)$, where c_i means that the user stayed at county i, and the tuple (c_i, c_{i+1}) is one pair of origin-destination counties in the user's movement. Note that a user may spend a long time in one county, and so if a user has multiple continuous records in a single county we only consider one entry. Finally, we extract all users' origin-destination counties from the stay county chain, and the matrix of origin-destination counties is represented as $M = (c_i, c_j, w)$, where c_i is the origin county, c_j is the destination county, and w is the intensity of flow between c_i and c_j.

To explore the BTHUG structure, we construct a spatial weighted-directed network based on the matrix of origin-destination counties M. First, each county is presented as a node (i.e., c_i is N_i), and the coordinate (x_i,y_i) of its center is regarded as the spatial location of the node. We then assign a directed edge e_{ij} to a pair of nodes (N_i,N_j) depending on whether there was human movement between them or not. The weighted W_{ij} of each edge e_{ij} is given by the intensity of flow w between c_i and c_j. A weighted-directed spatial network $G = (N, E, W)$ is thus obtained.

Figure 3 demonstrates the spatial interaction patterns using the movement flows on BTHUG. As the figure shows, each node in this graph corresponds to a geographic center of counties. The red lines indicate a higher flow between counties. We observe that slight amounts of counties displayed higher movement flows, including Beijing, Tianjin, Shijiazhuang, which represent the hubs of the whole network. Further, the figure shows that most of the counties in the same cities (e.g., Baoding, Xingtai, Handan, et al.) have higher spatial interaction with each other.

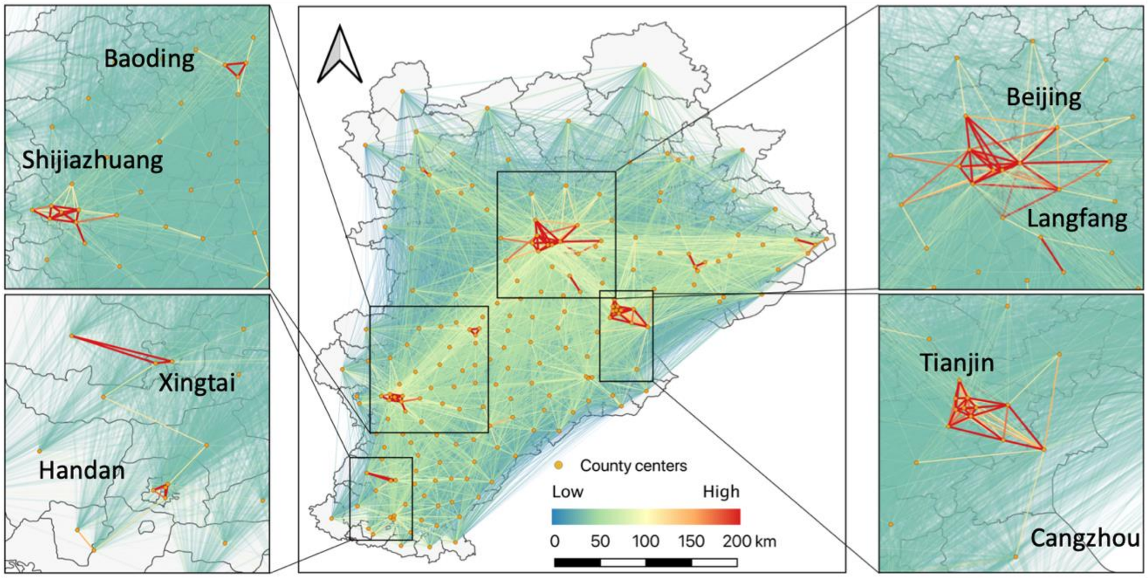

Figure 3. Spatial interaction patterns in BTHUG based on the movement flows.

4 Urban agglomeration structure analysis

In this study, a spatial structure that consists of four city levels are defined to better understand the agglomeration structure:

- Global (core) center: an area that comprises a number of cities (nodes) that have the *largest* volume of human flow and *strong* linkages among them.
- Local center: an area that comprises a number of cities (nodes) that has *large* human flow and *strong* linkages with the global center.
- Major cities: cities that have *low* human flow and *strong* linkages with global and local centers.
- Peripheral cities: cities that have *low* human flow and *weak* linkages with global and local centers.

Given the spatial network *G* constructed in section 3, where the administrative geographic units (i.e., county in *G*) are presented as a node, and intensity of human movements is directed-weighted link between a pair of nodes. A spatial network analysis framework is developed to explore the urban agglomeration structure (see figure 4). There are three steps in this framework: 1) exploring the spatial structure based on spatial network metrics, which includes degree, strength, rich-club coefficient and assortativity coefficient; 2) applying three different community detection algorithms to delineate the organization of urban agglomeration structure; 3) comparing the detected communities from the selected community detection algorithms. The detail of each step is illustrated as follows.

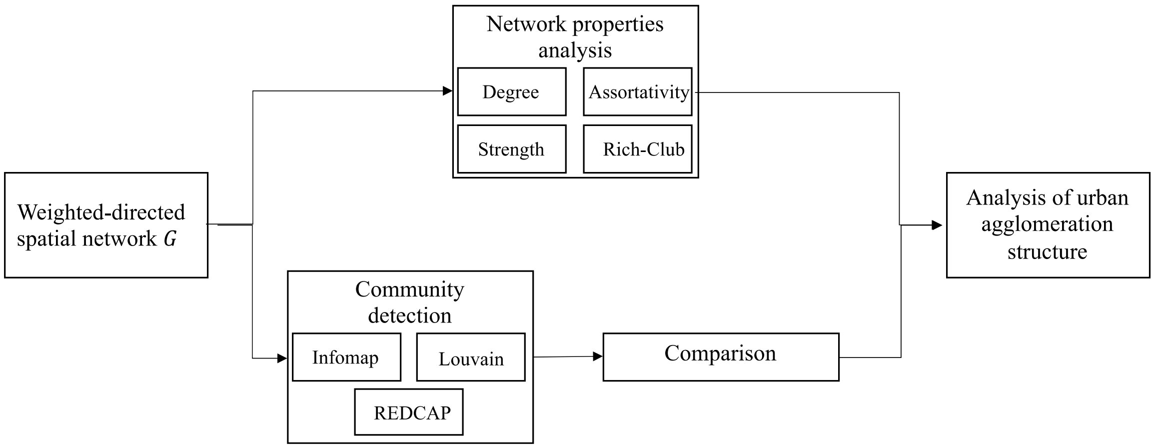

Figure 4. Workflow of empirical analysis.

4.1 Characterizing structure of spatial network

We characterize the BTHUG structure using the metrics of spatial network G. The degree, strength, rich-club coefficient, and assortativity coefficient are applied to reveal the *global (core) center* area and the linkages between this and other counties in the BTHUG.

(1) The *degree* d of a node refers to the number of edges connected to it. The degree distribution $P(d)$ is the proportion of nodes with degree d in the network. We further divide degree into out-degree and in-degree, according to the direction of human movement between each pair of nodes.

(2) The *strength* s of a node refers to the sum of the weights of all edges connected to it. The strength distribution $P(s)$ is the proportion of nodes with strength s. Again, the strength of each node is subdivided into out-strength and in-strength.

(3) The *rich-club coefficient* is an effective tool to measure the structural characteristics of a network based on *degree* d or *strength* s. This can reveal "rich-member" nodes, which are subgroups of powerful nodes that preferentially and intensely connect with each other, while they maintain connections with "poor" nodes (Colizza et al. 2007, Van Den Heuvel and Sporns 2011, Alstott et al. 2014). Nodes with a strength greater than a certain value of r are typically considered as rich nodes, and thus r can be used to define rich nodes (e.g., those with high d or s). The rich-club coefficient is measured using the global and the local rich-club coefficient value, ϕ_{local} (Opsahl et al. 2008, Opsahl 2009).

(4) The *assortativity re*presents how nodes tend to connect with other nodes that have similar degrees, whereas *disassortativity* is the characteristic of nodes tending to connect with others that have different degrees. It is widely used to measure the network structural characteristics and is defined as the Pearson correlation coefficient of degree between pairs of linked nodes (Newman 2002, Xu et al. 2010). The assortativity value is expressed as a scalar value, ρ, in the range from -1 to 1. A positive value of ρ means a network is characterized by assortativity, with a negative value indicating disassortativity. For a directed network, the assortativity can be measured with degree, in-degree and out-degree. Thus, a global assortativity and four directed assortativity (i.e., ρ(out, in), ρ(in, out), ρ(out,

out), and ρ(in, in) can be estimated to quantify the tendency of nodes with un-directed and directed methods in a network (Foster et al., 2010).

4.2 Community detection methods

As discussed in Section 2, the bottom-up border of a community's structure in a spatial network can be identified using community detection algorithms. In this work, the Infomap, Louvain and Regionalization methods are applied to the spatial network G to detect how communities are organized (i.e., *local centers*, *major cities*, and *peripheral cities*). The detail of each method is introduced as follows.

4.2.1 Infomap community detection algorithm

The Infomap algorithm proposed by Rosvall and Bergstrom (2008) is focused on revealing community structure in weighted and directed networks. The main idea is that the interaction flow can be measured based on the probability flow of random walks in a network, and the network can be decomposed into modules by making the interaction flow inside a community significantly larger than those between communities. Nodes in the network are first given a unique code according to the visiting frequency of the random walk. Huffman coding is then used to assign more frequently visited nodes a shorter code, and thus the random walk trajectory of a network can be described as the prefixed community code plus the suffixed code of nodes inside the communities. Finally, the communities are clustered by finding the minimum description code length. The average trajectory length of the code describing a step of the random walk is estimated using the following map equation:

$$L(M) = qH(Q) + \sum_{i=1}^{m} p_i H(p_i) \tag{1}$$

where $L(M)$ is the expectation of the average trajectory length of code spent on a random walk inside and outside communities, $qH(Q)$ is the entropy of movement among clusters, and $\sum_{i=1}^{m} p_i H(p_i)$ is the entropy of movement within clusters. Specifically, q is the probability that a random walker moves from one cluster to another, whereas p_i is the probability of movement within cluster i.

4.2.2 Louvain community detection algorithm

The Louvain algorithm proposed by Blondel et al. (2008) and extended by Leicht and Newman (2008) is a classic of modularity optimization methods on community detection algorithms for an undirected network. The algorithm was then adjusted by Dugué and Perez (2015) to compute communities for a directed network. Specially, the modularity optimization method provides a way to assess the existence of an edge between two nodes in a directed network by comparing it with the probability of have such an edge in a random model following the same degree distribution than the original network. For instance, if two nodes *i* and *j* have small in-degree/large out-degree and small out-degree/large in-degree, then having an edge from *i* to *j* should be considered more surprising than having an edge from *j* to *i*. In this study, the modularity Q of a partition C for the spatial network G is defined as follows:

$$Q = \frac{1}{W}\sum_{ij}(w_{ij} - \frac{s_i^{in} s_i^{out}}{W})\delta(c_i, c_j) \quad (2)$$

where W stands for the sum of weighted for edges in G. w_{ij} is the weight associated to the edge connecting the node i and the node j. s_i^{in} and s_j^{out} are the in-strength and out strength of node i and j, respectively. The function $\delta(c_i, c_j)$ is defined as 1, when nodes i and j belong to same community, and 0 otherwise.

4.2.3 Regionalization

Regionalization with dynamically constrained agglomerative clustering and partitioning (REDCAP) is a family of regionalization methods based on spatially constrained hierarchical clustering (e.g., single linkage (SKL), complete-linkage (CLK), and average-linkage (ALK) methods) (Guo 2011). To apply the regionalization method to the spatial network G, a similarity measure is defined for each pair of nodes (i.e., counties). This work applies the concept of modularity measures, which is defined as follows:

$$Modularity_{ij} = Obserflow_{ij} - Estflow_{ij} \quad (3)$$

$$Estflow_{ij} = \frac{p_i p_j}{d_{ij}^{\beta}} \quad (4)$$

where $Obserflow_{ij}$ is the observed flow between node i and j, and $Estflow_{ij}$ represents the expected flow between node i and j under the gravity model (i.e., equation (4)); p_i and p_j are the population in county i and j; d_{ij} is the spatial distance between node i and j, and the β is distance-decay parameter obtained by fitting a gravity model(Kang et al. 2013).

After estimating the modularity measure for each pair of counties in BTHUG, we have a weighted matrix for each link of counties, where the higher the modularity, the stronger the connection between the two counties is. Given the modularity matrix, an average-linkage clustering based regionalization method (i.e., full-order-ALK) is used to detect the community for the spatial network G. There are two steps in the method: (1) the first step builds a hierarchy of clusters from the bottom up via iteratively merging the most connected clusters based on the county to county modularity values, and the output is a spatially contiguous tree; (2) the second step partitions the spatially contiguous tree to obtain community while maximizing within-community modularity and enforcing the spatial contiguous constraints (i.e., the detected communities should consist of spatially contiguous counties).

4.3 Comparing the selected community detection methods

As this section 2 mentioned, many studies suggest that the structures of network communities in the geographic space generally correspond well with top-down administrative borders. One of the reasons is that the spatial structure is both shaped by "top-down" administrative planning and the "bottom-up" spatial patterns. Especially, it suggests that close spaces will be connected, in the form of network communities, follow the effects of spatial proximity. Because the interaction strength between two spaces decreases as the geographical distance between them increases. Hence, to quantitatively compare the quality of the communities detected by the selected community detection methods, we estimate the similarity of spatial distribution to evaluate how well the detected communities match the top-down administrative cities. We first use $link_{ij}$ to present whether counties *i* and *j* belong to the same community/city (i.e., $link_{ij} = 1$, when counties *i* and *j* belong to the same community/city, and 0 otherwise). Then, the outcomes from the selected community detection methods can be represented as various matrices M_{Info}, $M_{Louvain}$, and M_{REDCAP}. Note that the top-down administrative cities composed of counties are also represented as a matrix M_{city}. Last, bivariate Pearson correlation analysis is applied to investigate the similarity of spatial distribution between each matrix with each other. The higher the Pearson correlation coefficient is, the better the detected communities match with each other. It should be noted that the three selected community detection methods are sensitive to resolution parameters or multiple implementations, which indicates that the detected communities cannot always converge the same result under different parameters or after multiple implementations. Hence, only results that have the best match with top-down administrative cities after hundred implementations with different parameters will be selected to further compare with others.

5. Results and discussion

In this section, we explore the BTHUG structure characteristics based on the constructed spatial network of human mobility *G* = (*N*, *E*, *W*), which was introduced in subsection 3.2.

5.1 Spatial network construction and characterization

As introduced in subsection 4.1, we select degree, strength, the rich-club coefficient, and the assortativity coefficient to detect the *global (core) center* and the linkages between the global center area and other counties in the BTHUG. The statistical characteristics of degree and strength are first calculated to explore the human movement patterns. As Figure 5(a) shows, the cumulative probability curves of degree range from 225 to 406 and decay slowly. We also find that the nodes with a higher degree ($d \geq 370$) count for a large proportion, which is 73%. The results indicate that the majority of counties among the BTHUG have linkages with other counties. However, Figure 5(b) shows that the nodes with a higher strength ($s \geq 100000$) account for a tiny percentage, which is 6.9%. Thus, a minority of counties in the BTHUG have larger and more frequent flows with other counties. A few developed counties therefore dominate the BTHUG structure with a very large number of human movements.

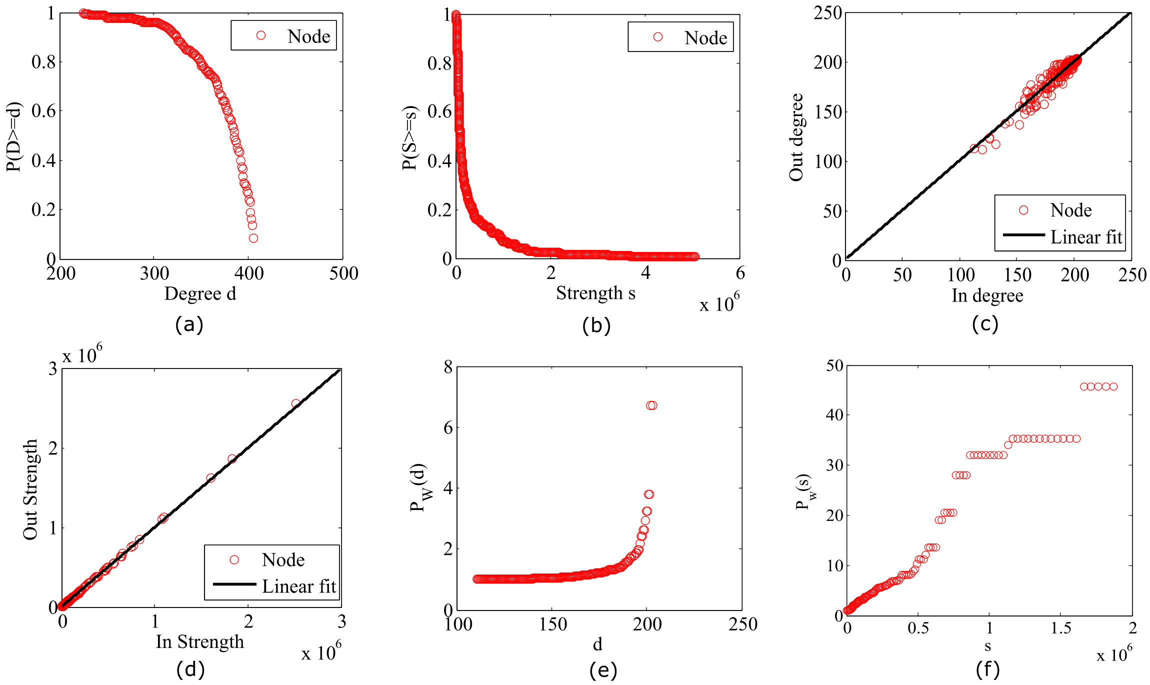

Figure 5. (a)-(b) Cumulative probability distributions of degree and strength; (c)-(d) Correlations between in/out degree and strength; (e)-(f) Rich-club coefficients when $r = d$ and $r = s$.

The constructed spatial network *G* is directed, so degree and strength can be further divided into in/out degree and strength, and we further explore their correlations. As shown in Figures 5(c) and 5(d), in-degree is strongly linearly correlated with out-degree, and the same true for in-strength and out-strength. All of the points are distributed on both sides of the line $y \sim x$, and the goodness of fit attained is above 0.92. These findings indicate that human movement is mutual among counties in the BTHUG. For two given counties, the human movement flows between them in both directions have no significant difference in terms of quantity in most cases.

The above results reveal that a few developed counties dominate the BTHUG structure with a very large number of human movements. The rich-club coefficient and assortativity coefficient are examined to further detect the dominant counties (i.e., the global center). Figures 5(e) and (f) present the global rich-club coefficients, and both $P_w(d)$ and $P_w(s)$ are greater than 1 and reveal a general upward trend, confirming the existence of an obvious rich-club effect in BTHUG. The results also reveal a significant "elbow point" feature in the changes for both the $P_w(d)$ and $P_w(s)$ curves.

In Figure 5(e), $d = 200$ is an important elbow point. When degree *d* is larger than 200 the curve rapidly rises. In Figure 5(f), the curve has a hierarchical upward trend. When $s < 6.3 \times 10^5$, the curve continues to rise smoothly. The curve then jumps when $s > 6.3 \times 10^5$ and reaches its peak when $s > 1.6 \times 10^6$. According to the results, we identify Chaoyang, Xicheng, Dongcheng, Fengtai, and Haidian (Figure 6 (a)) as the *"rich-club" counties* (i.e., the global center area) that have the most connected and powerful spatial interactions in the BTHUG. The rich-club counties are defined as those with $d > 200$ and

$s > 1.6 \times 10^6$. Note that all of these counties are in Beijing, and Chaoyang is the heart of the BTHUG with the maximum strength, $s = 5 \times 10^6$.

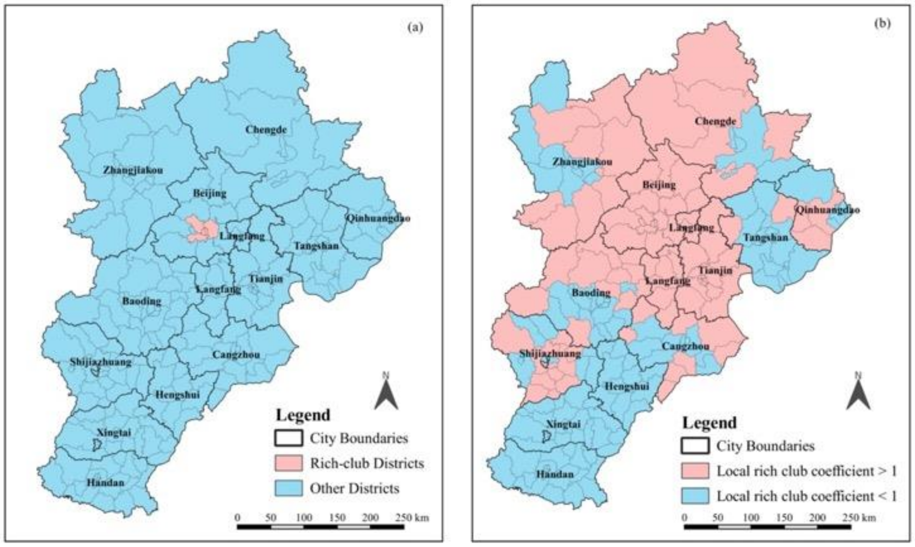

Figure 6. Rich-club coefficients of cities in BTH region: (a) Rich-club members in global rich-club coefficients; (b) Local rich-club coefficients.

The global assortativity coefficient is -0.183, and the directed assortativity coefficients are -0.183, -0.176, -0.179 and -0.179 for ρ*(out,in)*, ρ*(in,out)*, ρ*(out,out)*, and ρ*(out,out)*, respectively. The results are consistent with the strongly linear correction among in-out degree and in-out strength, which suggests that human movement is mutual among counties in the BTHUG. Further, the assortativity coefficients indicate that the BTHUG structure has relatively weak disassortativity. The non-rich nodes with low degrees tend to be connected to the rich nodes, and vice-versa (Colizza et al. 2006). This is consistent with the rich-club analysis; the rich nodes connect strongly with one another, and thus the human movement flows are highly central among rich-club counties in the BTHUG. The local rich-club coefficient of each node is inferred based on the method given in section 4.1. Figure 6(b) shows the result divided into two groups according to the values of the local rich-club coefficients: $\phi_{local} > 1$ and $\phi_{local} < 1$. The counties in the $\phi_{local} > 1$ group tends to connect with rich nodes, whereas the others tend to connect with non-rich nodes. The spatial distribution between the two groups obviously differs. Counties in the $\phi_{local} > 1$ group are mainly concentrated around Beijing. Specifically, it expands from the five rich-club counties (Chaoyang, Xicheng, Dongcheng, Fengtai, and Haidian), and includes four exclaves (i.e., eastern Chengde, northern Qinhuangdao and Cangzhou, and Shijiazhuang). A $\phi_{local} < 1$ group is situated in the southern part of the BTHUG. This indicates that direct interactions between southern counties and rich BTHUG members are weaker. Distance decay may account for this. People in the southern part of the BTHUG tend to go to their neighboring $\phi_{local} > 1$ region (e.g., Shijiazhuang), which has strong direct linkages with rich members, and thus form a hierarchical structure.

5.2 Results of community detection

Before performing the selected community detection methods, we first estimate the expected flow using the gravity model, and then calculate the modularity following the method given in section 4.2.3. We find that the expected flow estimated using the gravity model has a strong linear correlation (i.e., R^2=0.84 and *p-value* <0.01) with the observed flow (shown in Figure S1). This confirms that the intensity of human spatial interactions between two counties decreases as the geographic distance between them increases in BTHUG. Then, we find that the expected flow may underestimate among counties with large geographic distance (i.e., modularity > 0, shown in Figure S2) and overestimate among counties with smaller geographic distance (i.e., modularity <0, shown in Figure S3).

As section 4.2 introduced, we detected the community structure of the spatial network *G* with the three selected methods: Infomap, Louvain, and Regionalization (i.e., full-order-ALK). It should be noted that only results that have the best match with top-down administrative cities after hundred implementations with different parameters will be selected to further compare with others. Figure 7 shows the results of the community detection for the spatial network *G*. Each discovered community is represented by a unique and randomly assigned color. The results show that the detected communities from Infomap and Louvain generally correspond well with the administrative city boundary. The major different parts between infomap and Louvain is Tianjin city, it has been divided into northern and southern parts of Tianjin by infomap, while the Louvain divides Tianjin into four non-spatially adjacent communities. Meanwhile, the spatially adjacent counties are grouped into the same community by using infomap and Regionalization. It is reasonable that the results of infomap group the spatially adjacent counties into the same community since the intensity of human spatial interactions between two counties decreases as the geographic distance between them increases in BTHUG (as the above gravity model revealed). Further, using Regionalization method will generate a more geographically compact region. The results show that the detected community detection by different community detection methods for the same spatial network are mostly different. Therefore, it is worth comparing these methods with top-down administrative cities for investigating the urban agglomeration structure.

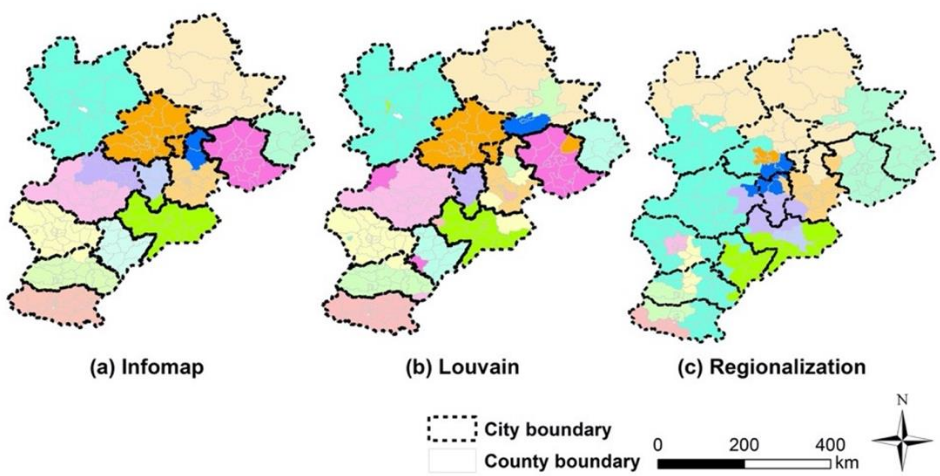

Figure 7. The results of different community detection and regionalization methods: (a) Infomap algorithm; (b) Louvain detection algorithm; (c) Regionalization method (i.e., full-order ALK).

To quantitatively compare the quality of the communities detected by the selected community detection methods, we use the similarity of spatial distribution to evaluate how well the detected communities match the top-down administrative cities. Specifically, we perform bivariate Pearson correlation analysis between each pair of them. Table 1 illustrates the results, which indicates that all of the pairs have significant correlations (i.e., p-value < 0.05). Especially, the Infomap and Louvain have higher (i.e., 0.87 and 0.79) Pearson correlation coefficient with top-down administrative cities, while the Regionalization method has the lowest one (i.e., 0.34). In addition, we apply adjusted Rand Index (RI) analysis to verify the results of bivariate Pearson correlation. RI analysis is a method to assess the similarity between two clusters (Rand 1971, Steinley 2004). It ranges from 0 to 1, and 1 stands for perfect match. Table S1 presents the results of RI analysis, which is consistent with results of bivariate Pearson correlation analysis.

The results indicate that using Regionalization method will generate more geographically compact regions which are quite different from top-down administrative cities in BTHUG. Moreover, since the spatial structure is both shaped by "top-down" administrative planning and the "bottom-up" spatial patterns. Thus, we select the detected communities from Infomap which has the highest association with top-down administrative cities to further analyze the urban agglomeration structure in the next section.

Table 1. The results of the bivariate Pearson correlation analysis between each pair of the methods

	City	***Infomap***	***Louvain***	***REDCAP***
City	-	0.87*	0.79*	0.34*
Infomap	-	-	0.83*	0.35*
Louvain	-	-	-	0.31*
REDCAP	-	-	-	-

**p-value* < 0.05.

5.3 Spatial structure identification

The above results reveal that the spatial interaction reforms the BTHUG into a hierarchical structure. Further, the comparison result of three selected community detection methods suggests that the detected communities from Infomap method are selected to further analyze the urban agglomeration structure. Figure 8 shows the distribution characteristics of 15 communities identified by the method in geographical and network space, and the share of the total flow of each community in the BTHUG (i.e., the sum of flow related to a community divided by the total flow in the BTHUG). Note that the sequence of communities' numbers follows the share of the flow of each community in the BTHUG (i.e., community 1 has the largest share of the flow). Figures 8(a) and (c) show that there is a hierarchical network structure. Figure 8(b) shows the distribution of the share of the flow of each community in the BTHUG. Beijing and northern Langfang (i.e., community 1) consist of the rich-club counties (i.e., the global center area) with the most connected and powerful spatial interactions, so we regard community 1 to be the *global center* of the BTHUG.

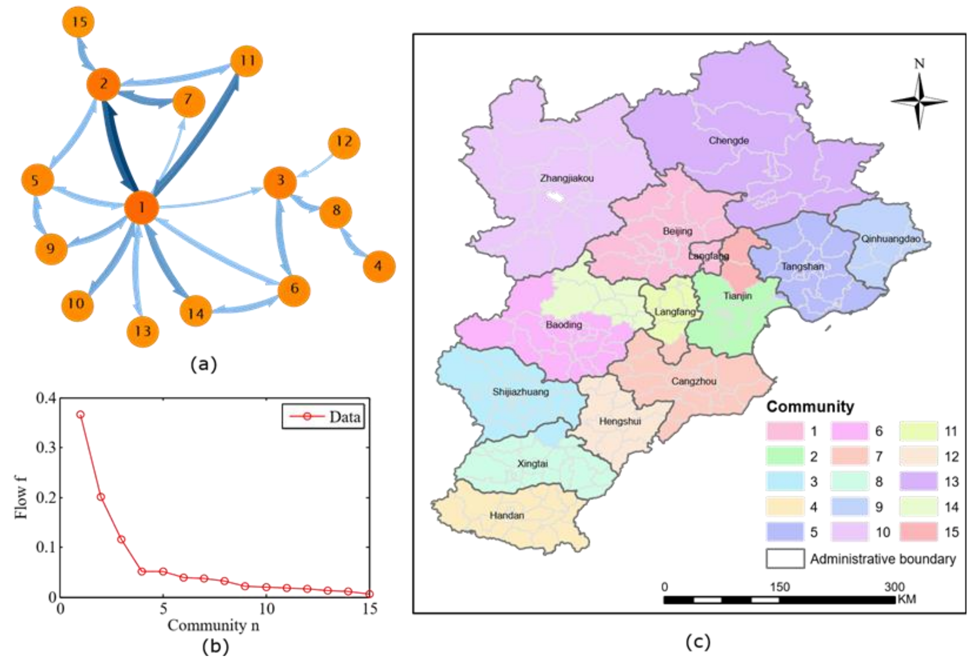

Figure 8. Community detection: (a) communities in network space; (b) percentage of total flow distribution in communities; (c) spatial distribution of communities.

Community 2 (southern Tianjin) and community 3 (Shijiazhuang) in Figure 8(c) are selected as *local centers* according to their share of the flow and their linkages with the global center in the BTHUG. The *major cities* include communities 5 (Tangshan), 6 and 14 (Baoding), 7 (Cangzhou), 9 (Qinhuangdao), 10 (Zhangjiakou), 11 (southern Langfang), 13 (Chengde), and 15 (northern Tianjin). Note that the major cities are defined as in the $\phi_{local} > 1$ group in section 4.1, and have strong and direct linkages with rich-club members (i.e., the global center). The *peripheral cities* contain communities 4 (Handan), 8 (Xingtai), and 12 (Hengshui). Note that the peripheral cities are located in the southern part of the BTHUG and have weaker linkages with the global center. Table 2 gives the details of the hierarchical BTHUG structure.

Table 2. Description of the 4-level hierarchical spatial structure of the BTHUG

City	***Community***	***Level***	***Share of the flow of each community in the BTH region***
Beijing and northern Langfang	***1***	Global center	38%
Southern Tianjin	***2***	Local center	20%
Northern Tianjin	***15***	Major city	1%
Shijiazhuang	***3***	Local center	13%
Handan	***4***	Peripheral cities	4%
Tangshan	***5***	Major city	4%
Baoding	***6***	Major city	3%
	14	Major city	1%
Cangzhou	***7***	Major city	3%

Xingtai	***8***	Peripheral cities	3%
Qinhuangdao	***9***	Major city	3%
Zhangjiakou	***10***	Major city	3%
Southern Langfang	***11***	Major city	2%
Hengshui	***12***	Peripheral cities	1%
Chengde	***13***	Major city	1%

Figure 8(b) indicates that the counties in three centers (i.e., the global center and two local centers) account for 71% of the human movement flows. The remaining counties account for 29% of the human movement flow in the BTHUG, and most movement in these counties is related to the three centers. Community 4 (i.e., the southernmost peripheral city, Handan) has a larger human movement flow but is defined as one of the *peripheral cities* in the hierarchical structure.

Figure 8(c) reveals that most of the 15 communities correspond well with the administrative city units, except that Tianjin and Baoding are divided into 2 distinct communities, and 3 counties of Langfang are merged with Beijing and Cangzhou. People from a county tend to go to neighboring counties, and particularly among those that have strong linkages in the same administrative city, as they share the same local "Hukou" system.

However, there are also strong incentives to cross the top-down administrative city units, and the center areas (i.e., the global center and two local centers) are attractive as they provide more employment opportunities and resources. The results suggest that human movement is generally constrained by the top-down administrative city units, but that there are strong incentives to break through this constraint and reshape the spatial structure.

Figure 9 illustrates the structure of the BTHUG from the perspective of network analysis. The colors and sizes of the icons indicate the roles of cities in the BTHUG structure. Note that the cities' positions are arranged according to their relative geographical location, and only strong linkages between cities/nodes are visualized, whereas the weak linkages are not shown.

Figure 9 clearly shows the BTHUG structure and the relative geographical location of the urban agglomeration defined in this study. The *global center* consists of the whole of Beijing and northern Langfang, as Beijing is the national capital and is therefore the economic, cultural, and educational center of the country. Northern Langfang is adjacent to Beijing, and the government promotes the integrated development of Beijing and northern Langfang in its top-down administrative urban planning approach. Tianjin is divided into northern and southern parts, and the latter is one of the two *local centers*. Southern Tianjin has dense and well-built commercial and trade-port areas, which have benefited from the policy resources provided by the central government (e.g., Binhai New Area). Although the top-down policy has encouraged the development of northern Tianjin, it has not formed an integrated development trend with southern Tianjin from the perspective of spatial interactions. Another *local center* is Shijiazhuang, the capital city of

Hebei province. Unlike the global center (i.e., Beijing and northern Langfang) and southern Tianjin, Shijiazhuang is located in the southwestern part of the BTHUG. It is an economic, cultural, and educational center in Hebei province, and has strong linkages with the global center and the southern cities of Hebei (i.e., Baoding, Xingtai, and Hengshui).

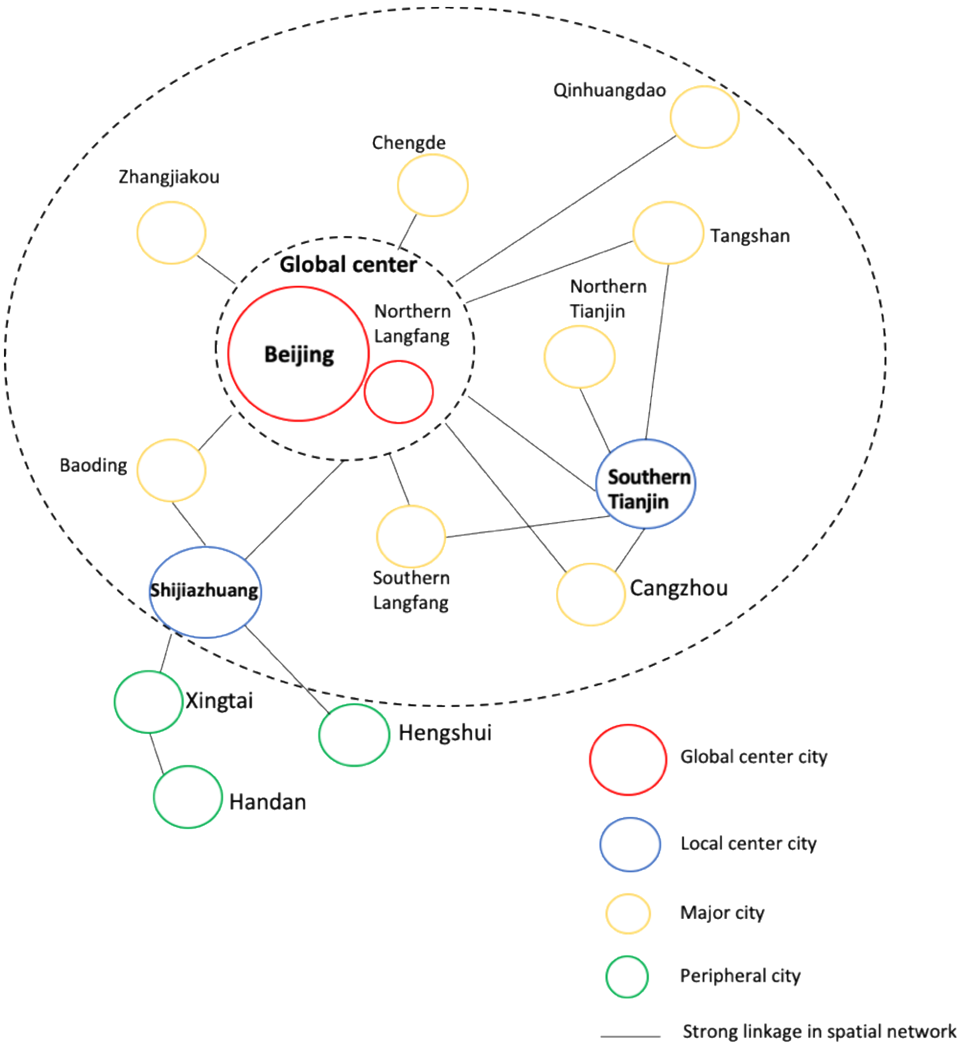

Figure 9. The hierarchical structure in the BTHUG.

Northern Hebei province contains the eight major cities of Tangshan, Baoding, Cangzhou, Qinhuangdao, Zhangjiakou, southern Langfang, Chengde, and northern Tianjin. These have strong connections with the global and local centers. The global center, local centers, and major cities form a hierarchical structure of integrated development in the BTHUG from the perspective of spatial interaction. The peripheral cities are those in the southern part of Hebei province (i.e., Handan, Xingtai, and Hengshui), and their linkages with the global center are weaker. The mobile phone big data-driven spatial interaction network plays a key role in the identification of the hierarchical structure of the BTHUG. Our results therefore provide new insights into the top-down urban agglomeration plans for the integrated development of the BTHUG.

6. Conclusion

In this study, we explore the spatial structure of the BTHUG using a proposed spatial network analysis framework based on a large-scale mobile phone dataset, which covers 20 million mobile phone users in the BTHUG. The spatial network framework can be applied to other spatial scales (e.g., intra-urban, inter-urban, nationwide) for spatial structure analysis, and other spatial interaction analyses of areas such as trade, public transportation, or urban freight logistics. During the past 40 years since the opening-up of China, the dramatic increases in flows of people, vehicles, goods, and capital have reshaped China's urban agglomeration structure along with the top-down administrative urban agglomeration planning approach of the central government. Thus, discussions about the urban agglomeration spatial structure for the BTHUG are rooted in the particular context of spatial interaction patterns in a spatial network.

We first construct a spatial weighted-directed network *G*, derived from over 20 million mobile phone users as they move among the counties in the BTHUG. By computing the spatial network degree, strength, rich-club coefficient, and assortativity coefficient, we observe a hierarchical urban structure shaped by human movement, and a "rich-member club" (i.e., the global center area) consisting of a few counties in Beijing, which are central to the BTHUG. In addition, three selected community detection algorithms are applied and compared to detect the organization of community structure in BTHUG. The results indicate using different community detection methods for a spatial network yield significantly different communities structure. Especially, using Infomap and Louvain algorithms will detect spatially similar community detection, which also corresponds well with top-down administrative cities. Meanwhile, using Regionalization method will generate more geographically compact regions which are quite different from top-down administrative cities in BTHUG.

Then, the detected community from Infomap which has the highest association with top-down administrative cities has been selected to further analyze the urban agglomeration structure. The Infomap algorithm identifies a hierarchical spatial structure consisting of 15 communities for the spatial network *G*, which are consistent with the top-down administrative city structure. The hierarchical spatial structure consists of one global center, two local centers, major cities with strong linkages with the centers, and peripheral cities that have weaker linkages with the centers. It should be noted the observed spatial structure of BTHUG is consistent with Zhu et al. (2020) by using social economical and transport dataset. The results also suggest that the top-down administrative city unit restricts human movement in terms of spatial interaction, but such movement tends to break through this constraint and reshape the spatial structure.

These empirical findings can remind policy-makers that it is necessary to rethink whether the administrative planning during the wave of urban agglomeration development is in fact rooted in spatial interaction patterns, or is only "a forced marriage" from the top-down. Especially, the results obtained in this study implies that the spatial structure of BTHUG is over-dependent on the Global center (i.e., Beijing and northern Langfang), which may lead to more series issues in Beijing urban sustainable development (e.g., traffic congestion and air pollution) (Xu et al. 2019, Zhao and Hu 2019). Further, although the top-down policy has encouraged the development of northern Tianjin, it has not formed an integrated development trend with southern Tianjin from the perspective of spatial interactions. It suggests that ignoring spatial interaction patterns in BTHUG development may lead to ineffective integrated development. Our research also identifies the necessity to consider spatial interaction patterns together with this top-down planning approach in future research into urban agglomeration integrated development.

Our study has the following limitations. First, there is much potential to further extend our case study on the evolution of the BTHUG structure. Unfortunately, we are limited by the availability of data. By collecting multi-year data, our spatial network method could be extended to investigate how human movements break down the constraints of city borders and reshape the structure of the BTHUG year by year. Meanwhile, our data source has been widely recognized as producing valuable material for large-scale (e.g., inter-county interaction) geographical research, as it covers the population with comprehensive demographic information, but mobile phone data only provide information about user movement without details of transportation behavior (e.g., driving, rail, or public transit). This data characteristic limits the correlation between the BTHUG structure and transportation networks. Further, in spatial terms, a comparative analysis of urban agglomeration structures could be conducted based on our method, if other data (e.g., from the Yangtze River Delta or the Guangdong-Hong Kong-Macao Greater Bay Area) can be collected.

Finally, since the Infomap and Louvain cannot always converge the same result, it is necessary to implement multiple times to generate a stable result (as we demonstrated in section 4.3). However, the Infomap and Louvain used in this study still can allow researchers to explore the spatial structure. Because the results obtained by Infomap and Louvain in this study are consistent with those obtained using other mobility data source (e.g., social media and smart card data), which report that the detected communities in the geographic space generally correspond well with top-down administrative borders (Zhong et al. 2014, Lengyel et al. 2015, Yin et al. 2017). Future studies would, of course, benefit from developing more stable and insensitive community detection algorithms.

Funding: This research was supported by a grant from RGC Early Career Scheme (Grant no. P0030875).

Appendix A

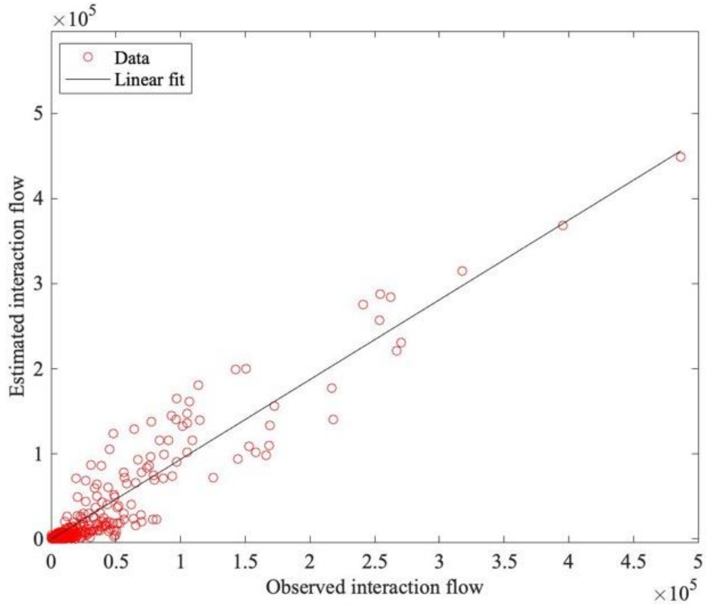

Figure S1. The observed interaction strength versus the estimated ones from the adopted gravity model with a fitted $\beta = 0.9$. The dark line indicates strong linear correlation between the estimated and observed interaction strength with $R^2=0.84$ and *p-value* < 0.01.

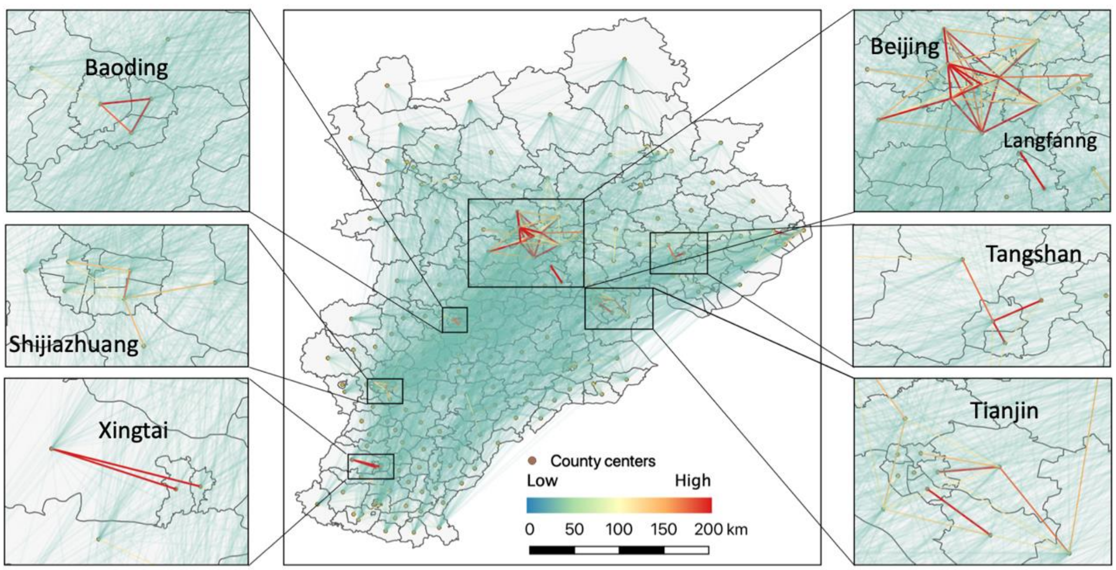

Figure S2. The observed interaction strength which is larger than the estimated ones from the adopted gravity model (i.e., ***Modularity*** **> *0***).

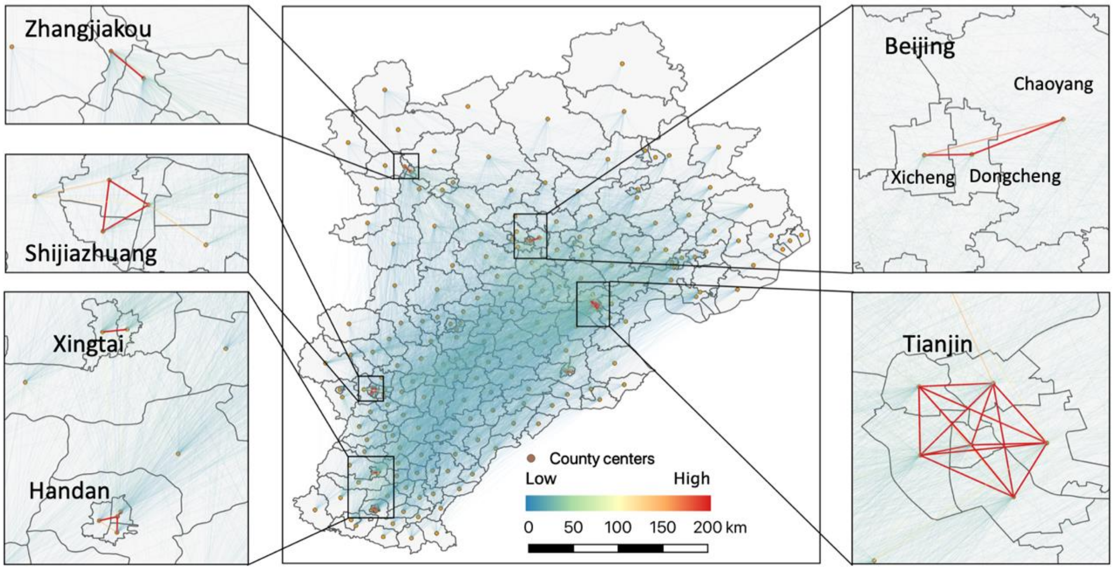

Figure S3. The observed interaction strength which is larger than the estimated ones from the adopted gravity model (i.e., ***Modularity < 0***).

Table S1. The results of the Rand Index (RI) analysis between each pair of the methods

	City	***Infomap***	***Louvain***	***REDCAP***
City	-	0.85	0.76	0.24
Infomap	-	-	0.80	0.24
Louvain	-	-	-	0.20
REDCAP	-	-	-	-

References

Alstott, J., Panzarasa, P., Rubinov, M., Bullmore, E.T. and Vértes, P.E. (2014). A unifying framework for measuring weighted rich clubs. Scientific reports, 4, 7258. doi: 10.1038/srep07258

Batty, M. (2013). The new science of cities. MIT Press.

Blondel, V.D., Guillaume, J.L., Lambiotte, R. and Lefebvre, E. (2008). Fast unfolding of communities in large networks. Journal of statistical mechanics: theory and experiment, 2008(10), 10008. doi: 10.1088/1742-5468/2008/10/P10008

Burgess E.W. (2008) The Growth of the City: An Introduction to a Research Project. In: Marzluff J.M. et al. (eds) Urban Ecology. Springer, Boston, MA. https://doi.org/10.1007/978-0-387-73412-5_5.

Cao, S., Hu, D., Hu, Z., Zhao, W., Chen, S. and Yu, C. (2018). Comparison of spatial structures of urban agglomerations between the Beijing-Tianjin-Hebei and Boswash based on the subpixel-level impervious surface coverage product. Journal of Geographical Sciences, 28(3), 306-322. doi: 10.1007/s11442-018-1474-0

Castells, M. (2011). The rise of the network society (Vol. 12). John wiley & sons.

Colizza, V., Flammini, A., Serrano, M.A. and Vespignani, A. (2006). Detecting rich-club ordering in complex networks. Nature physics, 2(2), 110-115. doi: 10.1038/nphys209

Colizza, V., Pastor-Satorras, R. and Vespignani, A. (2007). Reaction–diffusion processes and metapopulation models in heterogeneous networks. Nature Physics, 3(4), 276-282. doi: 10.1038/nphys560

De Montis, A., Caschili, S. and Chessa, A. (2013). Commuter networks and community detection: a method for planning sub regional areas. The European Physical Journal Special Topics, 215(1), 75-91. doi: 10.1140/epjst/e2013-01716-4

Ducruet, C., Cuyala, S. and Hosni, A.E. (2016). The changing influence of city-systems on global shipping networks: an empirical analysis. Journal of Shipping and Trade, 1(1), 4. doi: 10.1186/s41072-016-0006-2

Dugué, N. and Perez, A., 2015. Directed Louvain: maximizing modularity in directed networks. Université d'Orléans.

Fortunato, S. (2010). Community detection in graphs. Physics reports, 486(3-5), 75-174. doi: 10.1016/j.physrep.2009.11.002

Foster, J.G., Foster, D.V., Grassberger, P. and Paczuski, M. (2010). Edge direction and the structure of networks. Proceedings of the National Academy of Sciences, 107(24), 10815-10820. doi: 10.1073/pnas.0912671107

Fotheringham, A. S. (1981). Spatial structure and distance-decay parameters. Annals of the Association of American Geographers, 71(3), 425-436. doi: 10.1111/j.1467-8306.1981.tb01367.x

Gao, S., Liu, Y., Wang, Y. and Ma, X. (2013). Discovering Spatial Interaction Communities from Mobile Phone Data. Transactions in GIS, 17(3), 463-481. doi: 10.1111/tgis.12042

Guo, D. (2008). Regionalization with dynamically constrained agglomerative clustering and partitioning (REDCAP). International Journal of Geographical Information Science, 22(7), 801-823. doi: 10.1080/13658810701674970

Guo, D. (2009). Flow mapping and multivariate visualization of large spatial interaction data. IEEE Transactions on Visualization and Computer Graphics, 15(6), 1041-1048. doi: 10.1109/TVCG.2009.143

Guo, D., Jin, H., Gao, P. and Zhu, X. (2018). Detecting spatial community structure in movements. International Journal of Geographical Information Science, 32(7), 1326-1347. doi: 10.1080/13658816.2018.1434889

Harris, C.D. and Ullman, E.L. (1945). The nature of cities. The Annals of the American Academy of Political and Social Science, 242(1), 7-17.

Hoyt, H. (1939). The structure and growth of residential neighborhoods in American cities. US Government Printing Office.

Huang, J., Liu, X., Zhao, P., Zhang, J. and Kwan, M.P. (2019). Interactions between bus, metro, and taxi use before and after the Chinese Spring Festival. ISPRS International Journal of Geo-Information, 8(10), 445. doi: 10.3390/ijgi8100445

Huang, Q. and Wong, D.W. (2016). Activity patterns, socioeconomic status and urban spatial structure: what can social media data tell us?. International Journal of Geographical Information Science, 30(9), 1873-1898. doi: 10.1080/13658816.2016.1145225

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(3), 431-448. doi: 10.1080/13658816.2012.689838

Kempinska, K., Longley, P. and Shawe-Taylor, J. (2018). Interactional regions in cities: making sense of flows across networked systems. International Journal of Geographical Information Science, 32(7), 1348-1367. doi: 10.1080/13658816.2017.1418878

Kuang, W., Chi, W., Lu, D. and Dou, Y. (2014). A comparative analysis of megacity expansions in China and the US: Patterns, rates and driving forces. Landscape and urban planning, 132, 121-135. doi: 10.1016/j.landurbplan.2014.08.015

Leicht, E.A. and Newman, M.E. (2008). Community structure in directed networks. Physical review letters, 100(11), 118703. doi: 10.1103/PhysRevLett.100.118703

Lengyel, B., Varga, A., Ságvári, B., Jakobi, Á. and Kertész, J. (2015). Geographies of an online social network. PlOS ONE, 10(9), e0137248. doi: 10.1371/journal.pone.0137248

Li, Y. and Phelps, N.A. (2017). Knowledge polycentricity and the evolving Yangtze River Delta megalopolis. Regional Studies, 51(7), 1035-1047. doi: 10.1080/00343404.2016.1240868

Li, Y. and Phelps, N.A. (2019). Megalopolitan glocalization: the evolving relational economic geography of intercity knowledge linkages within and beyond China's Yangtze River Delta region, 2004-2014. Urban Geography, 40(9), 1310-1334. doi: 10.1080/02723638.2019.1585140

Liu, X., Derudder, B. and Wu, K. (2016). Measuring polycentric urban development in China: An intercity transportation network perspective. Regional Studies, 50(8), 1302-1315. doi: 10.1080/00343404.2015.1004535

Liu, X., Gong, L., Gong, Y. and Liu, Y. (2015). Revealing travel patterns and city structure with taxi trip data. Journal of Transport Geography, 43, 78-90. doi: 10.1016/j.jtrangeo.2015.01.016

Liu, Y., Kang, C., Gao, S., Xiao, Y. and Tian, Y. (2012). Understanding intra-urban trip patterns from taxi trajectory data. Journal of geographical systems, 14(4), 463-483. doi: 10.1007/s10109-012-0166-z

Liu, Y., Sui, Z., Kang, C. and Gao, Y. (2014). Uncovering patterns of inter-urban trip and spatial interaction from social media check-in data. PloS one, 9(1), e86026. doi: 10.1371/journal.pone.0086026

Louail, T., Lenormand, M., Picornell, M., Cantú, O.G., Herranz, R., Frias-Martinez, E., Ramasco, J.J. and Barthelemy, M. (2015). Uncovering the spatial structure of mobility networks. Nature communications, 6, 6007. doi: 10.1038/ncomms7007

Ma, L.J. (2005). Urban administrative restructuring, changing scale relations and local economic development in China. Political Geography, 24(4), 477-497. doi: 10.1016/j.polgeo.2004.10.005

Newman, M.E. (2002). Assortative mixing in networks. Physical review letters, 89(20), 208701. doi: /10.1103/PhysRevLett.89.208701

Opsahl, T. (2009). Structure and evolution of weighted networks, Doctoral dissertation, Queen Mary, University of London.

Opsahl, T., Colizza, V., Panzarasa, P. and Ramasco, J.J. (2008). Prominence and control: the weighted rich-club effect. Physical review letters, 101(16), 168702. doi: 10.1103/PhysRevLett.101.168702

Rand, W. M. (1971). Objective criteria for the evaluation of clustering methods. Journal of the American Statistical association, 66(336), 846-850.

Ratti, C., Sobolevsky, S., Calabrese, F., Andris, C., Reades, J., Martino, M., Claxton, R. and Strogatz, S.H. (2010). Redrawing the map of Great Britain from a network of human interactions. PloS one, 5(12), e14248. doi: 10.1371/journal.pone.0014248

Rodrigue, J.P. (2016). The geography of transport systems. Taylor & Francis.

Rosvall, M. and Bergstrom, C.T. (2008). Maps of random walks on complex networks reveal community structure. Proceedings of the National Academy of Sciences, 105(4), 1118-1123. doi: 10.1073/pnas.0706851105

Shaw, S.L., Tsou, M.H. and Ye, X. (2016). Human dynamics in the mobile and big data era. International Journal of Geographical Information Science, 30(9), 1687-1693. doi: 10.1080/13658816.2016.1164317

Steiger, E., Westerholt, R., Resch, B. and Zipf, A. (2015). Twitter as an indicator for whereabouts of people? Correlating Twitter with UK census data. Computers, environment and urban systems, 54, 255-265. doi: 10.1016/j.compenvurbsys.2015.09.007

Steinley, D. (2004). Properties of the Hubert-Arable Adjusted Rand Index. Psychological methods, 9(3), 386. doi: 10.1037/1082-989X.9.3.386

Taylor, P.J. and Derudder, B. (2015). World city network: a global urban analysis. Routledge.

Taylor, P.J., Catalano, G. and Walker, D.R. (2002). Measurement of the world city network. Urban Studies, 39(13), 2367-2376. doi: 10.1080/00420980220080011

Van Den Heuvel, M.P. and Sporns, O. (2011). Rich-club organization of the human connectome. Journal of Neuroscience, 31(44), 15775-15786. doi: 10.1523/JNEUROSCI.3539-11.2011

Wang, Y., Dong, L., Liu, Y., Huang, Z. and Liu, Y. (2019). Migration patterns in China extracted from mobile positioning data. Habitat International, 86, 71-80. doi: 10.1016/j.habitatint.2019.03.002

Wang, Z., Zhang, F. and Wu, F. (2017). Neighbourhood cohesion under the influx of migrants in Shanghai. Environment and Planning A: Economy and Space, 49(2), 407-425. doi: 10.1177/0308518X16673839

Wei, Y., Song, W., Xiu, C. and Zhao, Z. (2018). The rich-club phenomenon of China's population flow network during the country's spring festival. Applied Geography, 96, 77-85. doi: 10.1016/j.apgeog.2018.05.009

Wu, F. (2002). China's changing urban governance in the transition towards a more market-oriented economy. Urban Studies, 39(7), 1071-1093. doi: 10.1080/00420980220135491

Wu, F. (2016). China's emergent city-region governance: a new form of state spatial selectivity through state-orchestrated rescaling. International Journal of Urban and Regional Research, 40(6), 1134-1151. doi: 10.1111/1468-2427.12437

Wu, F. (2020). Emerging cities and urban theories: a Chinese perspective. In Theories and models of urbanization (pp. 171-182). Springer, Cham. doi: 10.1007/978-3-030-36656-8_10

Wu, W., Wang, J. and Dai, T. (2016). The geography of cultural ties and human mobility: Big data in urban contexts. Annals of the American Association of Geographers, 106(3), 612-630. doi: 10.1080/00045608.2015.1121804

Xing, Y., Lu, J. and Chen, S. (2016). Weighted complex network analysis of shanghai rail transit system. Discrete Dynamics in Nature and Society. 2016, 8. doi: 10.1155/2016/1290138

Xu, X.K., Zhang, J. and Small, M. (2010). Rich-club connectivity dominates assortativity and transitivity of complex networks. Physical Review E, 82(4), 046117. doi: 10.1103/PhysRevE.82.046117

Xu, Y., Jiang, S., Li, R., Zhang, J., Zhao, J., Abbar, S. and González, M.C. (2019). Unraveling environmental justice in ambient PM2. 5 exposure in Beijing: A big data approach. Computers, Environment and Urban Systems, 75, 12-21. doi: 10.1016/j.compenvurbsys.2018.12.006

Xu, Y., Li, J., Belyi, A. and Park, S. (2020). Characterizing destination networks through mobility traces of international tourists—A case study using a nationwide mobile positioning dataset. Tourism Management, 82, 104195. doi: 10.1016/j.tourman.2020.104195

Yeh, A.G.O. and Chen, Z. (2020). From cities to super mega city regions in China in a new wave of urbanisation and economic transition: Issues and challenges. Urban Studies, 57(3), 636-654. doi: 10.1177/0042098019879566

Yin, J., Soliman, A., Yin, D. and Wang, S. (2017). Depicting urban boundaries from a mobility network of spatial interactions: a case study of Great Britain with geo-located Twitter data. International Journal of Geographical Information Science, 31(7), 1293-1313. doi: 10.1080/13658816.2017.1282615

Zhang, B, (2016). Coordinated Development for the Beijing-Tianjin-Hebei Region: A Strategic Trade-off in State Spatial Governance. In Urban Planning Forum (No. 4, p. 5).

Zhang, W., Derudder, B., Wang, J. and Shen, W. (2018). Regionalization in the Yangtze River Delta, China, from the perspective of inter-city daily mobility. Regional Studies, 52(4), 528-54. doi: 10.1080/00343404.2017.1334878

Zhao, P. and Hu, H. (2019). Geographical patterns of traffic congestion in growing megacities: Big data analytics from Beijing. Cities, 92, 164-174. doi: 10.1016/j.cities.2019.03.022

Zhao, P., Liu, X., Shi, W., Jia, T., Li, W. and Chen, M. (2018). An empirical study on the intra-urban goods movement patterns using logistics big data. International Journal of Geographical Information Science, 34(6), 1089-1116. doi: 10.1080/13658816.2018.1520236

Zhong, C., Arisona, S.M., Huang, X., Batty, M. and Schmitt, G. (2014). Detecting the dynamics of urban structure through spatial network analysis. International Journal of Geographical Information Science, 28(11), 2178-2199. doi: 10.1080/13658816.2014.914521

Zhong, C., Zeng, S., Tu, W. and Yoshida, M. (2018). Profiling the spatial structure of London: from individual tweets to aggregated functional zones. ISPRS International Journal of Geo-Information, 7(10), 386. doi: 10.3390/ijgi7100386

Zhu, X., Wang, Q., Zhang, P., Yu, Y. and Xie, L. (2020). Optimizing the spatial structure of urban agglomeration: based on social network analysis. Quality & Quantity. doi: 10.1007/s11135-020-01016-3