Hindawi
Journal of Advanced Transportation
Volume 2021, Article ID 6646768, 14 pages
https://doi.org/10.1155/2021/6646768

Research Article

Analysis of Travel Hot Spots of Taxi Passengers Based on Community Detection

Shuoben Bi,[1] **Yuyu Sheng**,[1] **Wenwu He**,[2] **Jingjin Fan**,[3] **and Ruizhuang Xu**[1]

[1]*School of Geographical Sciences, Nanjing University of Information Science & Technology, No. 219 Ningliu Road, Nanjing 210044, China*
[2]*Mathematics and Physics Institute, Fujian University of Technology, No. 33 Xuefu South Road, University New District, Fuzhou 350001, China*
[3]*Institute for the History of Science and Technology, Nanjing University of Information Science & Technology, No. 219 Ningliu Road, Nanjing 210044, China*

Correspondence should be addressed to Shuoben Bi; bishuoben@163.com

Received 12 November 2020; Revised 6 March 2021; Accepted 16 March 2021; Published 26 March 2021

Academic Editor: Xinyue Xu

It is an important content of smart city research to study the activity track of urban residents, dig out the hot spot areas and spatial interaction patterns of different residents' activities, and clearly understand the travel rules of urban residents' activities. This study used community detection to analyze taxi passengers' travel hot spots based on taxi pick-up and drop-off data, combined with multisource information such as land use, in the main urban area of Nanjing. The study revealed that, for the purpose of travel, the modularity and anisotropy rate of the community where the passengers were picked up and dropped off were positively correlated during the morning and evening peak hours and negatively correlated at other times. Depending on the community structure, pick-up and drop-off points reached significant aggregation within the community, and interactions among the communities were also revealed. Based on the type of land use, as passengers' travel activity increased, travel hot spots formed clusters in urban spaces. After comparative verification, the results of this study were found to be accurate and reliable and can provide a reference for urban planning and traffic management.

1. Introduction

With the rapid development of information technology, spatial analysis driven by data forces geographic information science to face new challenges. Furthermore, visual analysis combined with geographic computing has greatly improved people's ability to mine new knowledge [1]. On the one hand, mobile information collection technology based on the global positioning system has become more mature; on the other hand, the flow space with urban residents' activities as the main carrier has become more extensive [2]. Although the intension of geographic information science has not changed, its content and form have become richer. Therefore, breaking through the traditional urban spatial research model is the key to discovering the law of urban residents' activities.

Time-space analysis based on residents' activities can explain the homogeneity of the influence of individual residents' behaviors on urban space, and the behaviors between different individuals can reflect that they are restricted by urban space and show their differences [3]. Therefore, as Harvey and Han [4] proposed the concept of geographic data mining and knowledge discovery, scholars have continued to explore knowledge in recent decades, and geography has experienced transition from an empirical paradigm to a system simulation paradigm and then to a data-driven paradigm [5]. Early research on the behavioral patterns of urban residents' activities mainly focused on extracting residents' activity points and on the correlation analysis of those points. For example, Veloso et al. [6] studied the strong association pattern of residents' activity locations, and Ahas et al. [7] studied the time difference and spatial distribution of

residents' activities. Recent research has mainly focused on the identification of urban hot spot functional areas, urban accessibility analysis, urban boundary division, polycentric evaluation, etc. For example, Scholz et al. [8] studied urban residents' behavioral patterns and the temporal and spatial development of urban hot spots, and Cui et al. [9] studied the accessibility of urban residential areas and the distribution of low-access residential areas. Zhong et al. [10] studied the overall spatial structure of changes in the center of the city boundary, and Huang et al. [11] studied the effects of urban traffic and pedestrian activities. For two-layer fine-grained networks, Guo et al. [12] studied the structure of different urban road networks and developed corresponding datasets. Hamedmoghadam et al. [13] studied the displacement index of individual travel granularity to simplify collective behavioral patterns. In addition, there are related studies on urban planning and environmental safety assessment. For example, Zheng et al. [14] studied the characteristics of cross connectivity between urban planning and taxi driving, and Wu et al. [15] studied the temporal and spatial patterns of urban road traffic accidents.

In summary, the early research model was relatively narrow in its scope, only considering residents' activities but ignoring the characteristics of urban space. In recent years, research has become relatively rich, mainly based on urban planning, which was based on the analysis of residents' historical activities, such as behavioral patterns. The landmark research achievements are the GN algorithm [16] and the Newman fast algorithm [17], both of which are classic community detection algorithms. These can fully reveal the different resident activities, the spatial pattern, and the impact of potential factors on decision-making. In addition, Qin et al. [18] studied the traffic intensity and edge weight of network nodes based on the network interaction characteristics of urban hot spots.

The movement trajectory as a type of multisource sensor data has been widely adopted by researchers. Through the movement trajectory, the travel mode of residents' activities can be understood more clearly, the hot spots of activities can be extracted more accurately, and the reasons for the resident movement can be analyzed. Research on moving trajectories in Nanjing mainly includes Xu et al. [19], who found that traffic hot spots in Nanjing have the spatial distribution characteristics of agglomeration from the surroundings to the center; Yang et al. [20], who found that the Nanjing public transport system has the characteristics of cascading failure congestion; and Jin and Xu [21], who showed that the traffic flow on the key nodes of different grades of road network in Nanjing has obvious hierarchical structure characteristics. Therefore, this study aims to use the passenger pick-up and drop-off points extracted from taxi movement trajectories to explore the travel rules of taxi passengers, analyze the time and space patterns of taxi pick-up and drop-off communities, establish passenger travel activity indicators based on community detection [16], combine the data of graded roads and points of interest, explore the temporal and spatial characteristics of taxi passenger travel hot spots, and examine the causes of the formation of spatiotemporal characteristics.

2. Data Description

Nanjing is located in the southwest of Jiangsu Province, China. The study area selected in this paper covers the main urban area of Nanjing, including Gulou, Xuanwu, Jianye, Qinhuai, and Yuhuatai, as shown in Figure 1.

The data used in this study includes two parts of the Nanjing taxi trajectory and feature dataset. The source of Nanjing taxi trajectory data was Datatang (https://www.datatang.com), which contains data from approximately 7,800 taxis with a sampling interval of 30 seconds. Data for the same period for three consecutive years were selected: January 25–31, 2015; February 13–19, 2016; and February 2–8, 2017. The source of road network data was Tianditu (https://www.tianditu.gov.cn), a national geographic information public service platform, which contains eight types of graded roads. Considering the nature of taxi services, railways, subways, light rails, and high-speed rails were excluded. Approximately 2,400 road sections classified as national roads, provincial roads, county roads, township and village roads, and other roads were used in the analysis. The data source of points of interest was Baidu POI (http://www.data-shop.net/tag/), which includes four types of land use: land for commercial use, residential land, land for public management and public service, and land for transportation. A total of approximately 26,000 points of interest were selected.

3. Methodology

This study used ArcGIS to perform map matching and geocoding preprocessing on taxi movement trajectories and the feature data of Nanjing City. A road network geographic database and a road network topology map were created using the complex network-modeling tool NetworkX to map the road intersections. The abstraction of the road intersections is a complex network node, the corresponding road section is abstracted as an edge, and community detection is performed on the taxi pick-up and drop-off points. Based on community detection, the passenger travel activity index is constructed, and hot spot mining is realized through spatial statistics. The technical process is shown in Figure 2.

3.1. Community Detecting. First, based on the concept of a dual graph [22], the road is defined as a generalized network composed of nodes and edges.

$$\begin{aligned} G &= \left\{\left(N_g, E_g\right) \middle| 1 \le g \le n_{\text{Graph}}\right\}, \\ \left|N_g\right| &= n_{\text{node}}, \\ \left|E_g\right| &= n_{\text{edge}}, \end{aligned} \tag{1}$$

where (N_g, E_g) represents any road segment, n_{Graph} represents the total number of road segments, n_{node} represents the number of nodes included in a road segment, and n_{edge} represents the number of edges included in a road segment.

The road is abstracted into a complex network, as shown in Figure 3: (a) is the original road graph, which contains 9 road sections and 14 nodes; (b) is the corresponding original

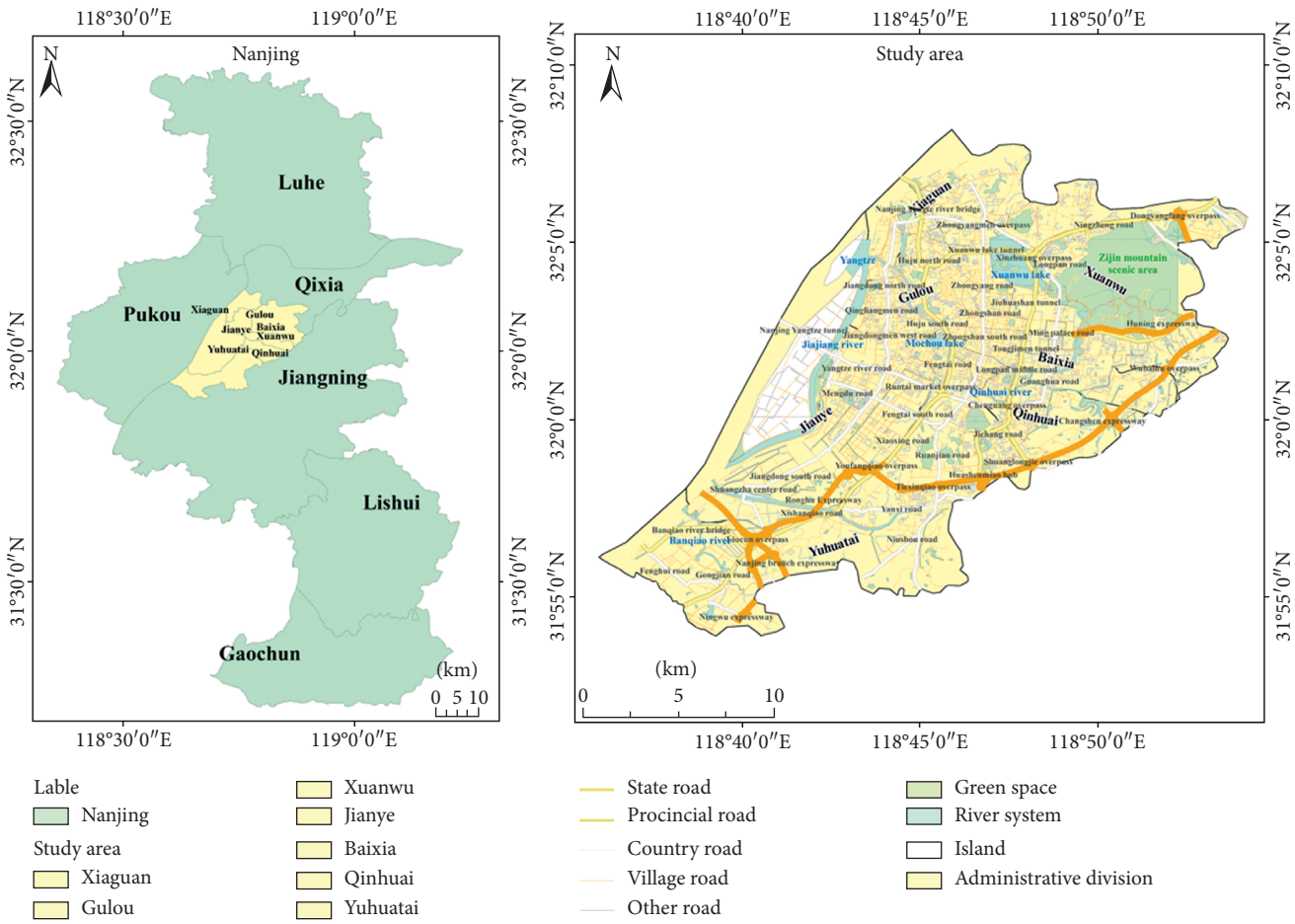


Figure 1: Study area map.

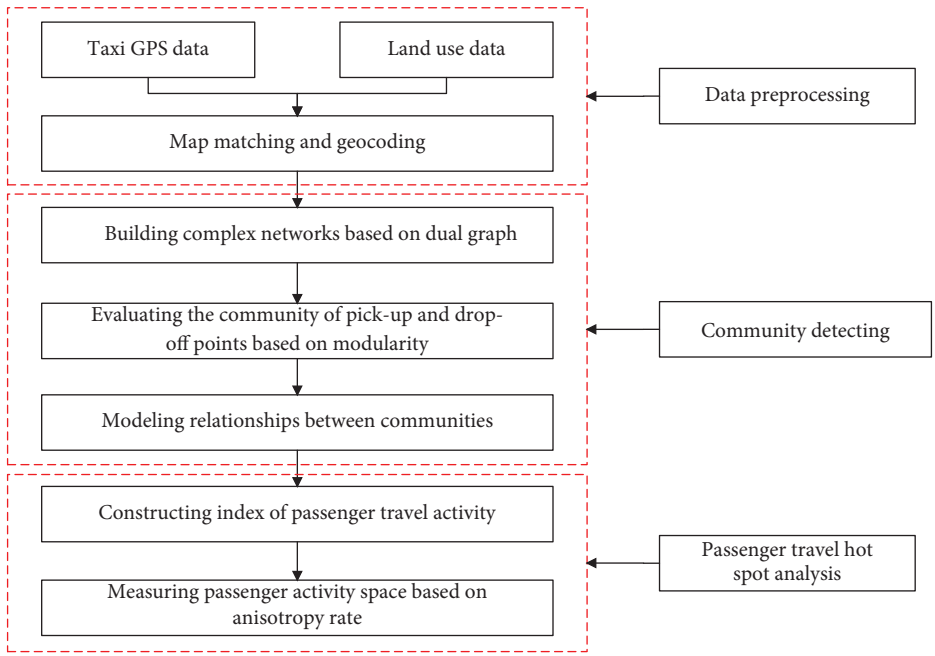


Figure 2: Technical flowchart.

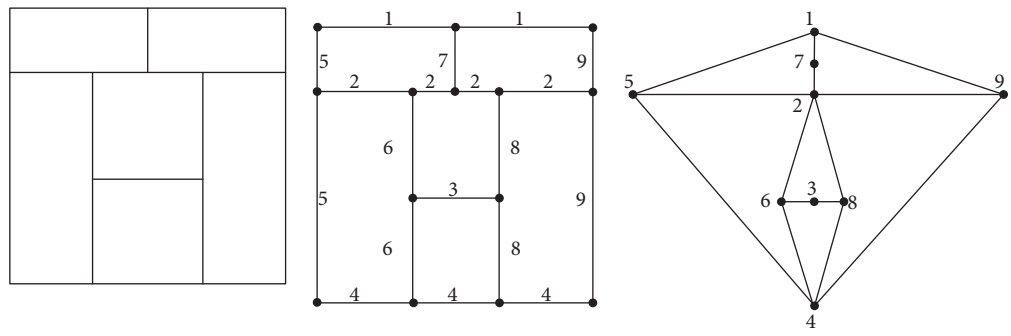


Figure 3: Diagram of a complex network of roads: (a) road network graph; (b) road network initialization graph; (c) road network dual graph.

graph, where the nodes represent road intersections and the edges represent road sections between nodes; (c) is the corresponding dual graph, in which nodes represent road sections and edges represent the intersecting relationship of road sections.

Figure 3(c) is used to abstract the road network and express intersections and road sections as nodes and edges, respectively. Before community detection, the stay points are mapped to the road network, which characterizes the geographic location using map matching technology. After abstraction, the road network not only retains the geographic location information but also characterizes the network connectivity. The connectivity of the road network changes with the degree centrality of each node in the abstract network. The more the pick-up points, the higher the exit degree of the node, and the more the drop-off points, the higher the entry degree of the node. Therefore, when the stay points change dynamically in the road network, they are aggregated into clusters according to Bayes' rule, and communities are established to represent the activities of the community's residents as hot spots.

Second, we define the stay point, corresponding to the stay point to the network, and use it as a separate atomic cluster [23].

$$S = \left\{ \left(S_i^O, S_i^D \right) | \quad 1 \le i \le N_S \right\}, \tag{2}$$

where (S_i^O, S_i^D) represents any pair of stay points, N_S represents the number of stay points included in the candidate dataset, S^O represents the pick-up point in a pair of stay points, and S^D represents the drop-off point in a pair of stay points.

Subsequently, the distance between the stay points is calculated according to the Euclidean metric. Taking the above stay points as an example (the same is true for the drop-off point), the two closest points are continuously merged into the same cluster, and the distance between the clusters is calculated according to the average distance measurement:

$$d_{\text{avg}}\left(C_i, C_j\right) = \frac{1}{n_i n_j} \sum_{Q_i \in C_i} \sum_{Q_j \in C_j} \left| Q_i - Q_j \right|, \tag{3}$$

where C_i and C_j are the clusters where the pick-up points S_i^O and S_j^O are located; n_i and n_j are the number of pick-up points contained in clusters C_i and C_j, respectively; and $|Q_i - Q_j|$ is the distance between clusters C_i and C_j. Q is the incremental matrix, which is the adjacency matrix that stores the nodes and edges within the cluster. Q is calculated as follows [24]:

$$Q = \frac{1}{E_g} - \frac{k_i k_j}{2E_g^2}, \tag{4}$$

where E_g is the total number of connected edges, k_i is the degree of node i, and k_j is the degree of node j.

For a network with N nodes, the execution process of the algorithm used in this study includes the following steps:

(1) *Initializing*. Treat each node as a cluster and set the increment matrix to $Q = 0$.

(2) *Merging and Updating*. Combine the two clusters C_i and C_j that have edges connected in such a way that maximizes d_{avg}, maximize d_{avg} using Bayes' rule, and update the combined cluster.

(3) *Terminating*. Continue the merging and updating process until there are no clusters that can be merged.

We then count the number of pick-up points in each cluster as the amount of information I and set the amount of information as the weight according to Bayes' rule, thereby establishing a community as

$$I = \frac{1}{n} \sum_{i=1}^{n} C_{n_i}^i,$$

$$P\left(\frac{n_i}{I}\right) = \frac{P\left(I/n_i\right) \times P\left(n_i\right)}{P\left(n_i\right) \times P\left(I/n_i\right) + P\left(\overline{n_i}\right) \times P\left(I/\overline{n_i}\right)}, \tag{5}$$

$$B_i = \frac{\sum_{i=1}^{n} \left(x_i, y_i\right) \times P\left(n_i/I\right)}{\sum_{i=1}^{n} P\left(n_i/I\right)},$$

where $C_{n_i}^i$ means that the cluster C^i in the community contains n_i pick-up points, and n is the total number of pick-up points in the community. $P(n_i/I)$ means that the pick-up

points are aggregated into the pick-up point community, and (x_i, y_i) denotes the coordinates of the pick-up points. B_i is the center of the mass coordinates of community (x_B, y_B).

Finally, the community is evaluated according to the degree of modularity M:

$$M = \frac{1}{2E}\sum_{ij}\left(C_{ij} - \frac{k_i k_j}{2E}\right)\partial\left(B_i, B_j\right), \tag{6}$$

where E is the number of edges in the network. C_{ij} takes the value 0 or 1; if $C_{ij} = 1$, there is an edge between the pick-up points S_i^O and S_j^O; otherwise, there is no edge; k_i and k_j are the degrees of the pick-up points S_i^O and S_j^O; B_i and B_j are the centroids of the communities where the pick-up points S_i^O and S_j^O are located. Only when $B_i = B_j$, $\partial(B_i, B_j) = 1$. The value range of M is $[0, 1]$; the larger the value, the more obvious the community structure.

3.2. Constructing an Indicator of Passenger Travel Activity. Based on the community detection results, the passenger travel activity point is set as $S_i = [S_i^O, S_i^D, B_i^O, B_i^D]$, where S_i^O is the coordinate of the taxi passenger pick-up point, S_i^D is the coordinate of the drop-off point, B_i^O is the centroid coordinate of the community to which the pick-up point belongs, and B_i^D is the centroid coordinate of the community to which the drop-off point belongs. Therefore, there are three situations of inclusion, intersection, and separation of the communities, in which the passenger board and drop-off points belong, as shown in Figure 4.

As shown in Figure 4, the center of the pick-up point community is O and the radius is r_O; the center of the drop-off point community is D and the radius is r_D; and the smallest circle center that contains the pick-up point community is C and the radius is R. No interaction between the pick-up point community and the drop-off point community is shown by the white area in the figure, while an interaction between the pick-up point community and the drop-off point community is shown by the shaded area in the figure.

The passenger travel activity index is a combination of outbound visit heat and arrival visit heat [25], with 1 h as the unit time for sampling and 1 km as the unit distance for calculation, defined as $A_i = [depart_i, arrive_i]$; the calculation is as follows:

$$\begin{aligned}
\text{scope}_i &= 1 - \frac{d_i}{2R_i},\\
p\left(S_i \in B_i\right) &= \begin{cases} 1, & S_i^O \notin \odot B_i^D \text{且} S_i^D \notin \odot B_i^O,\\ 0, & {}_i^O \in \odot B_i^D \text{或} S_i^D \in \odot B_i^O, \end{cases}\\
\text{depart}_i &= \sum_{S_i^O \in B_i^O}^{n_O} p\left(S_i^O \in \odot B_i^O\right) \times \text{scope}_i^O,\\
\text{arrive}_i &= \sum_{S_i^D \in B_i^D}^{n_D} p\left(S_i^D \in \odot B_i^D\right) \times \text{scope}_i^D,
\end{aligned} \tag{7}$$

where d_i is the distance between taxi passengers' pick-up and drop-off points, R_i is the smallest radius of the circle that contains the community of the pick-up point, n_O denotes all pick-up points included in the pick-up point community, and n_D denotes all drop-off points included in the drop-off point community. $scope_i$ is a probability density function that describes the distance between the pick-up point and the centroid of the community it belongs to; $depart_i$ represents the probability density estimation from the pick-up point to the pick-up point community, namely, the popularity of passenger outbound travel activities; $arrive_i$ represents the probability density estimation from the drop-off point to the drop-off point community, that is, the popularity of passenger arrival travel activities.

Given N nodes, there can be at most $N(N-1)/2$ edges, and a random network can be obtained by randomly selecting M edges from these edges. Obviously, a total of $C_{N(N-1)/2}^M$ random graphs are possible, each with the same probability. When the node's connection probability p exceeds the critical probability $p_c(N) = \ln N/N$, every random graph is connected. Therefore, a random graph Q with N nodes and connection probability $p = p(N)$ satisfies

$$\lim_{N\longrightarrow\infty} P_{N,p}(Q) = \begin{cases} 1, & p(N)/p_c(N) \longrightarrow \infty,\\ 0, & p(N)/p_c(N) \longrightarrow 0. \end{cases} \tag{8}$$

For the community of pick-up and drop-off points formed by the corresponding pick-up and drop-off points, when $p(N)/p_c(N) \longrightarrow \infty$, the random network is completely connected, forming a closed network with no isolated nodes. In other words, the pick-up point will not belong to the community of the drop-off points, and the drop-off point will not belong to the community of the pick-up points, namely, $S_i^O \notin \odot B_i^D$ and $S_i^D \notin \odot B_i^O$. When $p(N)/p_c(N) \longrightarrow 0$, the random network has a tree structure, and there are branch nodes belonging to other connected subgraphs, namely, $S_i^O \in \odot B_i^D$ or $S_i^D \in \odot B_i^O$.

Because passengers are not necessarily restricted to moving in certain pairs of communities, the standard deviation ellipse method [26] is used to measure the spatial distribution characteristics of passenger activities and the interaction of the community where the passengers travel activity points are evaluated according to the anisotropy rate, that is, an equal ellipse. The higher the anisotropy rate under the area, the more directional and purposeful the passenger activities in the community. The anisotropy rate α is calculated as follows:

$$\alpha = \frac{\sigma_{x'} - \sigma_{y'}}{\sigma_{x'}} \times 100\%, \tag{9}$$

where $\sigma_{x'}$ is the length of the major axis of the ellipse and $\sigma_{y'}$ is the length of the minor axis of the ellipse.

4. Results and Analysis

4.1. Spatial and Temporal Characteristics of Passenger Travel Activity. One hour was adopted as the unit time interval to summarize the passenger pick-up and drop-off points

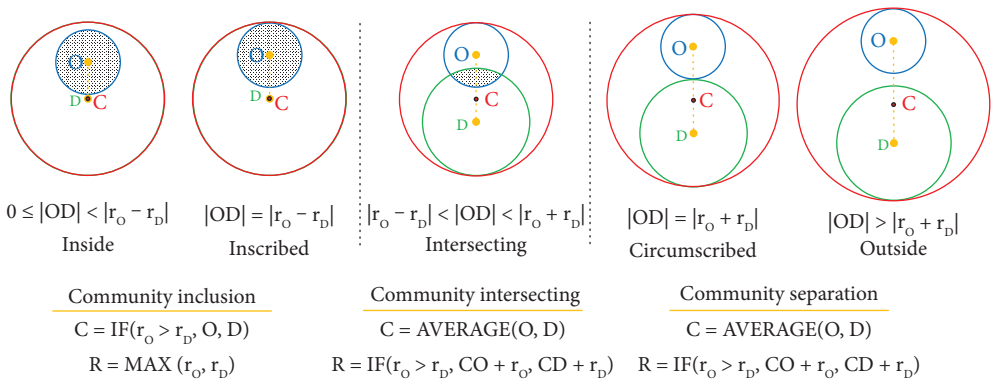


Figure 4: Schematic diagram of three community situations. Note. The blue circle represents the taxi pick-up community, its center is O, and its radius is r_O. The green circle represents the taxi drop-off community, its center is *D*, and its radius is r_D. The red circle represents the smallest circumscribed circle that contains two communities, its center is C, and its radius is R. In the case of community inclusion, if $r_O > r_D$, then C is O; otherwise, C is D. The value of *R* is maximum in r_O and r_D. In the case of community intersection or community separation, the value of C is average of O and *D* coordinates. If $r_O > r_D$, then the value of *R* is the sum of the values of CO (distance value between C and O) and r_O; otherwise, the value of *R* is the sum of the values of CD (distance value between C and D) and r_O.

recorded in the taxi trajectory data for the weeks included in the data analysis (Section 2), as shown in Figure 5.

It can be seen from Figure 5 that the number of taxi passengers getting on and off is consistent across days of the week, and there are fluctuations at different times of the day. The daytime is higher than the nighttime, and there is a significant increase during the morning rush hour. Moreover, there is also a certain increase during the evening rush hour. Thus, taxi passenger travel show more daytime activity, less nighttime activity, and frequent activity during the morning and evening peak hours.

Taking one hour as the unit time interval, the average modularity and anisotropy rate of the communities where taxi passengers were picked up and dropped off in 2015, 2016, and 2017 are shown in Figure 6.

It can be seen from Figure 6 that, during the morning and evening peak hours, the modularity is relatively high, and the anisotropy rate curve is relatively steep. When the modularity increases, the anisotropy rate also increases. In other periods, the modularity is relatively low, and the anisotropy rate curve is relatively flat. When the modularity decreases, the anisotropy rate increases. This shows that the community structure of taxi passengers' pick-up and drop-off points becomes closer as the purpose of passengers' travel increases. For example, during morning peak hours, passengers travel mainly from home to office; during evening peak hours, passengers travel mainly from office to home; and in other periods, residents' activities are affected by differences in travel motivation, thus showing randomness.

To clearly reflect the differences in residents' travel activities at different times, the morning peak hours were 8 : 00–9:00, working hours 13 : 00–14 : 00, evening peak hours 18 : 00–19 : 00, and rest period 22 : 00–23 : 00. We can conduct community detection at the points where taxi passengers board and alight, as shown in Figure 7.

It can be observed from Figure 7 that during the period of 8 : 00–9:00, the corresponding communities of the pick-up and drop-off points are separate, and during the period of 13 : 00–14 : 00, the corresponding communities of the pick-up and drop-off points are mainly intersecting. The pick-and-drop points, as shown in Community No. 1 (Figure 7(b)), are mainly gathered in the southeast of Gulou, southwest of Xuanwu, west of Qinhuai, and northeast of Jianye area. During the period of 18 : 00–19 : 00, the corresponding communities of the pick-up and drop-off points are mainly separate. During 22 : 00–23 : 00 (Figure 7(d)), the corresponding communities of the pick-up and drop-off points are mainly inclusive. The pick-up and drop-off points shown in Community No. 1 are mainly concentrated in the northwest of Xuanwu, and the pick-up and drop-off points shown in Community No. 2 are mainly concentrated in Jianye. In the northeast, the pick-up and drop-off points shown in Community No. 5 are mainly concentrated in the northeast of Yuhuatai, and the pick-up and drop-off points shown in Community No. 6 are mainly concentrated in the middle of Gulou.

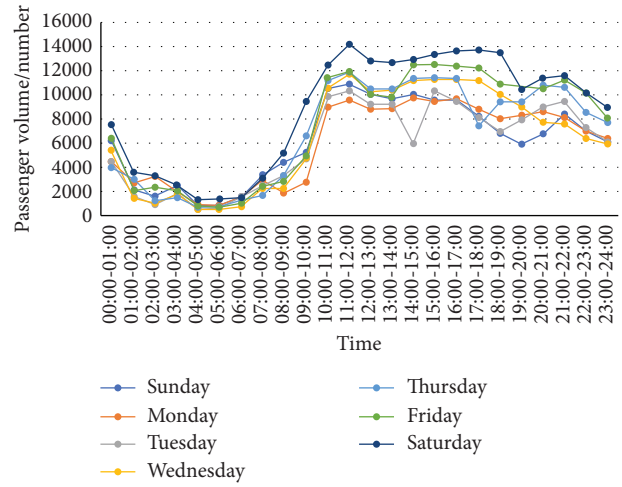


Figure 5: Graph of average passenger volume.

This shows that, during the same period, passenger travel activities are affected by the purpose of travel, showing the same behavioral pattern in the same community, obvious

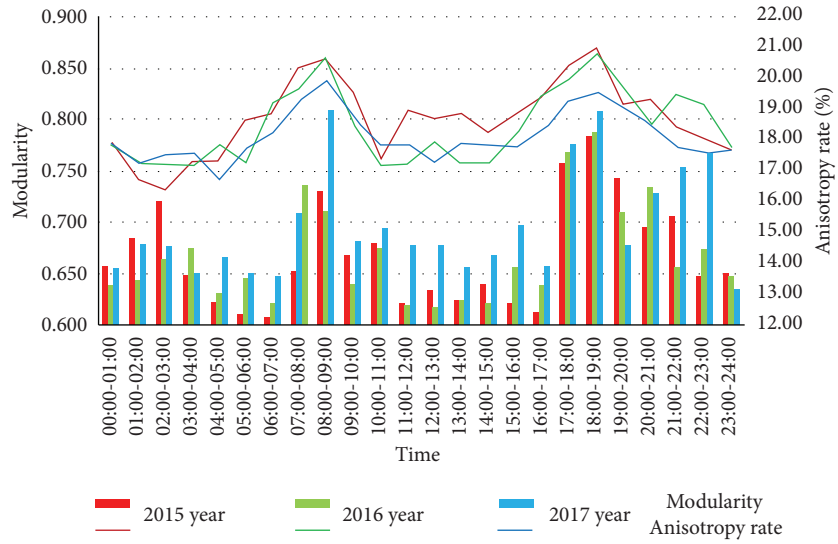
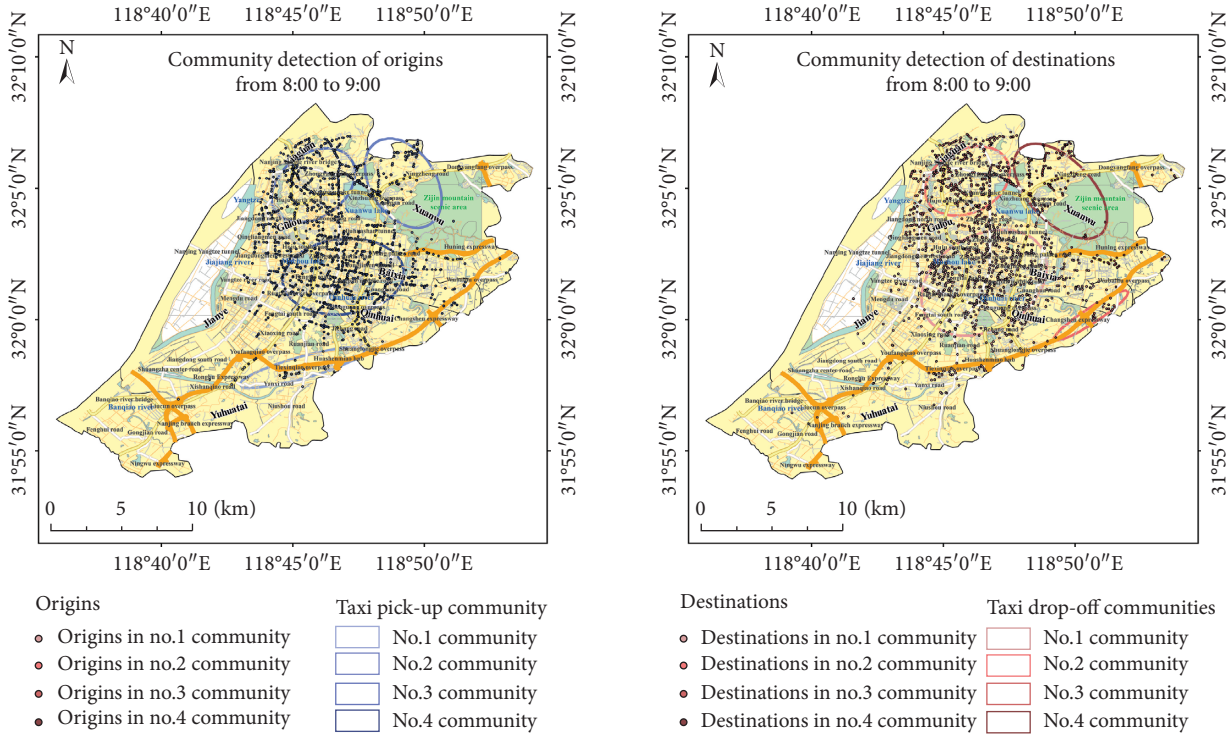
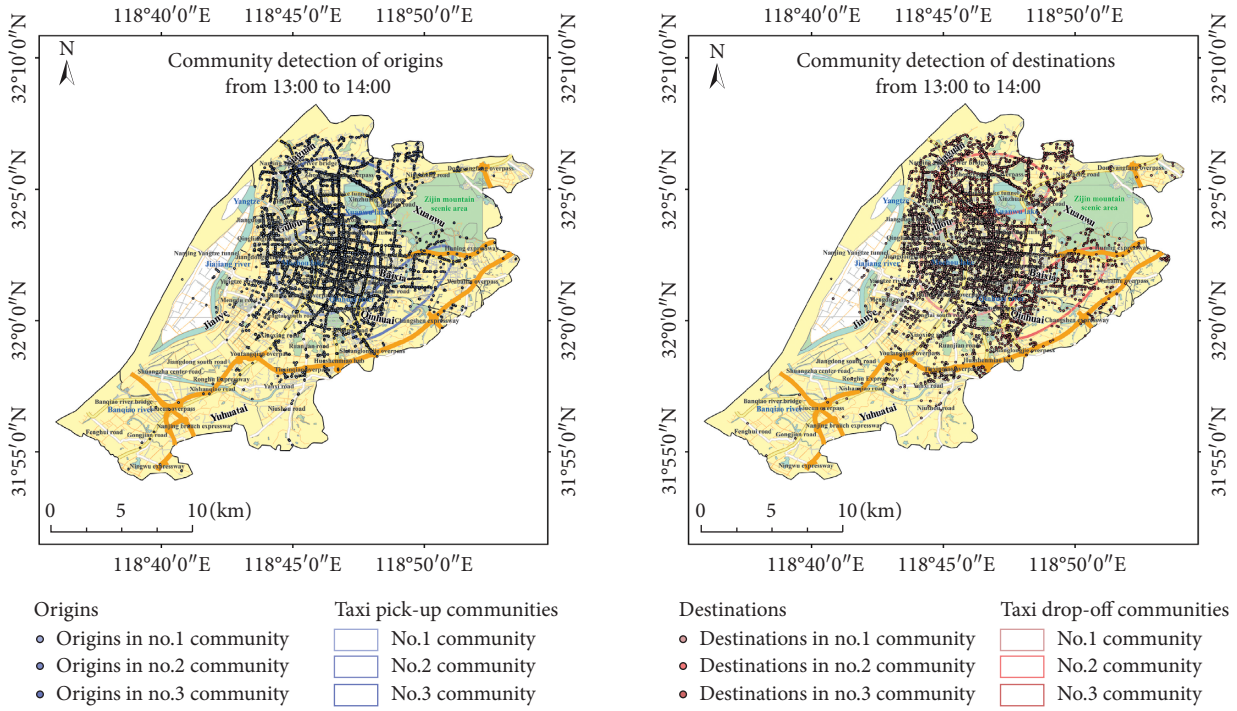


Figure 6: Graph of average community modularity and anisotropy rate.

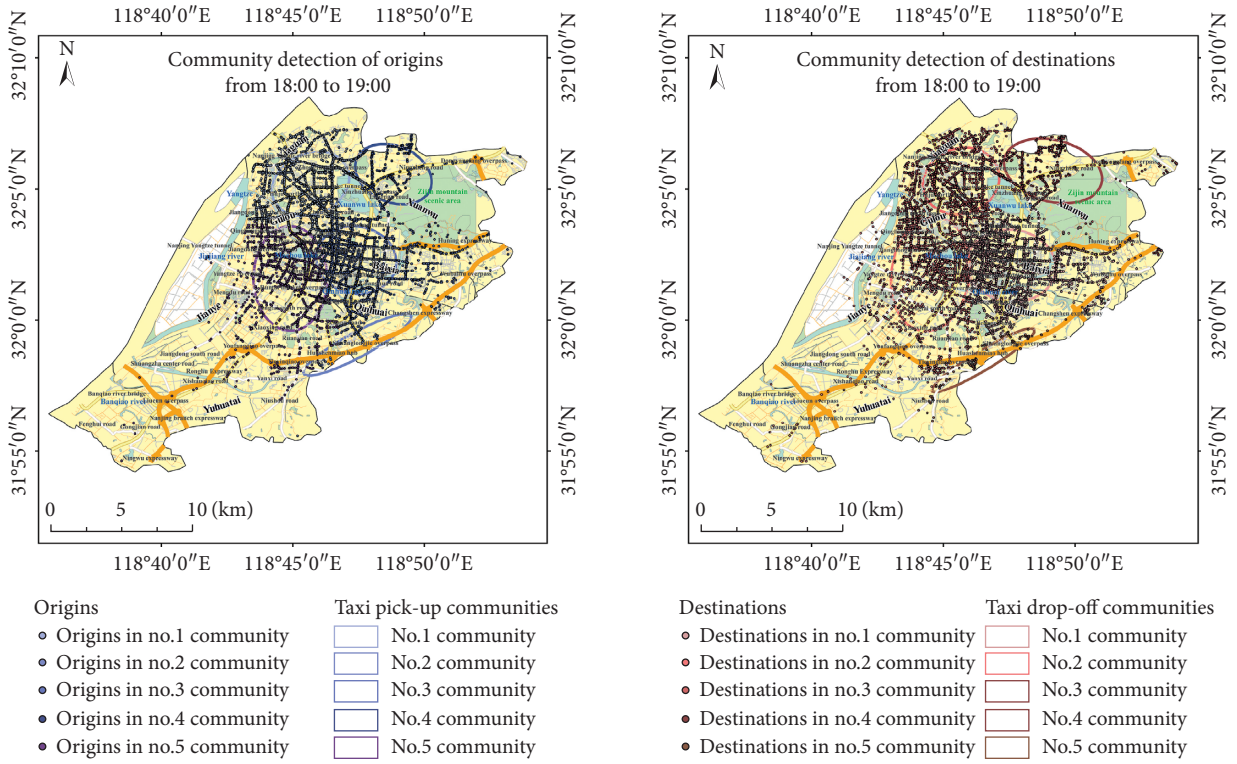


(a)

Figure 7: Continued.



(b)



(c)

Figure 7: Continued.

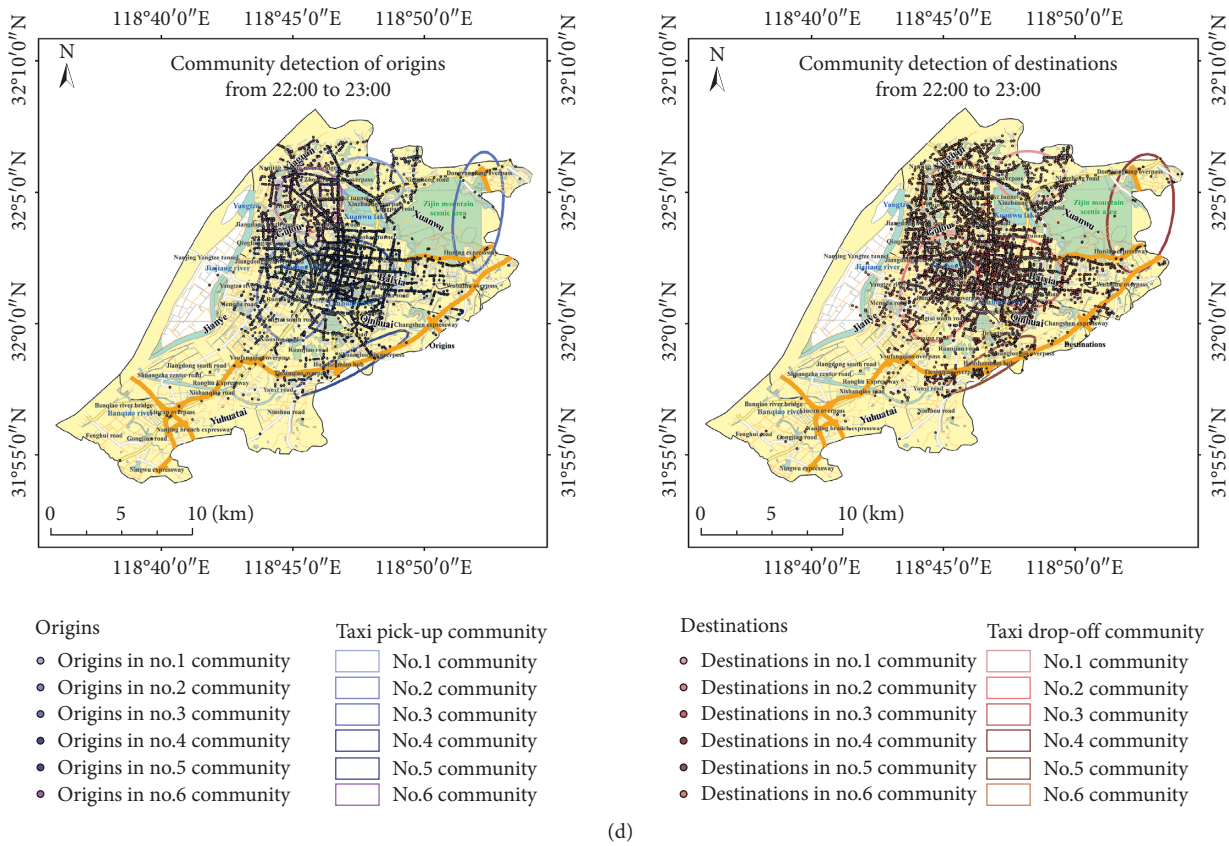


Figure 7: (a) Community detection results of origin and destination from 8 : 00 to 9 : 00. (b) Community detection results of origin and destination from 13 : 00 to 14 : 00. (c) Community detection results of origin and destination from 18 : 00 to 19 : 00. (d) Community detection result of origin and destination from 22 : 00 to 23 : 00.

spatial clustering, and differences between different communities.

In summary, the characteristics of the passenger travel activity time distribution show more daytime and fewer nighttime activities and frequent peak hours in the morning and evening. The characteristics of the passenger travel activity spatial distribution show concentrated urban centers and scattered peripheral areas. Affected by the purpose of travel and structure of the community, passenger travel activities behave in the same way in the same community at the same time, and there is an interaction between different communities at the same time.

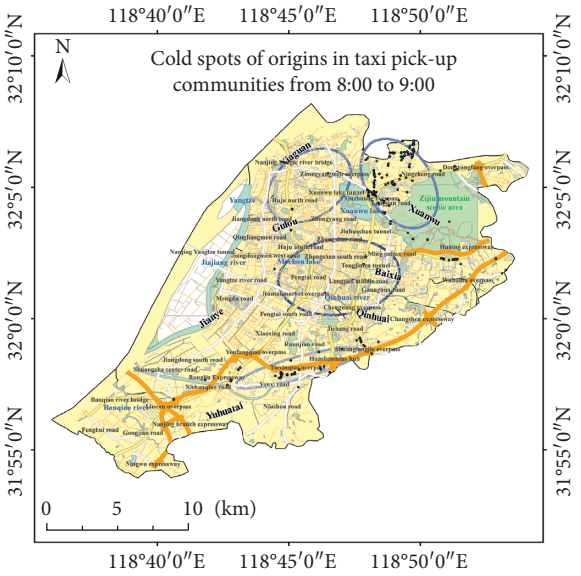
4.2. Passenger Travel Hot Spot Analysis. Taking the 8 : 00–9: 00 time period as an example, we considered the minimum circle radius of the community, including the pick-up and drop-off points, as the aggregation distance, and the outbound visit heat and arrival visit heat of the passenger's travel activity as the indicators. The corresponding communities were divided according to the first decile, and the pick-up and drop-off points were aggregated to extract hot spots. Furthermore, the pick-up and drop-off points in the corresponding community were aggregated according to the last decile to extract cold spots, as shown in Figure 8.

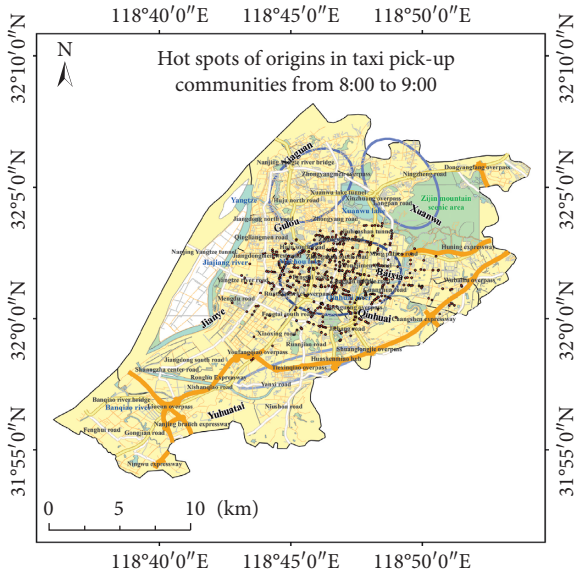
It can be clearly observed from Figure 8 that the pick-up points of the decile before the visit were distributed in Community No. 4, and the drop-off points of the decile before the visit were mostly distributed in Community No. 1: in the southeast of Gulou, southwest of Xuanwu, west of Qinhuai, northeast of Jianye, and north of Yuhuatai. These are adjacent areas of the main urban centers, and the pick-up or drop-off points in the decile after the visit heat and the visit heat were randomly distributed. Therefore, it can be concluded that passenger travel hot spots were clustered or dispersed as passenger travel activity increased or decreased.

Xu et al. [19] showed that the hot spots in Nanjing have a spatial distribution characteristic of clustering from the surroundings to the center and that Moran's I value around the clustering center is negative. The hot spots of passenger travel extracted in this study are consistent with the results of the previous study, and a more obvious spatial local agglomeration can be found based on community detection of pick-up and drop-off points.

Taking the period from 8 : 00 to 9 : 00 as an example, the average visit heat and average visit heat statistics were calculated on five graded road sections: national highway, provincial highway, county highway, township and village highway, and other roads, as shown in Table 1.

It can be seen from Table 1 that the average visit popularity ordered from high to low was township and village roads, provincial roads, other roads, county roads, and national roads. The average visit popularity ordered from





Cold spots of origins value of departure:

[0.050896, 0.085807]

Taxi pick-up communities

- No.1 community
- No.2 community
- No.3 community
- No.4 community

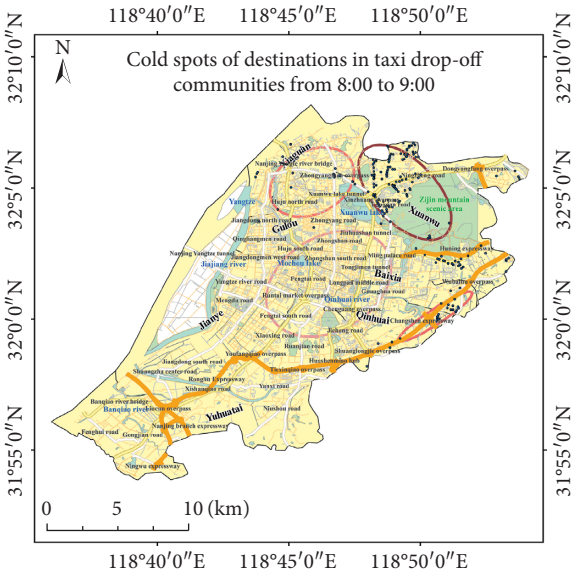
Hot spots of origins value of departure:

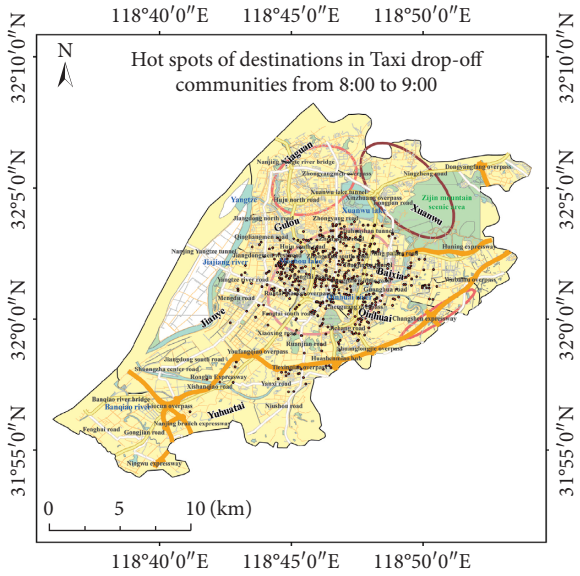
[0.730182, 0.80000]

Taxi pick-up communities

- No.1 community
- No.2 community
- No.3 community
- No.4 community

(a)





Cold spots of destinations value of arrival:

[0.039249, 0.075679]

Taxi drop-off communities

- No.1 community
- No.2 community
- No.3 community
- No.4 community

Hot spots of destinations value of arrival:

[0.734238, 0.80786]

Taxi drop-off communities

- No.1 community
- No.2 community
- No.3 community
- No.4 community

(b)

Figure 8: (a) Hot and cold spots of origins in pick-up communities from 8 : 00 to 9 : 00. (b) Hot and cold spots of destinations in drop-off communities from 8 : 00 to 9 : 00.

Table 1: Average activity of graded roads.

Graded road	Number of roads	Average value of departure	Average value of arrival
National roads	66	0.530	**0.598**
Provincial roads	84	0.648	0.474
County roads	181	0.594	0.546
Township and village roads	1191	**0.694**	0.582
Other roads	178	0.626	0.592

high to low was national roads, other roads, town and village roads, county roads, and provincial roads.

By searching the database, we found that the representative road sections with higher outbound visits on the graded roads were Yurun Street, Fengqi Road, Jiajiang Bridge, Jinsu'an Road, and Caodu Lane; representative sections with lower outbound visits were Fengwu Road, Binjiang Road, Jiangshan Street, Chuanjiang Street, and Houde Road. Representative road sections with higher arrival visits were Zhenxing Road, Shuangtang Road, Jiangshan Street, Xiaofenqiao, and Fanjiatang. Representative sections with lower arrival visits were Fengwu Road, Moxiang Road Overpass, Nanjing Yangtze River Tunnel, Lingyin Road, and Kuitou Alley. This shows that passenger travel activities are closely related to the traffic functions carried by the graded roads. The main function of expressways is to enable continuous traffic, of trunk roads is to enable transportation, of secondary trunk roads is to enable distribution traffic, and of branch roads is to enable service in local areas.

A study by Yang [20] showed that there is cascading failure and congestion in the traffic system of Nanjing. Travel conditions of people in an unbalanced road network load are affected by the coupling of sub-road networks. The results of passenger travel activity on graded roads in this study are consistent with the conclusions of that research.

Taking the 8 : 00–9:00 period as an example, based on the residents' walking considerations, the pick-up and drop-off points are the center of the circle with a radius of 300 m for coverage, covering commercial land, residential land, public management and public service land, and transportation land. The average outbound visit heat and average arrival visit heat were calculated for approximately 30 types of land use involving a total of 26,000 points of interest, as shown in Table 2.

It can be seen from Table 2 that the land use type with the highest average outbound visit heat was urban residential land, and the land use type with the lowest average outbound visit heat was commercial and financial land. The land use type with the highest average arrival visit heat was commercial and financial land. The lowest average arrival visited land use type was urban residential land.

By searching the database, the representative point of interest with higher average outbound visits was Yangzhuang Village, corresponding to Shiyang Road. The representative point of interest with lower average outbound visits was Flower Building, corresponding to Software Avenue. The representative point of interest with higher average arrival visits was Commercial Century Plaza, corresponding to the Xinjiekou commercial pedestrian area. The representative point of interest with lower average arrival visits was Sun Ye Village, corresponding to Longzang Avenue. This shows that passenger travel activities were closely related to the zoning functions carried by land use types.

A study by Jin and Xu [21] showed that the inflow and outflow on the key nodes of Nanjing's road network of different levels have an obvious hierarchical structure, and different points of interest play a certain role in the flow of tourists. The results of passenger travel activity at different points of interest in our study were consistent with the conclusions of the previous research.

In summary, urban roads contain information about the classification functions of expressways, arterial roads, secondary arterial roads, and branch roads and are affected by land use types. The pick-up and drop-off points with high passenger travel activity were concentrated near points of interest, forming hot spots. On the contrary, the pick-up and drop-off points with low passenger travel activity were concentrated near points of interest, and cold spots were formed. The hot spots of outbound visits were scattered on urban residential land, and the hot spots of arrival visits were concentrated on commercial and financial land.

5. Comparison and Discussion

5.1. Comparison. The GN algorithm [16] includes a splitting algorithm that uses the number of shortest paths passing through each edge in the network as a measurement index, and gradually deletes edges that do not belong to any community. Newman's fast algorithm [17] uses a cohesive algorithm, starting with each node occupying a community and continuously merging in the direction that maximizes the increase in modularity. Compared with the GN algorithm and the Newman fast algorithm, we use Bayes' rule to set the weight of the edge betweenness of the network, and the heap data structure to calculate the modularity; we also reduce the complexity of the algorithm and use the standard deviation ellipse to make the detected community structure clearer. For a complex network with n nodes and m connecting edges, the comparison results of the GN algorithm, Newman fast algorithm, and the algorithm in this paper are listed in Table 3.

Theoretically, if there are n communities, an $n \times n$ symmetric matrix F can be defined. The trace of the matrix (the sum of the diagonal elements of the matrix) is $Tr(F) = \sum_i f_{ii}$, which means the ratio of all edges connecting the nodes within the community to the total number of edges in the network. $Tr(F)$ value is in the range of $[0, 1]$. It is used to calculate modularity, and to a certain extent also characterizes the complexity of the network structure.

Table 2: Activities at different points of interests.

Land use types		Interests	Total	Value of departure	Value of arrival
Commercial	Retail land	Shopping malls, supermarkets, etc.	4381	0.722	0.578
	Dining land	Hotels, restaurants, etc.	5210	0.700	0.568
	Financial land	Office buildings, financial centers, etc.	462	0.276	0.648
	Other land	Banks, business halls, etc.	2284	0.738	0.590
Residential	Residential land	Apartments, villas, etc.	2161	0.790	0.288
Public	Agency land	Government agencies, etc.	810	0.674	0.574
	Education land	Schools, institutes, etc.	1376	0.614	0.508
	Medical land	Hospitals, pharmacies, etc.	1881	0.660	0.540
Traffic	Green land	Parks, gardens, etc.	74	0.554	0.450
	Street land	Parking lot, transportation station, etc.	324	0.700	0.576
	Highway land	Toll station, bus station, etc.	7206	0.308	0.444

Table 3: Method comparison.

Characteristic	GN algorithm	Newman fast algorithm	Algorithm of this paper
Algorithm complexity	$O(nm^2)$	$O(n^2)$	$O(m\log^2 n)$
Number of communities	Unknowable	Knowable	Knowable
Community structure	No overlap	Overlap	Overlap

When the network structure is abnormally chaotic, there are fewer edges connecting nodes within the community, and the value of $Tr(F)$ is minute. When the network structure is abnormally single, there are excessive number of edges connecting the nodes within the community, and the value of $Tr(F)$ is extremely large. When $Tr(F)$ value is in the range of $[0.4, 0.6]$, it can be assumed that the network structure is normal and that the value is not an abnormal value.

Therefore, another way of expressing modularity is $M = Tr(F) - F^2$, and F^2 is the modulus of matrix F^2. We compare the accuracy of community detection models using the GN algorithm, Newman fast algorithm, and the algorithm in this paper, as shown in Figure 9.

The community detection algorithm centered on the hierarchical structure is divided into split and aggregation types. The GN algorithm belongs to the split type, and the Newman fast algorithm and the algorithm proposed in this article belong to the aggregation type.

The GN algorithm gradually deletes edges that do not belong to any community (i.e., the edges connected between communities) according to the degree to which the edges do not belong to the community, until all edges are deleted. Because the edge betweenness of each connected edge needs to be recalculated every time an edge is removed, for complex network structures, the algorithm can be easily implemented by splitting it across more independent communities.

Newman's fast algorithm starts with each node occupying a community and continues to merge communities in the direction that maximizes the increase in modularity until the entire network merges into one community. Because the modularity needs to be increased every time the communities connected by edges are merged, when the network structure is simple, the execution of this algorithm will easily lead to the incorrect division of nodes.

The algorithm proposed in this paper introduces Bayes' rule and takes the amount of information as the increment of modularity, without calculating the adjacency matrix to ensure the increment of modularity. Therefore, when unconnected communities are merged, the degree of modularity remains unchanged; thus, the communities that are connected by edges and the corresponding internal nodes can be divided more accurately.

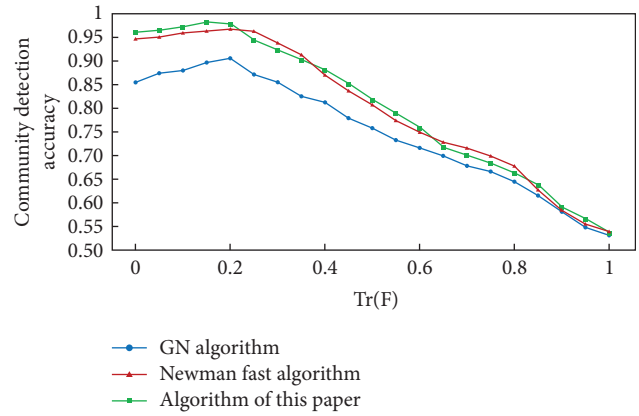


Figure 9: Algorithm performance comparison chart.

As shown in Figure 9, the abscissa is $Tr(F)$ and the ordinate represents the accuracy of community detection. The circle is the GN algorithm, the triangle is the Newman fast algorithm, and the square is the algorithm used in this study. It can be clearly observed from the figure that the accuracy of the algorithm in this study is significantly higher than that of the GN algorithm. Compared with the Newman fast algorithm, when $Tr(F)$ is $[0, 0.2]$, $[0.4, 0.6]$, and $[0.8, 1]$, the algorithm used in this study has higher accuracy. Therefore, according to Figure 9 and Table 3, the accuracy of the algorithm in this study is equivalent to that of the Newman fast algorithm, but the running time is faster, and thus the performance is better.

This demonstrates that when the network structure is abnormally single or chaotic, the community detection model using the algorithm proposed in this study can discover more complex community structures and has better interpretability for community detection results.

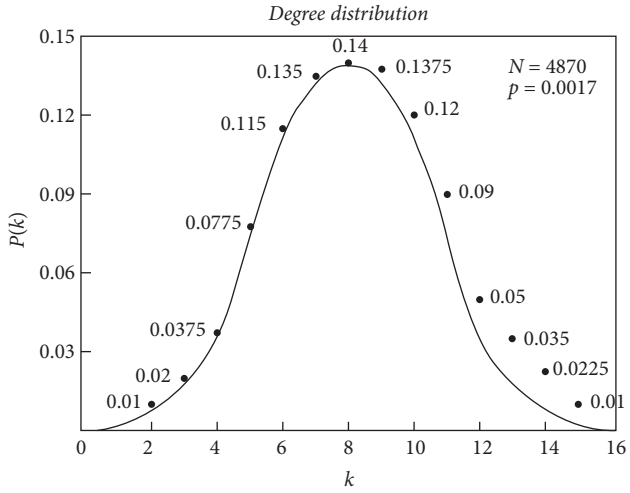


Figure 10: Degree distribution of the model.

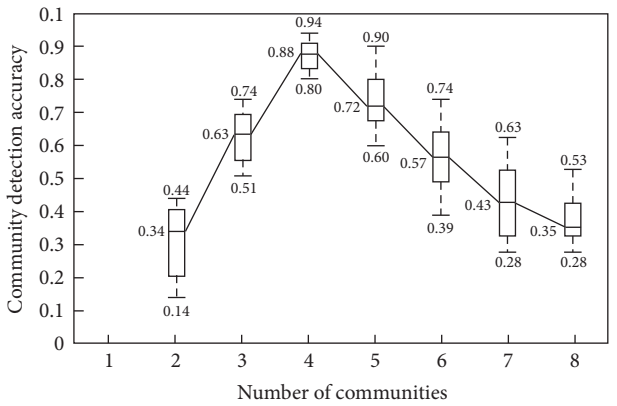


Figure 11: Box plot of model parameter sensitivity analysis.

5.2. Discussion. In order to explore the parameter sensitivity of the community detection model in this study, taking the period of 8 : 00–9:00 as an example, the degree distribution of the random network was calculated, as shown in Figure 10, and 21 simulation experiments were performed to compare the accuracy, as shown in Figure 11.

As shown in Figure 10, the abscissa represents the degree of the node, the ordinate represents the degree distribution probability, N represents the number of nodes, and p represents the connection probability of the nodes. It can be clearly observed from Figure 10 that the average degree of the node is eight, and the degree distribution follows the Poisson distribution.

As shown in Figure 11, the abscissa represents the number of communities, and the ordinate represents the accuracy of community detection. It can be clearly observed from Figure 11 that when the number of communities detected is four, the accuracy reaches its peak.

In summary, in a random network composed of 4870 key road network nodes, different communities are delineated based on taxi passengers' pick-up and drop-off points within a representative period, and the detected travel hot spots have reasonable spatial distribution characteristics.

Qin et al. [18] analyzed the intensity of node access degrees and edge weights based on the network interaction of urban hot spots, without considering the potential impact of land use on urban residents' travel decisions. This study combined the hierarchical road network and point of interest data to explore hot spots from the perspective of individual taxi passengers interacting with the community, which helped to explore the formation process of urban hot spots.

6. Conclusions

This study extracted the passenger pick-up and drop-off points from taxi movement trajectory data, constructed a taxi passenger travel activity index based on community detection, and extracted the hot spots of taxi passenger travel in the main urban area of Nanjing. The following three conclusions were drawn:

(1) The travel activities of taxi passengers showed a time distribution pattern of more daytime, less nighttime, and frequent morning and evening peak hours. Affected by the purpose of travel, the degree of community modularity and anisotropy rate of taxi passengers' pick-up and drop-off points were positively correlated during morning and evening peak hours and negatively correlated during other periods.

(2) The travel activities of taxi passengers presented a spatial distribution pattern, in which the central area of the city was concentrated and the outer areas were scattered. Affected by the structure of the community, passenger travel activities showed a consistent behavioral pattern within the community and had obvious spatial gathering characteristics. Furthermore, there was a significant interaction between different communities.

(3) The hot spots for taxi passengers' travel were scattered on urban residential land and concentrated on commercial and financial land. Affected by land use, passenger travel activity indicators were closely related to road grades and types of points of interest. Passenger travel hot spots were clustered as activity levels increased and dispersed as activity levels decreased.

Subsequent research needs to consider more sources of data, such as combining rental car trajectory data with bus trajectory data, analyzing the travel preferences of different groups of people, and further exploring the temporal and spatial patterns of urban traffic congestion by urban residents using the impact of different travel modes.

Data Availability

All data, models, and code that support the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors' Contributions

Yuyu Sheng, Shuoben Bi, and Wenwu He conceived and designed the experiments; Yuyu Sheng and Ruizhang Xu performed the experiments; Yuyu Sheng, Shuoben Bi, and Ruizhuang Xu wrote the Chinese paper; Shuoben Bi and Jingjing Fan translated the paper.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grants: 41971340 and 41271410).

References

[1] Q. Q. Li and D. R. Li, "Big data GIS," *Geomatics and Information Science of Wuhan University*, vol. 39, no. 06, pp. 641–644, 2014.

[2] Y. Zheng, "Introduction to urban computing," vol. 40, no. 01, , pp. 1–13, Geomatics and Information Science of Wuhan University, 2015.

[3] Y. Liu, X. Liu, S. Gao et al., "Social sensing: a new approach to understanding our socioeconomic environments," *Annals of the Association of American Geographers*, vol. 105, no. 3, pp. 512–530, 2015.

[4] J. M. Harvey and J. W. Han, *Geographic Data Mining and Knowledge Discovery*, CRC Press, London, UK, 2009.

[5] C. X. Cheng, P. J. Shi, C. Q. Song et al., "Geographic big-data: anew opportunity for geography complexity study," *Acta Geographica Sinica*, vol. 73, no. 08, pp. 1397–1406, 2018.

[6] M. Veloso, S. Phithakkitnukoon, and C. Bento, "Urban mobility study using taxi traces," in *Proceedings of the 2011 International Workshop on Trajectory Data Mining and Analysis – TDMA'11*, pp. 23–30, ACM, New York, NY, USA, 2011.

[7] R. Ahas, A. Aasa, Y. Yuan et al., "Everyday space-time geographies: using mobile phone-based sensor data to monitor urban activity in Harbin, Paris, and Tallinn," *International Journal of Geographical Information Science*, vol. 29, no. 11, pp. 2017–2039, 2015.

[8] R. W. Scholz and Y. Lu, "Detection of dynamic activity patterns at a collective level from large-volume trajectory data," *International Journal of Geographical Information Science*, vol. 28, no. 5, pp. 946–963, 2014.

[9] J. Cui, F. Liu, D. Janssens, G. Wets, and M. Cools, "Detecting urban road network accessibility problems using taxi GPS data," *Journal of Transport Geography*, vol. 51, no. 12, pp. 147–157, 2016.

[10] C. Zhong, S. M. Arisona, X. Huang, B. Michael, and G. Schmitt, "Detecting the dynamics of urban structure through spatial network analysis," *International Journal of Geographical Information Science*, vol. 28, no. 11, pp. 2178–2199, 2014.

[11] Q. Huang, Y. Yang, Z. Yuan et al., "The temporal geographically-explicit network of public transport in Changchun City, Northeast China," *Scientific Data*, vol. 6, no. 190026, pp. 1–10, 2019.

[12] F. Guo, D. Zhang, Y. Dong et al., "Urban link travel speed dataset from a megacity road network," *Scientific Data*, vol. 6, no. 61, pp. 1–8, 2019.

[13] H. Hamedmoghadam, M. Ramezani, and M. Saberi, "Revealing latent characteristics of mobility networks with coarse-graining," *Scientific Reports*, vol. 9, no. 1, pp. 1–10, 2019.

[14] Yu Zheng, Y. Liu, J. Yuan et al., "Urban computing with taxicabs," in *Proceedings of the 13th International Conference on Ubiquitous Computing - UbiComp'11*, pp. 89–98, ACM, Beijing, China, 2011.

[15] R. L. Wu, X. Y. Zhu, and W. Guo, "Spatiotemporal distribution patterns of urban road traffic accidents," *Geomatics & Spatial Information Technology*, vol. 41, no. 07, pp. 103–106, 2018.

[16] M. Gong and M. Girvan, "Finding and evaluating community structure in networks," *Physical Review E Statistical Nonlinear & Soft Matter Physics*, vol. 69, no. 2, pp. 26113–26120, 2004.

[17] M. E. J. Newman, "Fast algorithm for detecting community structure in networks," *Physical Review E*, vol. 69, no. 6, Article ID 066133, 2004.

[18] K. Qin, Q. Zhou, Y. Q. Xu et al., "Spatial interaction network analysis of urban traffic hotspots," *Progress in Geography*, vol. 36, no. 9, pp. 1149–1157, 2017.

[19] J. Xu, S. B. Bi, Y. Zhang et al., "PSO-SVM model based analysis on traffic flow of road intersections in Nanjing," *Modern Electronics Technique*, vol. 39, no. 17, pp. 128–213, 2016.

[20] X. X. Yang, "On the Model and Congestion Performance of Nanjing Urban Public Traffic Systems," *Nanjing University of Posts and Telecommunications*, 2018.

[21] C. Jin and J. Xu, "Study on the tourists flow among external transport nodes and hotels in Nanjing," *Human Geography*, vol. 31, no. 05, pp. 55–62, 2016.

[22] L. Zhao, M. Deng, D. L. Peng et al., "Structural property analysis of urban Street networks based on complex network theory," *Geography and Geo-Information Science*, vol. 26, no. 05, pp. 11–15, 2010.

[23] G. N. Wang, *Spatial-Temporal Data Mining Based on GPS Trajectory and Geo-Tagged Photo Trajectory*, Central South University, Changsha, China, 2013.

[24] S. Z. Guo and Z. M. Lu, *The Basic Theory of Complex Network*, Science Press, Beijing, China, 2012.

[25] Y. Batty, *Study on Human Activity Space Patterns and Network Spatial Temporal Characteristics in Urban Cities Using Taxi Trajectory Data*, Wuhan University, Wuhan, China, 2016.

[26] Y. L. An, Z. F. Huang, W. D. Chen et al., "Spatial evolution of county economy in Anhui Province during 2001-2010," *Progress in Geography*, vol. 32, no. 05, pp. 831–839, 2013.