Bicycle Sharing Demand Analysis and Operation Community Division based on Riding Spatial-Temporal Data -Taking Beijing as an Example

Muran Wu*

School of Architecture and Urban Planning, Shenzhen University, Shenzhen, China

* Corresponding Author Email: 2019022192@email.szu.edu.cn

Abstract. Under the background of vigorously building sustainable cities, shared bicycles have developed rapidly, and some operation and management problems have gradually become prominent. In recent years, the research based on spatial-temporal big data of shared bicycles mostly focuses on the analysis of hotspots in time and space, and the current management of shared bicycle is mostly divided into operating communities by administrative divisions. The construction of bicycle operation management is not carried out in combination with the characteristics of spatial-temporal data and actual needs. Based on the shared bicycle order data in Beijing, this paper analyzes the basic spatial and temporal characteristics of shared bicycles, and uses geographically weighted regression model to analyze the impact of various types of built environments on the demand distribution for shared bicycles. Finally using Fast Unfolding algorithm to identify communities with close demand for shared bicycles. The study finds that the overall utilization rate of bicycle resources is currently low, and various types of POIs have different ranges and degrees of impact on parking demand, as well as the division of bicycle communities does not fully coincide with administrative divisions. This study helps to understand the overall demand and impact characteristics of shared bicycles, provides a basis for the development of differentiated strategies for the delivery, management, and scheduling of shared bicycles, and helps urban managers to optimize the design of the shared bicycle system.

Keywords: Shared bicycle, GWR, community detection, differentiated operation.

1. Introduction

As a green and low-carbon travel mode, shared bicycles have developed rapidly under the background of the construction of sustainable city and the growth of chronic traffic demand with their characteristics of pile-free and mobile payment [1]. In recent years, with the gradual increase in the scale of shared bicycles and the continuous expansion of the scope of operation, the problem of shared bicycle transportation and management has become more and more prominent, such as the imperfect planning of parking points, the imbalance between the actual distribution of parking space and the demand, and the destruction on the public environment. Conventional management methods are difficult to ensure good operational efficiency. How to formulate more scientific and efficient operation strategies has become a hot topic in urban traffic research.

With the rise of spatial-temporal big data, the increasingly refined analysis of residents' travel characteristics also provides a reference for the scientific planning and efficient management of urban space. Fu et al. confirmed that bike-sharing travel has a high degree of temporal regularity by analyzing a week's Mobike data in Shanghai [2]. Cheng et al. found that the traffic flow of shared bicycles near different land uses was mainly concentrated in the business district, business district and near universities in descending order [3]. Guo conducted a visual comparative analysis of the OD intensity of different types of public bicycle stations under different travel purposes in Nanjing, and studied the path preference in combination with the shortest use path [4]. Lyu et al. analyzed the cycling traffic hotspots, traffic corridors and parking supply and demand contradictions of shared bicycles in Shanghai [5]. In generally, these studies focus more on the hot spots in the time and space of bicycle data, and how to combine the space-time big data of residents' behavior to carry out bicycle operation planning and construction is still not much explored.

At present, the operation and management of shared bicycles are mostly divided into operation areas based on administrative regions. However, such a division has a poor correlation with the travel characteristics and actual demand of bicycle users, which increases the scheduling cost and affects the operational effect. Many scholars have conducted comparative studies on the travel volume of shared bicycles in different areas of the city, and found that there are obvious differences in the travel volume of different regions of the city [6], and there are cluster characteristics [7]. For example, Patrick Vogel used cluster analysis to divide the pile-sharing bike sites into five types by the characteristics of taking and returning cars, so as to analyze the riding behavior of customers in the region [8]. Etienne Come et al. divided the bike-sharing sites by constructing a statistical model of the built environment and cycling behavior near the bike-sharing sites, and revealed the different patterns of cycling behavior [9]. By extracting bicycle travel data and aggregation areas and analyzing their demand and aggregation characteristics, it helps to provide a scientific basis and planning guidelines for rationally dividing urban bicycle operation areas [10].

In summary, in this context this paper takes Beijing as an example. Based on the analysis of the basic usage characteristics of shared bicycles, according to the actual travel demand of shared bicycles, the geographically weighted regression model is used to analyze the impact of various built environments on the demand distribution of shared bicycles. The Fast-Unfolding algorithm is used to identify communities with close demand for shared bicycles. These will provide a basis for optimizing the layout of shared bicycle systems and formulating differentiated application strategies.

2. Methods

2.1. Data Source and Description

2.1.1 Shared bicycle order data

Shared bicycle order data is provided by the 2017 Mobike Cup Algorithm Challenge, and about 1.2 million data records from the dataset of four working days from May 15 to 16, 2017 and May 18 to 19, 2017 in Beijing are selected for this paper. The original dataset fields are orderid, userid, bikeid, biketype, starttime, geohashed_start_loc, and geohashed_end_loc.

2.1.2 Point of interest data

This paper collects Point of Interest data (POI) of Beijing in 2017. POI data mainly reflects the land use characteristics of different regions. Combined with the results of previous related studies, this paper selects seven kinds of POI as the potential dominant factors, which are Restaurant, Company, Shopping, Traffic, Education, Entertain, Residence [11, 12]. The specific attributes of each class of POI are shown in Table 1.

Table 1. Specific attributes for each class of POI

Type	Attribute	Count
Restaurant	Includes various types of catering services, etc.	71008
Company	Including companies, corporations, office buildings, etc.	132613
Shopping	Including markets, shopping centers, shopping streets, etc.	7550
Traffic	Including parking lots, bus stops, subway stations, etc.	72247
Education	Including research organizations, schools, libraries, art museums, etc.	58500
Entertain	Including stadiums, amusement parks, movie theaters, etc.	31279
Residence	Including residence community, dormitories, and business residences	31455

2.1.3 Administrative division data

This paper collects Beijing 's administrative division data, including municipal administrative division data, county administrative division data, and township administrative division data.

2.2. Introduction to the Methods

2.2.1 Geographical weighted regression model

Geographical Weighted Regression (GWR) is a local linear regression method based on the modeling of spatial variation relationship, which is evolved from the traditional Linear Regression (LR). It generates a regression model describing the local relationship in each part of the study area, which can well explain the local spatial relationship and spatial heterogeneity of the variables.

Multiscale Geographically Weighted Regression (MGWR), on the other hand, considers different spatial relationship weighting models for different independent variables based on GMR. The method infers the respective bandwidth for each independent variable and generate a bandwidth vector for multiple independent variables. The spatial extent of influence of the respective variables is inferred from the magnitude of the bandwidth obtained by the model. As shown in Fig. 1

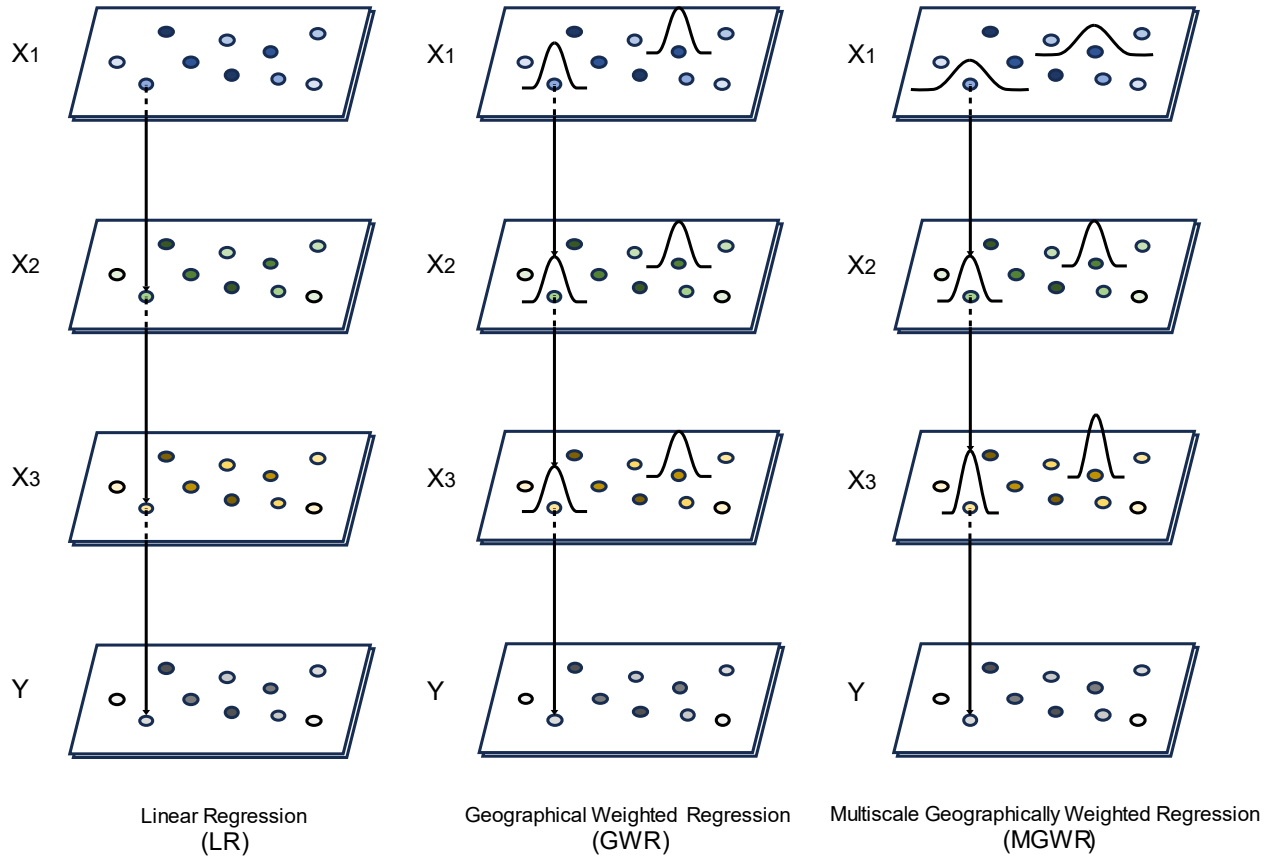


Fig. 1 Comparison of three regression methods

2.2.2 Community detection-fast-unfolding algorithm

Community detection is an important branch in the discipline of network analysis. With the help of community detection algorithms can decompose the network into different communities, and then study the spatial structure of the relationship.

In social network, users are equivalent to each point, and the structure of the whole network is formed by the relationship between users through mutual attention. In such a network, some users are more closely connected to each other, while others are more sparsely connected to each other. The more closely connected part of the network can be regarded as a community, where the nodes within it are more tightly connected to each other, while the connection between two communities is relatively sparse.

Modularity is an important standard to measure the quality of community division. The larger the value of modularity of the divided network, the better the effect of community division. Fast Unfolding algorithm is an iterative algorithm based on modularity to divide communities. The basic

idea is that the nodes in the network try to traverse the community labels of all their neighbors and select the community label that maximizes the increment of modularity. After maximizing the modularity, each community is viewed as a new node and repeated until the modularity no longer increases. As shown in Fig. 2

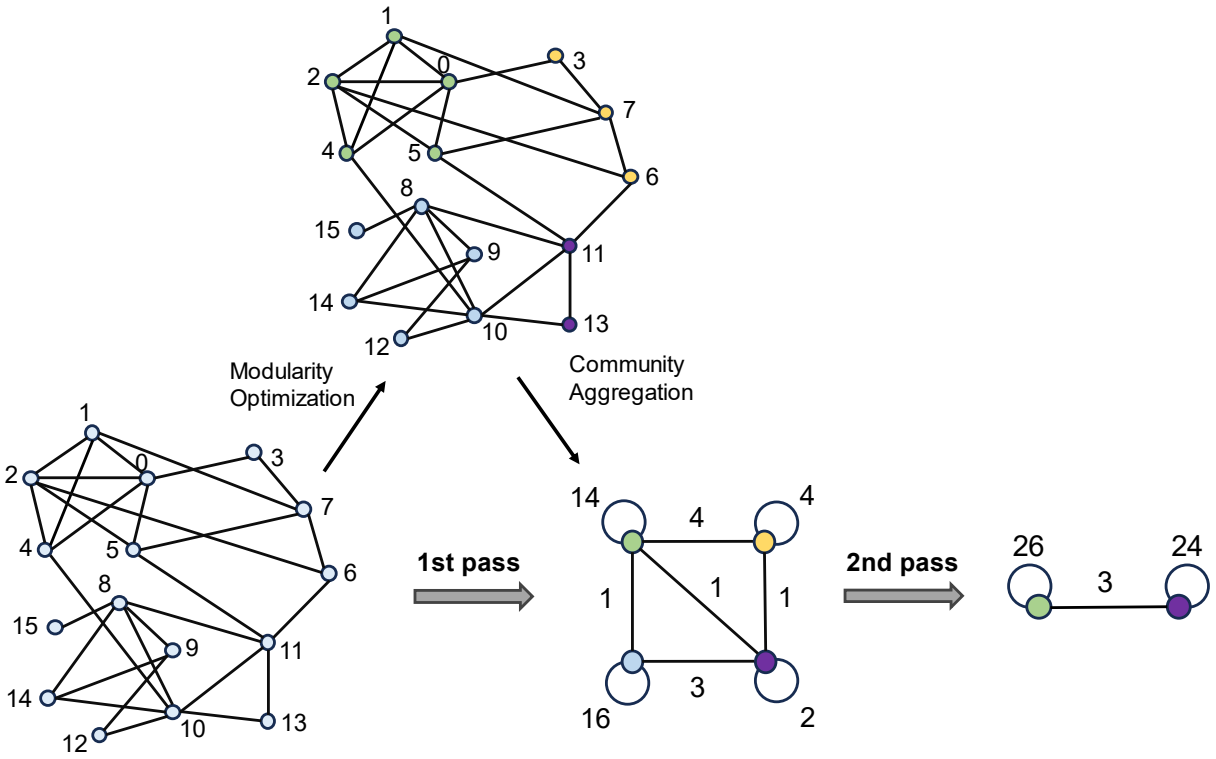


Fig. 2 Iterative process of Fast Unfolding algorithm

2.3. Data Preprocessing

The order data of shared bikes are preprocessed as follows:

Step 1. Filter out the orders for the required date range.

Step 2. Decode the start and end geohash coding system into latitude and longitude coordinates.

Step 3. Calculate the distance between the start and end locations.

Step 4. Cleaning the ride data and deleting data smaller than 100m or larger than 10km. The reason: the ride of smaller than 100m does not actually fulfill the user's travel demand, which may be caused by the user 's discovery of a single vehicle failure after unlocking the use of the lock and then locking. The ride of larger than 10km does not conform to the normal cycling distance range.

Step 5. Calculate the duration of the order according to the distance, assuming that the riding speed of the shared bicycle is 10km/h.

Step 6. Order the orders according to time.

Step 7. Filter out the required columns.

Examples of the raw data and preprocessed data of the bike shared order are shown in Table 2 and Table 3.

Table 2. Example of raw data for Shared Bike Orders

orderid	userid	bikeid	biketype	starttime	geohashed_start_loc	geohashed_end_loc
1893973	451147	210617	2	2017-05-14 22:16:50	wx4snhx	wx4snhj
4657992	1061133	465394	1	2017-05-14 22:16:52	wx4dr59	wx4dquz
2965085	546189	310572	1	2017-05-14 22:16:51	wx4d5r5	wx4fu5n

Table 3. Example of pre-processed data for Shared Bike Orders

UserID	BikeID	Stime	Slon	Slat	Elon	Elat	Distance	Duration
520822	196645	2017-05-15 00:00:05	116.347	39.719	116.351	39.718	466.3	167.9
729803	39606	2017-05-15 00:00:05	116.412	39.948	116.409	39.944	515.5	185.2
517344	462784	2017-05-15 00:00:08	116.332	39.893	116.345	39.891	1097.8	395.2

2.4. Indicator Selection and Description

Table 4 shows the full name, description and data type of the variables after data preprocessing.

Table 4. Indicator Selection and Description

Full Name	Description	Data type
UserID	User ID	Categorical data
BikeID	Shared bicycle ID	Categorical data
Stime	Order start date and time	Time data
Slon	Order starting longitude (WGS84 coordinates)	Geographic location data
Slat	Order starting latitude (WGS84 coordinates)	Geographic location data
Elon	Order ending longitude (WGS84 coordinates)	Geographic location data
Elat	Order ending latitude (WGS84 coordinates)	Geographic location data
Distance	Riding distance (unit/m)	Numeric data
Duration	Duration distance (unit/second)	Numeric data

3. Results and Discussion

3.1. Basic Usage Characteristics

The analysis of the basic use characteristics of shared bicycles can be divided into two aspects : time and space. The following discusses the basic usage characteristics of shared bicycles from the number of rides per bicycle per day, the duration of single ride, the length of use per bicycle per day, the single riding distance, distribution of the parking demand and distribution of the riding OD.

3.1.1 Time characteristics

Counting the number of single-day rides and the length of use of each bicycle is helpful to understand the overall situation of the use of bicycle resources. From the analysis of the usage time characteristics, it can be seen that most of the shared bicycles are used only once a day, and the vast majority of them are actually used a total of 1 hour per day. The peak time is about 15 minutes, and most of the time is in the parking state. The duration of single use is within 20min, with a peak at around 8min. As shown in Fig. 3, 4, and 5.

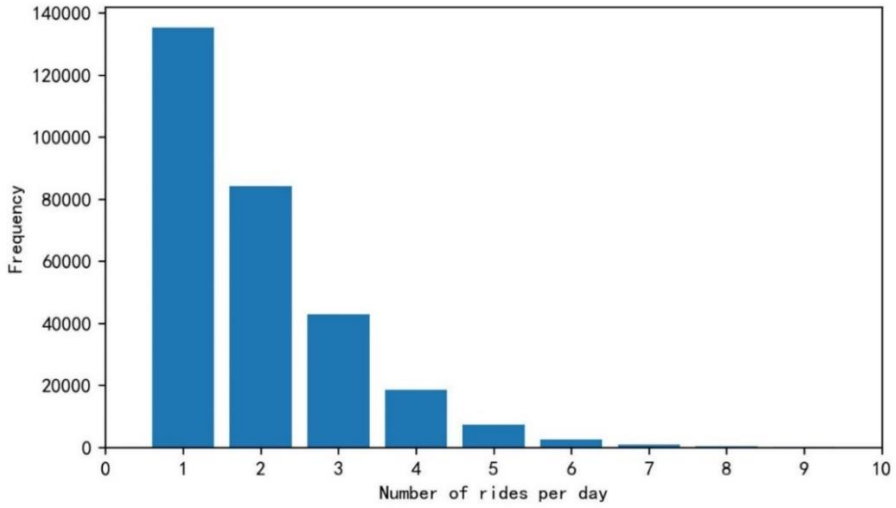


Fig. 3 Frequency distribution of the number of rides per bicycle per day

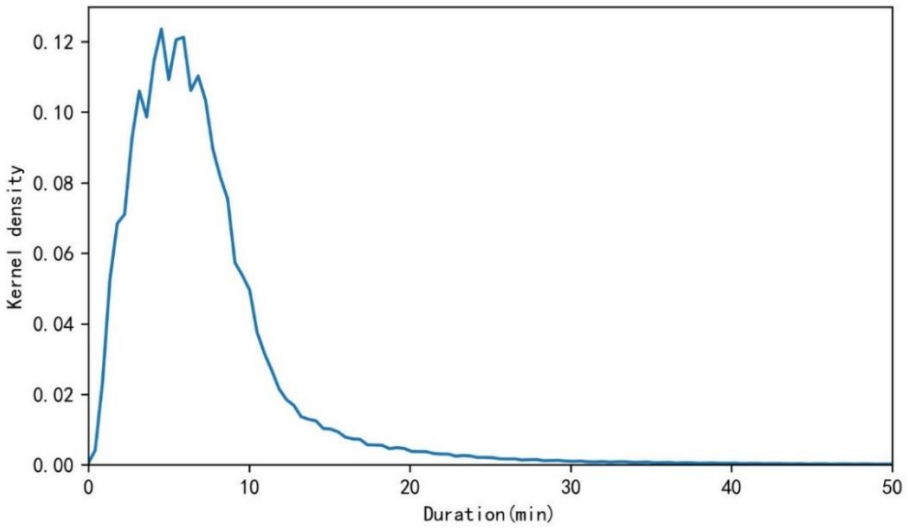


Fig. 4 Kernel density distribution of the duration of single ride

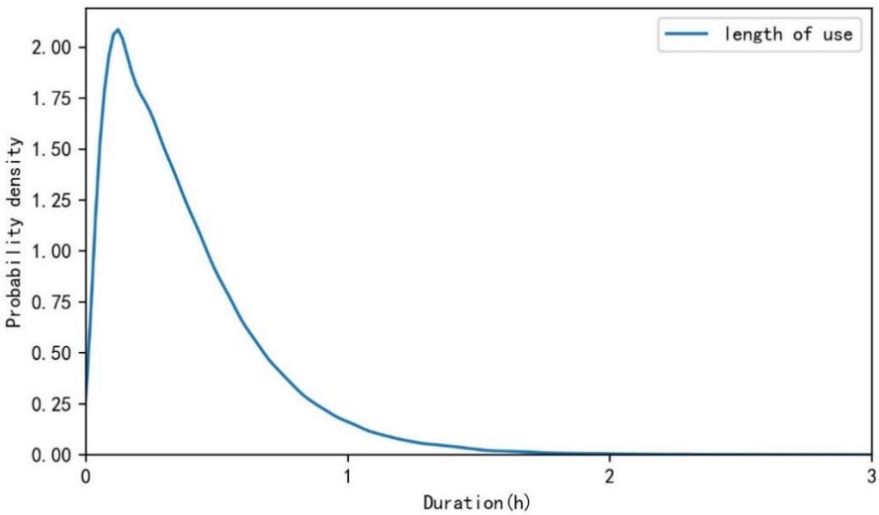


Fig. 5 Kernel density distribution of the length of use per bicycle per day

3.1.2 Spatial characteristics

Counting the spatial characteristics of each bicycle is helpful to understand the demand distribution of bicycle resources. From the kernel density distribution of the single riding distance, it can be seen

that shared bicycles are generally short-distance cycling. The vast majority of riding distance is within 2km, mainly distributed in 200-1000m, and the peak section is in 400-900m. As shown in Fig. 6.

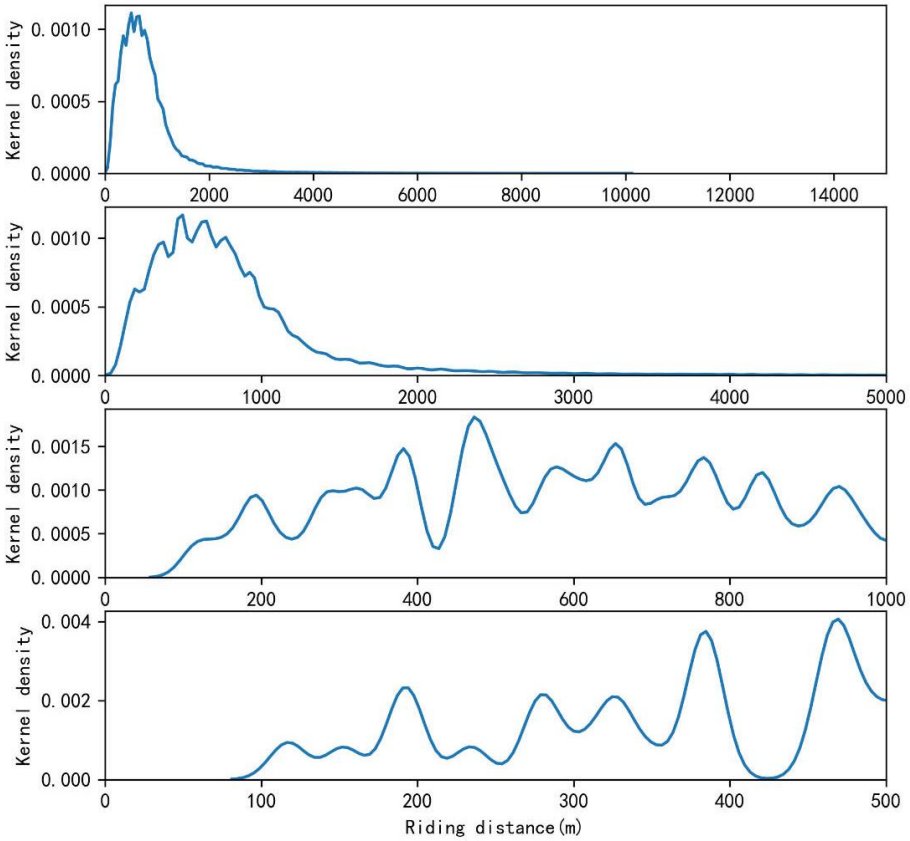


Fig. 6 Kernel density distribution of the single riding distance

By aggregating the demand for parking in each administrative division of Beijing, it can be seen that the demand for parking is high in the central area of Beijing (Second Ring Road to Fourth Ring Road), while the demand for parking is low in the suburbs (outside Fifth Ring Road). As shown in Fig. 7.

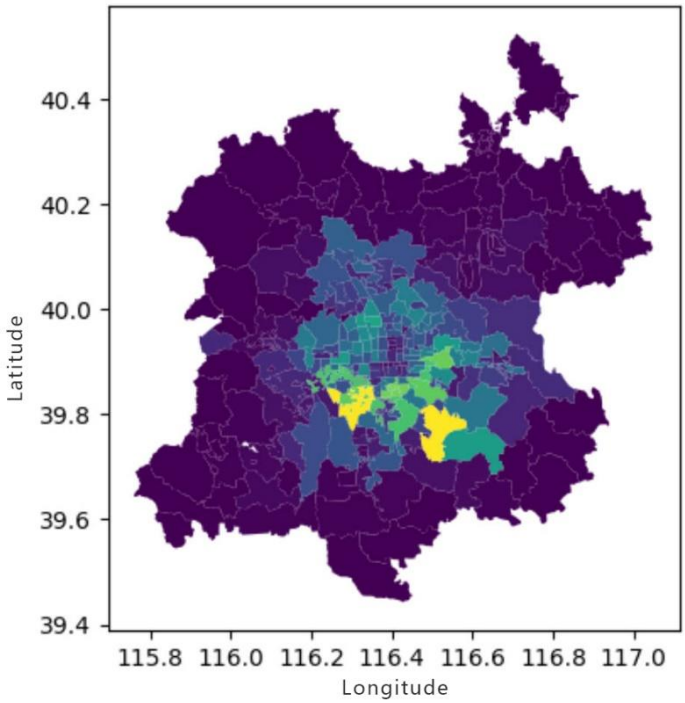


Fig. 7 Distribution of the parking demand

By identifying the riding OD information in the bicycle data and visualizing it, it can be seen that the distribution characteristics of the data count of OD are similar to the above parking demand distribution characteristics. In the city center area, the data count of OD is large and dense, and it is easy to form a continuous parking aggregation area. In the suburbs, the number is small and scattered, but it is more concentrated on the north-south central axis. In addition, it can be observed that the road direction in the center of Beijing is mostly south-north or west-east. As shown in Fig. 8.

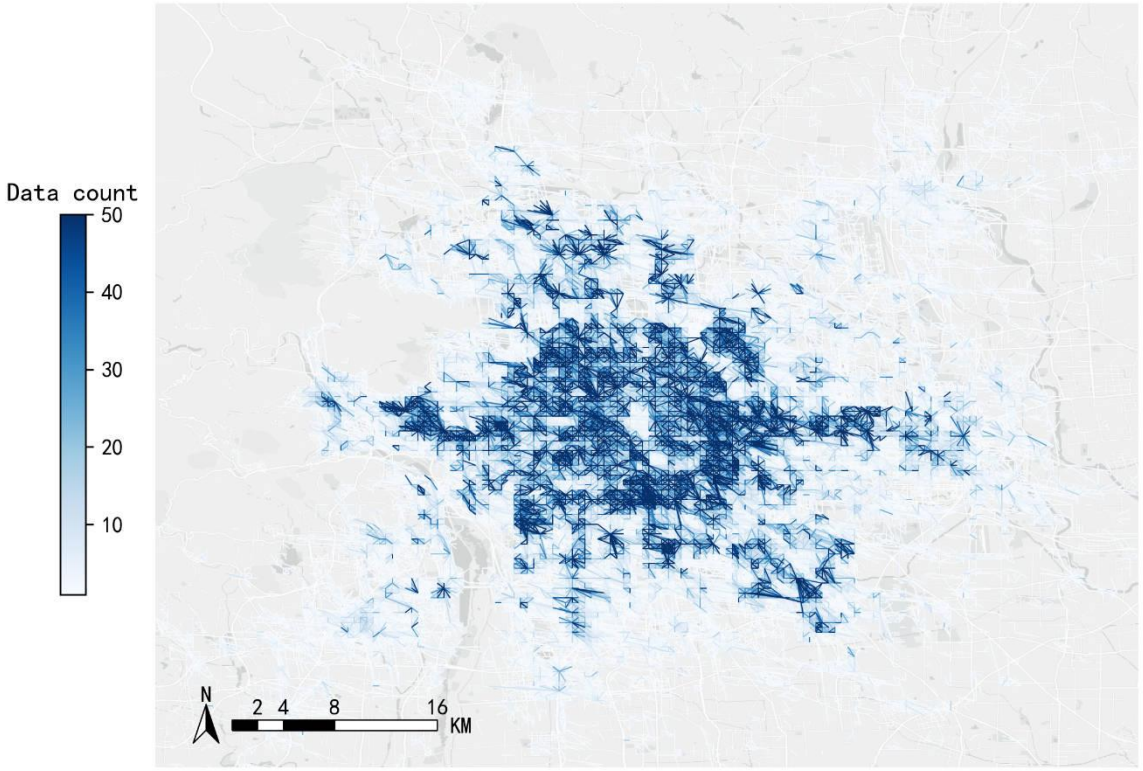


Fig. 8 Distribution of the riding OD

3.2. Results and Discussion of GWR Model

In this section, the parking demand of shared bicycles is taken as the dependent variable and the number of various types of POIs is taken as the independent variable. The data on May 15,2017 is selected and visualized using GWR and MGWR models respectively, and the analysis results of the two are compared. The coefficient is red when it is positive, indicating that this type of POI has a positive impact on parking demand. When the coefficient is negative, it is blue, which has a negative impact on parking demand. Intercept is understood as the demand level of shared bicycles without considering the impact of POI.

3.2.1 GWR

As shown in Fig. 9. From the intercepts, it can be seen as a whole that the parking demand for shared bikes is more in the city center and the south, and less in the north and some of the southern suburbs, which is similar to the characterized in Fig. 7 and Fig. 8 above.

The distribution of coefficients in the GWR fits for the various POIs varies significantly. Shopping brings an increase in parking demand only in the core area of the city center, while Entertain brings a significant increase in demand in the core area of the city center and the northeastern suburbs. Restaurant and Company, on the other hand, have more positive impacts on the periphery of the city. The impact of Education on parking demand has no obvious regularity, compared to Residence, which has a positive impact on parking demand in most areas of the city.

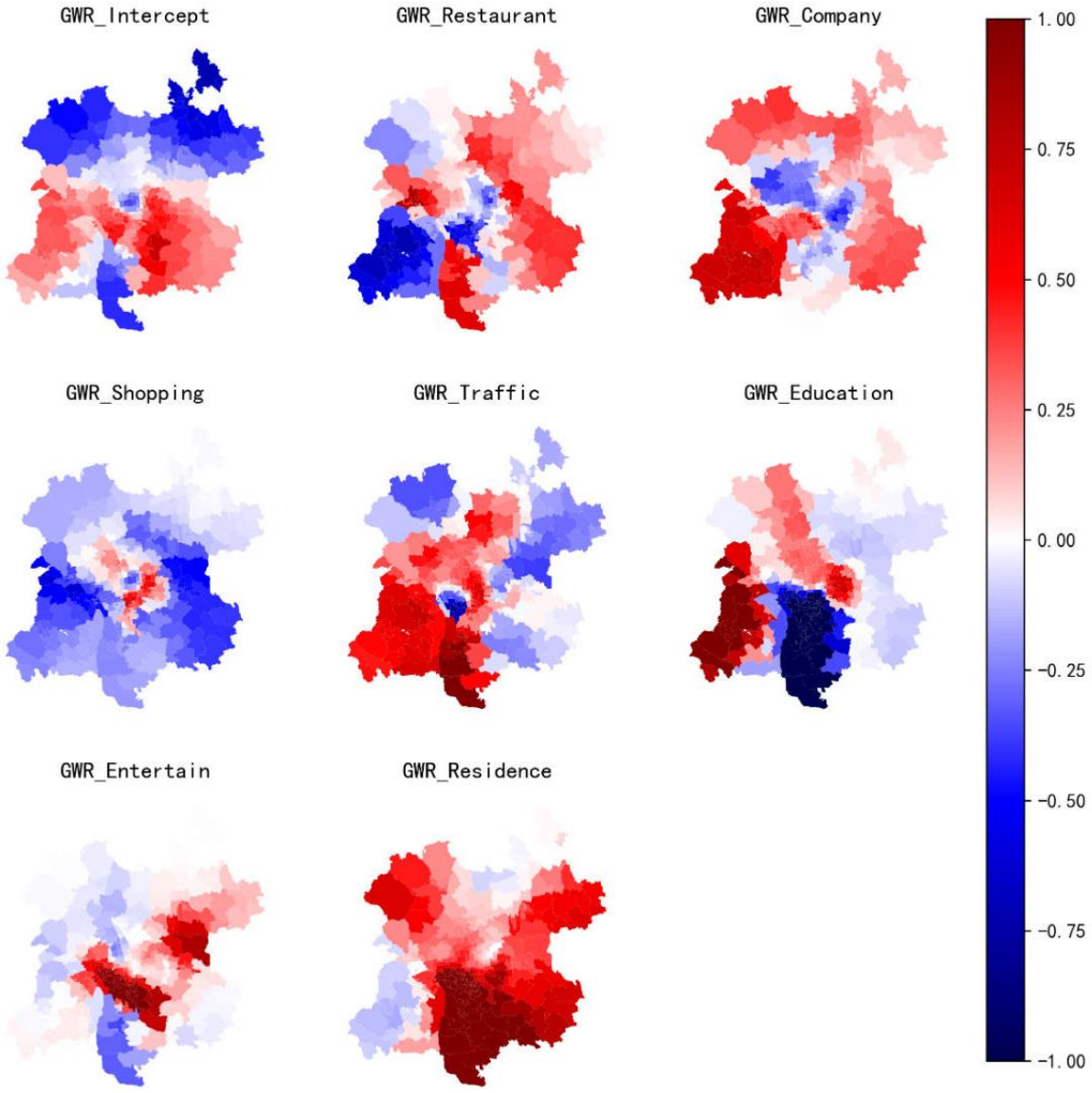


Fig. 9 Dependent Variable Coefficients for GWR

3.2.2 MGWR

As shown in Fig. 10, the difference between MGWR and GWR can be clearly seen. The bandwidth (BW) of the intercept is the smallest, and the coefficient values are more pronounced in different areas of the city, indicating that the MGWR model considers that the geographic location has a greater impact on the magnitude of parking demand levels, the impact of POI is smaller.

The Traffic and Residence coefficient values are high throughout the city and bandwidth is medium, indicating that an increase in the number of transit facilities and residences can significantly increase parking demand in a wide range of areas. The possible reason for this is that shared bikes can provide connection for transportation facilities such as subway stations and bus stations, and provide services for residents ' daily living. Education has a large bandwidth, which can increase parking demand in most areas of the city and decreases the demand for parking in parts of the northeast and south.

Restaurant and Company have large bandwidths and coefficient values close to zero, suggesting that these POI have little impact on the parking demand. Entertain and Shopping have large bandwidths. The coefficient value of the former is a small and positive value, which is close to our perception of bike-sharing services. However, the coefficient value of the latter is a small and negative value, indicating that there will be less demand for bicycle parking. The reason for this remains to be studied.

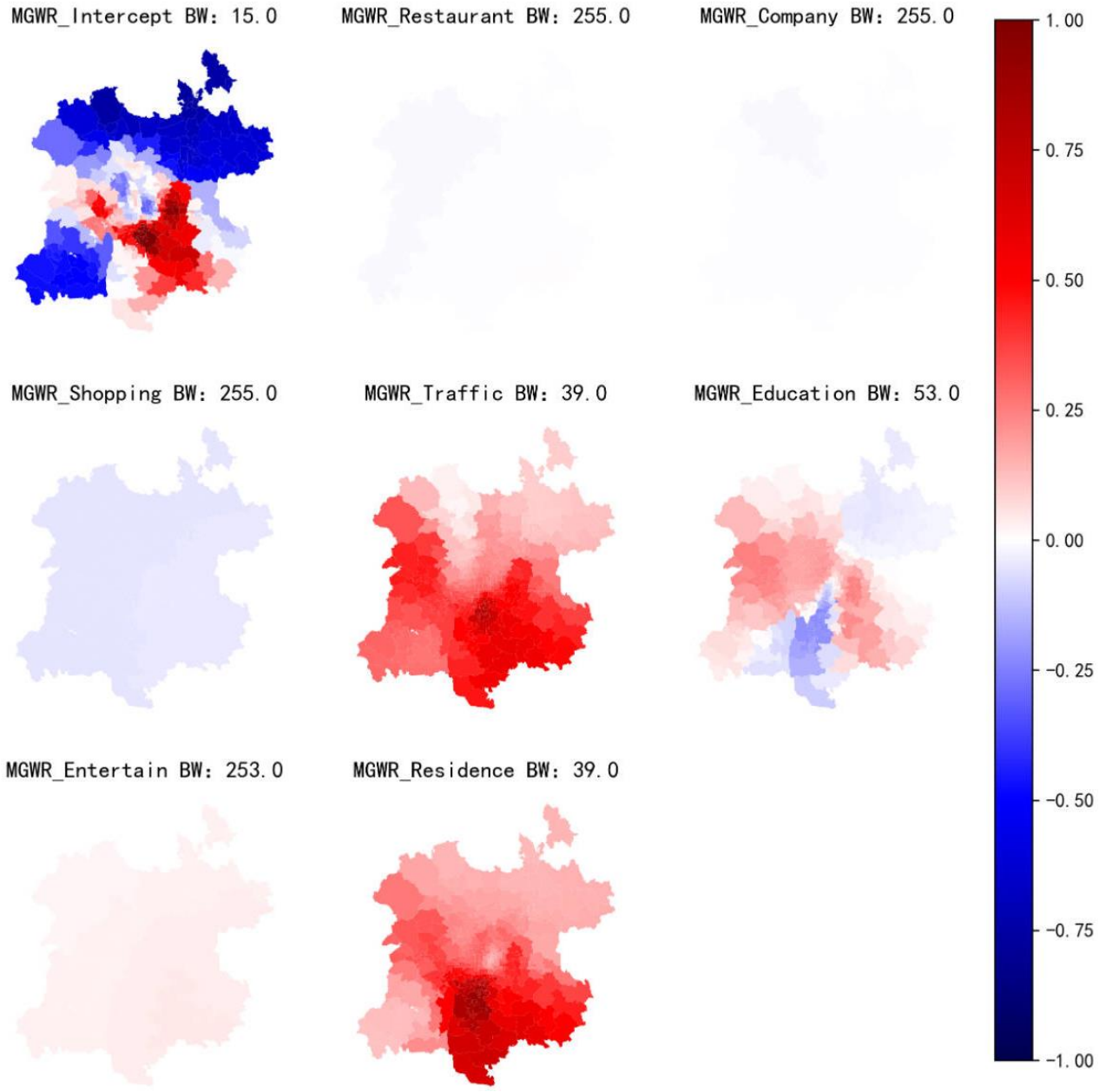


Fig. 10 Dependent Variable Coefficients for MGWR

3.3. Results of Community Detection-Fast Unfolding Algorithm

The idea of market-oriented zoning of shared bikes based on community detection is to divide the entire operation area into a number of closely linked operation areas based on the actual travel demand for shared bicycles, and to implement differentiated operation strategies in different operation areas. From the perspective of the use characteristics of shared bicycles, these operating areas have more strong bicycle demand trips, compared to fewer cross-area trips.

After only retaining the communities containing more than 10 grids, the preliminary visualization results are shown as Fig. 11. Different colors represent different groupings, and grids of the same color are mostly clustered together in geospatial proximity. This shows that the spatial connection of shared bicycles can bring geospatially proximate areas together in close proximity.

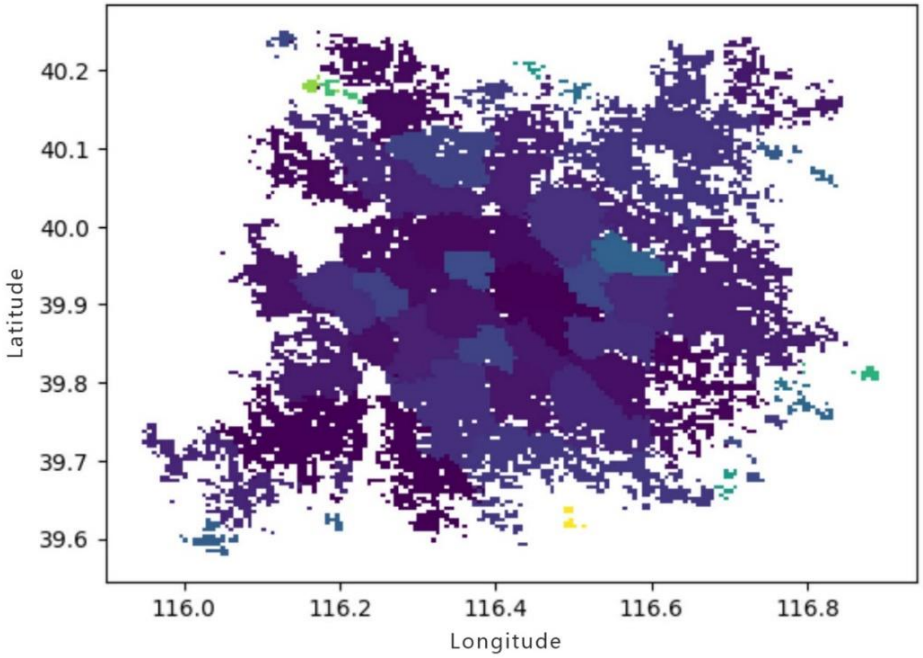


Fig. 11 Grid with community division information (preliminary visualization)

After eliminating the independently existing grids, the grids of the same group are merged into communities and different colors of the same brightness and saturation are set for each community. As shown in Fig. 12.

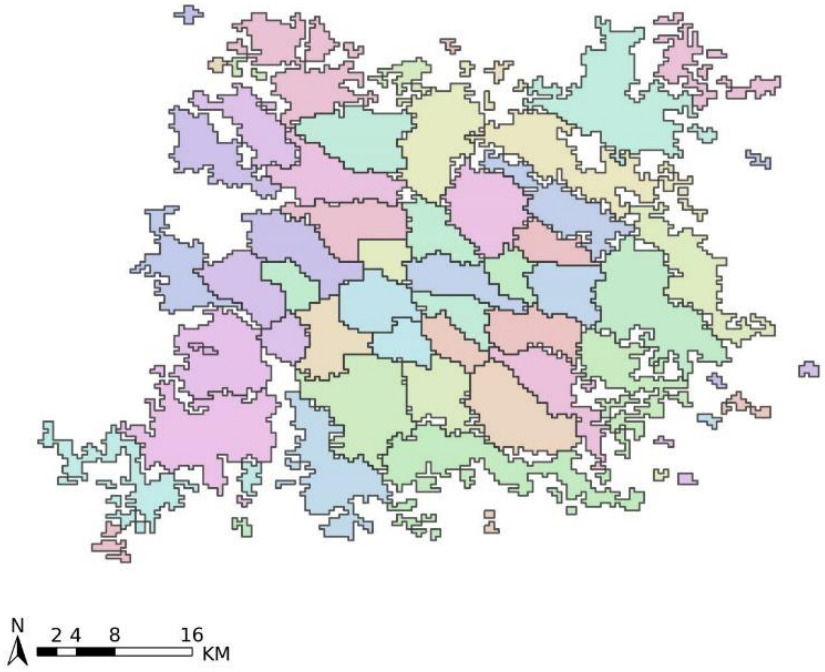


Fig. 12 Visualization of community detection results

The shared bicycle community boundary is compared with the county-level and township-level administrative boundary on the same map. Among them, red is the boundary of administrative regions, and black is the boundary of communities, as shown in Fig. 13 and Fig. 14. It can be observed that in some area community boundary coincide with administrative division boundary. In most areas of the city center, a community may span multiple administrative divisions. In suburban areas, a community may only occupy a portion of an administrative district; while in the outskirts of the city there may be no bike-sharing community.

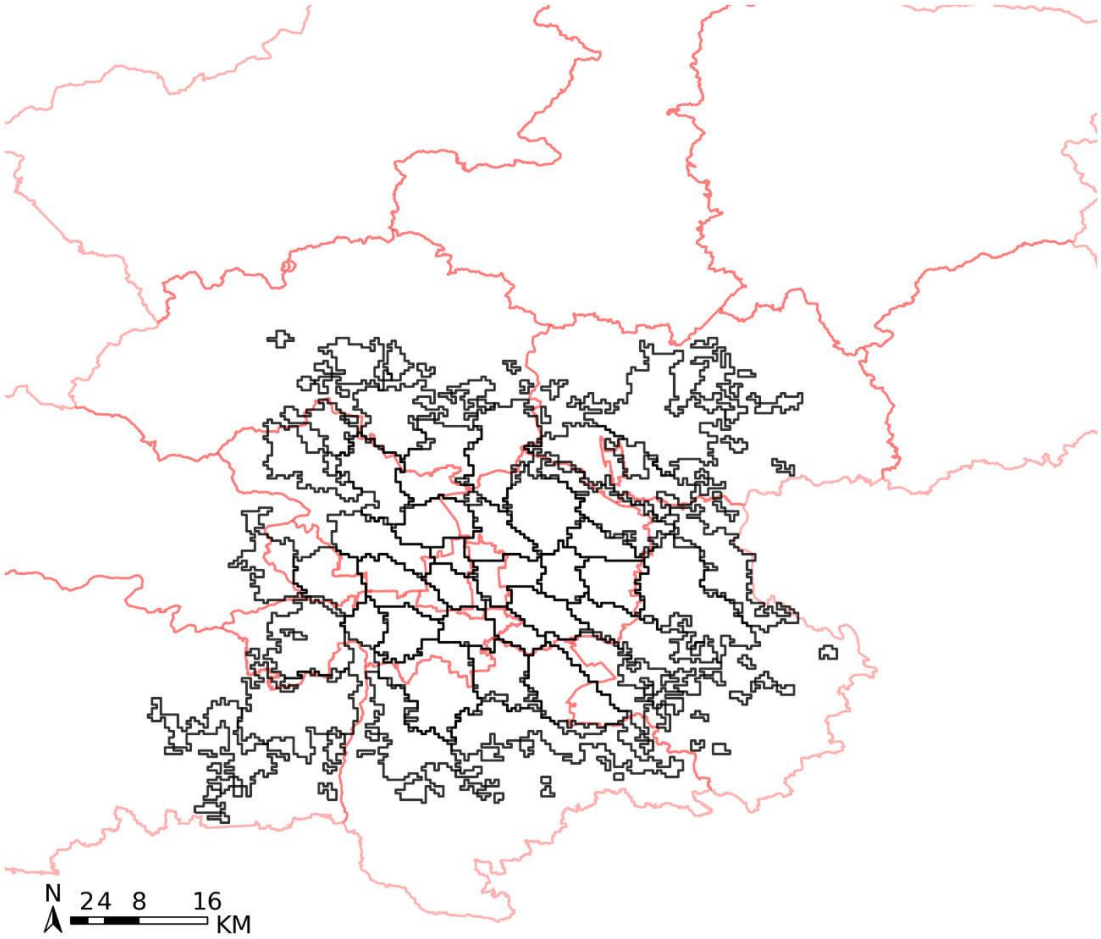


Fig. 13 Comparison of community boundary with county administrative boundary

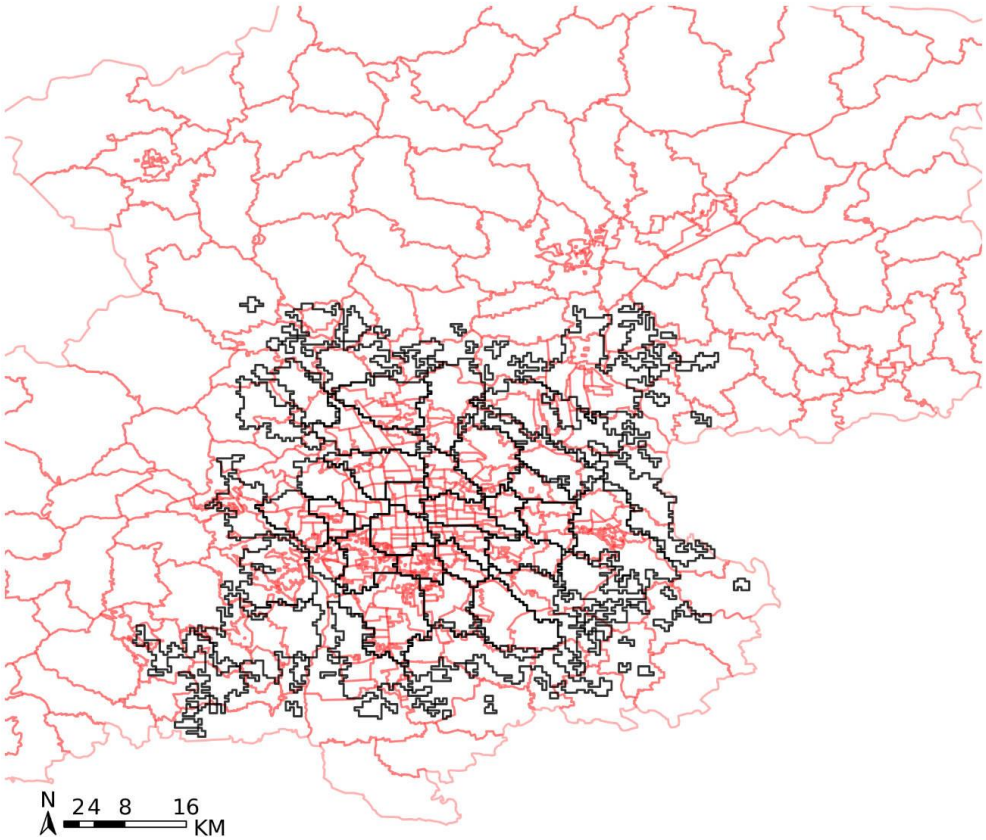


Fig. 14 Comparison of community boundary with township administrative boundary

4. Conclusion

4.1. Conclusion

By analyzing the basic spatial and temporal usage characteristics of shared bicycles, it is found that a large number of bicycle resources put in by bicycle companies have extremely low utilization rates in actual use, and parking occupies a large amount of public space is one of the drawbacks of the current operation of shared bicycles. Through geographically weighted regression analysis of the impact of various POI on the distribution of demand for shared bicycles, it is found that different POI have different ranges and degrees of impact on the distribution of demand for bicycles. Based on this analysis, relevant policies such as improving the layout of bicycle systems can be improved.

Through community detection-Fast Unfolding algorithm, it is found that there is a certain relationship between the demand for shared bicycles and administrative districts, but it is not completely consistent. If the operating community of shared bicycles is divided according to administrative divisions, it may be contrary to the actual bicycle riding demand.

4.2. Suggestions

Based on the research results, the following points are proposed. First, bicycle delivery. When bicycles are placed in these operating communities, the demand for bicycles can be quantitatively predicted according to the characteristics and needs of each operating community, and the bicycles can be placed according to the prediction results. For example, calculate the number of bikes that may be needed in each region and ensure that the bikes don 't flow to other regions in large quantities, thus ensuring the sufficient and efficient use of bikes.

Second, bicycle management. Identify the hotspots of parking demand in the operating community. On the one hand, corresponding parking infrastructure can be provided. On the other hand, it can also implement differentiated parking management measures in different regions, such as the implementation of electronic fences, dynamic pricing, etc. By identifying and solving parking problems, user experience and bicycle utilization can be improved.

Third, bicycle scheduling. Each operating partition may have different bicycle demand patterns. For example, there may be only a few key areas around a subway station in the suburbs to attract the surrounding bicycles, while in the city center, bicycles form an inner loop between multiple key areas. Therefore, different bicycle scheduling policies should be formulated for each operating area.

At the same time, it can also provide dynamic pricing for cross-regional travel and provide additional charging services for users of cross-regional travel. Ensure that the demand for bicycle travel will not be affected on a large scale, while avoiding the loss of bicycles and reducing the workload of bicycle scheduling.

References

[1] Cui Shuping. Planning and Enlightenment of Hangzhou Slow Traffic System. Urban and Rural Construction, 2010, (12): 36-37.

[2] Fu Xuemei, Jun Zhicai. Research on the cycling mode of shared bicycles in Shanghai in the time dimension. Transportation system engineering and information, 2020, 20(03): 219-226.

[3] Cheng Yuan, Xing Kai, Xu Xiaoyan. Research on the distribution of shared bicycle traffic and the characteristics of users' riding in Beijing. Proceedings of the 2018 Annual Conference on Urban Transportation Planning in China, 2018, 3724-3737.

[4] Guo Suping, Zhen Feng, Yin Qiuyi. Analysis of the spatial characteristics of urban public bicycle travel time. Planner, 2017, 33(06): 112-118.

[5] Lyu Eagle, Pan Tsunami. Analysis of the characteristics of bike-sharing in Shanghai based on Mobike open data. Shanghai Urban Planning, 2018, (02): 46-51.

[6] Chen Mian, Chen Long, Peng Xiang. Study on short-term prediction method of regional shared bicycle parking volume. 2019 China Urban Transportation Planning Annual Conference Papers, 2019, 2343-2353.

[7] Gu Zhizhi, Ye Fan, Cui Zhixiang, et al. Analysis of the factors affecting the structure and vitality of urban cycling communities based on Infomap algorithm-Taking Shenzhen as an example. Residential area, 2022, 107(01): 43-50.

[8] Vogel P, Greiser T, Mattfeld D C. Understanding Bike-Sharing Systems using Data Mining: Exploring Activity Patterns. Procedia - Social and Behavioral Sciences, 2011, 20(6): 514-523.

[9] Etienne C, Latifa O. Model-based count series clustering for bike sharing system usage mining: a case study with the Vélib'system of Paris. ACM Transactions on Intelligent Systems and Technology (TIST), 2014, 5(3): 1-21.

[10] Zhu D, Huang Z, Shi L, et al. Inferring spatial interaction patterns from sequential snapshots of spatial distributions. International Journal of Geographical Information Science, 2018, 32(4): 783-805.

[11] Zhao D, Ong G P, Wang W, et al. Effect of built environment on shared bicycle reallocation: A case study on Nanjing, China. Transportation Research Part A: Policy and Practice, 2019, 128: 73-88

[12] Zhang Y, Thomas T, Brussel M, et al. Exploring the im-pact of built environment factors on the use of public bikes at bike stations: case study in Zhongshan, China. Journal of Transport Geography, 2017, 58: 59-70.