



## Article

# Analyzing Urban Spatial Patterns and Functional Zones Using Sina Weibo POI Data: A Case Study of Beijing

Ruomu Miao <sup>1</sup>, Yuxia Wang <sup>2,\*</sup>  and Shuang Li <sup>3,\*</sup> 

<sup>1</sup> School of Media and Communication, Shanghai Jiao Tong University, Shanghai 200240, China; miaoruomu@sjtu.edu.cn

<sup>2</sup> School of Geographic Sciences, Key Lab of Geographic Information Science (Ministry of Education), East China Normal University, Shanghai 200241, China

<sup>3</sup> Institute of Chinese Historical Geography, Fudan University, Shanghai 200433, China

\* Correspondence: yxwang@geo.ecnu.edu.cn (Y.W.); li\_shuang@fudan.edu.cn (S.L.); Tel.: +86-021-5434-1218 (Y.W.); +86-021-5566-5097 (S.L.)

**Abstract:** With the development of Web2.0 and mobile Internet, urban residents, a new type of “sensor”, provide us with massive amounts of volunteered geographic information (VGI). Quantifying the spatial patterns of VGI plays an increasingly important role in the understanding and development of urban spatial functions. Using VGI and social media activity data, this article developed a method to automatically extract and identify urban spatial patterns and functional zones. The method is put forward based on the case of Beijing, China, and includes the following three steps: (1) Obtain multi-source urban spatial data, such as Weibo data (equivalent to Twitter in Chinese), OpenStreetMap, population data, etc.; (2) Use the hierarchical clustering algorithm, term frequency-inverse document frequency (TF-IDF) method, and improved k-means clustering algorithms to identify functional zones; (3) Compare the identified results with the actual urban land uses and verify its accuracy. The experiment results proved that our method can effectively identify urban functional zones, and the results provide new ideas for the study of urban spatial patterns and have great significance in optimizing urban spatial planning.



**Citation:** Miao, R.; Wang, Y.; Li, S. Analyzing Urban Spatial Patterns and Functional Zones Using Sina Weibo POI Data: A Case Study of Beijing. *Sustainability* **2021**, *13*, 647. <https://doi.org/10.3390/su13020647>

Received: 28 December 2020

Accepted: 8 January 2021

Published: 12 January 2021

**Publisher’s Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 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 (<https://creativecommons.org/licenses/by/4.0/>).

**Keywords:** urban spatial pattern; urban functional zones; Sina Weibo POI; cluster analysis; VGI

## 1. Introduction

In the era of Web2.0 and mobile Internet, people often use Weibo (equivalent to Twitter in Chinese), online comments, photo sharing, travel records, and social media to generate, process, and share a large amount of information [1–3]. With the popularization of global positioning systems (GPS) and wireless cellular positioning technology in mobile devices, most of the information spontaneously created by users automatically carries spatial information [4]. This kind of spatial information is called volunteered geographic information (VGI) in academia [5]. VGI’s real-time, diversity, and content creativity have huge application potential in the fields of spatio-temporal analysis, urban planning, environmental monitoring, disaster warning, and public information services [6–9]. These massive data are gradually being mined and analyzed, and thus people have truly entered the era of big data. Goodchild also pointed out that we are rapidly entering an era where ordinary citizens are both consumers and producers of geographic information [1].

The advent of the big data era has put forward new ideas for the study of urban spatial patterns. Currently, data based on location-based service (LBS) technology are the most widely used data in urban research, such as bus card records, taxi trajectory data, mobile phone call records, and login data based on social media [10–14]. These data can be interpreted as a description of the city, and their mining and analysis can lead to a more people-oriented urban spatial pattern [1,2]. Traditional surveying methods based on visual and statistical data have some limitations in the research process, such as being

expensive, not open to everyone, insufficient in real-time, and having data accuracy greatly affected by the environment [15,16]. They usually require more time and cost, but are more accurate, especially for small towns and when done by engaged people studying urban patterns [17,18]. Using easily accessible big data and location information can quickly obtain timely information that can describe the city, and thus is a meaningful research field in urban planning. In particular, the detailed urban functional zone classifications are of great importance for the landscape of urban spatial structure. Fine city function delineation using VGI data could depict the pre-planning layout for future cities, and this yields valuable insights for longstanding city economic development [19]. Taking the areas yet to be developed as an example, currently, they may be well equipped with public transport and industrial sites, however, balancing the provision between jobs and housing to facilitate healthy polycentric growth is crucial for sustainable and efficient urban development [20,21].

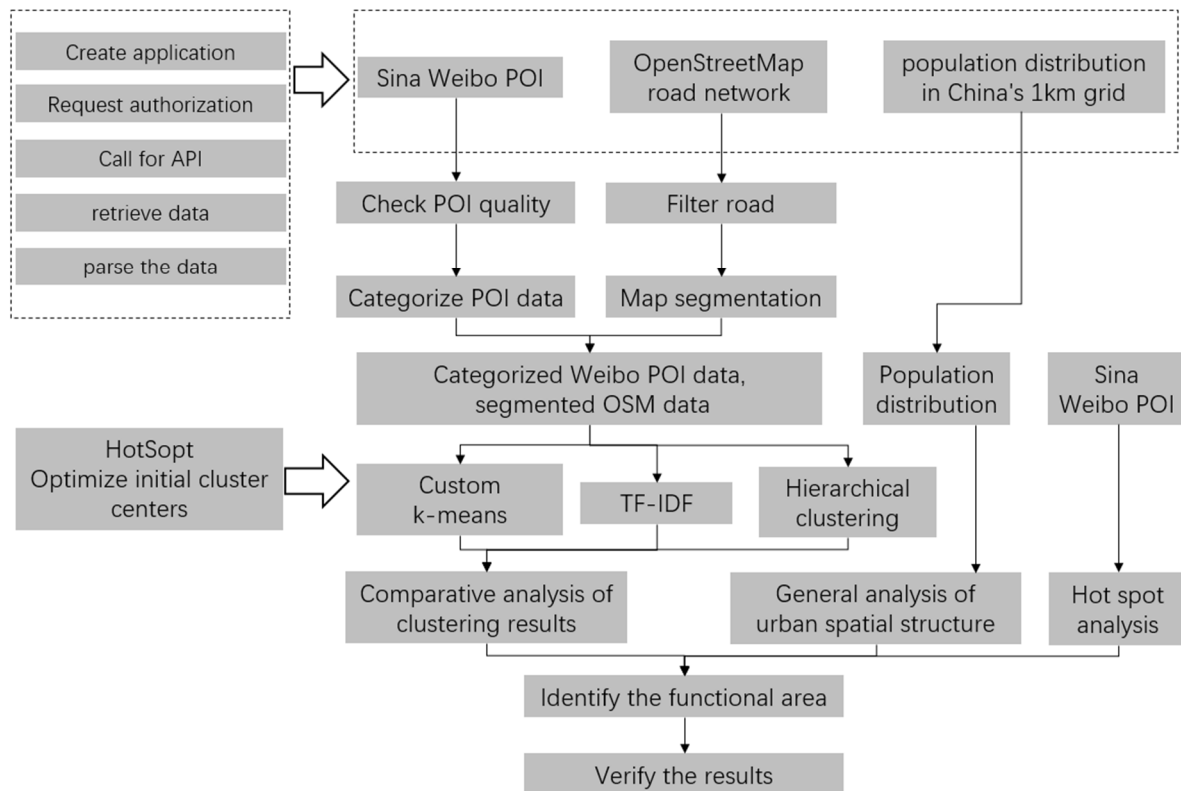
Many urban researchers have done related work using multi-source VGI data. Patrick Lüscher et al. [22] used an iterative online questionnaire to construct a city center cognitive model, and divided the city into different functional zones using the kernel density analysis method for the points of interests (POI) data obtained from the British Army Survey. Jameson L. et al. [23] used mobile phone data to analyze the spatio-temporal dynamics of the population, and analyzed land use patterns based on machine learning classification algorithms. John Steenbruggen et al. [24] comprehensively reviewed the research on mobile phone data and emphasized the feasibility of using digital data to optimize city management. Vincent Blondel [25] used more than 200 million communication data to study the corresponding areas and boundaries, and finally proposed geographic mobile communication based on the communication frequency and its average duration. Yang used [26] Baidu POI data to analyze the spatial composition of the urban network and divided it into 12 functional zones to analyze its aggregation mode. Liu et al. [27] used one-week taxi trajectory data in Shanghai, and used the source-sink model to quantify daily traffic characteristics and then discover the urban land use functions. Long et al. [28] used OpenStreetMap road network data and crowdsourced POI data to distinguish urban residential areas, and then they used census data to integrate population attributes. Rao et al. [29] used Shenzhen mobile phone data for one week to analyze user's spatio-temporal attributes and proposed a model to identify different zones of employment in Shenzhen.

It can be seen that a lot of research has been conducted on urban spatial patterns and functional zones using mobile phone data, taxi data, and other POI data. This kind of VGI data are large-volume, easily obtainable, time-saving, and more people-oriented than traditional datasets, and their application in delineating city functional zones could provide more detailed information. Therefore, taking Beijing (the capital of China) as the study area, the focus of this study was to automatically extract and identify the urban spatial patterns and functional zones in Beijing using Sina Weibo data, OpenStreetMap, and other data. Through the theoretical and experimental research in this article, the following contributions were made: Firstly, the feasibility of using user-generated social media data on investigating urban spatial structures was verified; Secondly, by dividing research units using the road network, we obtained the natural areas in Beijing; Finally, the automatically-identified urban functional zones using social media data provided more information than did generally-defined residential or employment areas. The results of this study can help people better understand the spatial composition of large cities and megacities, and can also assist urban planners in planning urban functional zones using POI data. Our study has new reference value for urban planning, and can also provide reference for the development and improvement of different urban functional zones.

The reminder of this paper is organized as follows: The Materials and Methods section describes the study area, data we collected, and the method used for analyzing urban spatial structure. The Results and Discussion section presents the experiments and results, and discusses next steps. Finally, the paper ends with the Conclusions section.

## 2. Materials and Methods

In this section, we describe our study area and present the data we collected (including data pre-processing). Then, we show how to conduct hotspot analysis based on these data and use clustering methods to identify urban functional areas. The specific research process is shown in Figure 1.



**Figure 1.** Flowchart of our study. (API is short for Application Programming Interface; POI is short for Point of Interest; OSM is short for OpenStreetMap; TF-IDF is short for Term Frequency–Inverse Document Frequency).

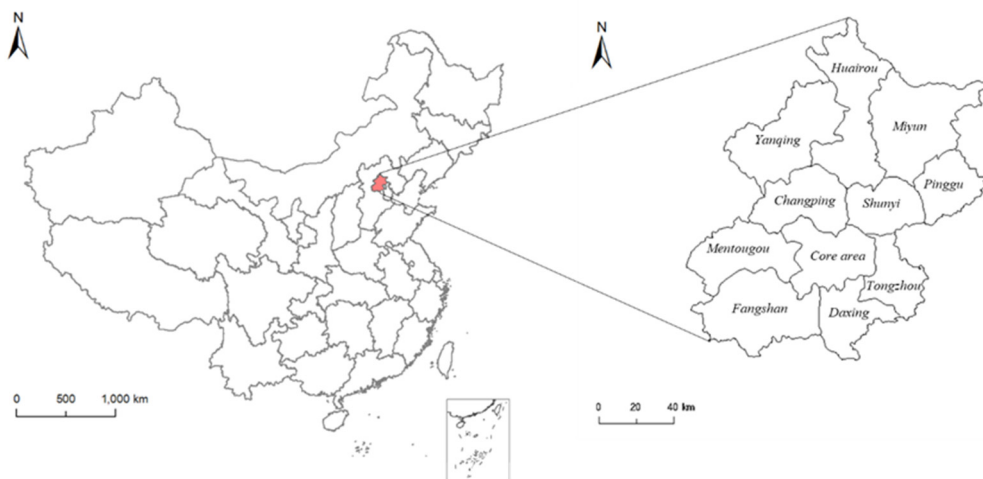
### 2.1. Study Area

The study area was Beijing, China ( $115^{\circ}24'39''$ – $117^{\circ}30'37''$  E,  $39^{\circ}26'9''$ – $41^{\circ}3'32''$  N, as shown in Figure 2). Beijing is the capital and also a typical megacity in China. With the rapid urbanization, its urban scale has been expanded 12 times in 55 years [30]. The total area is 16,410.54 km<sup>2</sup>, and the permanent population is 21.53 million. Considering the complexity of urban space, the large population (who can act as sensors), and even the increasingly prominent problem of big city diseases, Beijing is an ideal study area.

### 2.2. Data Collection

#### 2.2.1. Sina Weibo POIs and Data Categorization

As one of the most popular social media platforms in China, Sina Weibo has the characteristics of fast updates, a large number of participants, and widely distributed users [3,9]. Most of the information on Sina Weibo is closely related to urban life. Since the content and types of Weibo POI are very rich, it is best to determine its category before acquisition. Considering the research content and the special background of Beijing, we then divided Weibo POI data into 15 categories (Table 1, modified according to [31]).



**Figure 2.** The location of the study area in China: Beijing.

**Table 1.** POI categories and their descriptions.

Code	POI Category	Description
01	Hotel	Hotels, guesthouses, inns, etc.
02	Restaurants and drinking	Restaurants, KFCs, McDonald's, Pizza Huts, cafes, etc.
03	Shopping	Shopping malls, shopping centers, shops, convenience stores, supermarkets, specialty stores, pedestrian streets, etc.
04	Tourist attraction	Scenic spots, resorts, parks, squares, zoos, botanical gardens, churches, etc.
05	Healthcare	Hospitals, clinics, emergency centers, pharmacies, etc.
06	Building (including but not limited to companies)	Office buildings, villas, industrial parks, enterprises, companies, etc.
07	Financial and insurance	Banks, ATMs (Automated Teller Machine), insurance offices, security offices, finance offices, etc.
08	Residential	Residential, bathing, laundry, beauty salons, car washes, business halls, express services, etc.
09	Public facility	Newsstands, public telephones, public toilets, post offices, etc.
10	Government agency	Government agencies, embassies, institutions, procuratorates, courts, offices, etc.
11	Industrial site	Factories, farms, fisheries, forest farms, pastures, etc.
12	Public transport	Airports, railway stations, bus stations, subway stations, parking lots, etc.
13	Highway	Expressways, toll stations, gas stations, service areas, etc.
14	Sport and entertainment	Stadiums, football fields, tennis courts, basketball courts, badminton courts, fitness centers, entertainment centers, KTV (Karaoke TV), discotheques, bars, chess rooms, Internet cafes, movie theaters, etc.
15	Science and education	Universities, schools, libraries, research institutes, science and technology museums, historical museums, exhibition halls, conference centers, art galleries, cultural palaces, archives, television stations, newspapers, publishing houses, magazines, theaters, etc.

According to coding classifications, different categories of data were collected. The collection time was from 13 April to 17 April 2015, and a total of 335,234 pieces of data were collected. Among them, the data volumes for codes 01 through 15 were 12,491; 12,152; 29,001; 13,405; 60,863; 113,206; 10,725; 34,175; 658; 5341; 4017; 14,036; 1696; 6690; and 16,778, respectively (shown in Figure 3). However, there were problems such as duplication of data records and ambiguity in place names, which required further data cleaning. Next, we deleted duplicate records and deleted records that did not meet a specific classification.

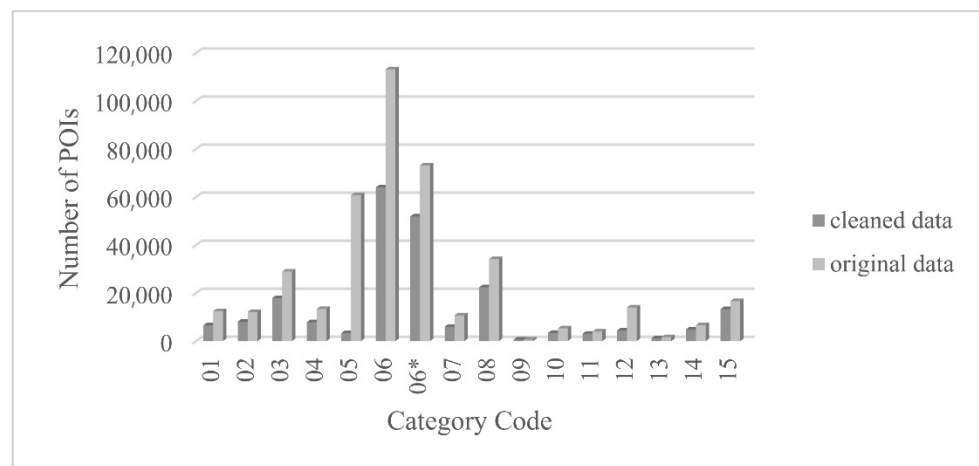


Figure 3. Statistics of the POIs.

It can be seen that in the process of categorization, some parks were also classified as tourist attractions. This is because Beijing has a large number of tourists, and they often visit parks. Therefore, the parks were classified as tourist attractions rather than public facilities. The number of companies in Beijing is high (73,224 companies out of 113,206 buildings). Therefore, in the following research, companies were not considered in the category of building, but the distribution of companies was studied separately. After data cleaning and processing, 51,916 company data points and 115,616 classified data points were finally obtained (company was categorized as 06\*). The POI data of each category is shown in Figure 3.

### 2.2.2. OpenStreetMap and Map Segmentation

We collected the road network data of Beijing on 14 April 2015 from OpenStreetMap (<https://www.openstreetmap.org>). The road data set included 50,816 roads with a total length of 24,877,717 m, and a railway with 5121 sections and a total length of 3,699,118 m. Then, we combined OpenStreetMap's road network classification and selected three levels: highway (motorway\_link), trunk (trunk\_link), and primary (primary\_link) as the research objects. There were 9655 lines with a total length of 6,077,240 m, as shown in Figure 4a. These three different road types constitute the natural division of Beijing. It can be seen intuitively that the outline of Beijing's road network meets the experimental requirements.

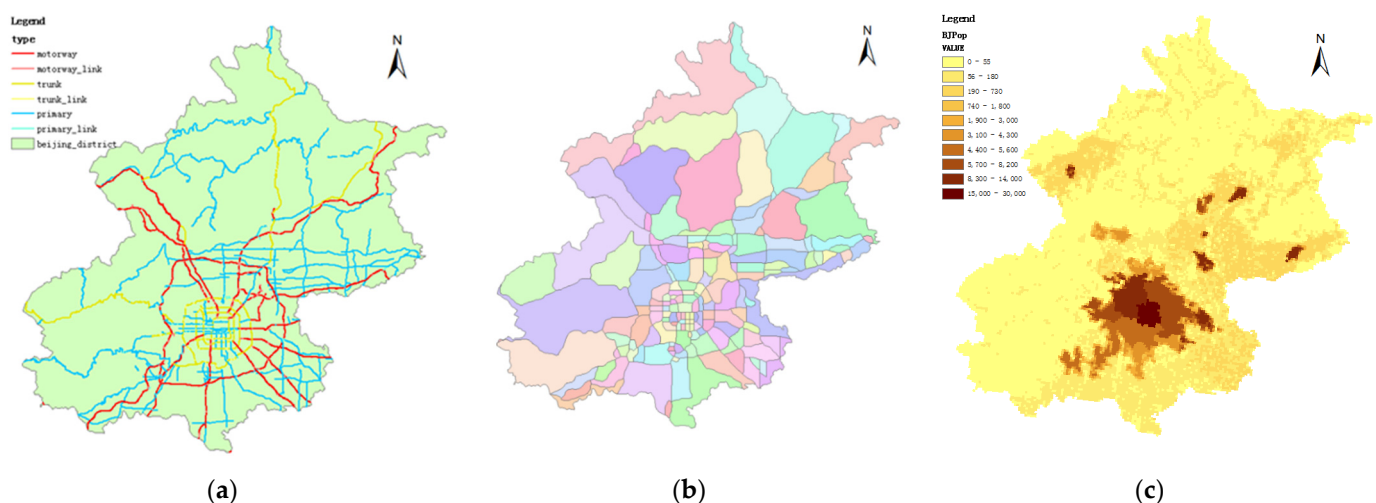


Figure 4. OpenStreetMap and population data of Beijing: (a) Selected OSM data; (b) Map segmentation results; (c) The spatial distribution density of the population in Beijing.



In order to better divide the research area into different zones, we needed to remove unnecessary details and ensure the topological relationships of the roads, including multi-lane merging, two-lane road centerline extraction, overpass deletion, and topological relationship correction. After checking the data, we segmented regions according to the center line of the road network (Figure 4b).

### 2.2.3. Population Data

China's 1 km grid population data set was based on land use type data and demographic data obtained from remote sensing data. The data set was used to establish a population spatial distribution model by using the spatial analysis function of the geographic information system to spatialize the statistical population data [32]. We extracted the population distribution data within the border of Beijing from China's 1 km grid population data set (2010). The generated population density distribution map is shown in Figure 4c.

It can be seen from the above analysis that the population density is the highest in central Beijing. As the urban center expands, the population density gradually decreases. However, in the suburbs, the population density shows a high-density distribution center in a small area, showing obvious characteristics of suburbanization. In addition, the population density in the core areas of suburban counties remains high. Generally speaking, the population density in the east is higher than that of the west, especially in areas where the population density distribution is expanding in the southeast and Langfang.

## 2.3. Analysis of Urban Spatial Structure

### 2.3.1. Analyzing Urban Hot Spots Based on Weibo POI Data

Weibo POI data can better describe the distribution of people in a city through the location information of volunteers. In order to further analyze the distribution characteristics, we selected a large number of checked-in POIs in Weibo for analysis, and used the checkin\_num of each POI point as a weight to analyze the kernel density.

The kernel density estimation (KDE) algorithm mainly uses a moving unit (equivalent to a window) to estimate the density of a point or line pattern [33]. It is defined as  $x_1 \dots x_n$  and is an independent and identically distributed sample drawn from the population of the distribution density function  $f()$ . To estimate the value of  $f()$  at a certain  $x$ , the Rosenblatt-Parzen kernel estimation is usually used:

$$f_n(x) = \frac{1}{nh} \sum_{i=1}^n k\left(\frac{x - x_i}{h}\right) \quad (1)$$

where  $k()$  is the kernel function;  $h > 0$  is the variable; and  $(x - x_i)$  represents the distance from the estimated point to the sample  $x_i$ . In KDE estimation, the determination or choice of variable  $h$  has a great influence on the calculation result. When  $h$  increases, the point density changes more smoothly in space, but it will hide the density structure; When  $h$  decreases, the estimated point density changes suddenly and unevenly [34].

In the KDE module of ArcGIS, the default bandwidth is automatically generated. The larger the search radius value, the smoother is the density grid generated and the higher is the generalization degree; therefore, the smaller the value, the more detailed is the information displayed in the generated grid. In order to obtain more detailed results, we changed the default search radius to 1500 m and the output cell size of the raster image to 100 m.

### 2.3.2. Identifying Urban Functional Zones

In this section, we used Sina Weibo POI data to analyze urban functional zones.

Cluster analysis is a statistical analysis method for studying classification problems, and it is also an important algorithm for data mining. In this research, we mainly used the k-means algorithm and hierarchical clustering algorithm.

- K-means: For a given data set, we made the following provisions: the set of  $n$  dimensional points was  $X = \{x_i, i = 1, \dots, n\}$ ; the set of  $k$  clusters was  $C = \{c_k, k = 1, \dots, k\}$ ; the mean value of  $c_k$  was  $\mu_k$ ; and the squared error was  $J(c_k) = \sum_{x_i \in c_k} \|x_i - \mu_k\|^2$ .

Therefore, the goal of K-means can be understood as a solution that minimizes  $J(c_k) = \sum_{x_i \in c_k} \|x_i - \mu_k\|^2$ .

- Hierarchical clustering algorithm: A hierarchical clustering method is used to construct and maintain a clustering tree formed by clusters and sub-clusters according to a given distance measurement criterion between clusters until a certain end condition is met. Hierarchical clustering algorithm is divided into condensed and split, from bottom-up and top-down, according to hierarchical decomposition. The default discussed in this article is cohesive.

TF-IDF (term frequency-inverse document frequency) is a statistical method used to evaluate the importance of a word to one of the documents in a document set or corpus [35]. The importance of a word is proportional to the number of times it appears in the document, but it decreases inversely proportional to the frequency of its appearance in the corpus. TF-IDF is  $TF \times IDF$ , where TF is term frequency (term frequency) and IDF is inverse document frequency (inverse document frequency). In a given document, TF refers to the frequency of a given word in the document,

$$tf_{ij} = \frac{n_{i,j}}{\sum_k n_{k,j}} \quad (2)$$

where  $n_{i,j}$  is the number of occurrences of the word in the file  $d_j$ , and the denominator is the sum of the number of occurrences of all words in the file  $d_j$ .

IDF is used to measure the universal importance of a word. The IDF of a specific word can be obtained by dividing the total number of documents in the research by the number of documents containing the word, and then taking the logarithm of the obtained quotient,

$$idf_i = \log \frac{|D|}{|\{j : t_i \in d_j\}|} \quad (3)$$

where  $|D|$  is the total number of documents in the corpus, and  $|\{j : t_i \in d_j\}|$  is the number of documents containing word.

Then, according to  $tfidf_{ij} = tf_{ij} \times idf_i$ , with a high word frequency in a particular file, and a low file frequency of the word in the entire file set, a high-weight TF-IDF can be generated.

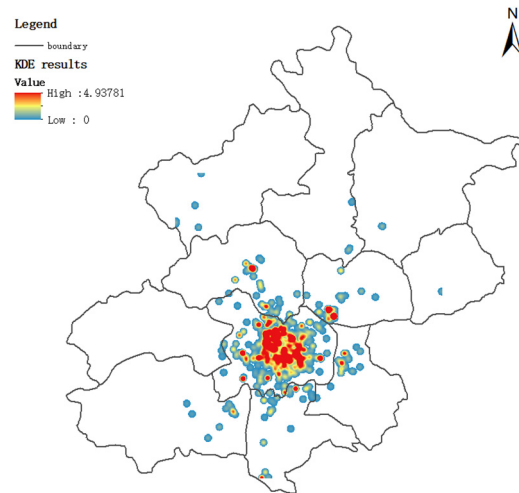
### 3. Results and Discussion

#### 3.1. Weibo Hot Spots Analysis Results

We selected a large number of Weibo POIs (with each category) for analysis, and used the checkin\_num kernel density analysis weight of each POI point to perform kernel density analysis, and obtained the following results.

It can be seen from Figure 5 that the spatial distribution of Weibo users in Beijing is large. In the urban area, it is mainly concentrated in science and education areas, commercial and entertainment areas, and diplomatic and political areas. It is not difficult to understand that there are a large number of universities in science and education areas. College students are an active group of Weibo users. At the same time, office people also like to use Weibo when commuting. In diplomatic and political areas, hot spots are mainly concentrated in tourist attractions of political significance with Tiananmen Square as the center. In commercial and entertainment areas, people mainly use Weibo to share information during leisure or entertainment activities. In addition to the urban area, the Capital International Airport District and Changping District are also hot spots for Weibo user activities. In Changping District, the campuses of some colleges and universities are relatively concentrated, and it is also the area where the Great Wall (Badaling Great Wall)

is located. It is a hot spot for people to sign in on Weibo. For the results of kernel density analysis, the data can be used for further interpretation. We selected the POI points with high *check-ins* numbers to display, as shown in the Table 2.



**Figure 5.** Kernel density analysis results of Weibo POI data.

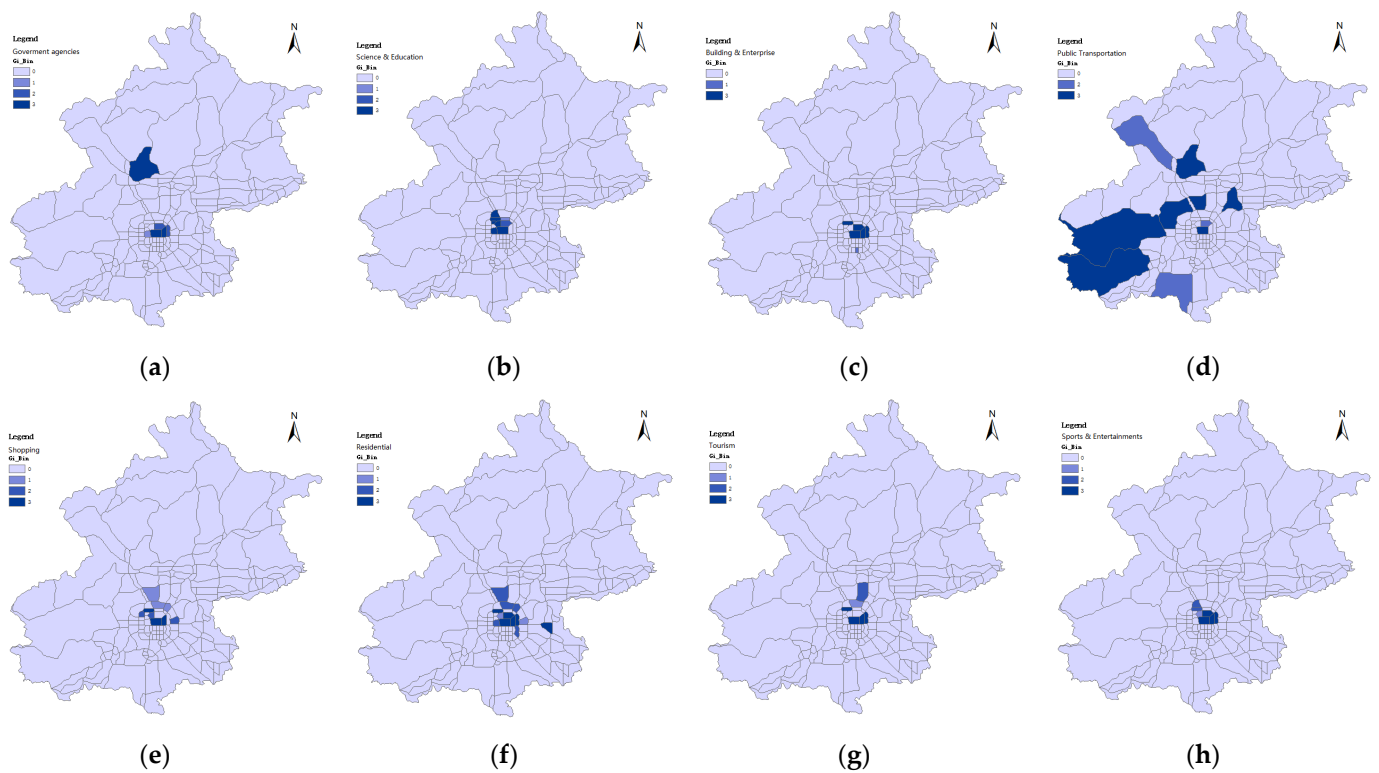
**Table 2.** POI points with high check-in times.

Checkin_num (Number of Checkin Points)	Title
150255	Capital Airport T3 Terminal
90515	Capital Airport T2 Terminal
76175	Weigong Village
69227	Beijing Normal University
67681	Beijing University
64146	Wangfujing
64146	Beijing University of Aeronautics and Astronautics
63287	Beijing Jiaotong University
62810	Tsinghua University
58960	Xidan
58136	University of Science and Technology
56575	Tiananmen Square
51035	Changxindian District
49570	Capital Airport
47521	Communication University of China

### 3.2. Identifying Urban Functional Zones

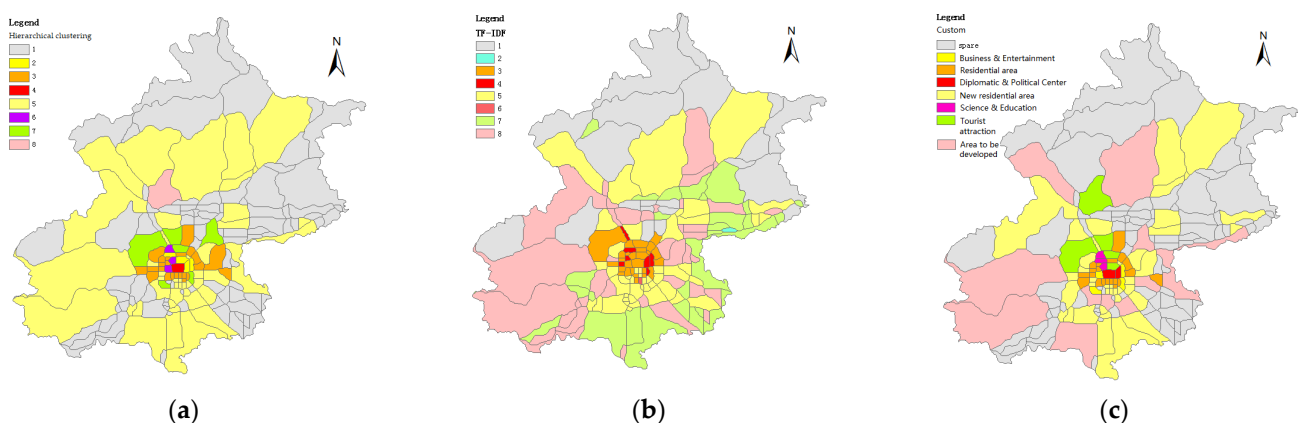
For the 15 categories of POI data we classified, we first used the spatial connection tool in ArcGIS to calculate the number of POI points in each divided area. Furthermore, the hot spot discovery tool was used to detect cluster centers. We selected eight typical categories of POI data to determine the clustering centers (as shown in Figure 6), and the distribution of hot spots obtained by ArcGIS.





**Figure 6.** Eight categories of POI hot spots: (a) government agencies; (b) science and education; (c) buildings; (d) public transport; (e) shopping; (f) residential; (g) tourist attractions; (h) sports and entertainments.

In specific experiments, we mainly used three methods, the hierarchical clustering method (Figure 7a), the TD-IDF method (Figure 7b), and the improved k-means clustering method (Figure 7c). The improved k-means method takes the aforementioned hotspot analysis results as the initial clustering center, thus expecting a better clustering result. The TF-IDF method compares the urban function exploring it as text-topic discovery, and this urban function similarity is further explored using a plain k-means method. We analyzed and compared these clustering results.



**Figure 7.** Clustering results: (a) hierarchical clustering; (b) TF-IDF; (c) custom k-means clustering.

By counting the POI data of each functional zone, we sorted the number of various categories of POIs in the functional zone, as shown in Table 3. We then comprehensively analyzed the three clustering results and statistical data, and finally determined eight categories of functional zones, including diplomatic and political centers, science and edu-

cation areas, mature residential areas, new residential areas, commercial and entertainment areas, tourist attraction areas, areas to be developed, and unclassified areas.

**Table 3.** POI category ranking value in each functional zone.

POI Category	1 (Functional Zone)	2	3	4	5	6	7
Restaurants and drinking	3	7	1	2	5	4	6
Highway	6	5	4	1	7	2	3
Industrial site	6	7	4	1	3	5	2
Public transport	5	7	3	1	6	4	2
Public facility	6	4	1	3	5	2	7
Shopping	3	4	1	2	7	5	6
Financial and insurance	3	6	1	2	5	4	7
Residential	4	7	1	2	5	3	6
Science and education	6	1	2	3	5	4	7
Tourist attraction	2	3	1	4	5	6	7
Sport and entertainment	2	6	1	3	5	4	7
Healthcare	3	7	1	2	5	4	6
Government agency	4	7	2	1	6	3	5
Hotel	3	6	1	2	5	4	7
Buildings (including but not limited to companies)	2	3	1	5	4	6	7

- Diplomatic and political zone

In these areas, a large number of embassies are gathered, and the number of POIs in tourist attractions, sports and entertainment, and buildings is large. Combining the fact that Beijing is also the capital, this area is not only the gathering place of embassies, but also the location of Tiananmen Square, the Forbidden City, and the Great Hall of the People, etc.

- Science and education zone

In these areas, POI data for science, education, culture, and publicity are the highest, and combined with the location of the area, it can be seen that there are a large number of universities in this area, such as Peking University and Tsinghua University. At the same time, Zhongguancun, China's earliest high-tech development center, has a large number of high-tech companies and scientific research institutes in these areas. Therefore, there are a large number of building and companies in these areas.

- Mature residential zone

In these areas, the number of residential POIs is the largest, and the number of restaurants, public facilities, shopping centers, financial and insurance, tourist attractions, sports and entertainments, and healthcare POIs are also the highest. It can be seen that in mature residential areas, all types of service facilities are the most complete. They are distributed around the core functional areas of the city. At the same time, very few areas in the suburbs have developed into mature residential areas.

- New residential zone

As seen in the mature residential areas, the residential POI of new residential areas is the largest of all categories, but the number of other categories in this functional zone is mostly lower than that of mature residential zones. Highway services, industrial sites, public transportation, and government agencies have the highest number of POIs. This is because this area is composed of many sub-regions with a large number of government agencies. In addition, it is located in the suburbs, and has more highway services and public transportation.

- Commercial and entertainment zone

This functional zone is located near the diplomatic and political zone, and next to the mature residential zone, which shows people's shopping habits. However, the number of

various categories in this zone is balanced, and the number in the same category is not high, which is mainly due to the small number of sub-regions.

- Tourist attractions zone

In this functional zone, there are many public transportation POIs. It can be seen from the distribution of sub-regions that the functional zone is basically distributed in the suburbs, but the number of tourist attractions POIs is not particularly large.

- Area to be developed

In these areas, all types of Checkin\_num are small, but the number of public transportation and industrial sites is large. At the same time, it can be seen from the distribution that they are adjacent to new residential areas and located in remote counties.

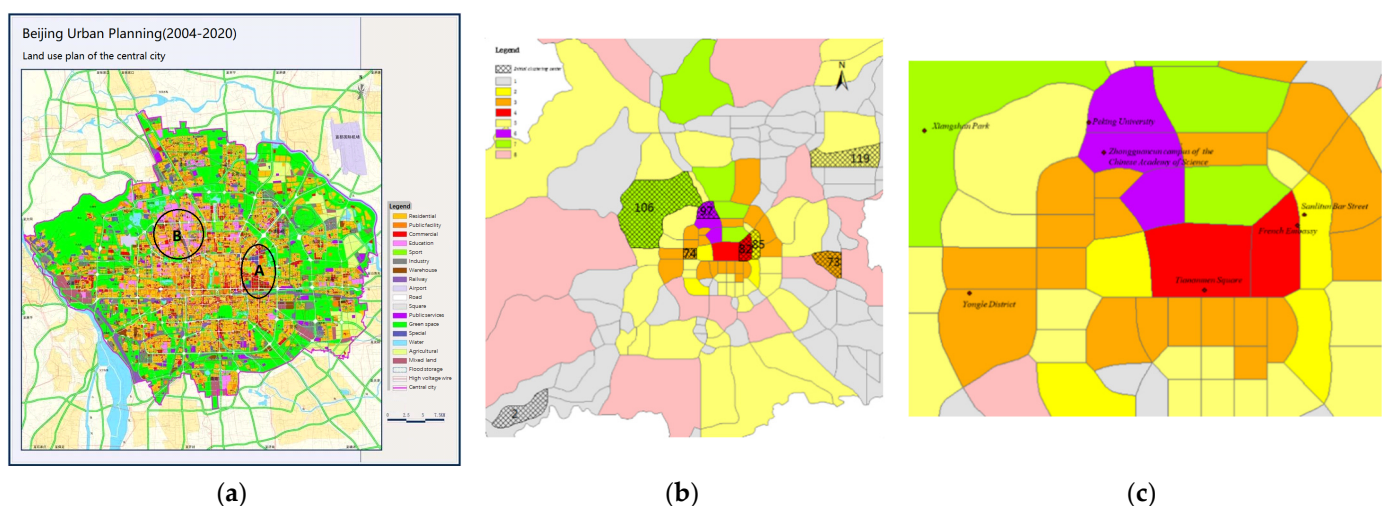
- Unclassified area

Since Weibo POI check-in data are essentially volunteer geographic information, the number of POIs in some areas is not high enough and they are not classified.

### 3.3. Verifying the Results

For the results of functional zones obtained by clustering, we evaluated them using the following three measures.

Firstly, we compared the clustering results with the Beijing City Master Plan (2004–2020, shown in Figure 8a), mainly comparing the downtown area. It can be clearly seen from the planning map that area A is land for commercial and financial use, and area B is land for science, teaching, and research, which is in full agreement with the results of this article. In addition, it can be seen from the planning map that the downtown area is a residential area, which is not inconsistent with functional zones such as the diplomatic and political area that we derived. Because residential areas have dominant functions in cities, and functions are established by human activities in the residential environment.



**Figure 8.** Verifying the results: (a) Beijing City Master Plan (2004–2020); (b) Verification of initial clustering centers; (c) Verification of typical areas.

Secondly, we compared the clustering results with the initial cluster centers of k-means (shown in Figure 8b). Because the clustering result of the k-means method itself depends on the initial clustering center, we started from the clustering method for comparative verification. It can be seen from the above that the eight initial clusters selected here fell within the corresponding local area, which can be seen in the selection of the initial cluster center. The method here is effective and the results are also reliable.

Finally, we selected some typical areas to verify the results (outer areas were not selected for comparison because they were mainly unclassified areas and areas to be developed). The results (shown in Figure 8c) show that Xiangshan Park is a tourist attraction in several typical areas selected at random. Yongle District is located in a mature residential area; Peking University and the Zhongguancun campus of the Chinese Academy of Sciences are located in the science, education, and cultural district. The French Embassy and Tiananmen Square are located in the diplomatic and political center. Sanlitun Bar Street is in the commercial and entertainment district.

Combining the above three verification methods, and considering the current situation of Beijing's highly mixed land use, it can be seen that the results of Beijing's functional zoning obtained by this method had great accuracy.

#### 3.4. Discussion

The significance of this work is the development a method to automatically identify detailed spatial functional zones. The fine distinction between mature and new residential zones, and the delineation of areas to be developed are of greater importance, except for the easily distinguishable zones (diplomatic and political, science and education, commercial and entertainment). Taking mature residential zones as reference, the new residential zones need to place more effort into promoting service-related facilities, including shopping, financial and insurance, sport and entertainment, and healthcare facilities. As for the areas to be developed, they distinguish themselves by highly ranked public transport and industrial sites, and they also next to the new residential zones in the suburban areas. This kind of functional zone is of great potential and places near center zones would enjoy better development if the infrastructures there were gradually improved.

We then combined the characteristics of the study area and the research results for further analysis.

Firstly, the central city of Beijing is showing a trend of suburbanization, and the spatial distribution structure presents a three-level structure of main center-sub center-town. However, despite the significant increase in population density in the suburbs, the construction of various infrastructures in the area is not yet complete, and the level of urbanization needs to be improved.

Secondly, Beijing is developing towards the southeast. It can be seen that the connection area between Beijing and Tianjin (Langfang, located in the southeast of Beijing) has a large population and spatial distribution density. At the same time, the distribution of Weibo POI also shows that the regional distribution density in the southeast direction is large.

Finally, diplomacy and politics; business and entertainment; and science, education, and cultural are the main service functions of the major urban areas. Mature residential areas are located near the city center. In the suburbs and counties of Beijing, there are new residential areas and areas to be developed. Commercial and entertainment areas are less distributed in suburban counties.

With the process of urbanization, the built-up area of Beijing has become larger and larger, and more and more people live in the suburbs. On the one hand, the process of suburbanization has eased the pressure on population, traffic, and housing in the major city, but at the same time many new problems have emerged. For example, many people live in the suburbs but work in the city center, which requires a long time to commute. On the other hand, it can be seen from the results of the analysis that the infrastructure in the suburbs is still not sound, so that the schooling and medical problems of children cannot be well addressed.

What should be done? In the process of ensuring the stable development of central urban areas, the development of emerging urban areas should also be balanced and attention should be paid to the equal distribution of resources, such as education, healthcare, and other supporting facilities. In addition, while optimizing the internal structure of the city, it is necessary to integrate Beijing's overall resources for external development, actively

drive the surrounding areas, and strive to achieve the coordinated development of the Beijing-Tianjin-Hebei metropolitan area.

#### 4. Conclusions

Currently, urban residents provide massive VGI, and understanding of the urban spatial pattern plays an increasingly important role in promoting urban spatial development. Using VGI and social media activity data, this article developed a method to automatically extract and identify urban spatial patterns and functional zones. We obtained a total of 167,532 Weibo POI data points in Beijing from 13 April to 17 April 2015, OpenStreetMap road network data on 14 April 2015, and China's 1 km grid population data set. Then, we used the hierarchical clustering algorithm, TF-IDF method, and improved k-means clustering algorithms and identified eight functional zones. The functional zones included the diplomatic and political zone, science and education zone, mature residential zone, new residential zone, commercial and entertainment zone, tourist attractions zone, areas to be developed, and unclassified areas. Finally, we verified the results of the study with the Beijing city master plan and typical areas, and the comparison shows that the clustering results had high accuracy.

The contributions of this work lie in three aspects. Firstly, the feasibility of using user-generated social media data on investigating urban spatial structures was verified. This kind of VGI data are large-volume, easily obtainable, more time-saving, and more people-oriented than traditional datasets, and their application in delineating city functional zones could provide more detailed information. Secondly, by dividing research units using the road network, we obtained the natural areas in Beijing. This street map segmenting method was more consistent with urban function division and was more effective in depicting city heterogeneities than was the urban uniform grid. Lastly, the automatically-identified urban functional zones using social media data provided more information than did generally-defined residential or employment areas. The advantage of mature residential zones over new residential zones provides us with useful information for the future planning of the newly developed areas and areas to be developed, so that sustainable development might be utilized for the creation of well-developed center zones. In general, the use of Weibo POI data and OpenStreetMap road network data combined with spatial clustering methods to analyze the urban spatial structure and explore functional areas, provides new ideas for the study of urban spatial structure.

**Author Contributions:** Conceptualization, R.M., Y.W., and S.L.; methodology, Y.W.; software, Y.W. and S.L.; validation, R.M., Y.W., and S.L.; formal analysis, Y.W.; investigation, R.M.; resources, S.L.; data curation, Y.W.; writing—original draft preparation, R.M.; writing—review and editing, R.M. and S.L.; visualization, R.M.; supervision, Y.W.; project administration, S.L.; funding acquisition, S.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by a grant from Beijing Key Laboratory of Spatial Development for Capital Region, the National Natural Science Foundation of China, No. 42001184, and the general project of "The Great Wall of Commerce of UFIDA Foundation", No. 2020-Y01.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data presented in this study are available on request from the corresponding author.

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

#### References

1. Goodchild, M.F. Citizens as sensors: The world of volunteered geography. *GeoJournal* **2007**, *69*, 211–221. [[CrossRef](#)]
2. Marti, P.; Serrano-Estrada, L.; Nolasco-Cirugeda, A. Social Media data: Challenges, opportunities and limitations in urban studies. *Comput. Environ. Urban Syst.* **2019**, *74*, 161–174. [[CrossRef](#)]
3. Weibo. Available online: <https://www.weibo.com> (accessed on 28 December 2020).



4. Peng, X.; Bao, Y.; Huang, Z. Perceiving Beijing's "city image" across different groups based on geotagged social media data. *IEEE Access* **2020**, *8*, 93868–93881. [[CrossRef](#)]
5. Jonietz, D.; Antonio, V.; See, L.; Zipf, A. Highlighting current trends in Volunteered Geographic Information. *ISPRS Int. J. Geo-Inf.* **2017**, *6*, 202. [[CrossRef](#)]
6. Noulas, A.; Scellato, S.; Lathia, N.; Mascolo, C. A Random Walk around the City: New Venue Recommendation in Location-Based Social Networks. In Proceedings of the 2012 International Conference on Privacy, Security, Risk and Trust and 2012 International Conference on Social Computing, Amsterdam, The Netherlands, 3–5 September 2012; IEEE: New York, NY, USA, 2012; pp. 144–153.
7. Khan, N.U.; Wan, W.; Yu, S. Location-based social network's data analysis and spatio-temporal modeling for the mega city of Shanghai, China. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 76. [[CrossRef](#)]
8. Sun, Y.; Fan, H.; Li, M.; Zipf, A. Identifying the city center using human travel flows generated from location-based social networking data. *Environ. Plan. B Plan. Des.* **2015**, *43*, 480–498. [[CrossRef](#)]
9. Zhang, X.; Sun, Y.; Zheng, A.; Wang, Y. A new approach to refining land use types: Predicting Point-of-Interest categories Using Weibo check-in data. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 124. [[CrossRef](#)]
10. Chen, W.; Liu, W.; Ke, W. The spatial structures and organization patterns of China's city networks based on the highway passenger flows. *Acta Geogr. Sin.* **2017**, *72*, 224–241.
11. Consterdine, E.; Everton, A. European migration network: Immigration of international students to the EU: Empirical evidence and current policy practice. *Science* **2012**, *290*, 1768–1771.
12. Fonte, C.C.; Minghini, M.; Patriarca, J.; Antoniou, V.; See, L.; Skopeliti, A. Generating up-to-date and detailed land use and land cover maps using OpenStreetMap and GlobeLand30. *ISPRS Int. J. Geo-Inf.* **2017**, *6*, 125. [[CrossRef](#)]
13. Liu, W.; Hou, Q.; Xie, Z.; Mai, X. Urban network and regions in China: An analysis of daily migration with Complex Networks Model. *Sustainability* **2020**, *12*, 3208. [[CrossRef](#)]
14. Liu, Y.; Sui, Z.; Kang, C.; Gao, Y. Uncovering patterns of inter-urban trip and spatial interaction from social media check-in data. *PLoS ONE* **2014**, *9*, e86026. [[CrossRef](#)] [[PubMed](#)]
15. Yang, G.; Han, Y.; Gong, H.; Zhang, T. Spatial-temporal response patterns of tourist flow under real-time tourist flow diversion scheme. *Sustainability* **2020**, *12*, 3478. [[CrossRef](#)]
16. Long, Y.; Shen, Z.J.; Mao, Q.Z. An urban containment planning support system for Beijing. *Comput. Environ. Urban Syst.* **2011**, *35*, 297–307. [[CrossRef](#)]
17. Lucchi, E.; Alonzo, V.D.; Exner, D.; Zambelli, P.; Garegnani, G. A density-based spatial cluster analysis supporting the Building Stock Analysis in Historical Towns. In Proceedings of the 16th IBPSA International Conference and Exhibition, Rome, Italy, 2–4 September 2019; pp. 3831–3838.
18. Wang, J.H.; Deng, Y.; Song, C.; Tian, D.J. Measuring time accessibility and its spatial characteristics in the urban areas of Beijing. *J. Geog. Sci.* **2016**, *26*, 1754–1768. [[CrossRef](#)]
19. Alex, A.; Richard, A.; Kenneth, A. Urban spatial structure. *J. Econ. Lit.* **1998**, *36*, 1426–1464.
20. Yang, T.; Jin, Y.; Yan, L.; Pei, P. Aspirations and realities of polycentric development: Insights from multi-source data into the emerging urban form of Shanghai. *Environ. Plan. B Urban Anal. City Sci.* **2019**, *46*, 1264–1280. [[CrossRef](#)]
21. Zhong, C.; Arisona, S.M.; Huang, X.F.; Batty, M.; Schmitt, G. Detecting the dynamics of urban structure through spatial network analysis. *Int. J. Geogr. Inf. Sci.* **2014**, *28*, 2178–2199. [[CrossRef](#)]
22. Patrick, L.; Robert, W.; Semantics, M. Cognitively Plausible Delineation of City Centres from Point of Interest Data. In Proceedings of the 13th Workshop of the ICA commission on Generalisation and Multiple Representation, Zürich, Switzerland, 12–13 September 2010; pp. 1–12.
23. Toole, J.L.; Ulm, M.; Bauer, D.; Gonzalez, M.C. Inferring Land Use from Mobile Phone Activity. In Proceedings of the ACM SIGKDD International Workshop on Urban Computing, Beijing, China, 12 August 2012; pp. 1–8.
24. John, S.; Emmanouil, T.; Peter, N. Data from mobile phone operators: A tool for smarter cities? *Telecomm. Policy* **2015**, *39*, 335–346.
25. Vincent, B.; Gautier, K.; Thomas, I. Regions and borders of mobile telephony in Belgium and in the Brussels metropolitan zone. *Brussels Stud.* **2010**, *42*, 1–12.
26. Yang, T. A study on spatial structure and functional location based on big data. *City Plan Rev.* **2018**, *42*, 28–38.
27. Liu, Y.; Wang, F.H.; Xiao, Y.; Gao, S. Urban land uses and traffic 'source-sink areas': Evidence from GPS-enabled taxi data in Shanghai. *Landsc. Urban Plan* **2012**, *106*, 73–87. [[CrossRef](#)]
28. Long, Y.; Shen, Z.J. Disaggregating heterogeneous agent attributes and location. *Comput. Environ. Urban Syst.* **2013**, *42*, 14–25. [[CrossRef](#)]
29. Rao, Z.H.; Yang, D.Y.; Duan, Z.Y. Resident mobility analysis based on mobile-phone billing data. *Procedia Soc. Behav. Sci.* **2013**, *96*, 2032–2041.
30. Wang, Y.; Xie, X.; Liang, S.; Zhu, B.; Yao, Y.; Meng, S.; Lu, C. Quantifying the response of potential flooding risk to urban growth in Beijing. *Sci. Total Environ.* **2019**, *705*, 135868. [[CrossRef](#)]
31. Get Points of Interest Data. Available online: <https://lbs.amap.com/api/ios-sdk/guide/map-data/poi/> (accessed on 31 December 2020).
32. Gao, Z.; Deng, X. Analysis on spatial features of LUCC based on remote sensing and GIS in China. *Chin. Geogr. Sci.* **2002**, *12*, 107–113. [[CrossRef](#)]



33. Okabe, A.; Satoh, T.; Sugihara, K. A kernel density estimation method for networks, its computational method and a GIS-based tool. *Int. J. Geogr. Inf. Sci.* **2009**, *23*, 7–32. [[CrossRef](#)]
34. Xie, Z.; Yan, J. Detecting traffic accident clusters with network kernel density estimation and local spatial statistics: An integrated approach. *J. Transp. Geogr.* **2013**, *31*, 64–71. [[CrossRef](#)]
35. Seo, Y.; Lim, D.; Son, W.; Kwon, Y.; Kim, J.; Kim, H. Deriving mobility service policy issues based on text mining: A case study of Gyeonggi Province in South Korea. *Sustainability* **2020**, *12*, 10482. [[CrossRef](#)]