Hindawi
Journal of Advanced Transportation
Volume 2020, Article ID 8835981, 15 pages
https://doi.org/10.1155/2020/8835981

Research Article

Mobile Phone Data in Urban Commuting: A Network Community Detection-Based Framework to Unveil the Spatial Structure of Commuting Demand

Qing Yu,[1,2] **Weifeng Li**,[1] **Dongyuan Yang**,[1] **and Haoran Zhang**[2]

[1]*Key Laboratory of Road and Traffic Engineering of the Ministry of Education, Tongji University, 4800 Cao'an Road, Shanghai 201804, China*
[2]*Center for Spatial Information Science, The University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa-shi, Chiba 277-8568, Japan*

Correspondence should be addressed to Weifeng Li; lwf0103@qq.com

Received 19 March 2020; Revised 28 October 2020; Accepted 29 October 2020; Published 16 November 2020

Academic Editor: Giuseppe Musolino

As the outcomes of rapid urbanization, the spatial separation of homes and workplaces extends the commuting distance and complicates the commuting demand of residents. To promote urban livability and sustainability, it becomes crucially important to understand the commuting patterns by decomposing and simplifying the diverse commuting demand. In this paper, a methodology framework is proposed to describe the spatial structure of commuting demand in a city using mobile phone data. Four steps are mainly included in the proposed methodology: the preprocessing of mobile phone data, the labeling of individuals and their activity points, the construction of the jobs-housing relationship network, and the network decomposition based on the community detection algorithm. To demonstrate the practical use of the proposed methodologies, a case study is carried out in Shanghai to explore the commuting patterns of Shanghai residents. The result indicates the regions with dense jobs-housing connections and cross-regional commuting demand. The result also finds that the administrative boundaries show a significant effect on the residential commuting behavior and the metro lines on the cross-regional commuting behavior. The results generated by the methodology proposed can be referenced by policymakers to support urban transportation planning and promote urban livability and sustainability.

1. Introduction

Commuting is defined as the regular travel between one's place of residence and place of work or full-time study. According to the comprehensive traffic surveys conducted in the major cities of China, commuting trips averagely accounts for as much as 40%–50% of weekday daily trips in a city. As a substantial component of urban transportation and individual mobility, commuting plays a very important role in the overall travel patterns of residents and determines urban livability and sustainability [1].

As the outcomes of rapid urbanization, large cities gradually form the metropolitan areas consisting of urban areas, subcity satellites, and intervening rural areas [2, 3]. In the urban space reconstruction, the spatial separation of home and workplace extends the distance of commute. Additionally, traffic congestion, environmental pollution, and the decline of life quality are also the consequences of jobs-housing separation [4]. Many studies find that commuting has a great impact on the residents' well-being [5, 6]. Commuting pattern is also widely regarded as an indicator of urban spatial structure [7]. Therefore, understanding the commuting patterns of residents and unveiling the spatial structure of commuting demand throughout the city are the prerequisites for the promotion of livability and sustainability.

However, in current practice, the discussion of commuting demand is mainly focused on the spatial distribution of commuting trips and to examine the spatial distribution of housing and job opportunities separately [8, 9]; such a way of consideration failed to capture the connection generated by the jobs-housing flow on the urban spatial structure.

Instead of seeing city as mere morphological entities with clear and detectable borders, in the recent discussion of urban development, the urban form of most urban regions is constructed by the functional network of commuting communities, which may be physically separated but connected through dense flows of commuting trips and other forms of daily mobility [10]. Under such a concept, to study cities, we should study the network and examine the "space of flow" [11].

In recent years, the newly arisen pervasive, geospatial data generated by individuals are widely used in studying individual mobility patterns [12], urban emissions [13], newly arisen transportation mode [14, 15], and city structure and city dynamic [16]. As a new travel survey tool, mobile phone data are more pervasive and accurate than the existing traditional methods, which provide a more complete track of the spatiotemporal movements at the individual level. It offers a new approach to study the jobs-housing relationship and urban commuting demand structure [17]. Mobile phone data can track individual travels and have been proven to provide the temporal and spatial resolution to human mobility in cities. It could be the potential data source to capture the commuting flows and study the urban commuting demand structure.

This paper proposes a methodology for describing the spatial structure of commuting flows in a city on the network connection aspect using mobile phone data. Four steps are mainly included in the proposed methodology: the preprocessing of mobile phone data, the labeling of individuals and their activity points, the construction of the jobs-housing relationship network, and the network decomposition based on the community detection algorithm. The primary outputs of the methodology are the nonoverlapping communities representing the division of spatial units with dense internal jobs-housing connection and the overlapping communities unveiling the association between commuting flows and other factors. A case study is conducted using mobile phone data collected over 15 days in September 2011 in Shanghai. The spatial structure of commuting flows in Shanghai is unveiled and analyzed based on the proposed framework.

After this section, this paper is organized as follows. Section 2 gives a literature review on the related work. Section 3 describes the problem of this paper. Section 4 introduces the methodology of this study. Section 5 conducts a case study of Shanghai using the mobile phone data and explores the spatial structure of commuting demand based on the proposed methodology. Finally, the contribution and future directions of this study are described in Section 6.

2. Related Work

Understanding the urban commuting demand and the jobs-housing relationship has long been considered as an essential research topic in urban studies. Many studies have provided evidence for the close relationship between commuting and the livability and sustainability of a city [7, 18]. These studies find that commuting behaviors not only result from life choices but also affect people's lives. Choi et al. examined the relative impacts of commuting time with the overall well-being and happiness of the residents by using survey data, suggesting that reduced congestion can improve the public subjective well-being [5]. Based on survey data, Zhu et al. compared the commuting pattern and effects of different groups of people and explored its relationship with the residents' overall well-being [19].

Dwelling and employment are the two fundamental elements of a city. The commuting behaviors of residents is closely related to the structure of a city. There are two main approaches to assessing the structure of city regions [20]. One is the morphological approach, which employs the attributes or internal characteristics of centers, such as the number of jobs [21]. The morphological approach assesses the city structure on the spatial pattern, with the balance in the size distribution or distribution of absolute importance of centers based on the data from field surveying, remote sensing, and policy consulting.

The other is the functional approach, classifying the metropolitan spatial structure based on the structure of flows within spatial systems [20]. The functional approach believes that the underlying structure of a city is determined by the flows of people, freight, money, and information, which connect the discrete places into an integrated system. More and more scholars are trying to capture the structure inside large cities or even the interaction between cities by studying flows using new sources of data. The new sources of data include public transportation card data [22], taxi trip data [16], and business services network data [23]. However, because of the limitations of data access, analytic tools, and computation capabilities, studies of human travel flows had limited development.

Traditionally, studies on the jobs-housing relationship in a city are usually based on survey data, which is called the small data [17]. Such a way of capturing commuting demand has several weak points: on the one hand, it is costly and inefficient [24], which makes it unable to easily cover large groups of the population. For example, the fifth travel survey of residents in Shanghai can only cover 0.8% of the residents [25], and on the other hand, survey data can only record the residents' commuting behavior in a short period with low accuracy [26].

In the past decade, the emergence of big geospatial data has triggered the opportunity of studying the human mobility pattern. Since 2005, Ahas and Mark [27] foresee that mobile phone data can be used for investigating the space-time behavior of society. In 2006, Ratti et al. [28] proposed that location-based services (LBS) data could become a powerful tool for urban analysis. Using mobile phone data, they studied the intensity of urban activities and their evolution through space and time at different times of the day. In 2010, Ahas et al. studied the daily commuting pattern of a subgroup of commuters and identified meaningful locations of mobile phone users [29]. Based on the research of Ahas et al. and Louail et al., researchers used LBS data to understand cities in various situations, including studying significant regions in cities by capturing flows of people or identifying activity hotspots [30, 31], studying the impact of

jobs-housing spatial mismatch on commuting behavior [32], understanding the spatial structure of urban commuting [33], and using nighttime light imagery and social media check-in map to identify the structure of polycentric cities [34].

In recent years, dozens of studies focus on using big geospatial data to capture user traveling behaviors in large cities or urban agglomeration areas [35]. Croce et al. dedicated to integrate the data fusion of traditional transport surveys data with big data and offer support for building transport system models. They also present formal criteria and thresholds to characterize and segment passenger mobility [36]. Harrison et al. pointed out in their paper that passively collected GPS-based "Track & Trace" datasets of individual mobility have great potential in enhancing transportation modeling and policy-making [37]. Zhang et al. investigated the temporal variations of trip-destination distributions and their association with city spatial structure using four types of inhomogeneous Poisson point process models [38]. Tang et al. proposed a method based on entropy-maximizing theory to model OD distribution in Harbin city using large-scale taxi GPS trajectories [39]. These studies validate the feasibility of using geospatial data to analyze the spatial-temporal features of urban travel patterns. Ghahramani et al. have explored the potential of using mobile phone data to study the inter and intra-interaction patterns of the urban community structure and identify activity hotspots, while they did not consider the overlapping community structure of urban interaction patterns [40–42].

In summary, studies are focusing on using a new source of geospatial data as a supplement of traditional survey data and a much more frequently updated data source for supporting urban planning. Given the above examples and features of big data, mobile phone data have great potential for examining the spatial structure of the commuting patterns in a city.

3. Problem Description

In this paper, the spatial structure of commuting demand concerns the spatial distribution of activities associated, characterized by the centralization and clustering of the associated activities. The spatial separation of home and workplace not only extends the commuting distance but also complicates the commuting patterns in the city. On the one hand, the commuting behavior varies from person to person. The commuting demand of residents in the city is an integrated comprise of the demand from different levels with different travel time, distance, frequency, and volume. On the other hand, commuting demand is influenced by many external factors. For example, according to the study conducted in Beijing, China, commuters who live along the expressways are more likely to have a long-distance commute 2019. Therefore, the main task of this paper is to clarify the spatial structure of commuting flows based on the massive input of commuting demand and reveal the relationship between commuting flows and other external factors.

Mobile phone data provided by the mobile operator are not initially collected for the analysis of human movement, but for the purposes of billing and operation. This paper tries to answer the following questions: how can we describe the commuting behaviors of residents in a city using mobile phone data? With the massive input of residents' commuting behaviors, how can we depict the spatial structure of commuting demand throughout the city?

In response to the questions mentioned above, first, the raw mobile phone data are preprocessed to mitigate the data noise. To describe the commuting behaviors, mobile users and their activity points are labeled according to some preset rules. To understand the spatial structure of the commuting demand, the commuting flows throughout the city are used to construct the network representing the spatial distribution of commuting flows. Network analysis is introduced in this paper as the tool to analyze the commuting flows in the city. The structure of the network can be a persuasive proof for the spatial structure of commuting demand.

4. Methodology

4.1. Framework. The framework of the methodology proposed in this paper is shown in Figure 1. In this paper, mobile phone data are used to unveil the spatial structure of commuting demand. The methodology can be divided into three steps: (1) data preprocessing: extract the human mobility information from the mobile phone dataset and mitigate the noise in the data by using the binning method. (2) Extracting user jobs-housing information: label mobile users as residents and commuters, label their activity points as home and workplace, and construct a jobs-housing relationship network to represent the commuting connection in a city. (3) Mining urban commuting demand structure: by using two types of network community detection methods, the spatial structure of commuting demand in a city can be depicted from two different aspects.

4.2. Data Preprocessing. Once a user can be captured by more than one BTS simultaneously, its signal will be handed over frequently between these BTSs and generate a significant number of records in a very short time. The frequent handover does not only lead to the waste of computational resources but also the misjudgment of spatial movement. Therefore, a binning method [43] was used to cope with this problem and reduce the volume of data. The resolution of spatial grids is set to 500 m ∗ 500 m, for not being too small to affect the activity intensity of users [44].

(1) A grid set is generated to cover the study area and reflect the spatial location

(2) The average positions of every user for every 10 minutes are calculated and define which grid it belongs to. The centroid of this grid will be regarded as the position of the user during the 10-minute period.

The binning will generate a set of control points for each phone user, formulated as the following equation:

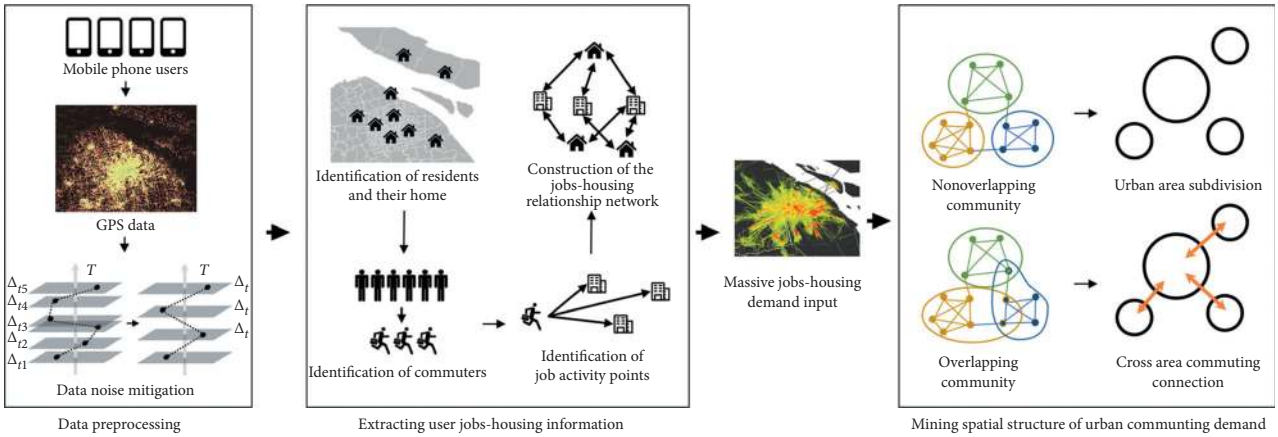


Figure 1: Framework of the method introduced.

$$c_i(\text{grid}_i, t_i), \quad i = 1, 2, \cdots, \tag{1}$$

where c_i refers to the i^{th} control point of a user; t_i represents the time when the user arrived at the control point.

4.3. Extracting User Jobs-Housing Information

4.3.1. Identification of Residents and Their Home. Due to the large data volume, a simple method proposed by Li et al. is utilized to identify the home locations of residents from mobile phone data [43]. The method includes two rules:

(1) From 9 p.m. to 9 a.m. of the next day, the user stays at a place for no less than 6 hours

(2) In our observation periods, the user stays in the place meeting the rule (a) for more than 2/3 days

If a user satisfies both the rules, the user can be considered as a resident, and the place will be considered as the location of the home.

4.3.2. Identification of Commuters. In previous studies, workplaces are often considered to be unique for every commuter. The identifying methods of work location always find a fixed place based on the regularity of individual travel patterns during the observation days [45]. In this way, the commuters who have multiple work locations are neglected. To avoid this defect, commuters are identified by the following method.

Obviously, if a resident appears in a place other than his home, he can be regarded as goes out. An assumption is proposed in this study that commuters should be the residents who stay significantly more time outside home on weekdays when comparing with noncommuters.

A simple method amounts to choose a threshold δ and to consider that the resident with the average stay time outside home over δ on weekday as a commuter; otherwise, he/she is a noncommuter. Here, the average stay time outside home of all residents is chosen as the threshold δ, which will split residents into two equal parts—commuters and noncommuters.

4.3.3. Identification of Job Activity Points. Given the i^{th} control point and the following k control points where $\text{grid}_{i+k} \neq \text{grid}_{i+k-1} = \cdots = \text{grid}_i \neq \text{grid}_{i-1}, k \geq 1$, the activity duration of the user who stays at grid_i can be calculated using the following equation:

$$d = t_{i+k-1} - t_i. \tag{2}$$

According to the household travel survey of Shanghai, 30 minutes can represent the critical station of an individual's daily movement and contribute to the comprehensive understanding of individual activities [46]. Thus, the location where an individual stays over 30 minutes is defined as an activity point.

However, the mobile phone data does not contain information of the activity purposes or activity types. In this paper, activity points at work time (9 a.m.–6 p.m. on weekdays) are defined as job activity points.

4.3.4. Construction of the Jobs-Housing Relationship Network. In this study, TAZ is chosen as the analysis unit. The grid with the centroid in a TAZ is regarded as belonging to the TAZ. The reasons why we do not adopt the grid points as the analysis unit are as follows.

First, the data noise of mobile phone data is a big problem in practical application (e.g., frequent handover between adjacent BTS). The data preprocessing can only mitigate the data noise but not completely eliminate them. In fact, we find that it is never possible to completely eliminate the data noise. When the noise occurs, the user's actual position will be lost. In such circumstances, choosing the grids as the analysis unit may generate results deviate from reality. On the other hand, choosing TAZ as the analysis unit can further mitigate the influence generated by the data noise.

Second, choosing the grid as the analysis unit has a major defect. The grids are usually too small to have enough data samples that can reflect the spatial structure of commuting demand. Choosing grids as the analysis unit, most of the jobs-housing connections between grids are at a very small value. On that basis, the community connection algorithm

will have a higher possibility to mistakenly classify the grids and generate unreliable results.

For the residents in TAZ i, the number of job activity points they have per hour in TAZ j can be defined as the number of connections from TAZ i to TAZ j, which is denoted as Q_{ij} here. By aggregating all job activity points and home location, a 403×403 matrix V can be obtained as the following equation:

$$V = \begin{bmatrix} Q_{11} & Q_{12} & \cdots & Q_{1j} \\ Q_{21} & Q_{22} & \cdots & Q_{2j} \\ \vdots & \vdots & \ddots & \vdots \\ Q_{i1} & Q_{i2} & \cdots & Q_{ij} \end{bmatrix}. \tag{3}$$

To build a network, each TAZ is represented as a node. Between every node i and j, two directed edge e_{ij} and e_{ji} will be constructed with the weight Q_{ij} and Q_{ji}.

4.4. Mining Spatial Structure of Urban Commuting Demand. The commute flows within a city connect discrete places into an integrated system. Among TAZs, commute trips can be aggregated to obtain spatial interactions between zones. Constructing a network and applying network analysis methods upon TAZs, we can further understand the urban commute interactions.

In network analysis, a community is a collection of highly interconnected nodes [47]. The nodes belonging to different communities are sparsely connected. In order to retrieve comprehensive information of the structure in the complex network, we decompose the network into different communities by using community detection. It can help us divide the city into subregions with intensely interactive jobs-housing relationships. The resulting meta-network, whose nodes are the communities, will then be used to visualize the city commuting demand structure.

There are mainly two types of community detection methods, nonoverlapping community detection and overlapping community detection. For the nonoverlapping community, every node in the network can only belong to one community. A huge variety of community detection techniques have been developed based variously on centrality measures, flow models, random walks, resistor networks, modularity optimization, and many other approaches [48, 49]. The other type of approach is overlapping community detection, which believes that communities in networks often overlap and nodes can simultaneously belong to several communities [50]. These approaches include the clique percolation method [51], local optimization of fitness function [52], and clustering link communities [50]. In recent studies, algorithms are also developed to detect the evolving tendency of the overlapping communities [53, 54].

In this study, we use both methods to decompose the jobs-housing network, as we find that both methods can describe the urban commuting demand structure in different aspects.

4.4.1. Nonoverlapping Community Detection. Nonoverlapping community detection can be implemented in many algorithms [55]. Here, the fast unfolding algorithm is adopted to decompose our network [56]. This algorithm is based on modularity optimization. The modularity of a partition is a scalar value between −1 and 1 that measures the density of links inside communities as compared to links between communities [57]. It is defined as the following equation:

$$M = \frac{1}{2m} \sum_{i,j} \left[Q_{ij} - \frac{k_i k_j}{2m} \right] \delta(c_i, c_j), \tag{4}$$

where Q_{ij} represents the weight of the edge e_{ij}, k_i is the sum of the weights of the edges attached to vertex i, and c_i is the community to which vertex i is assigned; the δ-function $\delta(u, v)$ is 1 if $u = v$ and 0 otherwise, and $m = 1/2 \sum_{ij} Q_{ij}$.

This algorithm includes the following two steps which are repeated iteratively until no increase of modularity is possible:

(1) Modularity optimization: optimized modularity by allowing only local changes of communities

(2) Community aggregation: the identified communities are aggregated in order to build a new network of communities

We adopted the fast unfolding toolkit provided in Python-igraph package in this study.

4.4.2. Overlapping Community Detection. As for overlapping community detection, we use the method based on link communities clustering [50]. The basic concept of this method is assuming that nodes in the network have multiple identities, and they will cluster in corresponding communities according to their identities. In another word, communities are depending on the attribute of the links between its members. Hence, this method clusters the links by measuring the similarity of links. The nodes connected by the links in the same cluster will be regarded as belonging to the same community. For that, there will be several links connecting a single node, with these links being clustered into different clusters; the node can simultaneously belong to several communities.

In this paper, we use the linkcomm package in R to conduct overlapping community detection. This algorithm chooses the Jaccard similarity coefficient to calculate the similarity matrix for links in the network and cluster the links using hierarchical clustering. The similarity between e_{ik} and e_{jk} is formulated as the following equation:

$$S(e_{ik}, e_{jk}) = \frac{\left| n_+(i) \cap n_+(j) \right|}{\left| n_+(i) \cup n_+(j) \right|}, \tag{5}$$

where $n_+(i)$ denotes the neighbors of node i. In order to determine the best cluster number, this algorithm also introduces the index of partition density to measure the connection inside communities. The detail of the algorithm is in [58].

5. Case Study and Results

5.1. Study Case. A case study is carried out in Shanghai, the economic center of China. By the end of 2011, the

administrative territory of Shanghai consisted of 16 districts and 1 county, covering an area of about 6340 km^2. According to the master plan of city of Shanghai, the central urban area is mainly located within the outer ring expressway. The Huangpu River divides Shanghai into two parts: Pudong on the east side and Puxi on the west side.

In this paper, the study area is supposed to cover all the administrative territories of Shanghai. However, after subdividing the territory of Shanghai into 447 traffic analysis zones (TAZs), we discover varying degrees of data missing existing in the raw dataset during the study period. As a result, 403 TAZs are selected as the study area after eliminating 44 TAZs with severe data missing (Figure 2). The remaining TAZs cover all the central urban areas and satellite towns of Shanghai.

Anonymous mobile phone data used in this paper were collected for billing and operational purposes from September 1 to September 15, 2011, in Shanghai, China. The dataset contains the basic information of the wireless communication between mobile stations and base transceiver stations (BTS), including the encrypted mobile phone identifier, the service time, the service type, the geographic location of the connected BTS, and the location area (LA). That is to say, the position of mobile phone users will be represented by the location of the BTS they are connected to. A record of mobile phone data will be generated when a call is placed or received, a text message is sent or received, the phone is switched on or switched off, or the phone signal is handed over from one BTS to the other BTS. The average number of records was 1 billion per day, covering 25 million active users. The coverage radius of a BTS is 500–800 meters.

5.2. Result of User Jobs-Housing Relationship Extraction. By the identification methods, we identified 9.86 million residents, accounting for 42% of the total population of over 23.47 million in Shanghai by the end of 2011 [59]. We compare the population density identified by mobile phone data with permanent residents in the sixth national census in 2010 (Figure 3). The correlation coefficient between them is 0.91. Although deviations inevitably exist, mobile phone data can generally cover residents in the area of Shanghai.

From the 9.86 million residents identified, we first eliminate the mobile phone users who never move during the observation period (1.13 million users in total). Then, for the remaining 8.73 million residents, we calculate the average stay time outside home on weekdays for each user and plot the probability density function as shown in Figure 4. Two peaks can be found: one is around 4 hours and another is around 11 hours. Staying outside home for 11 hours is rational for a commuter on weekdays, i.e., go out at 7 or 8 a.m. and return home at 6 or 7 p.m. The mean value of stay time outside home on weekdays for all 8.73 million residents is 7.93 hours. Choosing this value as threshold δ can divide the residents into two equal parts—commuters and noncommuters.

In order to verify the job activity points identified, we introduce the net inflow index to measure whether a TAZ tends to be a job center or a residential community. In the

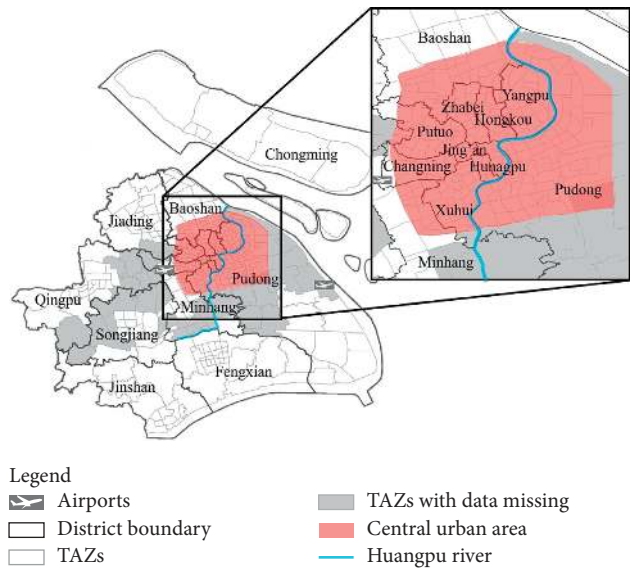


Figure 2: Study area.

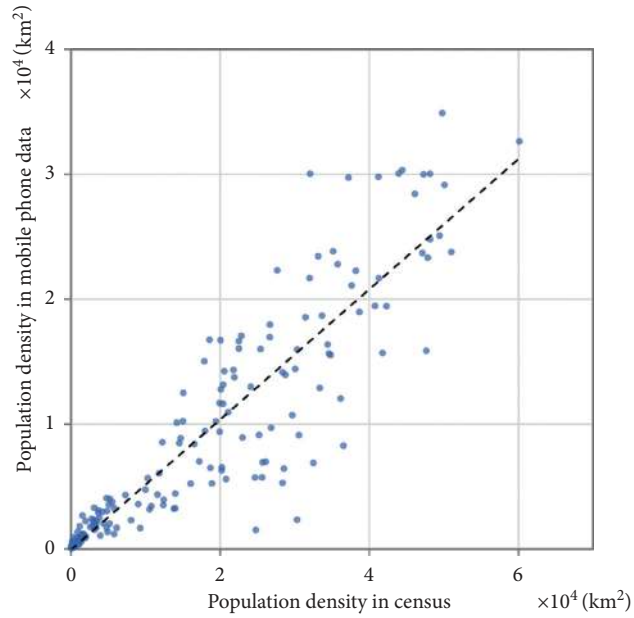


Figure 3: Comparison of residents identified with population in the census.

network we constructed, the connection between every two TAZs can be regarded as the commuting flow. The commuting flow from TAZ i to TAZ j can be considered as the outflow from TAZ i and the inflow to TAZ j. The net inflow index for TAZs is defined as the following equation:

$$\text{net}_i = \sum_i Q_{ji} - \sum_i Q_{ij}. \quad (6)$$

As the average value of net_i is 0, a TAZ with $\text{net}_i > 0$ means that more job activity points are attracted into it, which means that this TAZ is more likely to be a job center.

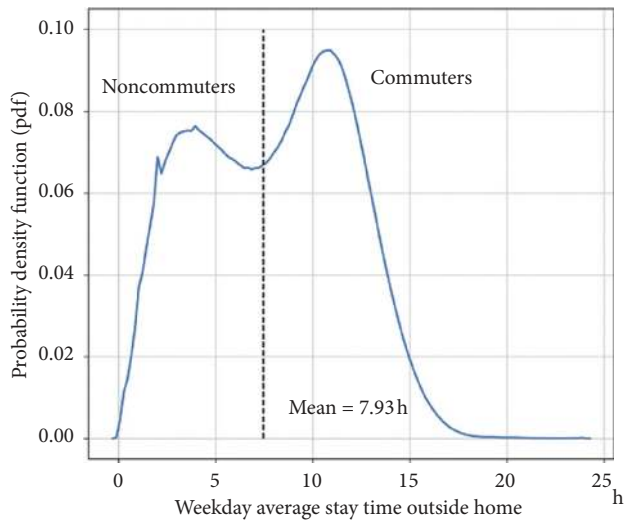


Figure 4: Identification of commuters.

While a TAZ with $\text{net}_i < 0$ means it is more likely to be a residential community.

We calculate the net inflow index for all TAZs and identify whether a TAZ is a job center or a residential community. As shown in Figure 5, TAZs 1–4 are top 4 central business districts (CBDs) in Shanghai, and TAZs 5-6 are two bases for high-tech industries; TAZs 7–10 are large residential communities. These results are in accordance with the actual land use. Therefore, the net inflow index can be used to characterize a TAZ as the job center or residential community. And it also verifies the job activity points we identified.

5.3. Result of Spatial Structure of Urban Commuting Demand

5.3.1. Result of Nonoverlapping Communities. After extracting the job activity points for commuters, we construct the jobs-housing network as the input of community detection. The nonoverlapping community detection algorithm iterates twice and finds a two-level hierarchical structure (Figure 6). In the two meta-networks constructed, whose nodes are the communities, we numbered communities in the descending order according to the number of job activity points inside them. Although we have never input any spatial relationship into the algorithm, it can still merge adjacent TAZs into the same community. The hierarchical subregional structure provides insights into how the city could be properly divided into closely related subregions based on jobs-housing relationship. Communities in the network represent regions with an intense jobs-housing connection.

One of the interesting findings is that in both the structures, the boundaries of communities perfectly coincide with administrative boundaries. In suburban districts, each community is an administrative unit. But in central urban areas, communities often involve several administrative units. The division of communities is related to the accessibility of job opportunities. In the suburban district, due to poor cross-regional traffic connections, cross-regional job opportunities are not easily accessible. But in central urban areas, the public transportation systems are well developed, which makes cross-regional employment accessible. This finding indicates that residential commuting behavior is highly restricted by administrative boundaries, especially in suburban areas. The reason can be traced back to transportation planning, which was based on the administrative division. The finding also proves the rationality of the city commuting demand structure uncovered.

In order to describe the commuting patterns between communities, we calculate the number of four types of job activity points for each community. N_1 is the total number of job activity points in the community; N_2 is the number of job activity points in the community produced by its own residents; N_3 is the number of job activity points produced by its residents but located outside the community; N_4 is the number of incoming job activity points from residents in other communities. The number of four types of activity points for each community is shown in Figure 7.

To further classify the communities, three indexes describing the numerical gaps between the four types of job activity points are proposed. The three indexes are $I_1 = N_2/N_1$, $I_2 = N_3/N_1$, and $I_3 = N_4/N_1$. Using the k-means clustering algorithm [60] and the three indexes as its input, the algorithm can easily classify communities into three clusters. The spatial distribution of communities is shown in Figure 8, and the average value of the indexes in each group is shown in Table 1. According to the characteristic of the communities, we name them as follows:

(i) Job center: Communities with a higher value of N_2 and N_4 but a lower value of N_3, which indicate that these communities contain much job opportunities and attract a great number of commuters from other communities.

(ii) Residential: Communities with a higher value of N_2 and N_3 but a lower value of N_4, which indicate that these communities are more likely as residential communities that a great part of residents has to seek job opportunities outside.

(iii) Isolated: Communities with a higher value of N_2 but a lower value of N_3 and N_4. These communities are rather isolated, for they do not attract commuters and their residents seldom work outside.

Concentric, sector, and multiple nuclei structure are the three generalizations of urban structure [61, 62]. From the result of classification, we can simplify the commuting demand structure of Shanghai into a combination of these three structures. On the city scale, we can see a multiple nuclei structure. The central urban area is the largest center, and there are several centers of isolated communities in suburban areas. In the central urban area, the Puxi area on the west side of the Huangpu River is a concentric structure, with Huangpu district (community 1 in level 1 structure) as the job center and several similar residential communities on the periphery (communities 2, 3, 4, 15, and 17 in level 1 structure). On the east side of the Huangpu River, Pudong

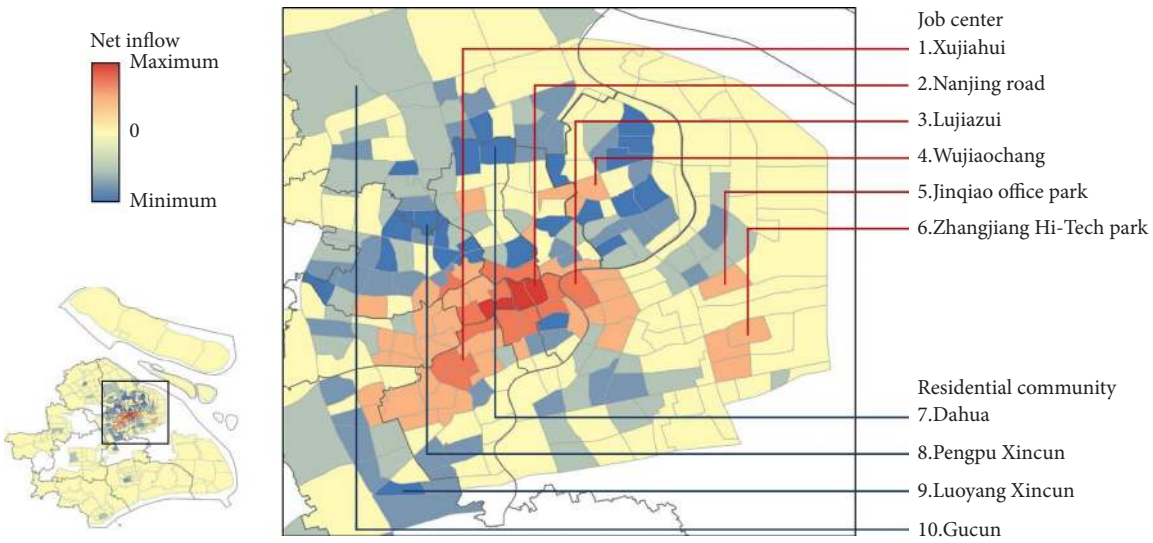


Figure 5: Net inflow index of TAZs.

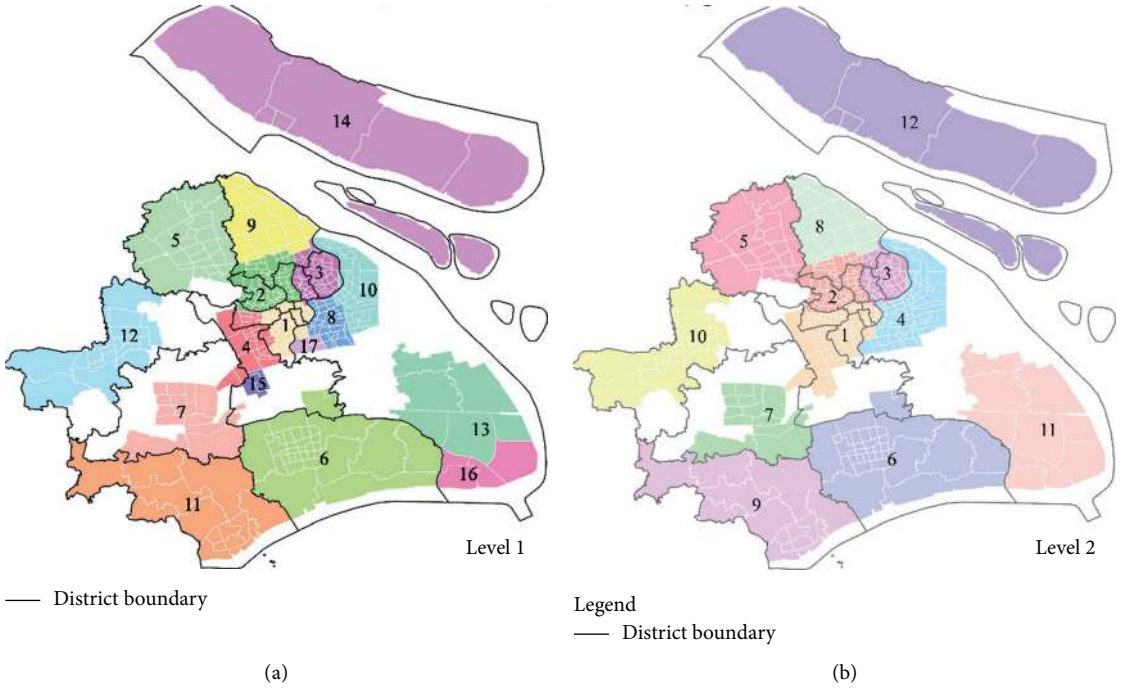


Figure 6: Result of nonoverlapping community detection: (a) level 1 (b) level 2.

district is a sector structure extended along the river. In the level 2 structure, we can clearly see the communities finally merging into a sector structure. The central urban district can be considered as a circle with four parts of areas (communities 1–4 in the level 2 structure) as sectors radiating out from the center of the circle. As a newly developed district, Pudong is a job center rather than a residential community, but the number of job activity points in the Pudong sector is still less than that of other sectors in central urban areas.

Further exploring the reason for forming the commuting demand structure, we compare the level 2 structure with the layout of the metro network (Figure 9). The metro network in Shanghai is shaped in a radial pattern, from the city center to suburban areas. In the level 2 community structure, communities in the central urban area are all extended outward along with radiating metro lines, with averagely three metro lines in one community. Communities are also formed at the end of the metro lines. From this structure, we can infer that the commuting behavior of residents living along the metro lines depends heavily on the metro line, and their workplaces aggregate along the metro line. For residents living at the end of metro lines, their workplaces are aggregated in

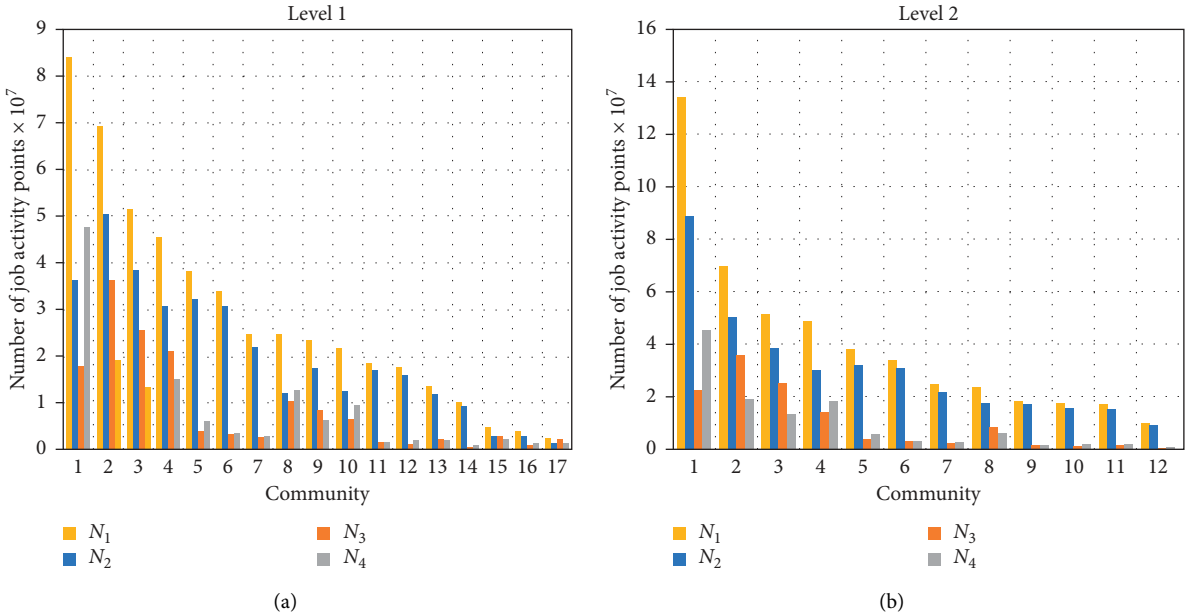


Figure 7: Number of job activity points: (a) level 1 (b) level 2.

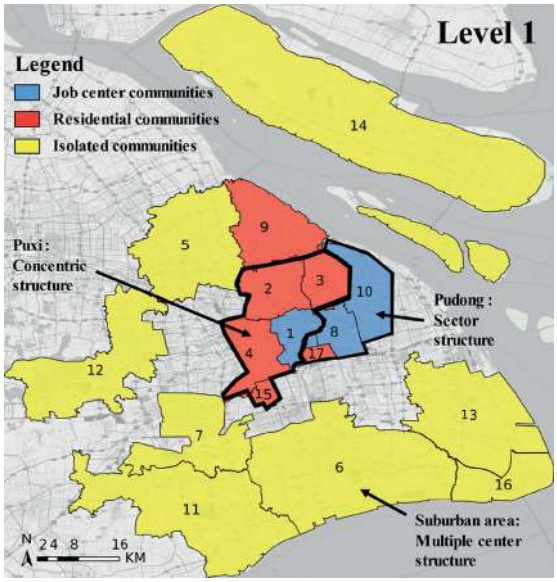


Figure 8: Three types of communities in Shanghai.

Table 1: Clustering result of communities.

Cluster	I_1 (N_2/N_1)	I_2 (N_3/N_1)	I_3 (N_4/N_1)
Cluster 1: job center	0.50	0.31	0.50
Cluster 2: residential	0.66	0.55	0.34
Cluster 3: isolated	0.87	0.10	0.13

suburban communities. This result demonstrates that, as the major traffic corridors, metro lines are playing important roles in forming the city commuting demand structure.

5.3.2. Result of Overlapping Communities. By applying overlapping community identification on the jobs-housing network, the algorithm segments the TAZs of Shanghai into 86 communities, with most of the communities merged by adjacent TAZs. Based on the shape of the communities, we classify the 86 communities into three types: large communities (17 communities in Figure 10(a), small communities (66 communities in Figures 10(b) and 10(c), and banded communities (4 communities in Figure 10(d)).

In the large communities in Figure 10(a), the first community is constructed by TAZs in the city center. The other 16 communities are all in the suburban area. These large communities show the area of the central urban district and the towns in the suburban area based on the jobs-housing relationship. In urban transportation planning, this result can help to determine the planning area. In suburban areas, the boundary of communities is mostly in correlation with the boundary of administrative districts. As comparing to nonoverlapping communities, the result is similar. In the central urban area, public transportation systems are well developed, which makes cross-regional employment easy. But in the suburban districts, due to poor cross-regional traffic connection, cross-regional job opportunities are difficult to reach.

When depicting the area of the central urban area and the towns in suburban areas, the overlapping communities also depict the small communities with the intense jobs-housing connection inside the large communities, as is shown in Figures 10(b) and 10(c). Small communities are mostly distributed in urban central areas, which indicate that there are a large number of short-distance commuters in the central area of the city. These commuters are commuting within a small area of 3–5 TAZs. In recent years, there are fewer immigrations moving into the central urban area of Shanghai. Most of the residents in the central urban area are

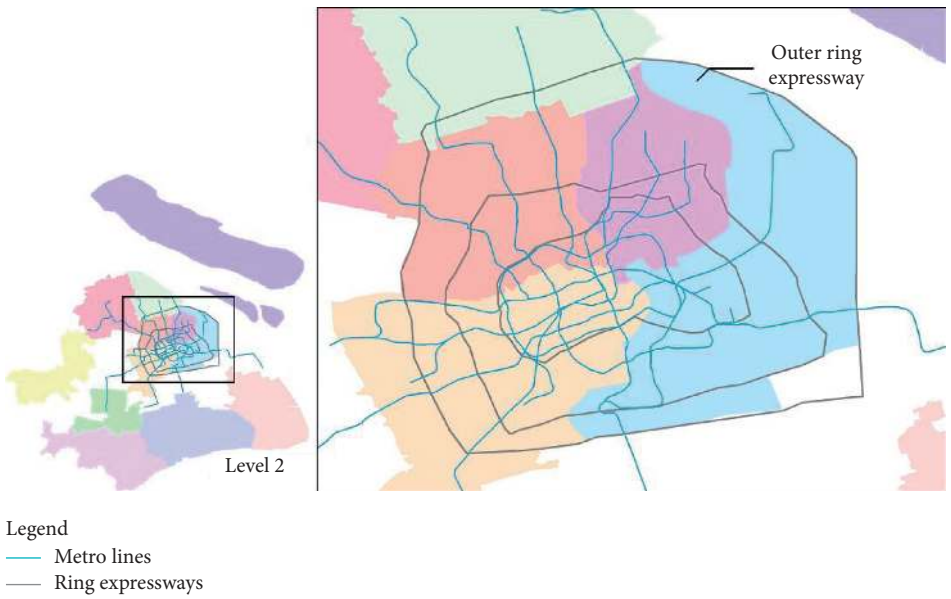
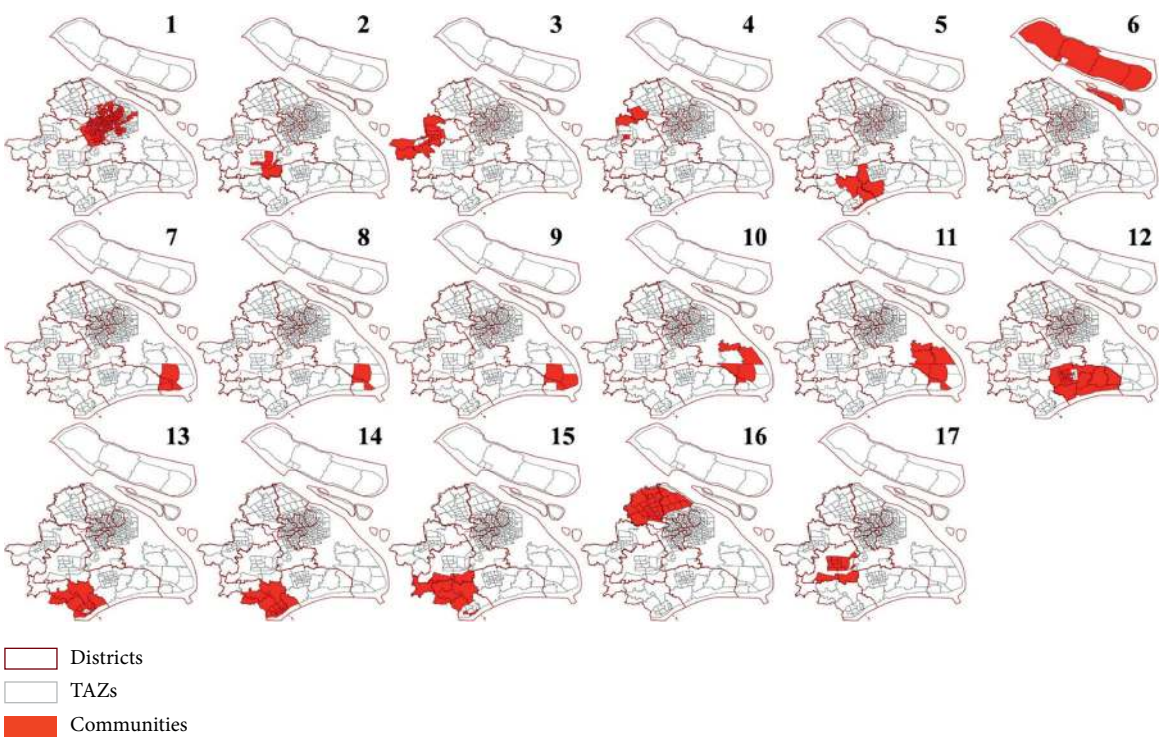
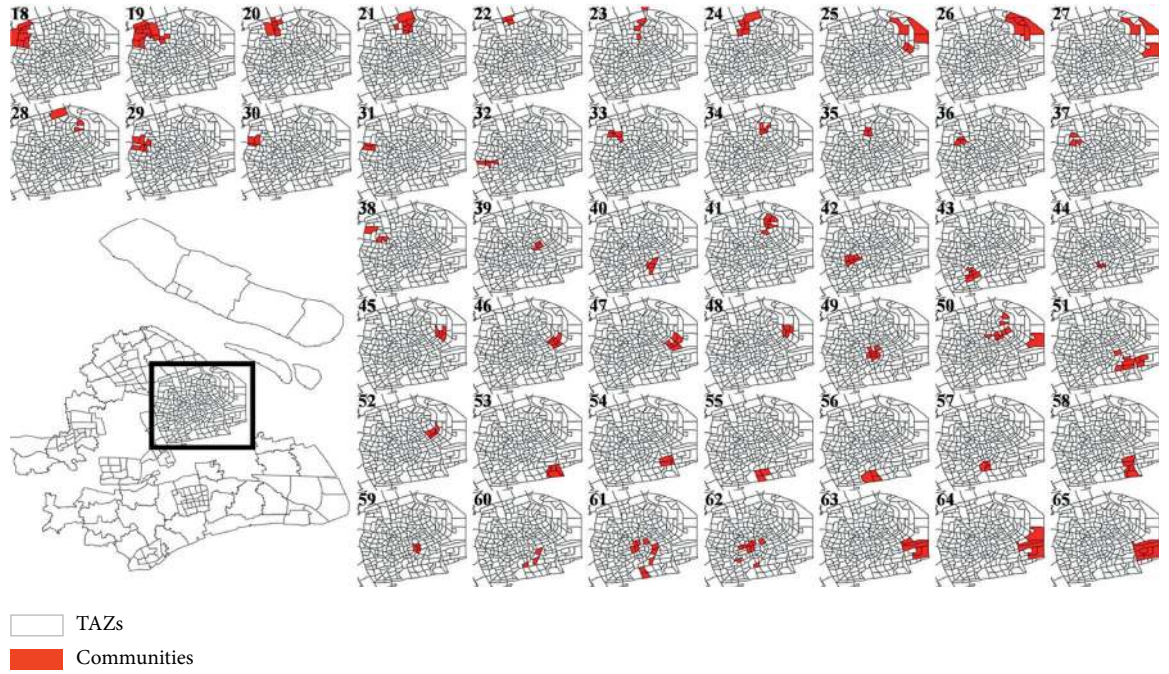


Figure 9: Comparison of level 2 community structure with metro lines and ring expressways.

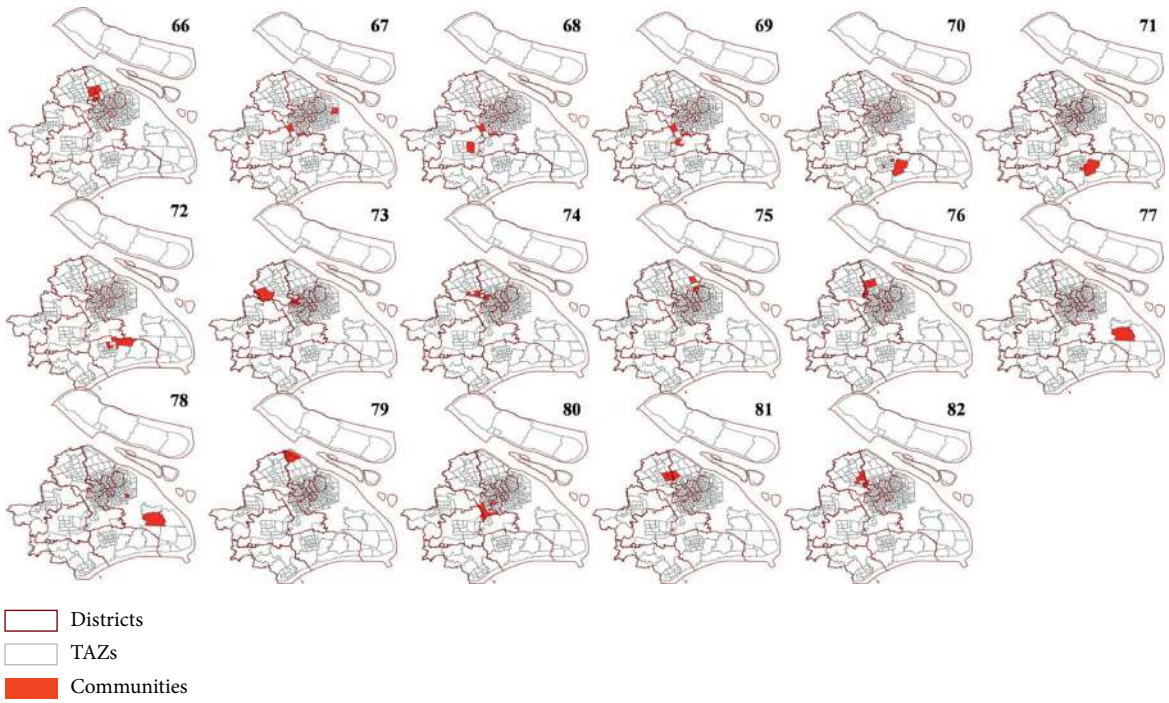


(a)

Figure 10: Continued.

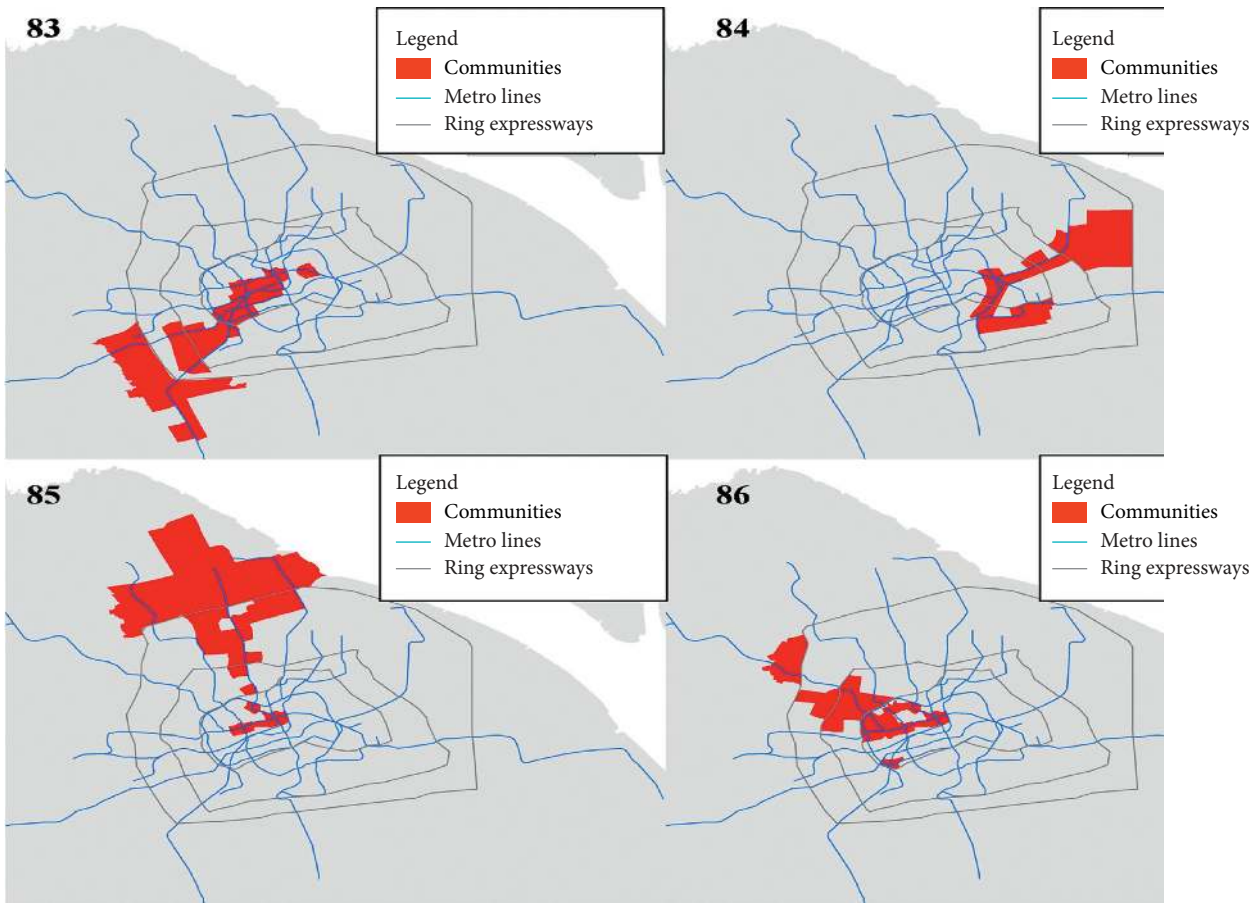


(b)



(c)

Figure 10: Continued.



(d)

Figure 10: Result of overlapping community detection. (a) Large communities, (b) small communities in the central urban area, (c) small communities in the suburban area, and (d) banded communities.

natives, and they have bought their house before the steep rising of the housing prices. Their living places are close to their working places, with small commute distance and forming the small communities.

At the same time, there are also 4 banded communities in the overlapping communities indicating the long-distance commuting demand (Figure 10(d)). At the fringe of the central urban area, the population consists of a high proportion of immigrations and forms into large-scale residential communities. The 4 banded communities are shaped as sector structure radiating out from the city center to the residential communities at the fringe of the central urban area. Comparing the communities with the metro lines in Shanghai, the banded communities are all extending outward along with radiating metro lines, with averagely two metro lines to form a banded community. From this structure, we can infer that the commuting behavior of residents living along the metro lines depends heavily on the metro line, with their workplaces aggregating along the metro line. This result demonstrates that, as the major traffic corridors, metro lines are playing important roles in forming the city commuting demand structure.

In summary, from the result of overlapping communities, we can describe the commuting demand structure of Shanghai as follows:

(i) On the city scale, there is a multiple nuclei structure, with the central urban area as the largest center and several centers of large communities in suburban areas.

(ii) Inside these multiple centers, there are many small communities with intense jobs-housing connections, and most of them are in the central urban area.

(iii) In the central urban area, there are several sector structures radiating out along the metro lines from the city center to the fringe, showing long-distance commuting demand.

5.4. Summary of Results. Comparing nonoverlapping and overlapping communities, we find that they describe the urban commuting demand structure in different aspects:

(i) For nonoverlapping communities, each node only belongs to one community, which forces them to be inside the community with the strongest connection. Thus, nonoverlapping communities are more suitable to describe the whole picture of the spatial structure of urban commuting demand.

(ii) On the other hand, each node in overlapping communities can belong to multiple communities at the same time, which allows it to describe the cross-regional commuting demand.

As for the spatial demand of urban commuting demand, it is found that

(i) By decomposing the jobs-housing network into nonoverlapping communities, according to the number of job activity points, communities can be classified into three types. The commuting demand structure in Shanghai can be simplified into a combination of concentric, sector, and multiple nuclei structure.

(ii) By decomposing the jobs-housing network into overlapping communities, a three-level urban commuting demand structure is discovered in Shanghai, which can be described by three types of communities: large communities indicating the multiple nuclei structure, small communities representing short-distance commuting communities, and banded communities indicating long-distance commuting demand.

The results in both community detection algorithms also have some similarities; it is found that

(i) The boundary of communities in nonoverlapping communities and the large communities in overlapping communities are mostly in correlation with the boundary of administrative districts, indicating that residential commuting behavior is highly restricted by the administrative boundaries, especially in suburban areas.

(ii) Level 2 structure in nonoverlapping communities and the banded communities in overlapping communities all extend along with radiating metro lines, demonstrating that metro lines are playing important roles in leading commuting demand and forming city commuting demand structure.

6. Contribution and Future Directions

As the focus has been shifted to designing demand from serving demand, it becomes increasingly important to depict the jobs-housing relationship and study the commuting patterns in a city. The better understanding of the jobs-housing relationship and commuting patterns enables us to gain an overall knowledge of commuting demand, city commuting demand structure, and even further, to promote urban livability and sustainability.

In this paper, a methodology framework is proposed to describe the spatial structure of commuting demand in a city from mobile phone data. Commuters and their job activity information is extracted to construct the jobs-housing network representing the commuting demand of the city. By using nonoverlapping and overlapping community detection to decompose the structure of the network, the commuting demand structure of the city is unveiled. To demonstrate the practical use of the proposed methodologies, a case study is carried out in Shanghai to explore the commuting patterns of Shanghai residents.

The main contributions of this study are as follows:

(i) The proposed methodology framework enables to decompose the commuting demand and extract the city structure from human mobility data, which has the potential to apply on different flow dataset to reflect urban structure in different aspects. For instance, applying the methodology on taxi trip flow data, cash flow data, and information flow data to reflect different flow connection structure.

(ii) The methodology can generate the result that describes the urban structure in a large city with multiple subcenters. The result generated is based on analyzes of current demand, which can be applied as the basis of the subareas division in practical transportation planning programs.

There are several further directions based on this study. There exists a debate about not only the residential density but also the commuting time that will affect the commuting behavior and urban commuting structure forming. Thus, how community time impacts an urban structure is a potential research topic. Furthermore, the methodology proposed in this paper only considers the spatial aspects of the jobs-housing connection in the city. In future studies, the temporal aspects of demand can also be considered to describe how the city commuting demand structure changes according to the change of time. Proposing a quantitative criterion to classify the communities according to their shape and location is a potential research topic. On the one hand, it requires to describe the spatial shape of communities using several numerical indexes, and on the other hand, it also requires to consider not only the spatial location but also the built environment of the communities. Further practical application based on the result of this study is also a research aspect. This study describes the commuting behavior of groups of people by aggregating the demand. Based on human group behavior, it will be exciting and meaningful to study and describe individual commuting behavior in a city and further our understanding of human traveling behavior [63].

Data Availability

The mobile phone data used to support the findings of this study have not been made available because of the privacy policy.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the National Key R&D Program of China (No. 2018YFB1601100).

References

[1] J. Zhu and Y. Fan, "Commute happiness in Xi'an, China: effects of commute mode, duration, and frequency," *Travel Behaviour and Society*, vol. 11, pp. 43–51, 2018.

[2] C. Fang and D. Yu, "Urban agglomeration: an evolving concept of an emerging phenomenon," *Landscape and Urban Planning*, vol. 162, pp. 126–136, 2017.

[3] G. D. Squires, *Urban Sprawl: Causes, Consequences, & Policy Responses*, The Urban Insitute, Washington, DC, USA, 2002.

[4] M. A. Niedzielski, M. E. O'Kelly, and E. E. Boschmann, "Synthesizing spatial interaction data for social science research: validation and an investigation of spatial mismatch in Wichita, Kansas," *Computers, Environment and Urban Systems*, vol. 54, pp. 204–218, 2015.

[5] J. Choi, J. F. Coughlin, and L. D'Ambrosio, "Travel time and subjective well-being," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2357, no. 1, pp. 100–108, 2013.

[6] D. Wheatley, "Travel-to-work and subjective well-being: a study of UK dual career households," *Journal of Transport Geography*, vol. 39, pp. 187–196, 2014.

[7] J. Sohn, "Are commuting patterns a good indicator of urban spatial structure?" *Journal of Transport Geography*, vol. 13, no. 4, pp. 306–317, 2005.

[8] M. Cai, Y. Liu, M. Luo, L. Xing, and Y. Liu, "Job accessibility from a multiple commuting circles perspective using baidu location data: a case study of Wuhan, China," *Sustainability*, vol. 11, no. 23, p. 6696, 2019.

[9] T. Li, Y. Chen, Z. Wang, Z. Liu, R. Ding, and S. Xue, "Analysis of jobs-housing relationship and commuting characteristics around urban rail transit stations," *IEEE Access*, vol. 7, pp. 175083–175092, 2019.

[10] M. Timberlake, *The Polycentric Metropolis: Learning from Mega-City Regions in Europe*, Earthscan, London, UK, 2009.

[11] M. Castells, *Rise of the Network Society*, Blackwell Publishers, Hoboken, NJ, USA, 2000.

[12] M. C. González, C. A. Hidalgo, and A.-L. Barabási, "Understanding individual human mobility patterns," *Nature*, vol. 453, no. 7196, pp. 779–782, 2008.

[13] X. Song, R. Guo, T. Xia et al., "Mining urban sustainable performance: millions of GPS data reveal high-emission travel attraction in Tokyo," *Journal of Cleaner Production*, vol. 242, Article ID 118396, 2020.

[14] Q. Yu, H. Zhang, W. Li et al., "Mobile phone data in urban bicycle-sharing: market-oriented sub-area division and spatial analysis on emission reduction potentials," *Journal of Cleaner Production*, vol. 254, Article ID 119974, 2020.

[15] H. Zhang, X. Song, T. Xia et al., "Battery electric vehicles in Japan: human mobile behavior based adoption potential analysis and policy target response," *Applied Energy*, vol. 220, pp. 527–535, 2018.

[16] X. Liu, L. Gong, Y. Gong, and Y. Liu, "Revealing travel patterns and city structure with taxi trip data," *Journal of Transport Geography*, vol. 43, pp. 78–90, 2015.

[17] P. Zhang, J. Zhou, and T. Zhang, "Quantifying and visualizing jobs-housing balance with big data: a case study of Shanghai," *Cities*, vol. 66, pp. 10–22, 2017.

[18] N. Ta, Y. Chai, Y. Zhang, and D. Sun, "Understanding job-housing relationship and commuting pattern in Chinese cities: past, present and future," *Transportation Research Part D: Transport and Environment*, vol. 52, pp. 562–573, 2017.

[19] Z. Zhu, Z. Li, H. Chen, Y. Liu, and J. Zeng, "Subjective well-being in China: how much does commuting matter?" *Transportation*, vol. 46, no. 4, pp. 1505–1524, 2019.

[20] N. Green, "Functional polycentricity: a formal definition in terms of social network analysis," *Urban Studies*, vol. 44, no. 11, pp. 2077–2103, 2007.

[21] E. Meijers, "Measuring polycentricity and its promises," *European Planning Studies*, vol. 16, no. 9, pp. 1313–1323, 2008.

[22] C. Roth, S. M. Kang, M. Batty, and M. Barthélemy, "Structure of urban movements: polycentric activity and entangled hierarchical flows," *PLoS One*, vol. 6, no. 1, Article ID e15923, 2011.

[23] M. Hoyler, T. Freytag, and C. Mager, "Connecting rhine-main: the production of multi-scalar polycentricities through knowledge-intensive business services," *Regional Studies*, vol. 42, no. 8, pp. 1095–1111, 2008.

[24] S. McLafferty, "Gender, race, and the determinants of commuting: New York in 1990," *Urban Geography*, vol. 18, no. 3, pp. 192–212, 1997.

[25] X.-M. Lu and X.-T. Gu, "The fifth travel survey of residents in Shanghai and characteristics analysis," *Urban Transport of China*, vol. 9, no. 5, pp. 1–7, 2011.

[26] D. Liang, N. Xinyi, and S. Xiaodong, "Identifying the commuting area of Shanghai central city using mobile phone data," *City Planning Review*, vol. 39, no. 9, pp. 100–106, 2015.

[27] R. Ahas and Ü. Mark, "Location based services-new challenges for planning and public administration?" *Futures*, vol. 37, no. 6, pp. 547–561, 2005.

[28] C. Ratti, D. Frenchman, R. M. Pulselli, and S. Williams, "Mobile landscapes: using location data from cell phones for urban analysis," *Environment and Planning B: Planning and Design*, vol. 33, no. 5, pp. 727–748, 2006.

[29] R. Ahas, S. Silm, O. Järv, E. Saluveer, and M. Tiru, "Using mobile positioning data to model locations meaningful to users of mobile phones," *Journal of Urban Technology*, vol. 17, no. 1, pp. 3–27, 2010.

[30] T. Louail, M. Lenormand, O. R. Cantu et al., "From mobile phone data to the spatial structure of cities," *Scientific Reports*, vol. 4, no. 2973, p. 5276, 2014.

[31] X. Yang, Z. Zhao, and S. Lu, "Exploring spatial-temporal patterns of urban human mobility hotspots," *Sustainability*, vol. 8, no. 7, p. 674, 2016.

[32] X. Zhou, X. Chen, and T. Zhang, "Impact of megacity jobs-housing spatial mismatch on commuting behaviors: a case study on central districts of Shanghai, China," *Sustainability*, vol. 8, no. 2, p. 122, 2016.

[33] X. Yang, Z. Fang, L. Yin, J. Li, Y. Zhou, and S. Lu, "Understanding the spatial structure of urban commuting using mobile phone location data: a case study of Shenzhen, China," *Sustainability*, vol. 10, no. 5, p. 1435, 2018.

[34] J. Cai, B. Huang, and Y. Song, "Using multi-source geospatial big data to identify the structure of polycentric cities," *Remote Sensing of Environment*, vol. 202, pp. 210–221, 2017.

[35] Q. Yu, W. Li, H. Zhang, and D. Yang, "Mobile phone data in urban customized bus: a network-based hierarchical location selection method with an application to system layout design in the urban agglomeration," *Sustainability*, vol. 12, no. 15, p. 6203, 2020.

[36] A. I. Croce, G. Musolino, C. Rindone, and A. Vitetta, "Transport system models and big data: zoning and graph building with traditional surveys, FCD and GIS," *ISPRS International Journal of Geo-Information*, vol. 8, no. 4, p. 187, 2019.

[37] G. Harrison, S. M. Grant-Muller, and F. C. Hodgson, "New and emerging data forms in transportation planning and policy: opportunities and challenges for "track and trace" data," *Transportation Research Part C: Emerging Technologies*, vol. 117, Article ID 102672, 2020.

[38] S. Zhang, X. Liu, J. Tang, S. Cheng, and Y. Wang, "Urban spatial structure and travel patterns: analysis of workday and holiday travel using inhomogeneous poisson point process models," *Computers, Environment and Urban Systems*, vol. 73, pp. 68–84, 2019.

[39] J. Tang, S. Zhang, X. Chen, F. Liu, and Y. Zou, "Taxi trips distribution modeling based on entropy-maximizing theory: a case study in Harbin city-China," *Physica A: Statistical Mechanics and Its Applications*, vol. 493, pp. 430–443, 2018.

[40] M. Ghahramani, M. Zhou, and C. T. Hon, "Extracting significant mobile phone interaction patterns based on community structures," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 3, pp. 1031–1041, 2018.

[41] M. Ghahramani, M. Zhou, and C. T. Hon, "Mobile phone data analysis: a spatial exploration toward hotspot detection," *IEEE Transactions on Automation Science and Engineering*, vol. 16, no. 1, pp. 351–362, 2018.

[42] M. Ghahramani, M. Zhou, and G. Wang, "Urban sensing based on mobile phone data: approaches, applications, and challenges," *IEEE/CAA Journal of Automatica Sinica*, vol. 7, no. 3, pp. 627–637, 2020.

[43] W. Li, X. Cheng, Z. Duan, D. Yang, and G. Guo, "A framework for spatial interaction analysis based on large-scale mobile phone data," *Computational Intelligence & Neuroscience*, vol. 2014, no. 9, p. 21, 2014.

[44] W. Li, *Analysis on Individuals' Activity Space Based on Mobile Phone Data*, Tongji University, Shanghai, China, 2018, in Chinese.

[45] C. Chen, L. Bian, and J. Ma, "From traces to trajectories: how well can we guess activity locations from mobile phone traces?" *Transportation Research Part C: Emerging Technologies*, vol. 46, pp. 326–337, 2014.

[46] X. Cheng, *Research on Travel Characteristics and Classification of Urban Residents Based on Mobile Phone Data*, Tongji University, Shanghai, China, 2015.

[47] S. Fortunato and C. Castellano, *Community Structure in Graphs*, Springer, New York, NY, USA, 2012.

[48] J. Chen, L. Chen, Y. Chen et al., "GA-based q-attack on community detection," *IEEE Transactions on Computational Social Systems*, vol. 6, no. 3, pp. 491–503, 2019.

[49] P. Schuetz and A. Caflisch, "Multistep greedy algorithm identifies community structure in real-world and computer-generated networks," *Physical Review E Statistical Nonlinear & Soft Matter Physics*, vol. 78, no. 2, Article ID 26112, 2008.

[50] Y.-Y. Ahn, J. P. Bagrow, and S. Lehmann, "Link communities reveal multiscale complexity in networks," *Nature*, vol. 466, no. 7307, p. 761, 2010.

[51] G. Palla, I. Derényi, I. Farkas, and T. Vicsek, "Uncovering the overlapping community structure of complex networks in nature and society," *Nature*, vol. 435, no. 7043, p. 814, 2005.

[52] A. Lancichinetti, S. Fortunato, and J. Kertész, "Detecting the overlapping and hierarchical community structure of complex networks," *New Journal of Physics*, vol. 11, no. 3, pp. 19–44, 2008.

[53] J. Cheng, M. Chen, M. Zhou, S. Gao, C. Liu, and C. Liu, "Overlapping community change point detection in an evolving network," *IEEE Transactions on Big Data*, vol. 6, 2018.

[54] J. Cheng, X. Wu, M. Zhou, S. Gao, Z. Huang, and C. Liu, "A novel method for detecting new overlapping community in complex evolving networks," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 49, no. 9, pp. 1832–1844, 2018.

[55] S. Fortunato, "Community detection in graphs," *Physics Reports*, vol. 486, no. 3–5, pp. 75–174, 2010.

[56] V. D. Blondel, J. L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of community hierarchies in large networks," *Journal of Statistical Mechanics Theory and Experiment*, vol. 10, 2008.

[57] M. Girvan and M. E. J. Newman, "Community structure in social and biological networks," *Proceedings of the National Academy of Sciences*, vol. 99, no. 12, pp. 7821–7826, 2002.

[58] A. T. Kalinka and P. Tomancak, "Linkcomm: an r package for the generation, visualization, and analysis of link communities in networks of arbitrary size and type," *Bioinformatics*, vol. 27, no. 14, pp. 2011-2012, 2011.

[59] Shanghai Statistical Publishing House, *Shanghai Statistical Yearbook*, Shanghai Statistical Publishing House, Shanghai, China, 2012.

[60] A. K. Jain, "Data clustering: 50 years beyond *K*-means," *Pattern Recognition Letters*, vol. 31, no. 8, pp. 651–666, 2010.

[61] C. D. Harris and E. L. Ullman, "The nature of cities," *The Annals of the American Academy of Political and Social Science*, vol. 242, no. 1, pp. 7–17, 1945.

[62] H. Hoyt, *The Structure and Growth of Residential Neighborhoods in American Cities*, US Government Printing Office, Washington, DC, USA, 1939.

[63] Beijing Transport Institute, *Commuter Travel Characteristics and Typical Areas in Beijing*, Beijing Transport Institute, Beijing, China, 2019, in Chinese.