MYU Tokyo

S & M 3811

Analysis of the Spatial Heterogeneity of Commuting Flows in Beijing: Perspectives from Mobile Phone Data

Sihui Guo,[1,2] Qiang Huang,[1,2] and Congcong Wen[3,4*]

[1]Beijing Institute of Surveying and Mapping,
No. 15 Yangfangdian Road, Haidian District, Beijing 100045, China
[2]Beijing Key Laboratory of Urban Spatial Information Engineering,
No. 15 Yangfangdian Road, Haidian District, Beijing 100045, China
[3]China Academy of Electronics and Information Technology,
No. 11 Shuangyuan Road, Shijingshan District, Beijing 100041, China
[4]Department of Electrical and Computer Engineering, New York University Abu Dhabi,
Saadiyat Island, Abu Dhabi, United Arab Emirates

(Received July 26, 2024; accepted October 11, 2024)

Keywords: commuting flows, spatial heterogeneity, mobile phone data, L function

Commuting flows refer to the regular movement of people from their homes to workplaces. The spatial heterogeneity of commuting flows indicates the uneven spatial distribution of places of residence, employment, or job–residence connections. Understanding the characteristics of spatial heterogeneity in commuting flows is vital for effective transportation planning. However, limited by the scarcity of flow data, traditional research on commuting flows predominantly focuses on the spatial distribution of employment, residence, and job–residence connections individually, which fails to unveil the spatial heterogeneity of commuting flows. In this study, we examined the spatial heterogeneity of commuting flows in Beijing using mobile phone data. We analyzed the degree of heterogeneity and the aggregation scale of commuting flows. The results showed that (1) the degree of spatial heterogeneity varies between different regional pairs, and (2) the aggregation scale of commuting flows varies with distance; moreover, the longer the distance, the larger the aggregation scale of commuting flow distribution, and dominant heterogeneous clusters expand from the center to the periphery. These findings enhance the strategic planning of public transit. In less densely populated areas such as those outside the Fifth Ring Road, efficiency can be boosted by new bus routes. In contrast, denser areas such as those within the Third Ring Road, may benefit from an integrated approach: establishing new rail systems alongside expanded bus services to optimize commuting.

1. Introduction

Commuting flows refer to the movement of individuals from their residences to their places of employment, constituting one of the most significant flows within urban areas.(1) Commuting accounts for the largest proportion of transportation demand within cities. Comprehensive transportation surveys in major Chinese cities indicate that commuting trips on average

*Corresponding author: e-mail: wencc1208@gmail.com
https://doi.org/10.18494/SAM5253

ISSN 0914-4935 © MYU K.K.
https://myukk.org/

constitute 40–50% of daily travel during the workweek; a substantial volume of these trips is concentrated during the morning and evening peak hours, leading to congestion that is a primary cause of traffic jams in urban peak periods.[2] Furthermore, as one of the most common types of daily travel, commuting flows are also significant pathways for transmitting infectious diseases.[3–5] Moreover, excessive commuting distances and increased travel times due to congestion significantly diminish the well-being of individuals. Therefore, research on commuting flows is vital for ensuring the efficient operation of urban transportation, safeguarding the health and safety of urban residents, and enhancing the overall sense of well-being among the populace.

The emergence of commuting flows is attributed to the phenomenon of job–residence separation among urban residents, where workplaces are located beyond the boundaries of residential communities.[6,7] On one hand, with the rapid pace of urbanization, urban spaces have been continuously expanding, leading to an imbalanced development of urban functions, particularly in the allocation of employment and residential spaces, resulting in a spatial mismatch between living and working areas. On the other hand, the "suction effect" of central business districts and surrounding large cities on employment functions and core infrastructures has diminished the significance of distance as the sole consideration for residents when choosing employment.[8] Consequently, the phenomenon of job–residence separation in cities has become increasingly pronounced, with residents' commuting distances and times gradually increasing, leading to a range of urban issues such as traffic congestion, environmental pollution, and a decline in the quality of life.

Traditional research on commuting flows has been constrained by the scarcity of flow data. For instance, studies on commuting flows based on census data are typically limited by the low spatial resolution of the data, allowing for analysis only from the macroperspective of cities or the mesoperspective of districts and counties, without capturing the microlevel individual variations in commuting behavior and their underlying mechanisms.[9,10] Research based on transportation survey data, while capable of obtaining detailed individual attribute information tailored to research needs, is fraught with issues such as high costs and limited representatives of the sample.[11]

The advent of the big data era has presented opportunities for the study of commuting flows. With the advancement of communication, networking, and location-aware technologies, along with the widespread application of various sensors and intelligent terminals in urban areas, large-scale human activity trajectory data are being recorded in different forms at the individual level. Examples are the Global Positioning System (GPS) trajectories of residents recorded by taxis or ride-hailing services,[12] the station information of people boarding and alighting from buses or entering and exiting subways recorded by smart transportation cards,[13–15] and the real-time or call locations of users recorded by mobile phones, social media check-in locations,[16,17] as well as the locations of internet and app requests.[18] These multisource, fine-grained, and large-scale trajectory data are experiencing explosive growth in cities, providing a rich source of flow data for the study of commuting flows and offering opportunities for in-depth research on the spatial distribution patterns and dynamic mechanisms of commuting from the perspective of commuting flows. Therefore, in this study, we aim to extract residential and employment locations from the trajectory information of residents in mobile phone signal data, express commuting flows in the form of origin–destination (OD) flows, and thereby study the spatial distribution characteristics of commuting flows.

The spatial heterogeneity of commuting flows refers to the uneven distribution of places of residence, employment, or job–residence connections in space. Investigating the characteristics of spatial heterogeneity in commuting flows can help identify areas within cities where the distribution of commuting flows is more concentrated, thereby addressing issues of localized traffic congestion.(19) Current research on the heterogeneity of commuting flow distribution primarily focuses on two aspects. On one hand, it examines the distribution characteristics of employment or residence individually and measures the spatial matching relationship between places of residence and employment through indicators such as the job-to-residence ratio and self-sufficiency within certain geographical units.(20–22) On the other hand, it pays attention to the distribution characteristics of commuting volume, distance, or commuting time formed by job–residence connections, characterizing the spatial differentiation of job–residence connections through visualization or network analysis methods. For instance, on the basis of points of interest (POI) and housing price data, Zhang *et al.* analyzed the commuting distance distribution of ten megacities in China with a resident population exceeding ten million using kernel density analysis and found that there is a clear spatial agglomeration of commuting distances in all cities, with residents in central areas and their vicinity typically having shorter commuting distances.(23) Yang *et al.* identified residents' places of residence and employment in Shenzhen using mobile phone data and constructed a commuting demand network, applying community detection algorithms to analyze the community structure of the commuting demand network, revealing that Shenzhen not only exhibits a polycentric structure overall but also contains polycentric structures within each community.(24)

However, current research on the heterogeneity of commuting flows predominantly focuses on the spatial distribution heterogeneity of employment, residence, and job–residence connections individually. There is a lack of studies that consider the spatial distribution heterogeneity of commuting flows holistically, such as the degree of spatial heterogeneity and the scale of aggregation. Understanding this heterogeneity is crucial for several reasons. First, it allows planners to identify areas with high commuting intensity, which are likely to experience traffic congestion. By pinpointing these areas, planners can develop targeted solutions, such as improving public transportation networks or creating incentives for alternative commuting methods. Second, the spatial distribution of commuting flows can guide the strategic placement of new transportation infrastructure. For instance, areas with large-scale aggregation of commuting flows may benefit from new rail systems, while those with smaller aggregation scales might require enhanced bus services. Third, recognizing the heterogeneity of commuting patterns can help in the planning of mixed-use developments that bring together residential, commercial, and recreational spaces. This can reduce the need for long-distance commuting and associated traffic congestion. Lastly, our research provides a granular analysis that moves beyond macrolevel assessments. By examining the spatial heterogeneity at a microlevel, we can uncover hidden patterns and nuances that traditional, aggregate studies might overlook. This detailed understanding can lead to more nuanced and effective urban planning strategies.

Therefore, in this study, we intend to investigate the geometric heterogeneity (overall distribution) and attribute heterogeneity (flow volume, flow length) of commuting flows by examining the scale of aggregation and spatial differences in flow volume and flow length, thereby providing references for urban and transportation planning.

2. Methods

The spatial heterogeneity of commuting flows refers to the degree to which the spatial distribution of these flows deviates from complete spatial randomness in a clustered form. We characterized the heterogeneity of commuting flows in two aspects. First, by using the heterogeneity metric *FA-w*, we delineated the degree of heterogeneity in the distribution of commuting flows, thereby understanding the imbalance in the distribution of commuting flows among different urban areas. Second, we characterized the scale of heterogeneity in the distribution of commuting flows (the maximum scale of spatial aggregation of commuting flows), thereby understanding the scope of concentrated commuting within the city and providing a reference for corresponding transportation planning.

2.1 Definition and expression of commuting flows

We define commuting flows as the movement between the home and the workplace of an individual. Commuting flows can be mathematically expressed as $f_i = \left(\left(x_i^O, y_i^O\right), \left(x_i^D, y_i^D\right)\right)$, in which $\left(x_i^O, y_i^O\right)$ is the coordinate of the location of residence and $\left(x_i^D, y_i^D\right)$ is the location of the workplace.

2.2 Definition of distance between flows

Since both the degree of heterogeneity and the scale of heterogeneity of flows involve the measurement of flow distance, We define flow distance in this section. According to Shu *et al.*, there are two definitions of flow distance, namely, maximum distance and additive distance.(25) The maximum distance refers to the greater value of the Euclidean distance between the origins and the destinations, and the additive distance refers to the sum of the Euclidean distances between the origins and the destinations. In this study, we used maximum distance to measure the degree and aggregation scale of the spatial heterogeneity of flows. The maximum distance between flows is calculated as

$$d_{ij}^f = \max\left(d_{ij}^O, d_{ij}^D\right), \tag{1}$$

where d_{ij}^f is the maximum distance between f_i and f_j, d_{ij}^O is the Euclidean distance between the origins, and d_{ij}^D is the Euclidean distance between the destinations. We chose the maximum distance because it is suitable for measuring the spatial extents of a flow cluster.(25) In this paper, the maximum distance is used to measure the aggregation scale of commuting flows, which is helpful for understanding the extent of residence or work place cluster, so as to put forward the corresponding traffic planning.

2.3 Measurement of the degree of spatial heterogeneity of flows

The spatial heterogeneity of commuting flows is defined as the degree of deviation from a random or uniform distribution. According to Shu *et al.*, *FA-w* has been demonstrated to be a robust statistic in quantifying the heterogeneity of flows.[26] Therefore, we used the *FA-w* statistic to quantify the heterogeneity in the spatial distribution of commuting flows. *FA-w* is defined as the ratio of the mean first-order nearest neighbor distance of the observed flow dataset to the expected first-order nearest neighbor distance of a completely spatially random flow dataset of the same intensity. The calculation formula is[26]

$$FA\text{-}w = \frac{\bar{W}}{E(W)} = \frac{\pi^{1/2}\lambda_F^{1/4}\sum_{i=1}^{n} W_i}{6^{1/4}\Gamma(5/4)n}, \tag{2}$$

where $\bar{W}$ is the mean first-order nearest neighbor distance of observed flows, and $E(W)$ is the expected first-order nearest neighbor distance of a completely spatially random flow dataset of the same intensity. For completely spatially random flows, *FA-w* follows a normal distribution, $N\left(1, \frac{\Gamma(3/2)-\Gamma^2(5/4)}{\Gamma^2(5/4)n}\right)$. Consequently, the *FA-w* of the observed flow dataset can be calculated using Eq. (1) to assess the degree of spatial heterogeneity: a higher *FA-w* indicates a lower degree of spatial heterogeneity in the flow distribution.

2.4 Measurement of the aggregation scale of flows

Aggregation is one of the most common heterogeneous patterns of geographical flows. The aggregation scale characterizes the radius of maximal aggregation.[25] Quantifying the scale of aggregation of commuting flows is important for transportation planning. For example, in areas with a small scale of aggregation, such as those with scales smaller than the distance that a bus can travel in 15 min during off-peak hours, commuting efficiency can be enhanced by adding bus lines; in areas with a large scale of aggregation, such as those with scales larger than the distance that a bus can travel in 15 min during off-peak hours, the commuting demands over a wide range can be met by a composite approach involving the construction of new rail transit systems and the addition of bus lines. Therefore, we aim to measure the aggregation scale of commuting flows on the basis of the L function of flows proposed by Shu *et al.* The L function is the derivative of the K function, which is calculated using[25]

$$K(r) = \sum_i \sum_j \sigma_{F,ij}(r) / n\hat{\lambda}_F, (i, j = 1,2,\ldots,n; i \neq j), \tag{3}$$

where r is the distance (i.e., the aggregation scale), n is the number of flows in the study area A, $\hat{\lambda}_f = n / V_A$, and V_A is the volume of the research flows being studied. When the distance between the flows f_i and f_j is not larger than r, then $\sigma_{f,ij}(r) = 1$; otherwise, on the basis of the K function of flows, the L function of flows can be calculated as

$$L(r) = \sqrt[4]{\frac{K(r)}{\pi^2}} - r = \sqrt[4]{\frac{\sum_i \sum_j \sigma_{F,ij}(r)}{n\hat{\lambda}_F \pi^2}} - r, (i, j = 1,2,...,n; i \neq j) \tag{4}$$

$$L_i(r) = \sqrt[4]{\frac{\sum_j \sigma_{F,ij}(r)}{n\hat{\lambda}_F \pi^2}} - r, (i, j = 1,2,...,n; i \neq j), \tag{5}$$

where $L(r)$ is the L function of flows and $L_i(r)$ is the local L function of each flow. The relevant variables in Eqs. (4) and (5) have the same meanings as those in Eq. (3).

On the basis of the above formula, we extract the dominant commuting flows of different levels through the L function and the derivative of the L function. These include three main steps. First, determine the dominant cluster's aggregation scale. Calculate the L function of the commuting flows and record the maximum of the L function as $[L_{max}]$; then, the first local minimum on the right of $L'([L_{max}])$ of the derivative of the L function is the aggregation scale r_m of the dominant cluster. Second, extract the dominant cluster. Calculate the local L function values of all flows under r_m and find the top 20 flows with local L function values and the flows within the r_m range of these flows, that is, the dominant cluster. Third, sequentially extract dominant clusters. Exclude the dominant clusters identified in step (2) and continue to repeat steps (1) and (2). Using these three steps, we can obtain different levels of aggregated commuting flows. Figure 1 shows the flow chart of the process of calculating aggregation scales and dominant clusters.

To illustrate the process of calculating aggregation scales and dominant clusters, we simulated a flow distribution and calculated its aggregation scales and extracted its dominant clusters. Figure 2 shows the simulated flow distribution [Fig. 2(a)] and the results of dominant clusters [Figs. 2(d) and 2(h)]. The simulated flows include 1000 flows, which are composed of 400 randomly produced flows, 400 concentrated flows with an aggregation scale of 0.2, and 200 concentrated flows with an aggregation scale of 0.1. The method of generating 400 aggregated flows with an aggregation scale of 0.2 is as follows: continuously generate a random flow, if the distance between the randomly generated flow and the predefined central flow of $f_1 = ((0.1,0.8),(0.8,0.1))$ is less than 0.1, it will be retained; otherwise, it will continue to generate random flows until the number of flows reaches 400. By the same method, the 200 concentrated flows with an aggregation scale of 0.1 are also produced, whose predefined central flow is $f_2 = ((0.3,0.2),(0.6,0.6))$. For the original flow distribution, we first calculated its $L(r)$, $L_i(r)$, and $L'(r)$. The results of $L(r)$ and $L'(r)$ are shown in Figs. 2(b) and 2(c), respectively. According to the curve of $L(r)$, we found that the local maximum of $L(r)$ is 1.5; therefore, $[L_{\max}]$ is 1.5. Further consider the curve of $L'(r)$, and find the first local minimum of $L'(r)$ to the right of $[L_{max}]$, which is 0.2; therefore, R_m is 0.2. On the basis of R_m, we then found the top 20 flows with

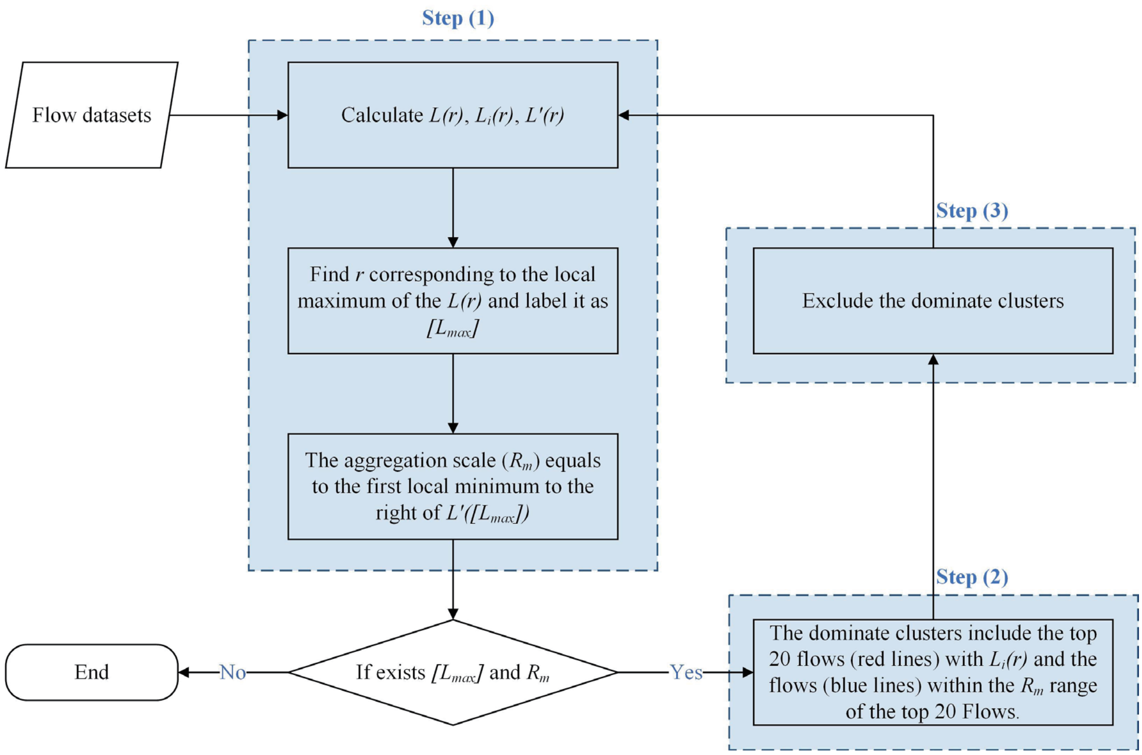


Fig. 1. (Color online) Flow chart of the process of calculating aggregation scales and dominant clusters.

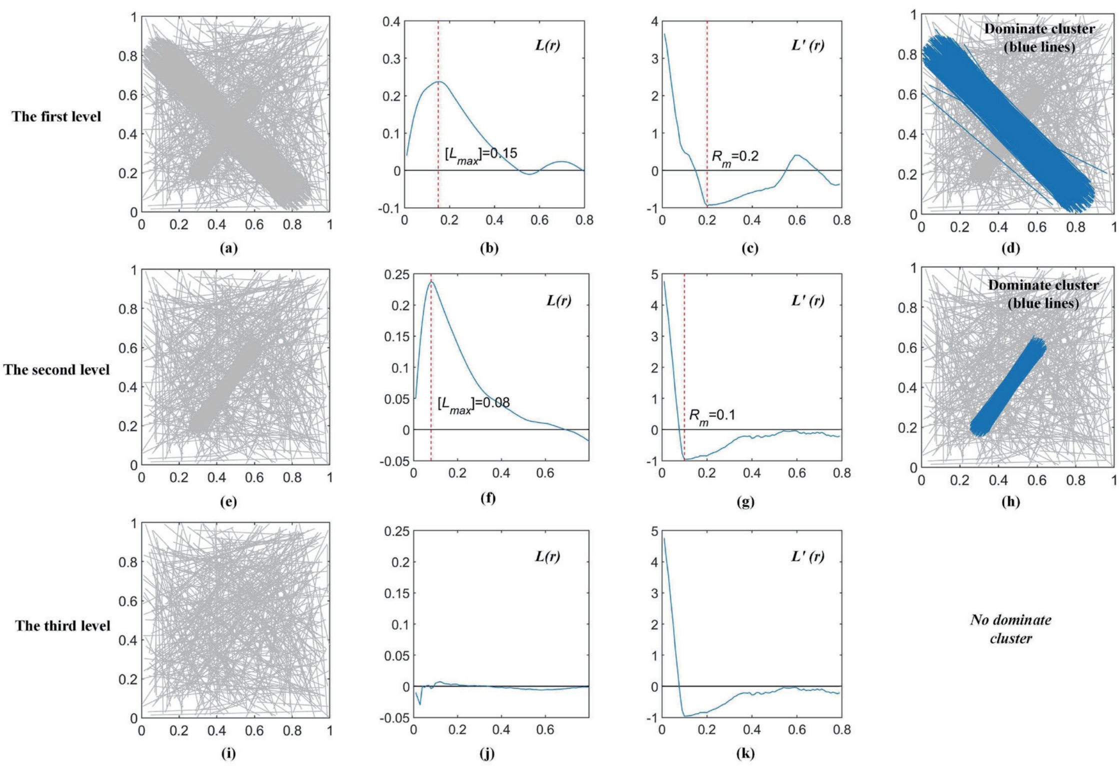


Fig. 2. (Color online) Illustration of calculating aggregation scales and dominant clusters.

$L_i(0.2)$ and the flows with a distance of less than 0.2 from the top 20 flows, and these flows composed the dominant cluster, which represents the statistically significant clustering flows. After excluding the first level of a dominant cluster, we repeated the above steps for the remaining flows [Fig. 2(e)] and obtained the second level of the dominant cluster [Fig. 2(h)]. After excluding the second dominant cluster, there is no local minimum of $L(r)$; therefore, there is no other dominant cluster.

3. Study Area and Datasets

We took Beijing as the study area. On the basis of the mobile phone data collected in Beijing, we identified the commuting flows of mobile phone users and quantified the heterogeneous characteristics of commuting flows.

3.1 Study area

Beijing is one of the megacities in China, with a permanent urban population exceeding 10 million. With the rapid urbanization and the continuously growing demand for commuting traffic, the traffic congestion problem in Beijing has become increasingly prominent, considerably reducing the efficiency of the city's normal functioning. Therefore, we took Beijing as an example to analyze the heterogeneous distribution characteristics of commuting flows.

When analyzing the spatial heterogeneity of commuting flows, we divided the area within the Sixth Ring Road of Beijing into different subzones and characterized the spatial heterogeneity of commuting flows between different zones. Specifically, on the basis of the boundary of the administrative district of Beijing, the adjacent small administrative districts were merged, and finally, 10 subdistricts were obtained. Figure 3 shows the study area.

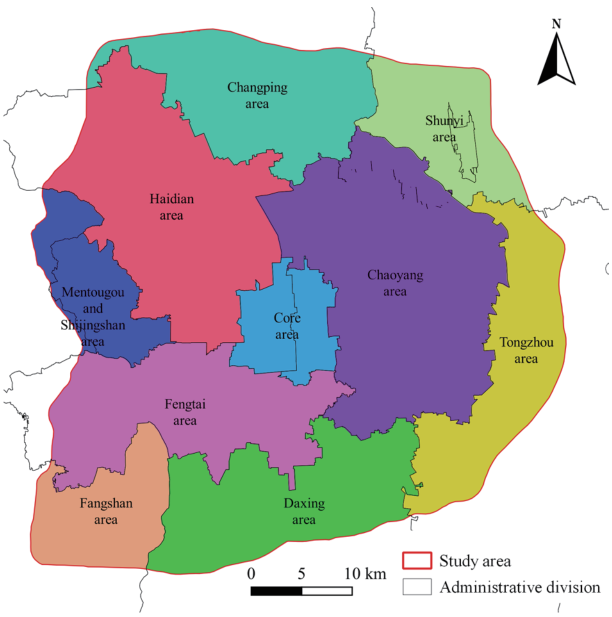


Fig. 3. (Color online) Study area.

3.2 Datasets

In this paper, we mainly used mobile phone signaling data from Beijing, which were collected from September 1 to September 30, 2019. Table 1 shows the location information recorded in the mobile phone signaling data. The data mainly include the encrypted unique identifier of the user, the longitude and latitude of the base station, and the timestamp when the base station was connected.

On the basis of location trajectory records in the mobile phone signaling data, we used the visit frequency method to identify the residence and workplace of mobile phone users, thereby obtaining the commuting flows of individual users. The identification principle of the visit frequency method is to regard the base station where the user appears most frequently between 22:00 and 06:00 as the location of home, and the base station where the user appears most frequently between 09:00 and 11:00 as well as between 14:00 and 17:00 on weekdays (Monday to Friday) as the workplace.

Since the peak congestion in Beijing is mainly concentrated within the Sixth Ring Road, in this study, we selected commuting flows within the Sixth Ring Road of Beijing for statistical analysis. Besides, the commuting that causes traffic congestion is mainly nonwalking distance commuting; therefore, we were mainly concerned with commuting flows with a distance (i.e., the length of the commuting flow) exceeding 800 meters.

4. Results and Discussion

We first analyzed the degree of the spatial heterogeneity of commuting flows among different subregions. Then, we compared the aggregation scales of commuting flows of different lengths and discuss the implications of the results.

4.1 Degree of spatial heterogeneity of commuting flows

We aggregated the commuting flows with each region shown in Fig. 1 and obtained the aggregated commuting flows between $10 \times 10 = 100$ regional pairs. The results of 1-*FA*-*w* statistics are shown in Table 2, in which the higher the value of 1-*FA*-*w*, the more heterogeneous the commuting flows. From Table 2, the heterogeneity of regional pairs shows significant differences, and the distribution of commuting flows shows asymmetric spatial heterogeneity. Among the regional pairs, the pairs with high heterogeneity (higher than 0.7) of commuting flow

Table 1
Sample of mobile phone signaling data.

User Id	Longitude of base station	Latitude of base station	Timestamp
6baf34***24069d	116.5***4	39.7***2	2019-09-16 00:00:09
6baf34***24069d	116.5***7	39.7***7	2019-09-16 00:00:19
…	…	…	…
6baf34***24069d	116.3***9	39.8***1	2019-09-16 23:59:17
6baf34***24069d	116.3***8	39.8***4	2019-09-16 23:59:31

Table 2
Heterogeneity of commuting flows among different regions in Beijing.

1-FA-w	Core area	Chaoyang	Haidian	Shijingshan	Fengtai	Changping	Shunyi	Tongzhou	Daxing	Fangshan
Core area	0.58	0.60	0.56	0.52	0.50	0.54	0.64	0.61	0.54	0.54
Chaoyang	0.58	0.61	0.56	0.53	0.42	0.58	0.62	0.54	0.51	0.66
Haidian	0.57	0.57	0.58	0.61	0.75	0.58	0.68	0.60	0.49	0.66
Shijingshan	0.55	0.60	0.56	0.59	0.49	0.57	0.64	0.61	0.52	0.74
Fengtai	0.49	0.59	0.62	0.56	0.61	0.64	0.71	0.64	0.62	0.61
Changping	0.55	0.62	0.56	0.57	0.58	0.66	0.62	0.59	0.51	0.69
Shunyi	0.52	0.58	0.54	0.56	0.59	0.64	0.62	0.51	0.50	0.72
Tongzhou	0.52	0.66	0.60	0.58	0.62	0.70	0.64	0.59	0.71	0.70
Daxing	0.52	0.61	0.59	0.51	0.59	0.65	0.67	0.65	0.64	0.70
Fangshan	0.65	0.60	0.58	0.56	0.60	0.67	0.63	0.67	0.58	0.72

Note: the columns of the table are the origins and the rows of the table are the destinations.

distribution are Fengtai–Haidian, Fangshan–Shijingshan, and Fangshan–Shunyi; the pairs with low heterogeneity (lower than 0.5) of commuting flow distribution are Dongcheng–Fengtai, Fengtai–Chaoyang, and Daxing–Haidian.

To understand the difference in the spatial heterogeneity of commuting flows between regions, we visualized the commuting flows between the above six regions with high and low heterogeneities. Figure 4 shows the result of visualization. From Fig. 4, the commuting flows with high heterogeneity are mainly distributed among regional pairs on the west side of the Sixth Ring Road. Among them, the high heterogeneity of commuting flow distribution in Fangshan–Shunyi may be related to the airport in Shunyi District, and the flows from other districts to the airport are relatively concentrated, resulting in high heterogeneity. However, the commuting flows with low heterogeneity are mainly distributed in the middle region of the Sixth Ring Road, and the distributions of their homes and workplaces are large and dispersed, so the distribution of flows is relatively random, and the heterogeneity is low.

4.2 Aggregation scales of commuting flows

To compare the aggregation scales of commuting flows of different lengths, we divided the commuting flows in the Sixth Ring Road into four categories according to different distances: short distance (≤1.5 km), median-short distance (1.5–5 km), median–long distance (5–10 km), and long distance (10–20 km). The aggregation scales of these four types of flow were obtained by using the L function defined in Eq. (4) and the derivative of the L function defined in Eq. (5). The results are shown in Fig. 5. In Fig. 5, grey and other colored lines represent commuting flows, the other colored lines are clusters identified by the local L function, and the thick red lines are the top 10 commuting flows ranked by the value of the local L function.

From Fig. 5, the aggregation scale of commuting flows varies with distance, and the longer the distance, the larger the aggregation scale of commuting flow distribution, and the dominant heterogeneous cluster expands from the center to the periphery. Among the dominant clusters, the aggregation scale of commuting flows in short distances (≤1.5 km) is 10.5 km, and the dominant heterogeneous cluster is within the Third Ring Road, especially in the northwest corner of the Third Ring Road. This result may be caused by the universities clustered in the

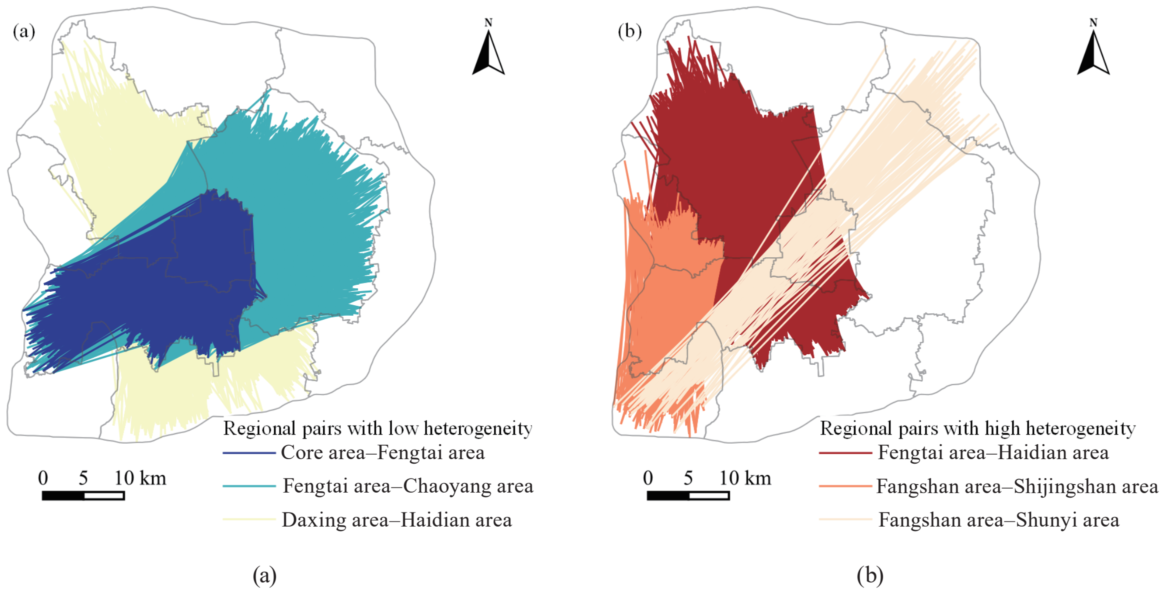


Fig. 4. (Color online) Distribution of commuting flows with (a) low and (b) high heterogeneities.

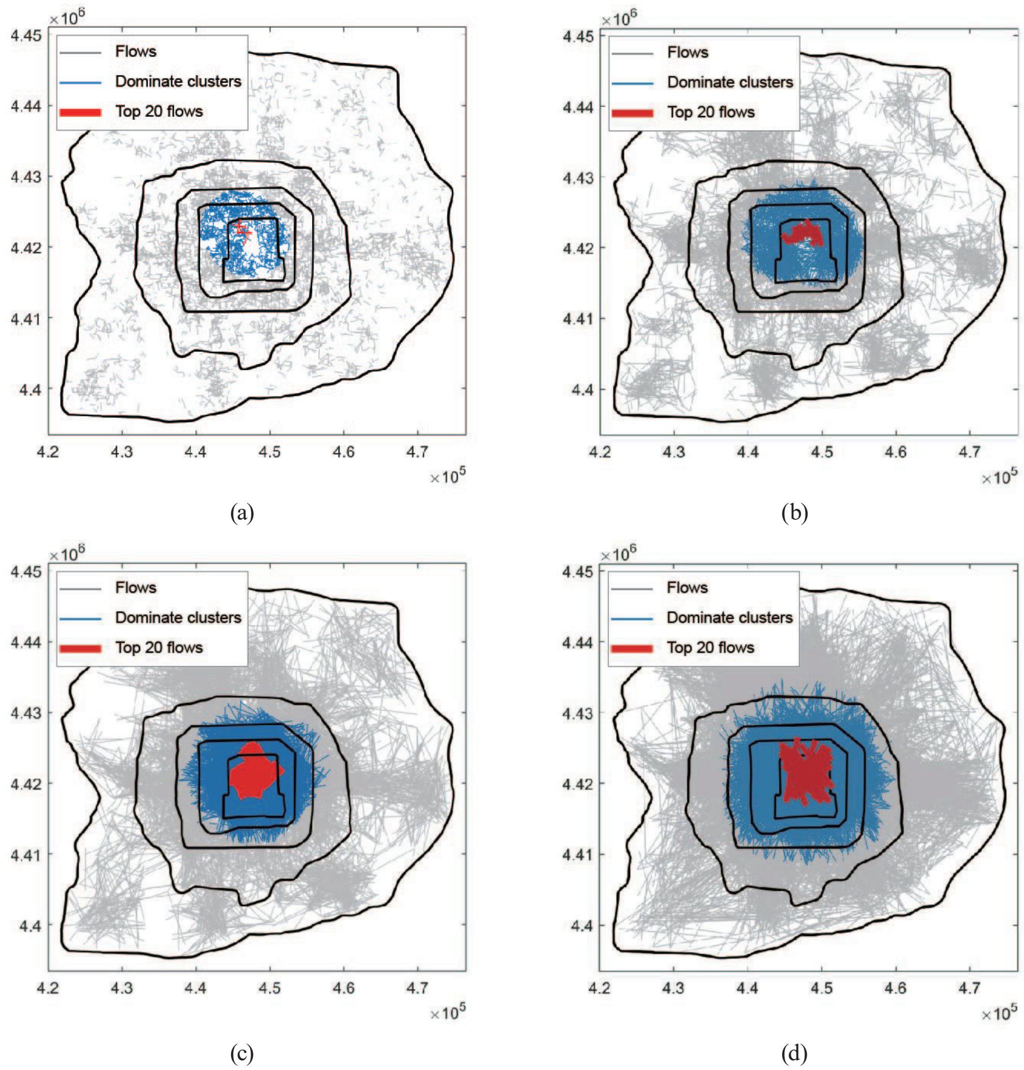


Fig. 5. (Color online) Aggregation scales of different distances of commuting flows: (a) short distance (10.5 km), (b) median-short distance (12.2 km), (c) median-long distance (14 km), and (d) long distance (17.2 km).

corner, such as Beijing Normal University, Beijing University of Aeronautics and Astronautics, Renmin University of China, Beijing Jiaotong University, Zhongguancun Campus of Beijing Institute of Technology, and University of Science and Technology Beijing. It shows that the clusters of short-distance commuting flows are mainly caused by the commuting of college students between the dormitory and the classrooms. The aggregation scales of commuting flows at median-short distances (1.5–5 km) and median-long distances (5–10 km) and long distances are 12.3 and 14 km, respectively, and their respective dominant cluster distributes at a similar range of space, both within the Third Ring Road or at the edge of the Third Ring Road. These areas are mainly the "unit courtyard" in Beijing. Before the 1880s, these regions were characterized by large compounds where living and working spaces were combined. In the post-1880s, the traditional structure of these institutional compounds began to disintegrate, yet many organizations retained the condition of proximity between the residential buildings and the workplaces for their employees. As a result, the commuting distances within these areas remained relatively short. For the long-distance commuting flows (10–20 km), their aggregation scale is 17.2 km, and their heterogeneous cluster is mainly in the Fourth Ring Road. This result may be due to the higher population density in the Fourth Ring Road than in other areas. Therefore, the commuting flows within the Fourth Ring Road form the cluster.

To further understand the clustering pattern of the commuting flows of different distances, it is necessary to unveil the hierarchical clustering of the commuting flows. Taking the commuting flows at median-short distances (1.5–5 km) and the commuting flows at long distances (10–20 km) as examples, the step-by-step clustering extraction method introduced in Sect. 2.3 is adopted. The first-level dominant clustering is first identified, then the first-level dominant clustering is eliminated, and the second-level dominant clustering is extracted for the remaining flow. This is repeated until a preset condition (such as a maximum level limit or a limit on the number of remaining streams) is met. The following introduces the hierarchical clustering patterns of median-short-distance and long-distance commuting flows.

4.2.1 Aggregation scales of median-short-distance commuting flows

Figure 6 shows the step-by-step cluster identification results of commuting flows at median-short distances. From Figs. 6(a) to 6(i), the dominant clusters from levels 1 to 9 are respectively shown by red lines. Dominant clusters at all levels gradually spread from the center to the periphery, the range of dominant clusters at all levels gradually decreased, and the flow density gradually decreased.

Figure 7 shows all the nine-level dominant clusters of commuting flows at median-short distances. It shows that the distribution of the dominant cluster at all levels basically covers all commercial centers in Beijing, such as the core area within the Third Ring Road, the central area between the Third Ring Road and the Fifth Ring Road, the extension line of Chang 'an Avenue between the Fifth Ring Road and the Sixth Ring Road, the subcenter of the city (Tongzhou), Fangshan, Daxing, Yizhuang, and other commercial districts.

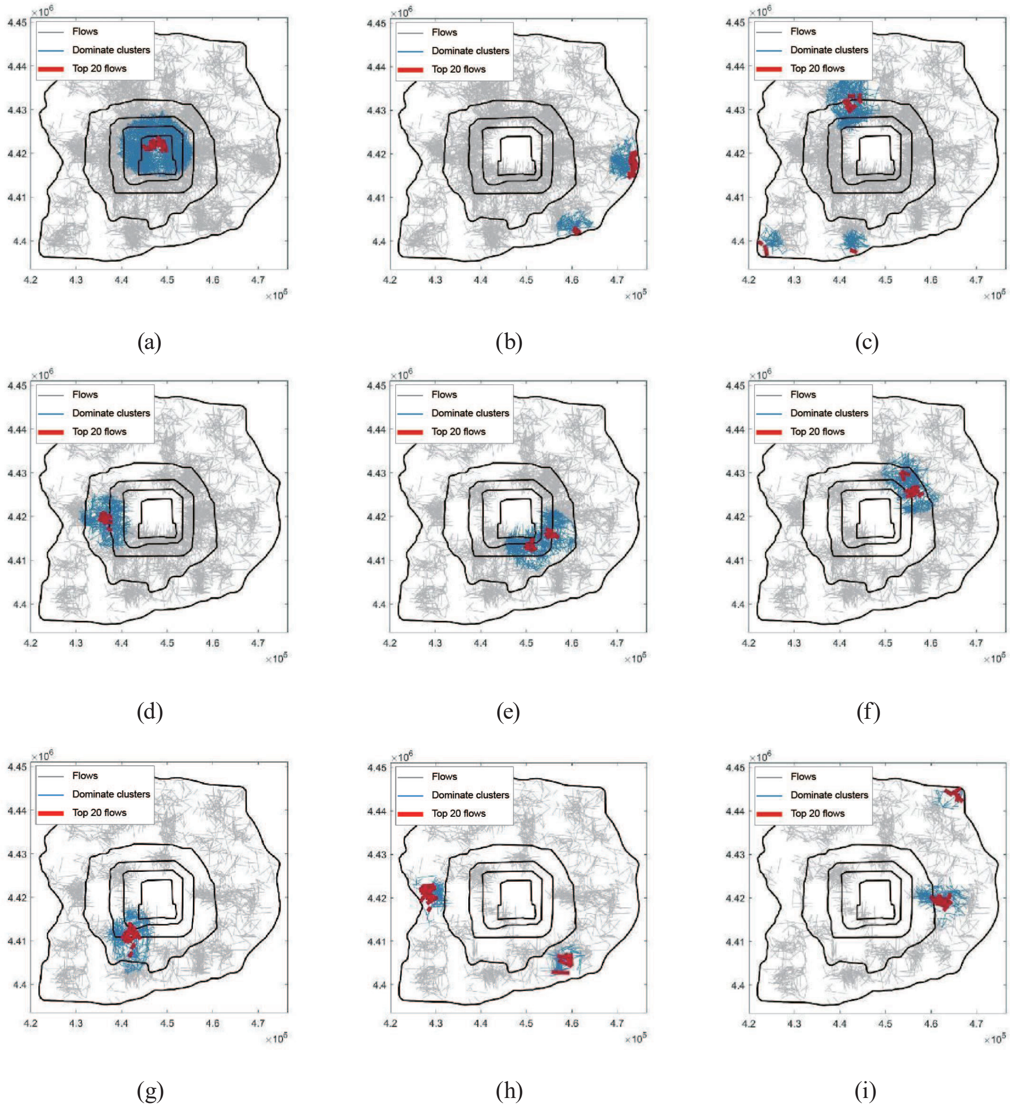


Fig. 6. (Color online) Step-by-step cluster identification results of commuting flows at median-short distances (1.5–5 km). (a) Level 1: 12.2 km. (b) Level 2: 8.9 km. (c) Level 3: 9.5 km. (d) Level 4: 9.5 km. (e) Level 5: 9.7 km. (f) Level 6: 8.7 km. (g) Level 7: 8.3 km. (h) Level 8: 7.0 km. (i) Level 9: 8.3 km.

4.2.2 Aggregation scales of long-distance commuting flows

Figure 8 shows the identification results of the first three levels of the dominant cluster for long-distance (10–20 km) commuting flows. As can be seen from Fig. 8, the scale of dominant clusters of the first three levels of long-distance commuting flows gradually decreases. These dominant clusters cover two important "commuter corridors" within the Sixth Ring Road, such as Tiantongyuan–downtown and downtown–Tongzhou. Another common commuter corridor, Huilongguan–Zhongguancun, was not recognized, possibly because its average commute distance is longer than 20 km.

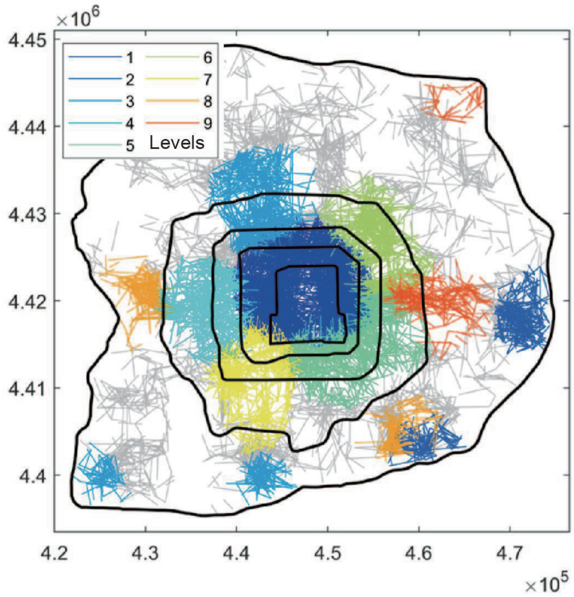
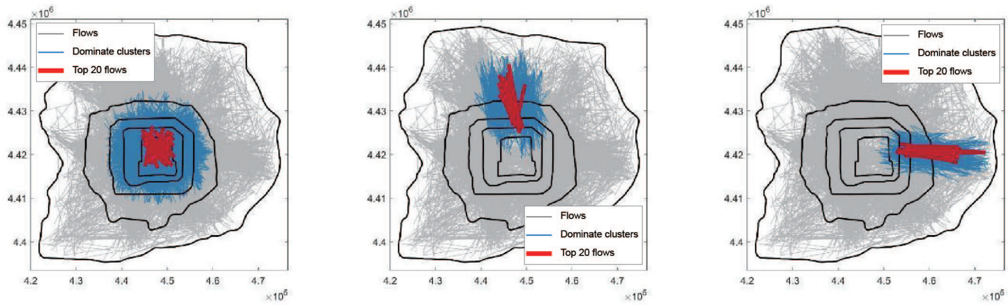


Fig. 7. (Color online) Distribution of hierarchical dominant clusters identified from median-short-distance (1.5–5 km) commuting flows.



(a) (b) (c)

Fig. 8. (Color online) Step-by-step cluster identification results of commuting flows at long distances (10–20 km). (a) Level 1: 17.2 km. (b) Level 2: 11.7 km. (c) Level 3: 9.7 km.

4.3 Implications for public transportation planning

The findings on the spatial heterogeneity and aggregation of commuting flows offer several key insights that can inform the rational planning of public transportation systems in Beijing. Given the varying levels of commuting flow aggregation across different regions and the nature of these clusters, we provide several recommendations for enhancing the public transportation system.

First, target infrastructure investment in high-heterogeneity areas is needed. In this study, we identified regions with high spatial heterogeneity in commuting flows, such as Fangshan–Shunyi and Fengtai–Haidian. These areas exhibit concentrated flow patterns that highlight bottlenecks in current transportation networks. Addressing these issues requires the following:

(1) High-capacity transit solutions: In regions such as Fangshan–Shunyi, where commuting flows are concentrated, investment in high-capacity transit solutions such as dedicated bus rapid transit (BRT) lanes or light rail can significantly alleviate pressure. The strategic placement of these systems along identified high-flow corridors will reduce congestion, especially during peak hours. Improving last-mile connectivity: high-heterogeneity areas often suffer from poor last-mile connectivity. (2) Expanding local bus services to feed into major transit hubs or adding bike-sharing systems near rail stations can improve accessibility. This will particularly benefit areas where the heterogeneity is driven by concentrated workplace destinations such as airports or business centers.

Second, a more flexible transportation strategy is needed for dispersed commuting in low-heterogeneity areas. We find that there are some areas with low spatial heterogeneity, such as Dongcheng–Fengtai regions. In such regions, expanding and optimizing bus routes will be more efficient than focusing solely on rail systems. Implementing dynamic, demand-responsive bus systems that adapt to real-time commuting data can offer flexible solutions. For example, microtransit services using smaller vehicles that adjust routes on the basis of commuting demand can serve dispersed residential and employment zones more effectively. Since commuting patterns are less predictable in low-heterogeneity areas, encouraging the use of integrated mobility platforms, where commuters can seamlessly transition between various transport modes (e.g., ride-sharing, public transit, bike-sharing), can improve transport efficiency and reduce reliance on single-mode transport solutions. These platforms can provide commuters with a range of options tailored to their specific routes.

Third, the transit system can be enhanced by considering the specific need of commuters with different commuting distances. For long-distance commuters, especially those traveling between suburban and central districts, three planning strategies deserve consideration: (1) Extending existing rail lines. Extending metro or suburban rail networks into densely populated outer zones, such as the Fourth Ring Road and beyond, will provide relief to existing road networks. This can reduce reliance on personal vehicles and enhance the overall sustainability of the transportation system. (2) Focusing on "commuter corridors". The study highlights specific commuter corridors such as downtown to Tongzhou. Investment in express rail lines or high-speed bus services along these routes can significantly reduce travel times and congestion. For example, dedicated buses or priority lanes for long-distance commuters can help avoid peak-hour bottlenecks. For short-distance commuters, dedicated short-distance transit services and pedestrian and cycling infrastructure should be enhanced. The results indicate that short-distance commuting flows are concentrated within specific regions, such as the areas surrounding major universities. Introducing frequent shuttle services or small electric buses that cater to short-distance commuters, such as university students or local employees, will be an effective solution. This can reduce the pressure on larger public transit systems by diverting short commutes to smaller, localized networks. Moreover, for areas where short-distance commuting is prevalent, enhancing pedestrian pathways and dedicated cycling lanes can promote active transport modes. For example, creating "bicycle highways" between densely packed residential and employment centers (such as university campuses) will support sustainable transport and reduce the need for short bus trips.

5. Conclusions

In this study, we investigated the spatial heterogeneity of commuting flows in Beijing using mobile phone data, analyzing the degree of heterogeneity and the aggregation scale of commuting flows. The three main findings are as follows:

(1) The degree of spatial heterogeneity varies between different regional pairs; while Fengtai–Haidian, Fangshan–Shijingshan, and Fangshan–Shunyi have high heterogeneity, Dongcheng–Fengtai, Fengtai–Chaoyang, and Daxing–Haidian have low heterogeneity.

(2) The aggregation scale of commuting flows varies with the distance, and the longer the distance, the larger the aggregation scale of commuting flow distribution, and the dominant heterogeneous clusters expand from the center to the periphery.

(3) The dominant cluster of commuting flows at median-short distances is mainly composed of commercial centers in Beijing, and the dominant cluster of commuting flows at long distances is mainly composed of "commuter corridors" within the Sixth Ring Road.

These findings can guide the rational planning of public transportation: in areas with a small scale of aggregation, such as scales smaller than the distance that a bus can travel in 15 min during off-peak hours, enhancing commuting efficiency can be achieved by adding bus lines; in areas with a large scale of aggregation, such as scales larger than the distance that a bus can travel in 15 min during off-peak hours, the commuting demands over a wide range can be met by a composite approach involving the construction of new rail transit systems and the addition of bus lines. However, this study also has some limitations. For example, we mainly focused on the spatial heterogeneity of commuting flows with distances shorter than 20 km, which may ignore the possible heterogeneity caused by longer commuting flows. In the future, longer commuting flows can be analyzed and compared to better understand the spatial distribution characteristics of commuting flows.

References

1 M. W. Horner: Prof. Geogr. **562** (2004) 160. https://doi.org/10.1111/j.0033-0124.2004.05602002.x
2 Q. Yu, W. Li, D. Yang, H. Zhang, and G. Musolino: J. Adv. Transp. **2020** (2020) 1. https://doi.org/10.1155/2020/8835981
3 J. Lee, B. Y. Choi, and E. Jung: J. Theor Biol. **454** (2018) 320. https://doi.org/10.1016/j.jtbi.2018.06.016
4 T. Mitze and R. Kosfeld: J. Geogr. Syst. **24** (2022) 5. https://doi.org/10.1007/s10109-021-00349-3
5 S. Zhou, S. Zhou, L. Liu, M. Zhang, M. Kang, J. Xiao, and T. Song: Int. J. Environ. Res. Public Health. **16** (2019) 5013. https://doi.org/10.3390/ijerph16245013
6 Y. Hu, F. Wang, and C. G. Wilmot: Pap. Appl. Geogr. **3** (2016) 14. https://doi.org/10.1080/23754931.2016.1248577
7 M. Li, M. P. Kwan, F. Wang, and J. Wang: J. Transp. Geogr. **72** (2018) 201. https://doi.org/10.1016/j.jtrangeo.2018.09.004
8 P. Zhao and Y. Cao: Transp. Policy. **92** (2020) 20. https://doi.org/10.1016/j.tranpol.2020.03.006
9 S. Angel and A. M. Blei: Cities **51** (2016) 21. https://doi.org/10.1016/j.cities.2015.11.031
10 A. Antipova: Sustainability **12** (2020) 1209. https://doi.org/10.3390/su12031209
11 B. Meng: Acta Geogr. Sin. **64** (2019) 1457. https://doi.org/10.11821/xb200912006
12 Y. Liu, F. Wang, Y. Xiao, and S. Gao: Landscape Urban Plann. **106** (2012) 73. https://doi.org/10.1016/j.landurbplan.2012.02.012
13 D. Huang and J. Wang: Sci. Total Environ. **612** (2018) 1293. https://doi.org/10.1016/j.scitotenv.2017.09.017
14 P. Lin, J. Weng, D. Alivanistos, S. Ma, and B. Yin: Sustainability **12** (2020) 5010. https://doi.org/10.3390/su12125010

15 M. Zheng, F. Liu, X. Guo, and X. Lei: Sustainability **11** (2019) 5346. https://doi.org/10.3390/su11195346
16 Y. Liu, F. Fang, and Y. Jing: Sustainable Cities Soc. **53** (2020) 101914. https://doi.org/10.1016/j.scs.2019.101914
17 Q. Yu, W. Li, H. Zhang, and D. Yang: Sustainability **12** (2020) 6203. https://doi.org/10.3390/su12156203
18 D. Ruths and J. Pfeffer: Sci. **346** (2014) 1063. https://doi.org/10.1126/science.346.6213.1063
19 C. Song, T. Pei, T. Ma, Y. Du, H. Shu, S. Guo, and Z. Fan: Int. J. Geogr. Inf. Sci. **33** (2019) 134. https://doi.org/10.1080/13658816.2018.1516287
20 K. Kelobonye, F. Mao, J. Xia, M. Swapan, and G. McCarney: Sustainability **11** (2019) 1488. https://doi.org/10.3390/su11051488
21 P. Zhang, J. Zhou, and T. Zhang: Cities **66** (2017) 10. https://doi.org/10.1016/j.cities.2017.03.004
22 X. Zhou and A. G. O. Yeh: Transp. **48** (2020) 1267. https://doi.org/10.1007/s11116-020-10094-z
23 H. Zhang, S. Xu, X. Liu, and C. Liu: J. Transp. Geogr. **91** (2021) 102929. https://doi.org/10.1016/j.jtrangeo.2020.102929
24 X. Yang, Z. Fang, L. Yin, J. Li, Y. Zhou, and S. Lu: Sustainability **10** (2018) 1435. https://doi.org/10.3390/su10051435
25 H. Shu, T. Pei, C. Song, X. Chen, S. Guo, Y. Liu, J. Chen, X. Wang, and C. Zhou: Int. J. Geogr. Inf. Sci. **35** (2020) 689. https://doi.org/10.1080/13658816.2020.1749277
26 H. Shu, T. Pei, C. Song, T. Ma, Y. Du, Z. Fan, and S. Guo: Int. J. Geogr. Inf. Sci. **33** (2019) 1355. https://doi.org//10.1080/13658816.2019.1577432

About the Authors

Sihui Guo received her B.S. degree from China University of Geosciences, Wuhan, China, in 2016 and her Ph.D. degree from the Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, China, in 2023. Since 2023, she has been an engineer at Beijing Institute of Surveying and Mapping. Her research interests are in big geodata mining, geographical information sciences, and science of cities. (guosihui@bism.cn)

Qiang Huang received his B.S. and M.S. degrees from Chongqing Jiaotong University, China, in 2016 and 2019, respectively, and his Ph.D. degree from the Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, China, in 2023. Since 2023, he has been an engineer at the Beijing Institute of Surveying and Mapping. His research interests include GIS, geographic big data, and sensors. (qiang12031@foxmail.com)

Congcong Wen received his B.S. degree in geographic information system from China University of Petroleum, China, in 2016 and his Ph.D. degree from the Aerospace Information Research Institute, Chinese Academy of Sciences, China, in 2021. He is currently a postdoctoral associate with the Department of Electrical and Computer Engineering at New York University and New York University Abu Dhabi. His research interests include computer vision, remote sensing, and data mining. (wencc1208@gmail.com)