



## Article

# Spatiotemporal Patterns of Population Mobility and Its Determinants in Chinese Cities Based on Travel Big Data

Zhen Yang <sup>1,2</sup> , Weijun Gao <sup>1,2,\*</sup> , Xueyuan Zhao <sup>1,2</sup>, Chibiao Hao <sup>3</sup> and Xudong Xie <sup>3</sup>

<sup>1</sup> Innovation Institute for Sustainable Maritime Architecture Research and Technology, Qingdao University of Technology, Qingdao 266033, China; z8dbb007@eng.kitakyu-u.ac.jp (Z.Y.); a9dbb410@eng.kitakyu-u.ac.jp (X.Z.)

<sup>2</sup> Faculty of Environmental Engineering, The University of Kitakyushu, Kitakyushu 808-0135, Japan

<sup>3</sup> College of Architecture and Urban Planning, Qingdao University of Technology, Qingdao 266033, China; hao5633852@163.com (C.H.); 13626396981@163.com (X.X.)

\* Correspondence: weijun@kitakyu-u.ac.jp; Tel.: +81-93-695-3234

Received: 16 April 2020; Accepted: 13 May 2020; Published: 14 May 2020



**Abstract:** Large-scale population mobility has an important impact on the spatial layout of China's urban systems. Compared with traditional census data, mobile-phone-based travel big data can describe the mobility patterns of a population in a timely, dynamic, complete, and accurate manner. With the travel big dataset supported by Tencent's location big data, combined with social network analysis (SNA) and a semiparametric geographically weighted regression (SGWR) model, this paper first analyzed the spatiotemporal patterns and characteristics of mobile-data-based population mobility (MBPM), and then revealed the socioeconomic factors related to population mobility during the Spring Festival of 2019, which is the most important festival in China, equivalent to Thanksgiving Day in United States. During this period, the volume of population mobility exceeded 200 million, which became the largest time node of short-term population mobility in the world. The results showed that population mobility presents a spatial structure dominated by two east–west main axes formed by Chengdu, Nanjing, Wuhan, Shanghai; and three north–south main axes formed by Guangzhou, Shenzhen, Shanghai, Wuhan, and Chengdu. The major cities in the four urban agglomerations in China occupy an absolute core position in the population mobility network hierarchy, and the population mobility network presents typical “small world” features and forms 11 closely related groups. Semiparametric geographically weighted regression model results showed that mobile-data-based population mobility variation is significantly related to the value-added of secondary and tertiary industries, foreign capital, average wage, urbanization rate, and value-added of primary industries. When the spatial heterogeneity and nonstationarity was considered, the socioeconomic factors that affect population mobility showed differences between different regions and cities. The patterns of population mobility and determinants explored in this paper can provide a new reference for the balanced development of regional economy.

**Keywords:** spatiotemporal patterns; population mobility; travel big data; socioeconomic factors; social network analysis (SNA); semiparametric geographically weighted regression (SGWR)

## 1. Introduction

Since the reform and opening up in 1978, the rapid economic development and the process of social modernization in China have made population mobility between cities more common. Population mobility is considered as the re-allocation of production factors in space; the mobility of population in a specific space promotes the reaggregation and diffusion of social and economic factors, thus reshaping patterns of population distribution [1]. From 2000 to 2010, China's floating population

increased by 109% [2]. In 2016, China's floating population reached 245 million, accounting for 18% and 76% of the total population in China and the United States, respectively (according to the Hukou system in China, the floating population refers to those who live away for work from their original Hukou system place). As China is one of the countries with the most extensive population mobility in the world, cross-regional and large-scale population mobility will become an important force to promote urbanization.

Many scholars have carried out in-depth studies on the origin, development, and internal mechanism of population mobility and migration, including the "immigration law," the "push and pull theory," the "macro-neoclassical theory," the "micro-neoclassical theory," and the "equilibrium law theory." Ravenstein's "immigration law" first proposed clear regulations related to population migration [3]. Bogue tried to explain the internal mechanism of population mobility by using the push-pull theory [4]. Macro-neoclassical theory holds that the income level of a future destination is the main force driving population mobility and migration [5,6]. Micro-neoclassical theory abstracts the mobility and migration of the population as a form of capital investment to maximize individual interests [7]. Equilibrium theory argues that destination amenities (environment, services, and living) play a key role in population mobility and migration, rather than the expected difference in benefits between regions [8,9].

Based on the above theories, China's population mobility and migration have been studied. Among them, Liu, Shen and Fan analyzed the spatial characteristics and patterns of migration [10,11], Fang and Dewen explored the determinants of population migration [12], Gu, Wang and Yu studied the formation mechanism of the spatial pattern of the floating population [13–15], and Fan discussed the impact of population mobility and migration on regional development [10]. The above research methods were based on gravity models or push-pull theories. Meanwhile, differences in regional economic development and market forces are always considered to be key factors causing population mobility [10,16]. Since the economic reform, China's coastal regions have rapidly improved the regional economy, relying on location advantages and institutional advantages [17,18]. This region has attracted a large number of populations under the influence of a significant increase in national productivity and the reform of the household registration system [19,20]. With the implementation of an investment environment and central policy, the central and western regions of China have accelerated the pace of urban development and played an increasingly important role in population mobility. Therefore, the analysis of the pattern of population mobility and the discussion of the role of socioeconomic factors among different cities can help in understanding demographic change and promote regional sustainable development.

However, previous studies on China's floating population have some limitations. On the one hand, all the studies are based on the census data; due to problems such as low data accuracy and long update time, they cannot describe the spatiotemporal patterns of population mobility in a timely, accurate, and dynamic manner [21–24]. On the other hand, the existing research is concentrated mainly on the Yangtze River Delta, the Pearl River Delta, the Beijing–Tianjin–Hebei urban agglomeration and megalopolis, whereas small and medium-sized cities are not often mentioned [25–28]. However, with the trend of inland development of China's economy, the central and western regions are increasingly becoming an important force to promote the development of national urbanization; therefore, the research on the abovementioned cities are equally important.

In recent years, with the rapid development of 3S technology, comprehensive and continuous observation of human spatiotemporal behavior data including geographic location, social attributes, movement trajectories, migration processes, and interaction patterns has become possible. Urban research based on Sina Weibo's check-in data [29], public comment data [30], residents' geographic behavior data [31], bus credit card data [32], and traffic travel big data has become a reality. These methods are real-time, objective, easy to analyze, and predictable, thus making up for the shortcomings of traditional survey methods (such as questionnaire, sampling, and census data). Therefore, they can provide sufficient and more accurate real-time data.

Meanwhile, with the in-depth research in the field of sociology, some quantitative analysis methods such as the social network analysis method have been widely used. Social network analysis (SNA) is the process of investigating social structures through networks and graph theory [33]. It characterizes networked structures in terms of nodes and the ties, edges, or links that connect them. It is a complex network based on nodes and connections to measure and map various aspects or relationships among people, organizations, and groups. Now, this method has emerged as a key technique in the modern social sciences—it is widely used in anthropology, biology, demography, communication studies, economics, geography, history, and information science—and has achieved remarkable results [34]. Therefore, it can provide sufficient technical support for a quantitative analysis of population mobility.

The Spring Festival is the biggest and most important festival in China. During the Spring Festival, there will be a large number of people traveling between their work place and hometown; this special phenomenon is called “ChunYun”. According to the statistics from the Ministry of Transport of the People’s Republic of China, the national transportation system had a total of 1.395 billion passengers during the 2019 ChunYun, which provides us with a good opportunity to study the spatial and temporal distribution of population mobility. Compared with the tourism flow during the National Day Golden Week, the population flow during the Spring Festival is characterized more by return flow and student flow. The national transportation system encounters enormous challenges during the Spring Festival travel rush and the various social problems that occur not only cause public concern, but also increase government spending. Therefore, it is of great practical significance to explore the factors influencing population mobility patterns related to it (unless otherwise noted, the Spring Festival referred to in this paper is that of 2019).

Based on the above analysis, this paper makes three main contributions:

1. Based on the dataset with high spatial and temporal resolution, a more accurate population mobility pattern was found during the Spring Festival, avoiding the spatial mismatch caused by the impact of the census dataset and the Hukou system. (The Hukou system is a special basic administrative system in China. It is a household management policy implemented by the People’s Republic of China for its citizens. The Hukou system shows the legitimacy of the person’s natural life in the local area. There are two main aspects of China’s Hukou system. First, agricultural and nonagricultural Hukou. Due to extensive disputes, the state gradually cancelled the division of agricultural and nonagricultural household registration. The second is the location of the Hukou. When a person is born in a certain place, he or she will be automatically registered to the local Hukou. In this era of high population mobility, this system leads to a serious problem of “the separation of people and Hukou”, which leads to a series of political and social problems; this kind of population is called the floating population.)

2. Based on the changes of net population mobility in different cities before and after the Spring Festival, this paper described the population activities in prefecture-level cities in China and analyzed the characteristics of the population mobility network through the social network analysis method, including the spatial structure, with key cities as nodes and typical “small world” features.

3. Furthermore, the semiparametric geographically weighted regression (SGWR) model was applied to explore the determinants of population mobility. The results showed that average wage, urbanization rate, foreign capital, value-added of primary industry, and value-added of secondary and tertiary industry are closely related to population mobility. When the spatial heterogeneity and nonstationarity was considered, the socioeconomic factors that affect population mobility were different in different regions and cities. The spatial disparity of these social economic factors was further discussed and development strategies among cities were analyzed.

The research data and area are discussed in Section 2. Section 3 outlines the methodology; Section 4 presents the results, including the spatial and temporal patterns of population mobility during the Spring Festival, and socioeconomic factors associated with population mobility. The discussion and conclusions are given in Section 5.

## 2. Research Data and Area

Location-based services (LBS) obtain the geographical location of a mobile user through the wireless communication network or the external positioning method of network operators. When users allow various mobile applications to call LBS services, their movement trajectories will be accurately recorded in real-time through the positioning information. The movement of a single user in geographical space seems to be random, but may take on a specific pattern when a large population group is accessed. According to the statistical report on the development of the Internet in China, by the end of 2018, the number of instant messaging users had reached 792 million, and the number of mobile Internet users had exceeded 817 million, accounting for 98.6% of Internet users using mobile phones [35]. In this context, every smartphone user can be seen as a mobile sensor, reflecting social characteristics and allowing for the collection of a massive amount of individual movement information, in real time and in an efficient manner.

The dataset we used in our research is the “Migration Map” section of Tencent Location Big Data [36], with the time interval set to a day and the accuracy able to be traced back to the individual level. The website counts the number of changes in the location of the smart terminal within a certain time interval to filter, summarize, and count the data. In consideration of user privacy, the website only provides the total amount of population inflows and outflows in a day, with the city as the basic unit (the intensity of inflows, source, and outflows limited to the destination of a single city on a certain day). The website provides a free Application Programming Interface (API) for researchers and programmers, allowing the above data to be obtained and used in scientific research. We used the API with Python programming language to obtain population mobility data during the Spring Festival of 2019 and store it in the SQL database.

The population mobility data we obtained contained the following content (after excluding the user’s private information) and were added to a city as the basic units instead of the individual by Tencent company: the source city name and its coordinates, the target city name and its coordinates, time, mobility intensity, and mobility type, which was consistent with the content displayed on the website. After manual filtering and sorting, it contained a total of 40,591 pieces of information, each of which covered eight aspects (source city name and its coordinates, target city name and its coordinates, time, and mobility intensity). Based on this, we constructed a data table of  $40,591 * 8$ , as shown in Table A1 (Appendix A). Other variables, such as total population, average wage, gross regional product (GRP), urbanization rate, unemployment rate, and other socioeconomic indicators, were obtained from the 2018 Urban Statistical Yearbook on the website of the National Bureau of Statistics. In this paper, the Geographic Information System (GIS), GeoDa, Gephi, and GWR4 were used to process, analyze, and calculate the data.

It should also be noted that the error factors and representativeness of the dataset cannot be ignored. Despite the huge amount of travel data obtained through location services, it is undeniable that there are still some groups who do not use any apps through smartphones, so their travel trajectories cannot be collected. It is predictable that this dataset is representative of specific regions and groups, such as developed regions and young middle-aged groups, which coincided with our research objects. Nonetheless, considering that this dataset provided a larger, more dynamic, and more efficient record of population mobility, combined with fine temporal and spatial resolution, it proved to be reliable in related research [37–40].

In this study, we considered 290 prefecture-level and above administrative units in China as the research focus, including four municipalities, two special administrative regions, 15 sub-provincial cities, and 269 prefecture-level cities. Due to the lack of data, some prefecture-level cities in Hainan province, Taiwan, and some ethnic minority autonomous prefectures in western China were not included in the study. Figure 1 shows the research areas in this paper.

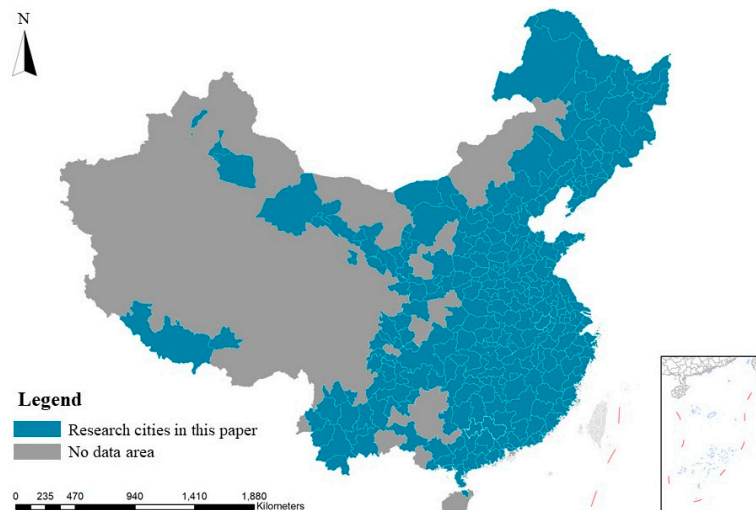


Figure 1. Research areas.

### 3. Methodology

In this study, we first used the social network analysis method to analyze the characteristics of the population mobility network. Then, spatial autocorrelation analysis was used to validate the spatial dependence of population mobility, and the ordinary least squares (OLS) method and correlation test were employed to identify correlated factors of population mobility. After that, three types of regressions analysis, including ordinary least squares (OLS), geographically weighted regression (GWR), and semiparametric geographically weighted regression (SGWR), were conducted to reveal the correlated factors of population mobility. Figure 2 gives a flowchart of this research.

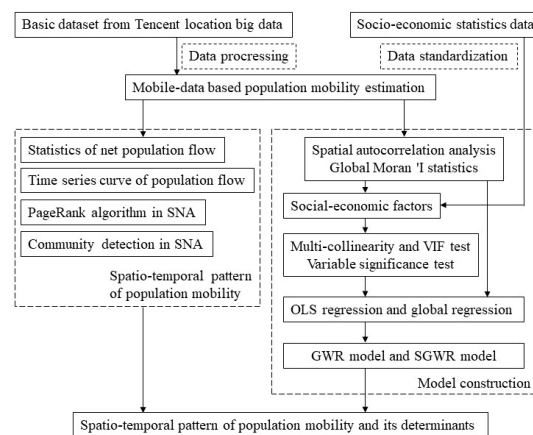


Figure 2. Flowchart of this research.

#### 3.1. Social Network Analysis (SNA)

Using the social network analysis method and taking the intensity of population flow between cities as the weight, we established a  $290 \times 290$  directed weighting matrix  $P = (P_{ij})$  to characterize population mobility within 14 days.  $P_{ij}$  represents the intensity of population flow from city  $i$  to city  $j$ . We studied the network characteristics of population flow by using the social network analysis method. The population flow network is a small world, scale-free network between a fully regular network and a completely random network. Network characteristics are usually measured by the PageRank algorithm and “community” detection indicators.



$$P_{ij} = \begin{bmatrix} 0 & P_{12} & \cdots & P_{1(n-1)} & P_{1n} \\ P_{21} & 0 & \cdots & P_{2(n-1)} & P_{2n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ P_{(n-1)1} & P_{(n-2)2} & \cdots & 0 & P_{(n-1)n} \\ P_{n1} & P_{n2} & \cdots & P_{n(n-1)} & 0 \end{bmatrix}. \quad (1)$$

PageRank is an algorithm used by Google search engines to rank the importance of web pages. It is then applied to network analysis in many fields, such as bibliometrics, social network analysis, and road networks [41,42]. Compared with other centrality indices for evaluating nodes in a network, such as degree, betweenness, and closeness, the PageRank algorithm not only considers the number of connections, but also measures the quality of connections, which means that if a node has fewer connections but all the important nodes are connected, it is still important. We believed that the mobility network formed during the Spring Festival was similar to the Internet and cities with higher importance attract more population and routes. Based on the intensity of population flow, we used the PageRank algorithm to rank the importance of urban nodes and get the hierarchical structure of population mobility. The formula of the PageRank algorithm is as follows:

$$PR_i = \frac{1-d}{N} + d \sum_{j \in B_j} \frac{PR_j}{L_j}, \quad (2)$$

where  $PR_i$  is the PageRank value of city  $i$ ;  $d$  is a constant, usually set to 0.85;  $N$  is the number of all cities;  $B_j$  is a collection of cities with all the population flow from city  $i$ ; and  $L_j$  is the number of links from city  $i$ , which is weighted by the intensity of the population flow.

Many methods have been used for community detection testing, especially fast algorithms for large-scale networks, such as the Girvan–Newman algorithm, the CNM algorithm, SCAN algorithm, and so on [43–45]. In this paper, the multilevel algorithm was used for the community detection test, which is a bottom-up algorithm proposed by Blondel et al. through optimizing modularity [46].

### 3.2. Semiparametric Geographically Weighted Regression (SGWR) Model

This paper applied the geographically weighted regression (GWR) model [47] to reveal the relationship between population mobility changes and socioeconomic factors at the global and local levels. Compared with the traditional geographically weighted regression model, semiparametric geographically weighted regression [48,49] pays more attention to the analysis of the diversity and nonstationarity of geospatial data. The model was proven to be applicable in the fields of geography, environmental science, and economics [50,51]. This model has better performance than the traditional geographically weighted regression model because of changes in the geographic parameters. Therefore, the SGWR model was constructed based on the statistical yearbook data, and the formula is expressed by Equation (3):

$$y_i = \sum_{k=1} \beta_k x_k + \sum_{j=1} \beta_j(u_i, v_i) X_{ij} + \varepsilon_i, \quad (3)$$

where  $k$  and  $j$  represent the global variables and local variables, respectively.  $(u_i, v_i)$  represents the coordinates of location  $i$ ,  $\beta_j(u_i, v_i)$  represents the local regression coefficient for the explanatory variable  $X_j$  at location  $i$ , and  $\varepsilon_i$  represents the error term.

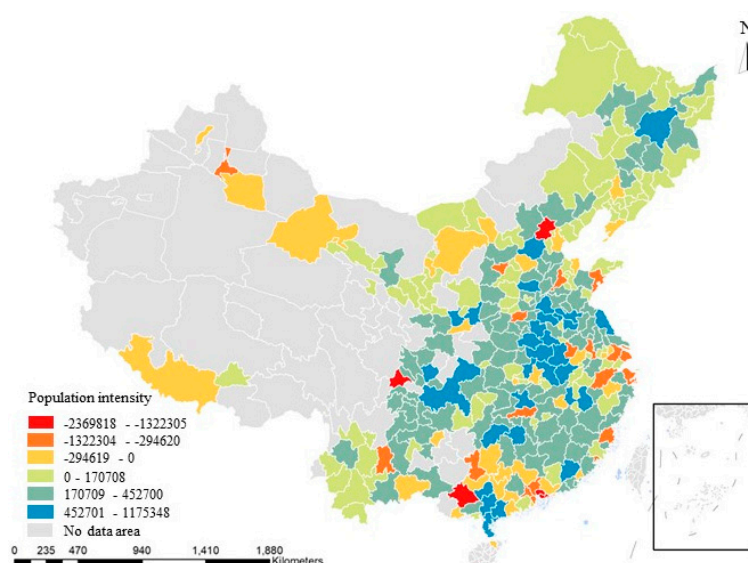
For the GWR model, the spatial weight matrix is critical. The selection of the spatial weight function has a great influence on the parameter estimation of the geographically weighted regression model [47]. In this paper, we used an adaptive bi-square kernel to calculate the weight matrix. We used adaptive bi-square kernels instead of fixed kernels based on two considerations. First of all, the regression points (the center of each city) appeared to be randomly distributed in the study area, and the adaptive kernel made the dataset large enough for each local regression [52]. Secondly, the adaptive

bi-square kernel can reduce the bandwidth in the data-intensive place and expand the bandwidth in the scattered place of the dataset and have a clear-cut range when the kernel weight is not 0. It is widely used in studies taking the city as the unit [53,54]. Meanwhile, the accuracy of the GWR model is greatly affected by the bandwidth of the weight function. Akaike information criterion (AICc) and cross-validation (CV) are two methods commonly used to determine the bandwidth. Compared with the latter, the former can quickly and efficiently resolve differences in the degrees of freedom in various models [48]. Therefore, the AICc was selected to determine the appropriate bandwidth when constructing the GWR model.

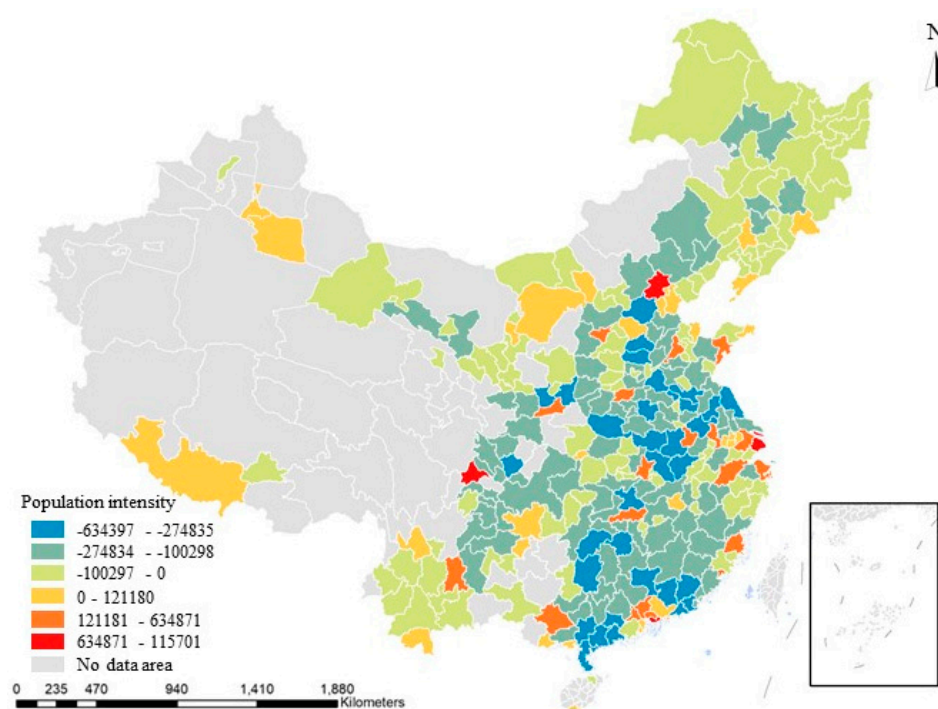
### 3.3. Mobile-Data-Based Population Mobility Variation (MBPMV)

Due to data acquisition, we got only 14 days of Tencent location data from 290 cities, but considering the sample size and data accuracy, this dataset could be used as the basic data for in-depth research. The first day of the dataset was 29 January and the last day of the dataset was 11 February. We divided the 14 days into two phases, before the Spring Festival and after the Spring Festival. Therefore, the net population mobility in all cities during the two time periods was considered to be representative of the population distribution during the Spring Festival. Through changes in time series and differences in population mobility between the two time periods, we found differences in the distribution of human activity.

Figures 3 and 4 show the inflow and outflow statistics of all cities before and after the Spring Festival, respectively. We selected cities with the most obvious population inflow and outflow in the two time periods, namely Beijing, Chongqing, Shenzhen, and Hengyang, and plotted their main population flow direction and intensity, as shown in Figure A1 (Appendix A). Combining Figures 3 and 4, we found that the cities with large population mobility were all located in the four major urban agglomerations in China, and the central region experienced a significant population inflow before the Spring Festival and a large population outflow after the Spring Festival. This pattern of population flow revealed the differences in China's regional development, so it was particularly important to explore the causes of this phenomenon, which may be geopolitical, economic, social, and geographical, etc. Meanwhile, the global spatial autocorrelation analysis of the population mobility changes in all cities after the Spring Festival resulted in a Moran index of 0.702, Z score of 50.128 and  $p$  value of 0.01, indicating that the probability of randomly generating the above spatial distribution pattern was less than 5%. Based on this, a SGWR model could be constructed to explore the socioeconomic factors related to the formation of the above population mobility pattern.



**Figure 3.** Population inflow and outflow of cities before the Spring Festival. (A positive value represents the intensity of the city's inflow population and a negative value represents the intensity of the city's outflow population.).



**Figure 4.** Population inflow and outflow of cities after the Spring Festival. (A positive value represents the intensity of the city's inflow population and a negative value represents the intensity of the city's outflow population.).

### 3.4. Independent Variables Selection and Model Construction

In order to analyze the determinants of population mobility patterns, we used three steps to determine the independent variables in the GWR model: (1) Select for socioeconomic factors related to urban development. The wage level of employees is a very important factor, because the difference in expected benefits is the main force driving population mobility [55]. Secondly, GRP and average gross regional product are direct reflections of the economic development of a city, which may affect population mobility [56]. At the same time, depending on different types of work, the three industries can also affect population mobility. Finally, considering that labor-intensive industries can absorb large amounts of labor, we considered foreign capital as a candidate variable [57]. In addition, we added several candidate variables related to urban development and social economy, such as urban total population, urbanization rate, and urban worker unemployment rate [58]. (2) Exclude multicollinearity between variables. We performed an OLS regression to detect multicollinearity between the variables. After all variables were normalized to fit the normal distribution, the variance inflation factor (VIF) of each independent variable was calculated, and then the independent variables with  $VIF > 7.5$  were eliminated from the final model. In this process, the VIF values of the nine independent variables we selected were all less than 7.5, indicating that there was no multicollinearity among the variables (Table 1). (3) Perform a correlation analysis of variables, excluding variables that are not related to population mobility at a 95% confidence level. In this process, the unemployment rate (UER) was eliminated. The results showed that there were no redundant and uncorrelated problems in the remaining variables. Therefore, after the above three processes, total population (TP), average wage (AW), gross regional product (GRP), Avg\_GRP, urbanization rate (UR), foreign capital (FC), value-added of primary industry (VAPI), and value-added of secondary and tertiary industry (VASTI) were used for GWR model analysis. Table 2 shows the details of the above variables.



**Table 1.** Variance inflation (VIF) value and correlation coefficient of all independent variables.

	TP	AW	GRP	Avg_GRP	UR	UER	FC	VAPI	VASTI
VIF	1.236	3.456	6.596	3.408	2.815	1.247	2.641	3.302	2.281
Coefficient	0.163 **	0.577 **	0.462 **	0.405 **	0.572 **	−0.590	0.582 **	−0.503 **	0.538 **
Sig.	0.003	0.000	0.000	0.000	0.000	0.361	0.000	0.000	0.000

Total population (TP), average wage (AW), gross regional product (GRP), average gross regional product (Avg\_GRP), urbanization rate (UR), foreign capital (FC), value-added of primary industry (VAPI), and value-added of secondary and tertiary industry (VASTI); \*\* represents that the variable is significant at the 0.05 level.

**Table 2.** Details of the dependent and independent variables.

Class	Variable	Notation	Explanation
Dependent Variable	MBPM variation	MBPMV	net inflow population of all the cities after the Spring Festival
Independent Variable	Total population	TP	total population at year end (1,000,000 persons)
	Average wage	AW	average wage of employees on duty (yuan/person)
	Gross region product	GRP	annual gross regional product (100,000,000 yuan)
	Average gross region product	Avg_GRP	annual gross regional product per capita (10,000 yuan/person)
	Urbanization rate	UR	proportion of urban population to total population (%)
	Unemployment rate	UER	proportion of unemployed population to working population (%)
	Foreign capital	FC	actual utilization of foreign investment (100,000,000 yuan)
	Value-added of primary industry	VAPI	annual value-added of primary industry (10,000 yuan)
	Value-added of secondary and tertiary industry	VASTI	annual value-added of secondary and tertiary industry (10,000 yuan)

After the OLS regression and correlation test, a total of eight variables were selected for the GWR model. In this paper, the significance ( $p < 0.05$ ) of all variables was defined as the pseudo  $t$  ( $\text{Est}/\text{SE}$ )  $> 1.96$  or  $< -1.96$  [48,59]. Considering that local models can improve accuracy, the SGWR model was further used to explore the spatial stationarity and non-stationarity of parameters affecting population mobility. An iterative process was used to determine whether the parameters were global or local variables. The most suitable model was judged based on AICc and the model with the smallest AICc value was selected as the best [60,61].

## 4. Results

### 4.1. Spatiotemporal Patterns of Population Mobility

Figures 3 and 4 show the inflow and outflow statistics of all cities before and after Spring Festival, respectively. We see that: (1) There were significant differences in population mobility in different cities in the two time periods. The population flow showed a high consistency with the city level—that is, the higher the city's development level, the greater its population flow during the Spring Festival, such as in Beijing, Shanghai, Guangzhou, Shenzhen, and Chengdu. (2) The inflows and outflows of population cities were also different in the two time periods. Before the Spring Festival, there was an obvious population inflow in the central and western regions, and the core cities in the four major urban agglomerations had a relatively obvious population outflow. After the Spring Festival, this phenomenon reversed. This might be because the regional core cities attract a large number of migrant laborers. Before the Spring Festival, the migrant laborers return home to be with their families. This is the commonly known as “returning flow.” After the Spring Festival, they return to the original workplace to continue their job, which will lead to the so-called “migrant flow,” resulting in a surge of migrant laborers to the workplace.

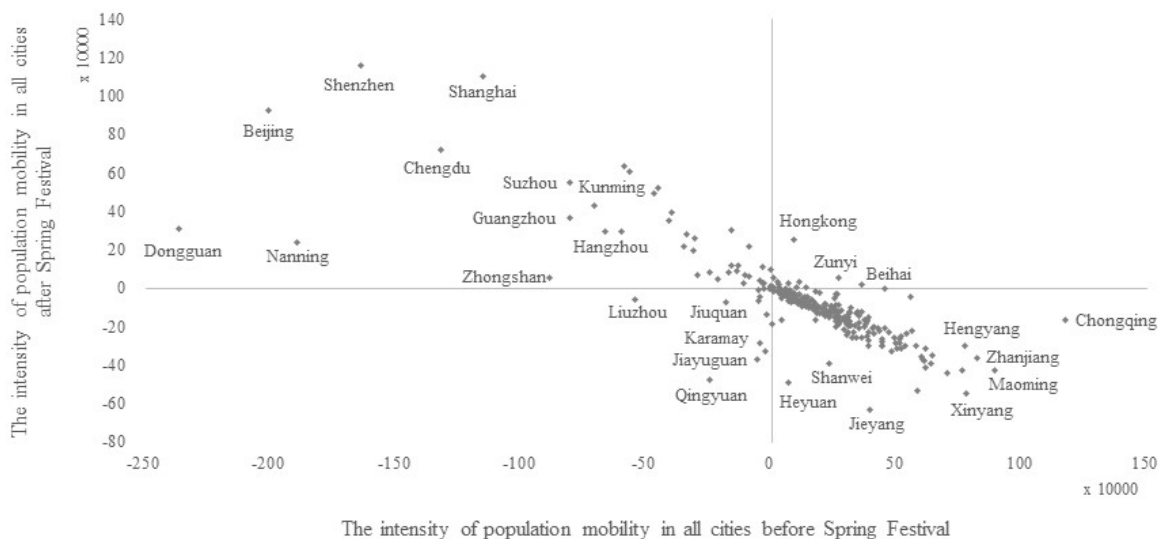
Based on the statistics of population inflows and outflows in the two periods (Table 3 and Figure 5), we divided all cities into four categories: continuous population inflows (II), continuous population

outflows (OO), population inflows then outflows (IO), and population outflows then inflows (OI). Table 3 lists the four different types of cities. We can see that first-tier cities and provincial capital cities located in southeastern China belong to the OI type; second and third-tier cities located in central and western China and small cities around the regional core cities belong to the IO type. Most of the cities with continuous population outflows were located in northwestern China, which has low population attractiveness for economic, environmental, and geographical reasons. The problem of irreversible population loss should attract the attention of the relevant city managers. The same situation also occurred in the Pearl River Delta urban agglomeration. Guangzhou and Shenzhen absorbed a large number of human and material resources from the surrounding areas, resulting in a continuous outflow of the population from the surrounding small cities, which to some extent destroyed the sustainable development of the region. With the improvement of people's living standard, traveling for the New Year became common. Therefore, a tourism-oriented city such as Sanya, Zunyi, Lijiang, or Beihai can continue to attract visitors to some extent during the Spring Festival.

**Table 3.** Different city types based on population inflow and outflow statistics.

City Type	Number	Cities
IO	223	Chongqing, Wenzhou, Xining, Lanzhou, Taizhou, Luoyang, Yangzhou, Xuchang, Shaoxing, and 214 other cities
OI	42	Beijing, Shanghai, Guangzhou, Xiamen, Chengdu, Zhengzhou, Suzhou, Changsha, Jinan, Kunming, Hefei, and 31 other cities
II	12	Hong Kong, Macau, Sanya, Xishuangbanna, Lijiang, Changzhou, Zunyi, Weihai, Shennong, Langfang, Baishan, and Beihai
OO	13	Karamay, Jiayuguan, Jiuquan, Yangjiang, Qiangjiang, Yunfu, Liuzhou, Jinhua, Haikou, Wuzhou, Chaozhou, Yangjiang, Qingyuan, and Tongchuan

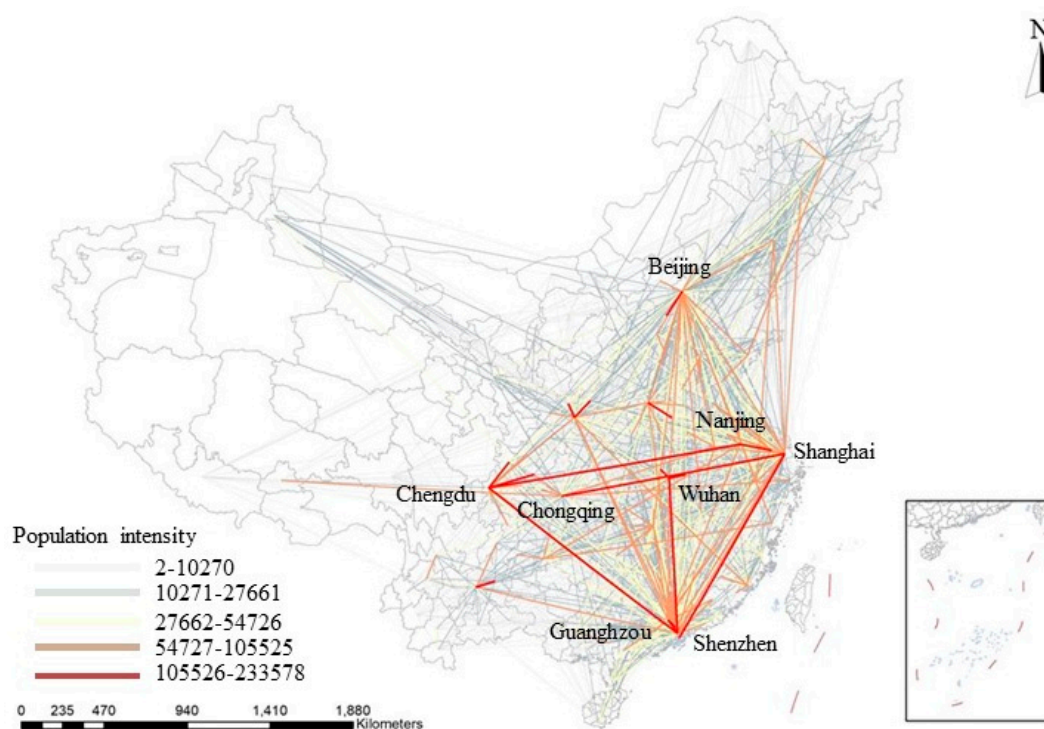
IO represents population inflow the before Spring Festival and outflow after the Spring Festival. OI represents population outflow before the Spring Festival and inflow after the Spring Festival. OO represents continuous population outflow during the Spring Festival; II represents continuous population inflow during the Spring Festival.



**Figure 5.** The bivariate relationship between the intensity of population mobility in the two time periods of all the cities.

Figure 6 is the grading map of the of population flow during the Spring Festival, from which we can clearly see the spatial pattern. First, unlike the diamond-shaped structure formed by the population mobility during the National Day Golden Week [62], the population flow during the Spring Festival presents a spatial pattern of two east–west main axes and three north–south main axes. The two east–west main axes are Shanghai–Nanjing–Chengdu and Shanghai–Wuhan–Chongqing, and the

three north–south main axes are Shenzhen–Chengdu, Shenzhen–Wuhan and Guangzhou–Shanghai, all located in the four major urban agglomerations in China. At the same time, we noted that, although Beijing is not prominent in this structure, its coverage covers most areas of China and it is also a distributing center for population. The Shandong peninsula is not obvious in this structure, which is badly out of line with its position in the national development strategy. Second, compared with the population flow during the National Day Golden Week, the population flow boundary of the major cities during the Spring Festival is relatively obvious. Large cities have a typical spatial orientation, while medium-sized cities show strong spatial proximity.



**Figure 6.** Grading map of the net flow of population during the Spring Festival.

For a further understanding of the spatiotemporal pattern characteristics of population flow, we first established a directed weighted matrix of population inflow and outflow between cities, then explored it by using the PageRank algorithm and community detection test in SNA. The PageRank algorithm was used to rank the importance of cities in the population mobility network, and the hierarchical structure of population flow was obtained. Figure 7 shows a hierarchical map of all cities in the population mobility network, which was classified according to the PageRank value by the natural break classification (NBC); the results are summarized in Table 4. We found that there are six cities in the nationwide network center, namely Beijing, Shanghai, Chongqing, Guangzhou, Shenzhen, and Chengdu, all located in the four major urban agglomerations of China, which is similar to the results in Figure 5. The nationwide network subcenter consists of 16 cities, which are either sub-provincial cities, provincial capitals, or developed cities in southeast coastal areas. To a certain extent, the above cities have a clear connection to population mobility during the Spring Festival. Compared with the cities in the southeast coastal areas, the cities in the central and western regions are mostly regional network centers or local network centers, which are not prominent in the whole population mobility network, indicating that the above areas show extremely weak attraction or radiation force in both the population inflow and outflow. From this we can find obvious differences between regions.

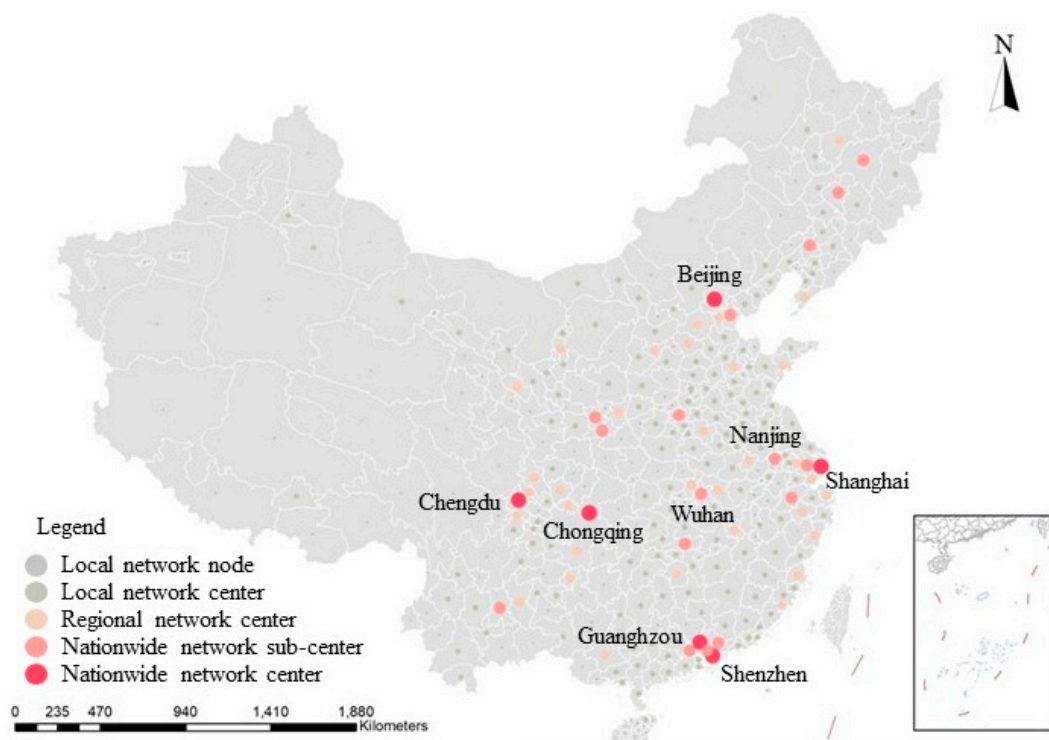


Figure 7. Hierarchical map of cities in the population mobility network.

Table 4. Summary of the city hierarchy in the population mobility network.

Level (PageRank Value of Network)	Cities
Nationwide network center	Beijing, Shanghai, Chongqing, Guangzhou, Shenzhen, and Chengdu
Nationwide network subcenter	Xi'an, Hangzhou, Wuhan, Changsha, Harbin, Zhengzhou, Nanjing, Suzhou, Tianjin, Kunming, Foshan, Huizhou, Shenyang, Changchun, Dongguan, and Xianyang
Regional network center	Baoding, Langfang, Guiyang, Xiamen, Jinan, Ningbo, Nanchong, Sanya, Nanning, Qingdao, Guang'an, Wenzhou, Shijiazhuang, Deyang, Weinan, Hefei, Mianyang, Hongkong, Zunyi, Fuzhou, Meishan, Huanggang and 14 other cities
Local network center	Yueyang, Dali, Shaoxing, Xuzhou, Baoji, Zhaoqing, Yancheng, Shaoguan, and 158 other cities
Local network node	The remaining 66 cities

Through the network analysis method to calculate the matrix of directed weighted population mobility, we found that the clustering coefficient of the population mobility network in the Spring Festival was 0.375, and the average path length was 2.792, which was much higher than the random network composed of 290 nodes (the clustering coefficient was 0.112, and the average path length was 2.075), indicating that the population mobility network during the Spring Festival conformed to the scale-free network characteristics and presented a typical “small world” network structure, which was different from Li’s results at the provincial scale [37]. With the help of the community detection test, we further revealed the relationship between the cities hidden in the population mobility network. Nodes belonging to the same community tended to be more closely linked, indicating that cities within the same community have more frequent population mobility than other cities. Figure 8 gives the distribution map of the network community structure and Table 5 summarizes the community structure of all cities.

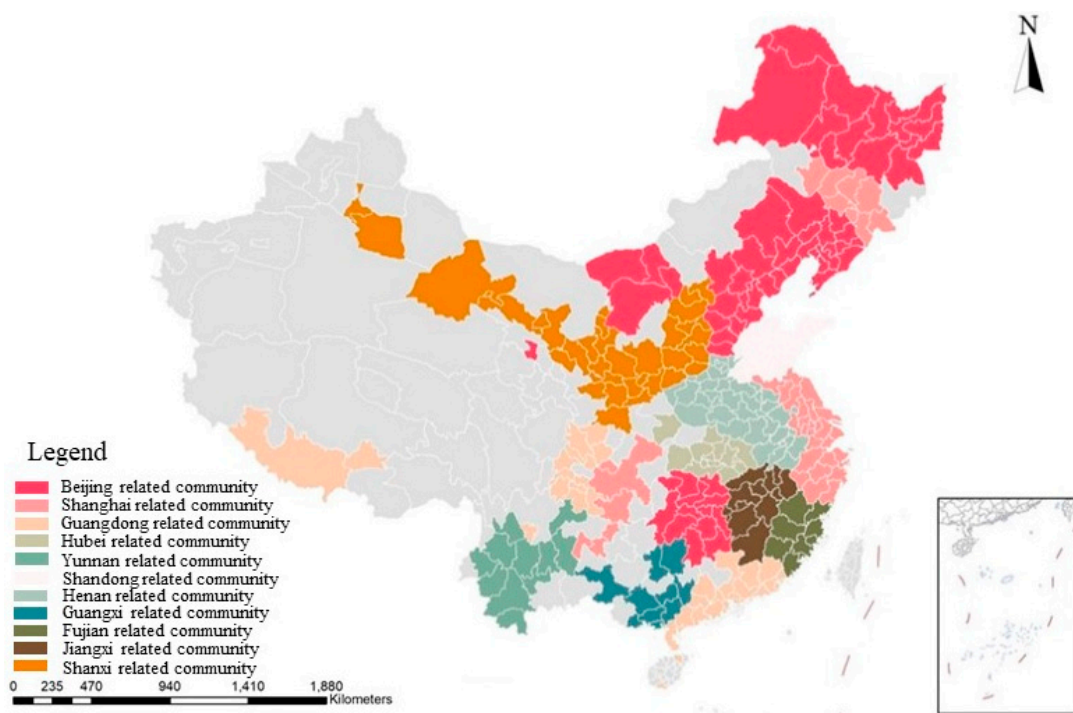


Figure 8. Distribution map of the network community structure.

Table 5. Summary of the city community structure.

ID	Major Covering Provinces	Key Cities Included	Number of Cities
0	Beijing, Tianjin, Hebei, Heilongjiang, Liaoning, Hunan	Beijing, Tianjin, Shijiazhuang, Harbin, Shenyang, Changsha	62
1	Shanghai, Jiangsu, Zhejiang, Chongqing, Jilin	Shanghai, Nanjing, Hangzhou, Chongqing, Changchun	44
2	Tibet, Sichuan, Guangdong	Lhasa, Chengdu, Shenzhen, Guangzhou	38
3	Hubei	Wuhan, Xiangyang	16
4	Yunnan	Kunming, Dali, Qujing	13
5	Shandong	Jinan, Qingdao, Yantai, Weifang	17
6	Henan, Anhui	Zhengzhou, Kaifeng, Luoyang, Hefei	33
7	Guangxi	Nanjing, Liuzhou, Guilin, Wuzhou	10
8	Fujian	Fuzhou, Xiamen, Zhangzhou, Quanzhou	9
9	Jiangxi	Nanchang, Jiujiang, Shangrao, Fuzhou	11
10	Shanxi, Shannxi, Ningxia, Gansu	Xi'an, Xianyang, Urumqi	37

Based on the analysis of the population mobility network during the Spring Festival, 11 different community structures were identified. According to the spatial composition of the community, we divided the 11 communities into three categories: the first is the cross-regional community, such as the community composed of Shanghai, Jiangsu, Zhejiang, Chongqing, and Jilin; the second is the neighborhood community, such as the community composed of Shanxi, Shannxi, Ningxia, and Gansu; the third is independent provinces, such as the community composed of all cities in Shandong province. We found that the second and third community structures accounted for a large proportion of the 11 communities, indicating that large-scale population mobility is still affected by the geographical and geospatial environment. However, like the first community structure, the spatial span was large and distributed across several independent spaces, so it can be seen that, with the improvement of the transportation infrastructure and economic level of the target city, large-scale, cross-regional, and high-density population mobility will become a future development trend, and the space–time distance in the traditional sense will be severely compressed. This reflected the special structure of the population mobility network during the Spring Festival, but we still needed to obtain longer time series data for a more general analysis.



#### 4.2. Semiparametric Geographically Weighted Regression (SGWR) Model Results

Table 6 summarizes the basic parameters of the OLS, GWR, and SGWR model outputs. We can see that the constructed SGWR model made significant improvements over the normal regression model and the GWR model. Compared with the traditional regression model, the SGWR model had a smaller AICc (472.83) and a larger adjusted  $R^2$  (0.751), which indicated better overall performance. Also, the F value (2.97) was much higher than the standard value (1.26), which meant the null hypothesis that the SGWR model does not improve upon the traditional regression model could be rejected at the 95% confidence level.

**Table 6.** Summary of ordinary least squares (OLS), geographically weighted regression (GWR), and semiparametric geographically weighted regression (SGWR) model outputs.

	OLS	GWR	SGWR		
AICc	534.54	480.49	472.83		
Adjusted R-squared	0.633	0.748	0.751		
Bandwidth	-	98.72	89.70		
Residual squares	101.40	53.60	50.01		
ANOVA					
Source	SS	DF	MS	F	F Criterion
Global Residuals	101.40	279			
GWR Improvements	47.80	64.37	0.74		
GWR Residuals	53.60	214.63	0.25	2.97 <sup>a</sup>	1.26

<sup>a</sup> Statistically significant at a confidence level of 95%.

Tables 7 and 8 illustrate the statistics of the SGWR model and global regression model outputs. The results showed that AW, UR, FC, VAPI, and VASTI were significant at the 95% confidence level, while TP, GRP, and Avg\_GRP were not significant at the 0.05 confidence level. Among them, VASTI and MBPMV had the strongest positive correlation; VAPI and MBPMV had the strongest negative correlation; and FC, UR, and AW also had a strong positive correlation with MBPMV. TP, GRP, and Avg\_GRP were not significantly correlated with MBPMV. Meanwhile, UR was finally selected as a global parameter after an iterative process in GWR4.

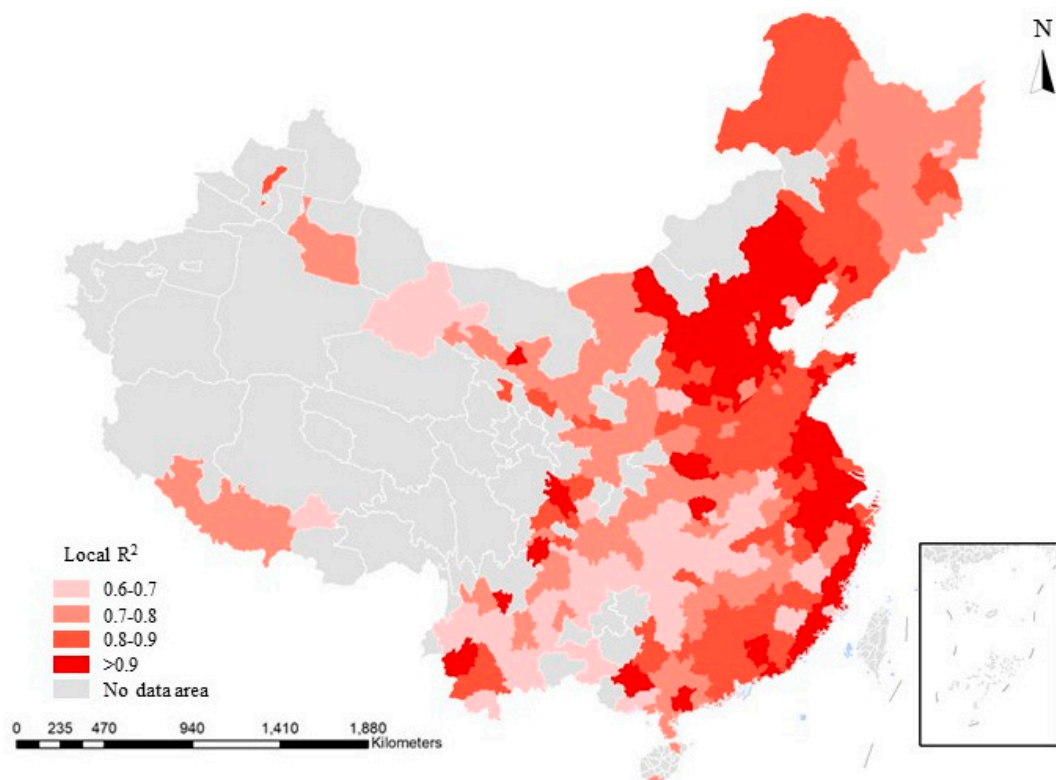
**Table 7.** Summary of the global regression model output.

Variable	Estimate	Standard Error	Pseudo <i>t</i>	<i>t</i> -Test	Spatial Stationarity
Intercept	0.000	0.103	0.000	-	-
TP	0.008	0.038	0.217	$p > 0.05$	-
GRP	0.021	0.063	0.339	$p > 0.05$	-
Avg_GRP	-0.093	0.063	-1.472	$p > 0.05$	-
UR	0.138	0.060	2.301	$p < 0.05$	global
AW	0.140	0.058	2.422	$p < 0.05$	local
FC	0.280	0.056	4.967	$p < 0.05$	local
VAPI	-0.814	0.070	-11.505	$p < 0.05$	local
VASTI	1.127	0.106	10.622	$p < 0.05$	local

Figures 9–13 visualize the spatial variation and estimation coefficients of all explanatory variables. Figure 9 shows that there is a large spatial difference in the value of local  $R^2$ , which indicates that, with the change in urban spatial location, explanatory variables have different interpretation forces on dependent variables, further reflecting the spatial nonstationarity between variables. In addition, the standard residual of the model was analyzed and the model presented a random distribution pattern in space, indicating that the constructed SGWR model had better performance.

**Table 8.** Summary of the SGWR model estimation coefficients.

Variable	Mean	Std.	Min	Lwr Quartile	Median	Upr Quartile	Max
Intercept	−0.320	1.366	−5.810	−0.215	−0.093	0.004	2.199
TP	−0.598	0.431	−1.339	−0.906	−0.643	−0.146	0.036
AW	2.120	1.257	0.099	1.010	1.935	3.237	4.645
GRP	0.621	1.025	−1.549	−0.140	0.586	1.406	3.265
Avg_GRP	−0.034	0.126	−0.251	−0.130	−0.039	0.044	0.296
FC	1.65	2.451	−3.98	−0.20	2.10	3.58	5.88
VAPI	−2.649	1.758	−6.133	−4.011	−2.451	−1.309	0.339
VASTI	1.02	2.312	−3.233	−0.64	0.17	2.315	5.971

**Figure 9.** Local  $R^2$  based on the SGWR model.

According to the statistical results of the model, the added value of the secondary and tertiary industries, the wage level of employees, the urbanization rate, and foreign capital are positively correlated to the population mobility. The added value of the primary industry is negatively correlated with the population mobility. There is no significant correlation between the total population, unemployment rate, GRP, and population mobility. In addition to the urbanization rate, other variables have different effects on different regions. These results are basically consistent with reality, as explained in the following.

The strongest positive correlation between VASTI and MBPMV indicates that the added value of secondary and tertiary industries has a significant effect on population mobility. This is because our research focused on the Spring Festival, during which the workflow is in an absolute position in the population mobility, representing the transfer of labor. Therefore, with the rapid development of the secondary and tertiary industries, the city will provide a large number of jobs, able to absorb the labor force in the surroundings and even farther afield. The population of underdeveloped areas will shift to developed areas, and the population of poor areas will shift to less developed areas. This progressive relationship affects population mobility in all areas.

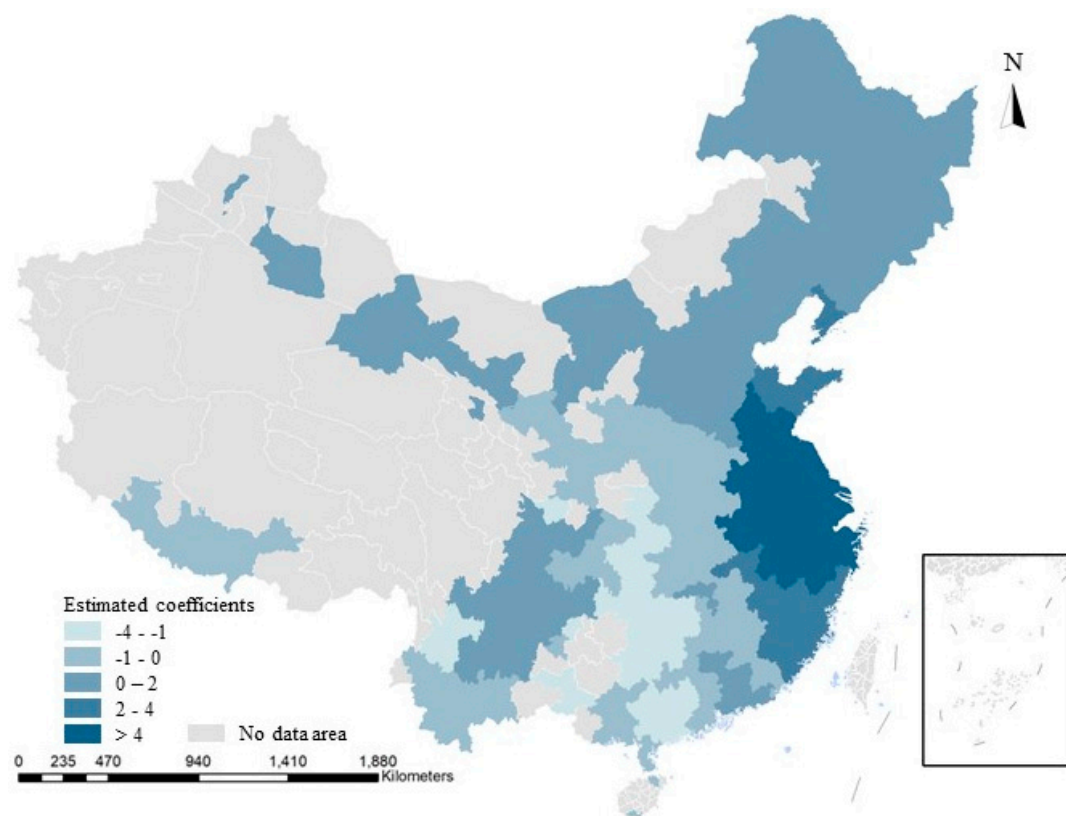


Figure 10. Local estimated coefficient of value-added secondary and tertiary industry.

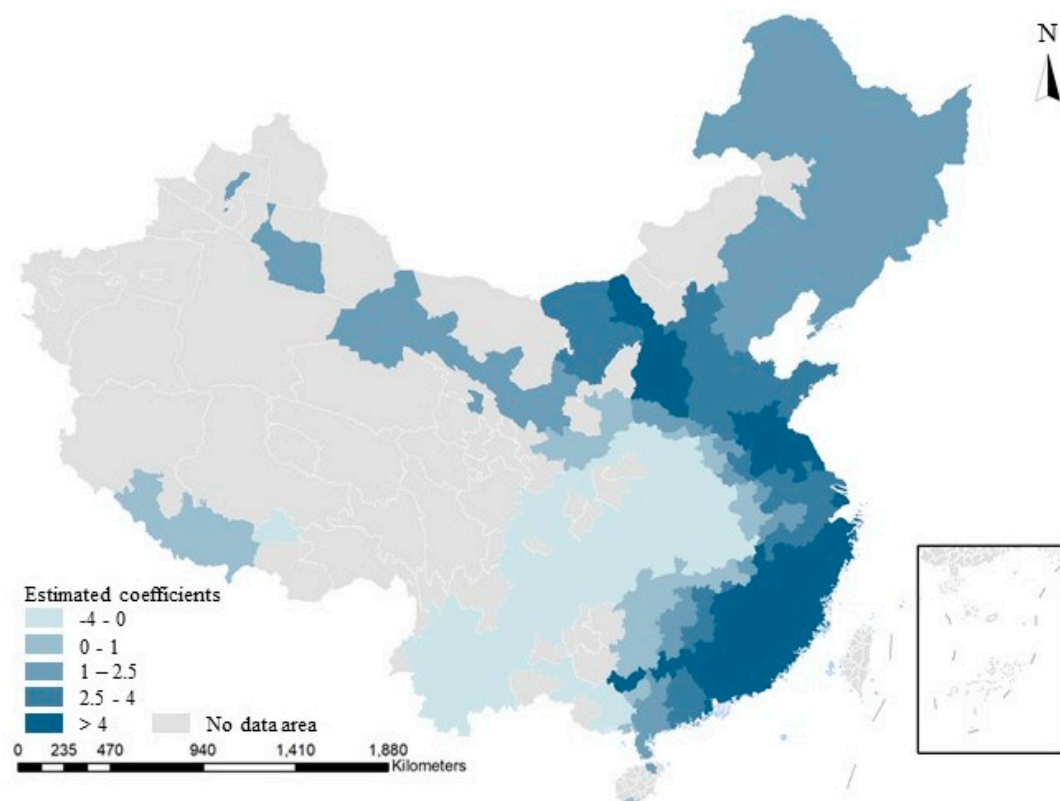


Figure 11. Local estimated coefficient of foreign capital.

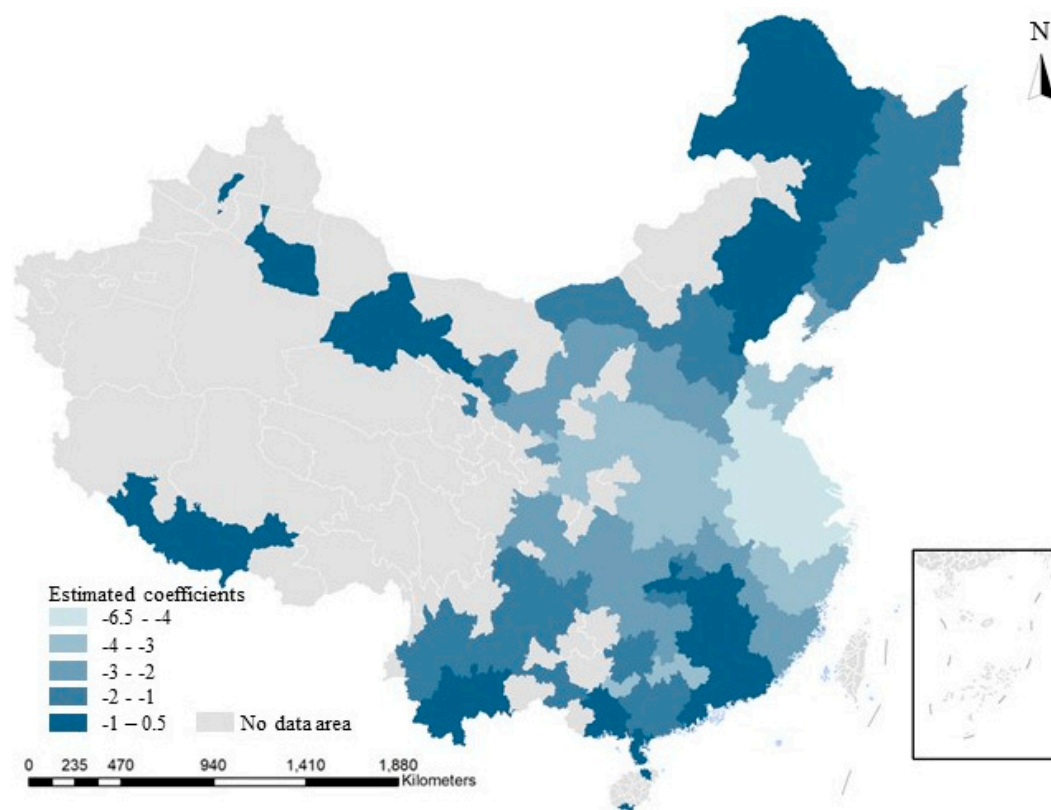


Figure 12. Local estimated coefficient of value-added primary industry.

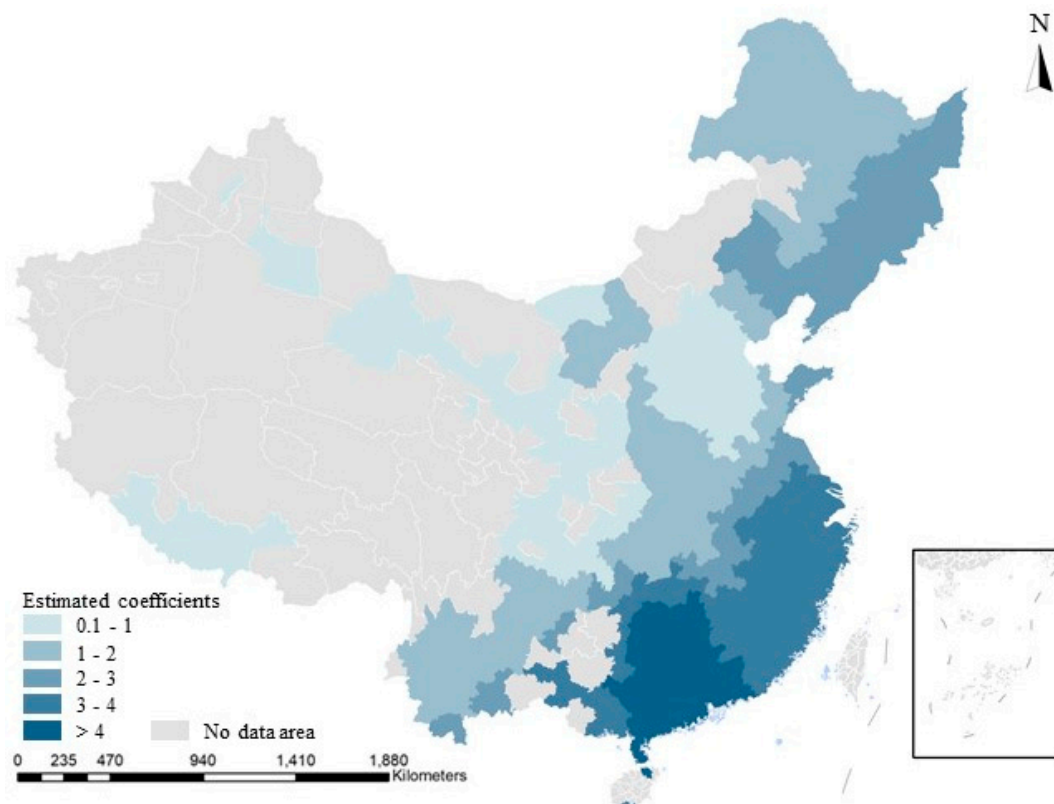


Figure 13. Local estimated coefficient of average wage.

The average wage of employees also has a positive correlation with MBPMV. This is because, as neoclassical theorists explain, the income level of the intended destination is the main driver force of the migration process. Therefore, when other costs are constant and incomes increase, more laborers will choose higher-paying areas for employment, which is similar to the impact of the added value of secondary and tertiary industries on MBPMV.

Total foreign capital also has a positive impact on population mobility. In most cases, overseas investment aims at the development of secondary and tertiary industries in the city, combined with the construction of labor-intensive enterprises, directly creating a large number of positions in the city, so this economic factor also increases population mobility.

Urbanization is also positively correlated with MBPMV. With the increase in the urbanization level, on the one hand, industrial industry can be effectively developed and more employment opportunities will be created through the intensive use of infrastructure. On the other hand, this accelerates residents' socialization and promotes developments in the service industry, which will also create a large number of employment opportunities.

There is a significant negative correlation between the added value of the primary industry and MBPMV, which indicates that, with the increase in primary industry, the population outflow will be intensified. This is determined by the nature of the work in primary industry. In China, agriculture, forestry, animal husbandry, and fisheries are classified as primary industries. In the context of mechanization, they cannot provide a large amount of labor or even absorb local surplus labor, resulting in population movement elsewhere.

In the above paragraph, we explained the explanatory variables related to MBPMV. Considering that the model takes into account the non-stationarity of space, we will focus on explaining the variation of variables in space as follows.

From Figure 10, we see that the development of the secondary and tertiary industries in the eastern coastal areas has a positive correlation with population mobility, which indicates that, if investment in the secondary and tertiary industries is increased in Jiangsu, Zhejiang, and Shanghai, they could attract more of the floating population. Meanwhile, there is a weak negative correlation between the Beijing–Tianjin–Hebei region and the Pearl River Delta region, which indicates that these two regions will not attract people through the development of secondary and tertiary industries. Some cities in the central and western regions have relatively obvious negative coefficients, especially those in Hunan, Hubei, Yunnan, Guangxi and Henan, all of which are major labor-outputting provinces, and the development of secondary and tertiary industries will not be attractive to the population mobility. Sichuan and Chongqing are also major labor-outputting regions, but their population mobility has a positive correlation with the city's secondary and tertiary industries, which means that population outflow could be slowed by increasing the proportion of secondary and tertiary industries.

Figure 11 implies that there are both positive and negative correlations between foreign capital and population mobility. The Beijing–Tianjin–Hebei region, as well as Zhejiang and Fujian, are the most obvious, which means that the increase in foreign capital can not only reduce the local population outflow, but also attract a large population inflow. Considering that the above regions are the most developed regions of the country, as well as the places where talent concentrates, the high-tech industries directly funded by foreign capital can further absorb the talent in the surrounding areas, resulting in a large population inflow. Large negative regression coefficients exist in Henan, Anhui, Sichuan, and Chongqing, indicating that increasing foreign capital might not be a good measure to attract population inflow for provinces with a large labor force output. There are weak positive and negative regression coefficients in the northeast and southwest, which may mean that they are already saturated with foreign investment and an increase will not attract further migration.

From Figure 12, we find that the value-added of the primary industry is negatively correlated with the population mobility in all cities studied, which is consistent with a recent study that examined the effects of rising agricultural productivity on migration [63]. The nature of the primary industry means that it can solve the local surplus labor to a certain extent, but it cannot attract external



population. The correlation between the central and eastern coastal areas is much higher than that of other regions, which indicates that the abovementioned regions, especially those of Anhui, Jiangsu, and Zhejiang, should focus on reducing the development of the primary industry in the hopes of attracting external population.

From Figure 13, we see that the average wage of employees is positively correlated with population mobility in all the cities studied. The coefficients of all cities in southern China are higher than those of the north, which means that the average wage of employees in the southern region, especially in Hunan, Guangdong, and Fujian, is more closely related to population mobility. These regions can attract population inflows by increasing the income of employees. The Beijing–Tianjin–Hebei region and central Shaanxi, Sichuan, Hubei, and other cities have a weak positive correlation coefficient. The former may mean that, even with the further increase in wages, it has not been able to attract a population inflow, while the latter group is mostly labor-outputting cities, perhaps due to the increase in local wages still not reaching the level of developed regions.

## 5. Discussion and Conclusions

Traditional census data cannot reveal the spatial patterns of population mobility and relevant socioeconomic factors within a specific period or even track people's trajectories because of the slow updating frequency and other shortcomings. Secondly, China's published population distribution statistics often have problems such as low granularity and a poor refinement level, which means daily population movement cannot be described with high spatial and temporal resolution. The spatiotemporal location big data explores the route and direction of population mobility in a relatively continuous time interval, which provides new data support for the study of population distribution and population mobility. Different from the macro model under the long-term evolution rule of statistical data, research based on travel big data can not only reflect the new characteristics of population inflow and outflow between cities and describe the agglomeration and diffusion of population flow, but also analyze the increasingly complex relationships between cities from the perspective of flow through space.

Based on the Tencent application dataset, this study first used the social network analysis method to explore the spatial and temporal distribution patterns and characteristics of population mobility during the Spring Festival, then constructed a SGWR model to reveal the socioeconomic factors related to population mobility. Different from the diamond-shaped structure proposed by Pan and Lai [62], the population mobility network during the Spring Festival presents a typical structure of two east–west main axes and three north–south main axes. The vertices of the structure are all located in the four major urban agglomerations of China, which reflects the great attraction of the above areas. The social network analysis method not only identified different community structures, but also classified all cities hierarchically, reflecting the status of different cities in the population mobility network. The results of the SGWR model show that population mobility is significantly correlated with regional average wage level, urbanization rate, foreign capital, value-added of primary industry, and value-added of secondary and tertiary industry, which is consistent with findings of Zhong [64] and Li [37].

Using refined datasets with high temporal and spatial resolution, this study explored the structural characteristics of population mobility networks and the heterogeneity of different regions in attracting populations, thus surpassing previous studies. We found that the population tends to shift from low-wage, low-input and low-development-level areas to high-wage, high-input and high-development-level areas. On the one hand, this supports the early analysis conclusion based on census data—that economic differences between regions are the main forces driving population mobility [10,11]. On the other hand, it validates the neoclassical migration theory that migrants make their decision to move as a response to interregional or rural–urban wage differentials and rational cost–benefit calculations [6,7]. The imbalance between developments in this region is due to the dual urban–rural development in the planned economy era and the priority development of the eastern region after the reform and opening-up [16,37]. The imbalance in regional development is directly

manifested in the imbalance of economic development, which intensifies the scale of population mobility and ultimately leads to a growing gap between regions. Meanwhile, such large-scale movement is also a huge challenge to the transportation system. A series of social problems such as left-behind children, poor living conditions, and urban diseases are often hidden behind the large-scale and long-term population mobility. Although we are unable to solve these problems, we can propose some positive development and management strategies through the analysis of population mobility. First, using location big data to track population activity trajectories and investigate population distribution is critical to the good operation of urban systems. Second, different economic and social development strategies should be implemented according to the development status and conditions to avoid excessive population loss and the “shrinking city” phenomenon. For major labor-outputting regions such as Sichuan and Shandong, the former can push industrial transformation to speed up the development of secondary and tertiary industries; the latter can increase the intensity of attracting foreign investment, and the central region should consider increasing the income level of local employees.

The limitations of this paper cannot be ignored. First, there will inevitably be problems such as data deviation, data discontinuity, and data loss. As mentioned earlier, Tencent has hundreds of millions of users, but there are still some people who do not use products developed by Tencent, so their travel behaviors will not be recorded. Due to privacy issues, the original dataset does not provide the social attributes of population (gender, age, occupation, and purpose), so we cannot accurately assess the purpose of the population movement (the majority of it is migrant worker flow, but there is still some student and tourism flow). The support of multi-source data and the cross-application of multi-disciplinary fields are the keys to studying cities and population in the era of big data. Second, population mobility is restricted and influenced by many complex factors. In this paper, socioeconomic factors were explored with the help of the SGWR model. Recent studies investigated amenity-led, public policy-led and tourism industry-led population mobility and indicated that the above two factors play an increasingly important role in attracting population mobility [57,65–67]. In the next stage, the relationships between the quality of life factor, amenity factor, and public factor and differences in regional population mobility need to be explored. Third, we analyzed only the 14-day population mobility in this paper because of the confidentiality of the data. It is well known that there is also a peak period of population mobility before and after the Lantern Festival (the 15th day of the 1st lunar month), which was not reflected in this paper. Therefore, in future studies, population mobility data with longer time series should be obtained and analyzed, and the results may be more representative and instructive. This paper combined the social network analysis method and the SGWR model to explore the spatiotemporal patterns and characteristics of population mobility, thereby revealing the socioeconomic factors related to population mobility. The research results can provide data support for urban policy makers and researchers, while also promoting progress in studies on population mobility.

**Author Contributions:** Conceptualization, Z.Y. and X.Z.; methodology, W.G.; formal analysis, Z.Y.; writing—original draft preparation, Z.Y.; writing—review and editing, Z.Y., X.Z. and W.G.; supervision, C.H.; software—X.X. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Appendix A



**Figure A1.** Main population direction and intensity in four sample cities before and after the Spring Festival. (a–d) represent Beijing, Chongqing, Shenzhen and Hengyang, respectively. Beijing is the outflow direction before the Spring Festival, Chongqing is the inflow direction before the Spring Festival, Shenzhen is the inflow direction after the Spring Festival, and Hengyang is the outflow direction after the Spring Festival. The closer to red line color, the higher the intensity of population flow, and vice versa.

Table A1. Sample data in this research.

Sample Data about Outflows (A Case in Beijing)							
Source	Longitude	Latitude	Target	Longitude	Latitude	Date	Mobility Intensity
Beijing	116.4093	40.1841	Chongqing	107.8643	30.0553	2019/02/04	94193
Beijing	116.4093	40.1841	Changsha	113.1529	28.2295	2019/02/04	47416
Beijing	116.4093	40.1841	Baoding	115.1686	39.0222	2019/02/04	47404
Beijing	116.4093	40.1841	Langfang	116.5358	39.1117	2019/02/04	46473
Beijing	116.4093	40.1841	Harbin	127.9629	45.6369	2019/02/04	46416
...	...	...	...	...	...	...	...
Sample Data about Inflows (A Case in Beijing)							
Source	Longitude	Latitude	Target	Longitude	Latitude	Date	Mobility intensity
Tangshan	118.3405	39.7358	Beijing	116.4093	40.1841	2019/02/04	10621
Shenyang	123.1366	42.0936	Beijing	116.4093	40.1841	2019/02/04	12807
Shanghai	121.4040	31.0844	Beijing	116.4093	40.1841	2019/02/04	69782
Hangzhou	119.4847	29.9049	Beijing	116.4093	40.1841	2019/02/04	28783
Nanchang	116.0244	28.6633	Beijing	116.4093	40.1841	2019/02/04	19356
...	...	...	...	...	...	...	...

## References

- De Haas, H. Migration and development: A theoretical perspective 1. *Int. Migr. Rev.* **2010**, *44*, 227–264. [\[CrossRef\]](#)
- Ma, H.; Chen, Z. Patterns of inter-provincial migration in China: Evidence from the sixth population census. *Popul. Res.* **2012**, *36*, 87–99.
- Ravenstein, E.G. The laws of migration. *J. Stat. Soc. Lond.* **1885**, *48*, 167–235. [\[CrossRef\]](#)
- Bogue, D.J. The use of place-of-birth and duration-of-residence data for studying internal migration. *Economics* **1960**, *9*, 97–101.
- Lewis, W.A. Economic development with unlimited supplies of labour. *Manch. Sch.* **1954**, *22*, 139–191. [\[CrossRef\]](#)
- Ranis, G.; Fei, J.C. A theory of economic development. *Am. Econ. Rev.* **1961**, *51*, 533–565.
- Sjaastad, L.A. The costs and returns of human migration. *J. Political Econ.* **1962**, *70*, 80–93. [\[CrossRef\]](#)
- Knapp, T.A.; Gravest, P.E. On the role of amenities in models of migration and regional development. *J. Reg. Sci.* **1989**, *29*, 71–87. [\[CrossRef\]](#)
- Mueser, P.R.; Graves, P.E. Examining the role of economic opportunity and amenities in explaining population redistribution. *J. Urban Econ.* **1995**, *37*, 1–25. [\[CrossRef\]](#)
- Fan, C.C. Modeling interprovincial migration in China, 1985–2000. *Eurasian Geogr. Econ.* **2005**, *46*, 165–184. [\[CrossRef\]](#)
- Liu, Y.; Shen, J. Modelling skilled and less-skilled interregional migrations in China, 2000–2005. *Popul. Space Place* **2017**, *23*, e2027. [\[CrossRef\]](#)
- Fang, C.; Dewen, W. Migration as marketization: What can we learn from China's 2000 census data? *China Rev.* **2003**, *3*, 73–93.
- Gu, C.; Cai, J.; Zhang, W.; Ma, Q.; Chan, R.; Li, W.; Shen, D. A study on the patterns of migration in Chinese large and medium cities. *Acta Geographica Sin.* **1999**, *3*, 66.
- Wang, G.; Pan, Z.; Lu, Y. China's inter-provincial migration patterns and influential factors: Evidence from year 2000 and 2010 population census of China. *Chin. J. Popul. Sci.* **2012**, *5*, 2–13.
- Yu, T. Spatial-temporal features and influential factors of the China urban floating population growth. *Chin. J. Popul. Sci.* **2012**, *4*, 47–58.
- Shen, J. Increasing internal migration in China from 1985 to 2005: Institutional versus economic drivers. *Habitat Int.* **2013**, *39*, 1–7. [\[CrossRef\]](#)
- Liao, F.H.; Wei, Y.D. Space, scale, and regional inequality in provincial China: A spatial filtering approach. *Appl. Geogr.* **2015**, *61*, 94–104. [\[CrossRef\]](#)
- Wei, Y.D.; Ye, X. Beyond convergence: Space, scale, and regional inequality in China. *Tijdschr. Voor Econ. En Soc. Geogr.* **2009**, *100*, 59–80. [\[CrossRef\]](#)

19. Liu, Y.; Stillwell, J.; Shen, J.; Daras, K. Interprovincial migration, regional development and state policy in China, 1985–2010. *Appl. Spat. Anal. Policy* **2014**, *7*, 47–70. [\[CrossRef\]](#)
20. Zhong, F.-N.; Qing, L.; Xiang, J.; Jing, Z. Economic growth, demographic change and rural-urban migration in China. *J. Integr. Agric.* **2013**, *12*, 1884–1895. [\[CrossRef\]](#)
21. He, J. The regional concentration of China's interprovincial migration flows, 1982–90. *Popul. Environ.* **2002**, *24*, 149–182. [\[CrossRef\]](#)
22. Liu, C.; Otsubo, K.; Wang, Q.; Ichinose, T.; Ishimura, S. Spatial and temporal changes of floating population in China between 1990 and 2000. *Chin. Geogr. Sci.* **2007**, *17*, 99–109. [\[CrossRef\]](#)
23. Morrill, R. Fifty years of population change in the US 1960–2010. *Cities* **2012**, *29*, S29–S40. [\[CrossRef\]](#)
24. Wang, X.-R.; Hui, E.C.-M.; Sun, J.-X. Population migration, urbanization and housing prices: Evidence from the cities in China. *Habitat Int.* **2017**, *66*, 49–56. [\[CrossRef\]](#)
25. Cao, G.; Liu, T. Rising role of inland regions in China's urbanization in the 21st century: The new trend and its explanation. *Acta Geogr. Sin.* **2011**, *66*, 1631–1643.
26. Ding, J.; Liu, Z.; Cheng, D.; Liu, J.; Zou, J. Areal differentiation of inter-provincial migration in China and characteristics of the flow field. *Acta Geogr. Sin.* **2005**, *60*, 106–114.
27. Shen, J. Changing patterns and determinants of interprovincial migration in China 1985–2000. *Popul. Space Place* **2012**, *18*, 384–402. [\[CrossRef\]](#)
28. Wang, Y.; Li, H.; Yu, Z.; Luo, B. Approaches to census mapping: Chinese solution in 2010 rounded census. *Chin. Geogr. Sci.* **2012**, *22*, 356–366. [\[CrossRef\]](#)
29. Bo, W.; Feng, Z.; Hao, Z. The dynamic changes of urban space-time activity and activity zoning based on check-in data in Sina web. *Sci. Geogr. Sin.* **2015**, *35*, 151–160.
30. Qin, X.; Zhen, F.; Zhu, S.; Xi, G. Spatial pattern of catering industry in Nanjing urban area based on the degree of public praise from internet: A case study of dianping. com. *Sci. Geogr. Sin.* **2014**, *34*, 810–817.
31. Xiao, Q.; Feng, Z.; Lifang, X.; Shoujia, Z. Methods in urban temporal and spatial behavior research in the big data era. *Prog. Geogr.* **2013**, *32*, 1352–1361.
32. Long, Y.; Zhang, Y.; Cui, C. Identifying commuting pattern of Beijing using bus smart card data. *Acta Geogr. Sin.* **2012**, *67*, 1339–1352.
33. Otte, E.; Rousseau, R. Social network analysis: A powerful strategy, also for the information sciences. *J. Inf. Sci.* **2002**, *28*, 441–453. [\[CrossRef\]](#)
34. Ivaldi, M.; Ferreri, L.; Daolio, F.; Giacobini, M.D.L.; Tomassini, M.; Rainoldi, A. We-Sport: From Academic Spin-off to Data-Base for Complex Network Analysis; An Innovative Approach to a New Technology. In Proceedings of the 3rd National Congress of Italian Society of Movement and Sport Sciences, Verona, Italy, 29 September–1 October 2011; p. 53.
35. China Internet Report. 2019. Available online: <https://www.abacusnews.com/china-internet-report> (accessed on 20 March 2019).
36. Tencent Location Big Data. Available online: <https://heat.qq.com/index.php> (accessed on 12 March 2019).
37. Li, J.; Ye, Q.; Deng, X.; Liu, Y.; Liu, Y. Spatial-temporal analysis on Spring Festival travel rush in China based on multisource big data. *Sustainability* **2016**, *8*, 1184. [\[CrossRef\]](#)
38. Wang, Y.; Dong, L.; Liu, Y.; Huang, Z.; Liu, Y. Migration patterns in China extracted from mobile positioning data. *Habitat Int.* **2019**, *86*, 71–80. [\[CrossRef\]](#)
39. Xu, J.; Li, A.; Li, D.; Liu, Y.; Du, Y.; Pei, T.; Ma, T.; Zhou, C. Difference of urban development in China from the perspective of passenger transport around Spring Festival. *Appl. Geogr.* **2017**, *87*, 85–96. [\[CrossRef\]](#)
40. Zhu, D.; Huang, Z.; Shi, L.; Wu, L.; Liu, Y. Inferring spatial interaction patterns from sequential snapshots of spatial distributions. *Int. J. Geogr. Inf. Sci.* **2018**, *32*, 783–805. [\[CrossRef\]](#)
41. Jiang, B.; Zhao, S.; Yin, J. Self-organized natural roads for predicting traffic flow: A sensitivity study. *J. Stat. Mech. Theory Exp.* **2008**, *2008*, P07008. [\[CrossRef\]](#)
42. Gupta, P.; Goel, A.; Lin, J.; Sharma, A.; Wang, D.; Zadeh, R. Wtf: The who to follow service at twitter. In Proceedings of the 22nd International Conference on World Wide Web, Rio de Janeiro, Brazil, 13–17 May 2013; pp. 505–514.
43. Girvan, M.; Newman, M.E. Community structure in social and biological networks. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 7821–7826. [\[CrossRef\]](#)
44. Clauset, A.; Newman, M.E.; Moore, C. Finding community structure in very large networks. *Phys. Rev. E* **2004**, *70*, 066111. [\[CrossRef\]](#)



45. Xu, X.; Yuruk, N.; Feng, Z.; Schweiger, T.A. Scan: A structural clustering algorithm for networks. In Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Jose, CA, USA, 12–15 August 2007; pp. 824–833.
46. Blondel, V.D.; Guillaume, J.-L.; Lambiotte, R.; Lefebvre, E. Fast unfolding of communities in large networks. *J. Stat. Mech. Theory Exp.* **2008**, *2008*, P10008. [\[CrossRef\]](#)
47. Brunson, C.; Fotheringham, A.S.; Charlton, M.E. Geographically weighted regression: A method for exploring spatial nonstationarity. *Geogr. Anal.* **1996**, *28*, 281–298. [\[CrossRef\]](#)
48. Fotheringham, A.S.; Brunson, C.; Charlton, M. *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*; John Wiley & Sons: Chichester, UK, 2003.
49. Nakaya, T.; Fotheringham, A.S.; Brunson, C.; Charlton, M. Geographically weighted Poisson regression for disease association mapping. *Stat. Med.* **2005**, *24*, 2695–2717. [\[CrossRef\]](#) [\[PubMed\]](#)
50. Bitter, C.; Mulligan, G.F.; Dall’erba, S. Incorporating spatial variation in housing attribute prices: A comparison of geographically weighted regression and the spatial expansion method. *J. Geogr. Syst.* **2007**, *9*, 7–27. [\[CrossRef\]](#)
51. Harris, P.; Brunson, C. Exploring spatial variation and spatial relationships in a freshwater acidification critical load data set for Great Britain using geographically weighted summary statistics. *Comput. Geosci.* **2010**, *36*, 54–70. [\[CrossRef\]](#)
52. Feuillet, T.; Salze, P.; Charreire, H.; Menai, M.; Enaux, C.; Perchoux, C.; Hess, F.; Kesse-Guyot, E.; Hercberg, S.; Simon, C. Built environment in local relation with walking: Why here and not there? *J. Transp. Health* **2016**, *3*, 500–512. [\[CrossRef\]](#)
53. Jin, C.; Xu, J.; Huang, Z. Spatiotemporal analysis of regional tourism development: A semiparametric Geographically Weighted Regression model approach. *Habitat Int.* **2019**, *87*, 1–10. [\[CrossRef\]](#)
54. Lee, K.H.; Heo, J.; Jayaraman, R.; Dawson, S. Proximity to parks and natural areas as an environmental determinant to spatial disparities in obesity prevalence. *Appl. Geogr.* **2019**, *112*, 102074. [\[CrossRef\]](#)
55. Liu, T.; Qi, Y.; Cao, G. China’s floating population in the 21st century: Uneven landscape, influencing factors, and effects on urbanization. *Acta Geogr. Sin.* **2015**, *70*, 567–581.
56. Cao, Z.; Zheng, X.; Liu, Y.; Li, Y.; Chen, Y. Exploring the changing patterns of China’s migration and its determinants using census data of 2000 and 2010. *Habitat Int.* **2018**, *82*, 72–82. [\[CrossRef\]](#)
57. Liu, Y.; Shen, J. Jobs or amenities? Location choices of interprovincial skilled migrants in China, 2000–2005. *Popul. Space Place* **2014**, *20*, 592–605. [\[CrossRef\]](#)
58. Zhang, K.H.; Shunfeng, S. Rural–urban migration and urbanization in China: Evidence from time-series and cross-section analyses. *China Econ. Rev.* **2003**, *14*, 386–400. [\[CrossRef\]](#)
59. Nakaya, T.; Fotheringham, S.; Charlton, M.; Brunson, C. Semiparametric Geographically Weighted Generalised Linear Modelling in GWR 4.0. In Proceedings of the GeoComputation 10th International Conference on GeoComputation, Sydney, Australia, 30 November–2 December 2009.
60. Bozdogan, H. Model selection and Akaike’s information criterion (AIC): The general theory and its analytical extensions. *Psychometrika* **1987**, *52*, 345–370. [\[CrossRef\]](#)
61. Burnham, K.P.; Anderson, D.R. Multimodel inference: Understanding AIC and BIC in model selection. *Sociol. Methods Res.* **2004**, *33*, 261–304. [\[CrossRef\]](#)
62. Pan, J.; Lai, J. Spatial pattern of population mobility among cities in China: Case study of the National Day plus Mid-Autumn Festival based on Tencent migration data. *Cities* **2019**, *94*, 55–69. [\[CrossRef\]](#)
63. Yang, J.; Wang, H.; Jin, S.; Chen, K.; Riedinger, J.; Peng, C. Migration, local off-farm employment, and agricultural production efficiency: Evidence from China. *J. Product. Anal.* **2016**, *45*, 247–259. [\[CrossRef\]](#)
64. Zhong, C.; Arisana, S.M.; Huang, X.; Batty, M.; Schmitt, G. Detecting the dynamics of urban structure through spatial network analysis. *Int. J. Geogr. Inf. Sci.* **2014**, *28*, 2178–2199. [\[CrossRef\]](#)
65. Cai, G.; Xu, L.; Gao, W.; Hong, Y.; Ying, X.; Wang, Y.; Qian, F. The Positive Impacts of Exhibition-Driven Tourism on Sustainable Tourism, Economics, and Population: The Case of the Echigo–Tsumari Art Triennale in Japan. *Int. J. Environ. Res. Public Health* **2020**, *17*, 1489. [\[CrossRef\]](#)

66. Wu, J.; Yu, Z.; Wei, Y.D.; Yang, L. Changing distribution of migrant population and its influencing factors in urban China: Economic transition, public policy, and amenities. *Habitat Int.* **2019**, *94*, 102063. [[CrossRef](#)]
67. Liu, Y.; Shen, J. Spatial patterns and determinants of skilled internal migration in C hina, 2000–2005. *Pap. Reg. Sci.* **2014**, *93*, 749–771. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).