

## Article

# Exploring the Spatiotemporal Patterns of Residents' Daily Activities Using Text-Based Social Media Data: A Case Study of Beijing, China

Jian Liu <sup>1</sup>, Bin Meng <sup>2,\*</sup>, Juan Wang <sup>2</sup>, Siyu Chen <sup>2</sup>, Bin Tian <sup>2</sup> and Guoqing Zhi <sup>2</sup>

<sup>1</sup> College of Resource Environment and Tourism, Capital Normal University, Beijing 100048, China; 2190902031@cnu.edu.cn

<sup>2</sup> College of Applied Arts and Sciences, Beijing Union University, Beijing 100191, China; wangjuan@buu.edu.cn (J.W.); 20201070510119@buu.edu.cn (S.C.); 191070510112@buu.edu.cn (B.T.); 191070510111@buu.edu.cn (G.Z.)

\* Correspondence: mengbin@buu.edu.cn

**Abstract:** The use of social media data provided powerful data support to reveal the spatiotemporal characteristics and mechanisms of human activity, as it integrated rich spatiotemporal and textual semantic information. However, previous research has not fully utilized its semantic and spatiotemporal information, due to its technical and algorithmic limitations. The efficiency of the deep mining of textual semantic resources was also low. In this research, a multi-classification of text model, based on natural language processing technology and the Bidirectional Encoder Representations from Transformers (BERT) framework is constructed. The residents' activities in Beijing were then classified using the Sina Weibo data in 2019. The results showed that the accuracy of the classifications was more than 90%. The types and distribution of residents' activities were closely related to the characteristics of the activities and holiday arrangements. From the perspective of a short timescale, the activity rhythm on weekends was delayed by one hour as compared to that on weekdays. There was a significant agglomeration of residents' activities that presented a spatial co-location cluster pattern, but the proportion of balanced co-location cluster areas was small. The research demonstrated that location conditions, especially the microlocation condition (the distance to the nearest subway station), were the driving factors that affected the resident activity cluster patterns. In this research, the proposed framework integrates textual semantic analysis, statistical method, and spatial techniques, broadens the application areas of social media data, especially text data, and provides a new paradigm for the research of residents' activities and spatiotemporal behavior.



**Citation:** Liu, J.; Meng, B.; Wang, J.; Chen, S.; Tian, B.; Zhi, G. Exploring the Spatiotemporal Patterns of Residents' Daily Activities Using Text-Based Social Media Data: A Case Study of Beijing, China. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 389. <https://doi.org/10.3390/ijgi10060389>

Academic Editor: Wolfgang Kainz

Received: 16 March 2021

Accepted: 1 June 2021

Published: 5 June 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/>).

## 1. Introduction

The continuous advancements of globalization and informatization have profoundly affected people's daily lives and behavioral activities, causing tremendous changes in the traditional patterns of residents' activities. On the one hand, the emergence of network activities has had numerous effects, including the substitution, complementarity, and enhancement of residents' daily activities [1], thereby affecting the use of urban physical space. On the other hand, the rapid development of information and communication technologies (ICTs) has changed the temporal and spatial relationships of residents' daily activities so that some activities are no longer subject to specific temporal and spatial constraints, thereby allowing better flexibility and coordination [2]. In this context, research into residents' activities has received extensive attention in many disciplines, such as geography, urban planning, transportation, computers, and public health [3–5]. By exploring the differences in the distribution scales and types of various resident activities, the urban

function and spatial structure can be better understood, the temporal and spatial laws of urban dynamics can be grasped, and the relationships between residents' activities and the objective environment in different temporal and spatial scenarios can be effectively revealed. This is of great significance for improving human health, guiding transportation and planning, and promoting the scientific understanding of human behavior [6–9].

With the rapid development of ICTs, while people are enjoying the convenience of ICTs, there has been an explosion of information on user activities and access records, either actively posted by users or passively recorded by devices and networks. It provides a wealth of convenient data to support research into human activities, and thus the exploration of new knowledge and methods of human activity patterns [10–12]. Some scholars have suggested that a new field is emerging that can utilize the capacity to collect and analyze data at a certain scale to reveal patterns of individual and group behavior [13], with sensible data mining algorithms making practical movement predictions that reveal trends and patterns that are difficult for humans to detect [14]. In this context, there has been tremendous progress in conducting resident activity research based on various types of big data, with an increasing variety of research results. However, the understanding of the interactions between residents' behavioral decisions and activity dynamics based on general big data (GPS data, mobile phone data, smart card data, inter-floating data, etc.) is limited due to the lack of information on the destinations that instigate population movements [15].

Fortunately, this limitation does not exist for text-based social media data based on user-initiated posts. Generally speaking, social media data contain detailed spatiotemporal, textual, image, social, and other multidimensional information about the user. It can be used to infer the user's activities and deeply study the characteristics of residents' activities and the influence mechanism of their choice of activities. However, among the existing studies on residents' activities via the use of social media data, most have primarily focused on using the spatial information in social media data, but less on the textual semantic information that contains rich activity content [16]. Specifically, the textual semantic information not only directly reflects the purpose and type of individual activities at a fine-grained level; furthermore, the quantity of data can also indicate the intensity of an individual's activity, and, combined with the spatial location, can efficiently reveal the behavioral activity characteristics of individual users [17]. Therefore, social media data (especially textual information) deserves more attention in the field of resident activity research. Meanwhile, it is necessary to organically combine spatiotemporal information with semantic information to improve the comprehensive utilization efficiency of social media data, and then to comprehensively and truly uncover the spatiotemporal characteristics of the users' activities and enhance the understanding of the dynamic characteristics of cities and residents.

In view of this, this study aims to introduce the current advanced natural language processing (NLP) technology into the field of resident activity research in order to efficiently extract the rich semantic information from the social media data. Then, the textual semantic information is combined with spatiotemporal information to improve the efficiency of social media data utilization in residential activity research, thus providing a high-quality data base for residential activity research. On this basis, the spatiotemporal characteristics and related patterns of residents' daily activities were explored, and the driving forces were then investigated to better examine and analyze the spatial structure of the city.

The remainder of this paper is organized as follows. In Section 2, existing research related to resident activities is presented from three perspectives. In Section 3, the study area and data collection method are presented. In Section 4, the process of classifying the residents' daily activities information is introduced, and the main methods used in this study are elaborated. In Section 5, the semantic characteristics, spatiotemporal patterns and attribution results of various residents' activities are specifically analyzed. Finally, the conclusions of this paper are drawn and future research directions are proposed in Section 6.

## 2. Related Work

### 2.1. Study of Resident Activity under the Traditional Perspective

The study of residents' activities can be traced back to the time–geography theory proposed by Hagerstrand [18], which was later praised and applied by related scholars [19–22]. In traditional resident activity research, travel surveys, questionnaires, interviews, and other means are primarily used to construct resident activity-diary surveys, and to carry out various related studies involving residents' activities and travel behaviors [23–25]. These efforts demonstrate that the daily activity patterns of residents have great regularity and are closely related to land use and urban built environments [26–28]. However, the questionnaire and interview-based approaches to the collection of activity information are costly in terms of time and money. Meanwhile, there is variability and risk to the reliability of the data and findings with the limitation of questionnaire design, interview rules, spatiotemporal scale, and the subjective nature of the respondents [29–31].

### 2.2. Study of Resident Activity in the Era of Big Data

With the rapid development of ICTs, the information storm triggered by the era of big data is transforming our lives, work, and thinking, and is initiating a major transformation of the era [32]. In this context, scholars have conducted a series of studies on residents' activities in the era of big data with the help of various types of big data. Specifically, big data-based research has been mainly focused on the use of the user activity records collected from various data platforms such as GPS devices, mobile phones, smart cards, floating vehicles, social media, wearable devices, etc., to explain the movement patterns of individuals or groups of people, reveal the spatiotemporal patterns of various residents' activities (travel, work, leisure etc.) or specific groups and dynamic changes [7,33–37]. For example, transit or travel smart card data are used to reveal the residents' daily activity patterns and laws [34,37] by extracting useful mobility information from the mobile phone location and call data to identify where residents live and work [38], investigating individual mobility patterns within cities [39]. Moreover, based on the extension of geographical information system (GIS) spatial models and analysis methods, and combined with data fusion, machine learning, and other means, the spatiotemporal patterns of human behavior can be extracted and the geospatial characteristics of human and socioeconomic elements can be inverted, which has become a hot research topic in recent years [40].

### 2.3. Study of Resident Activity Based on the Social Media Data

With the widespread adoption of mobile devices and location-based services, social media data have increasingly attracted the attention of scholars due to their large user base, rich spatiotemporal and semantic information, and low cost of access [12,17]. Social media data incorporating spatiotemporal and textual semantic multidimensional information has greatly enhanced the role of understanding human behavior and complex social dynamics in geographic space. Some scholars even argue that data generated based on internet communication and interaction may revolutionize our understanding of collective human behavior [41].

However, among the existing studies on residents' activities via the use of social media data, most have primarily focused on using the spatiotemporal information. These investigations include scalable and efficient spatiotemporal analyses via large-scale, location-based social media data [42,43], the modeling and prediction of user behavior and activity patterns [44,45], and the revelation of the functions, dynamics, and spatial structures of cities [46,47]. Specifically, the digital footprints collected from social media platforms are clustered through various spatiotemporal analysis methods and their variants to identify various types of residents' daily activities (e.g., living, working, entertainment, and eating) [48]. Alternatively, activity types are inferred based on the geographical location of each data record linked to the type of place in combination with other types of data, such as land-use data, points of interest (POI), and street-view imagery [49–51]. Nevertheless, most clustering methods consider only the temporal or spatial distribution characteristics

of travel activity points while ignoring their geographical context. It would result in the clustering of different types of activity data into the same cluster. This problem also exists in the use of place types to infer activity types. For example, a check-in at a residential building may label the location as “home”, however, the place may not be the user’s home, but rather their friend’s home, and the user’s behavior at this location should instead be labeled as “social” or “party”. Similarly, a place marked as a place of entertainment may also be the user’s workplace. In such cases, the results should be interpreted with caution [52].

In addition, as mentioned earlier, using textual information from social media data to conduct research on residents’ activities is a very effective approach, but due to previous technical and algorithmic limitations, it is difficult to fully extract semantic information from text, and thereby the relevant literature is lacking. In the only relevant studies, activity information or activity topics were mainly identified and extracted via feature word extraction (e.g., Word2vec model) or some clustering methods (e.g., density-based spatial clustering of applications with noise (DBSCAN), Latent Dirichlet Allocation (LDA) model, etc.) [16,49,53,54]. However, these studies lack a comprehensive consideration of the semantic content, resulting in some bias in the authenticity of the obtained activity data. Especially for social media data with word limits like Twitter and Weibo, the sparse nature of short textual features makes it riskier to rely on feature words alone for semantic classification (e.g., “apple is ripe” and “Apple Inc.” have two completely different meanings). With the significant breakthroughs in natural language processing in recent years [55], existing technologies have been able to support the efficient classification of large-scale text data. Therefore, it is necessary to introduce advanced natural language processing techniques into the study of residents’ activities, by conducting comprehensive and in-depth mining of textual resources in social media data to obtain a high-quality dataset of residents’ activities, which is important for improving the understanding of residents’ activities and urban dynamics [48,56,57].

### 3. Study Area and Data

#### 3.1. Study Area

Beijing, the capital of China, is also the political and cultural center of the country. As of the end of 2019, the city had a total area of approximately 16,410 km<sup>2</sup>, including 16 districts, and a resident population of 21,536,000 [58]. To fit well with the other datasets used in this research, the study area was divided into more than 16,000 one km grids; thus, the one km grid was the basic research unit for this study.

#### 3.2. Data

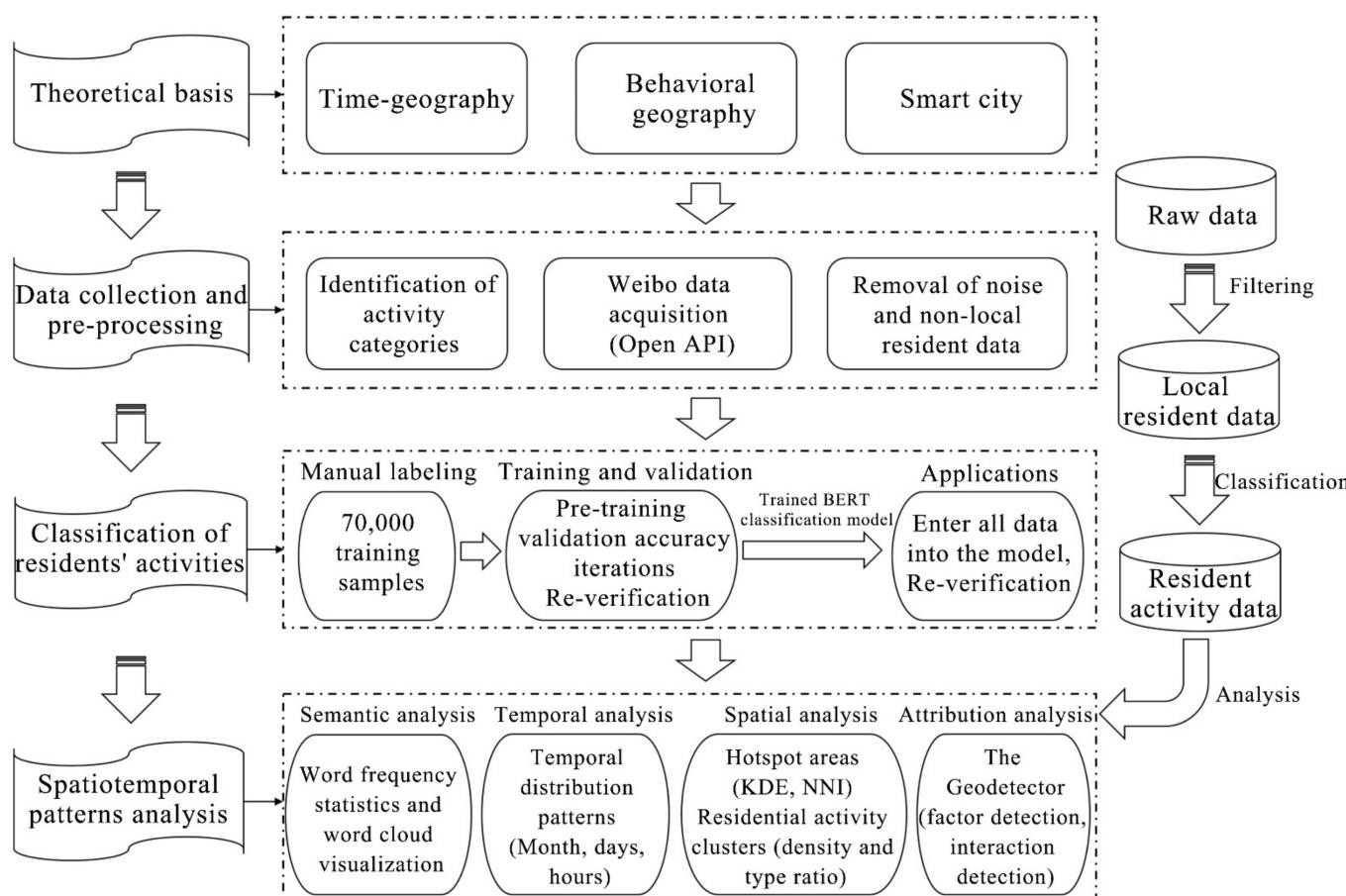
The social media data in 2019 were obtained from Sina Weibo API using web crawler tools. In total there were 11,500,105 pieces of Weibo data involving more than one million users covering Beijing City. The attributes included the user ID, text, time, latitude, and longitude. According to the Weibo User Trends Report in 2020 by Weibo Data Center (<http://data.weibo.com/datacenter/recommendapp>, accessed on 30 September 2020), as of September 2020, the number of monthly active users on Weibo had increased to 511 million, with an average of 224 million daily active users. These data indicate the further strengthening of Sina Weibo’s position as the leading social media platform in China. However, we are also aware of the problems of sample bias and representativeness in social media data. Specifically, the social media platforms are used by a relatively young group, and the users on social media may be varied by socio-economic attributes (ages, gender, occupation, etc.) and individual behavior differences [59]. However, many studies have shown that social media data still play an important role for the extraction of human activities, emotions and experiences associated with a place with the advantage of rich contextual content and geographical location information [17,60]. On this basis, large amounts of social media data can be integrated in order to profile groups of users and their activity patterns, thereby providing insight into the dynamics of cities and people on a

larger scale [16,61]. Therefore, it is reasonable and effective to select the massive Weibo data to identify the spatiotemporal patterns of residents' daily activities at the group scale.

2019 Beijing POIs data were also sourced from the Gaode Map platform, and included 12 main categories (restaurants, shopping, accommodation, science, education, culture, etc.). The 2019 WorldPop dataset was also used as one of the base datasets for this study.

#### 4. Methodology

This study constructs a framework for studying residents' activities by integrating natural language processing, statistical analysis and spatial analysis (Figure 1), and innovatively introduces textual multiclassification techniques, combined with machine learning methods into the study field of residents' daily activities in order to fully exploit the rich textual and spatiotemporal information in social media data. Specifically, first of all, the main types of residents' daily activities are identified based on time-geography and behavioral geography theories. Secondly, machine learning algorithms and BERT models are used to perform text multiclassification on the collected large-scale social media datasets to identify the specific types of users' activities, and then a high-quality spatiotemporal dataset of residents' daily activities is formed by combining the posting locations. Based on this, spatiotemporal patterns of residents' daily activities and related laws are explored from three perspectives: semantic, temporal and spatial.



**Figure 1.** Framework of residents' activities research based on social media data.

#### 4.1. Identification of Activity Categories

The study of residents' daily activities has traditionally been an important part in the fields of time-geography and behavioral geography, involving a range of activities such as commuting, shopping, and leisure [62]. There exists a wealth of existing methods for the classification of daily activities, with the number of activity types ranging from four to hundreds [63,64]. However, residents' daily activities are habitual, stable and highly repetitive, overly broad or trivial classifications of activities are not conducive to research, and the randomness of related activities may make it difficult to extract valuable regular features. Moreover, some scholars have found that social media-based daily activities include those such as at-home activities, working, eating, shopping, learning, leisure, and entertainment, with an activity coverage rate of 94.5% [15,52]. However, given the broad scope of at-home activities, the spatial characteristics are not obvious, and the Weibo data themselves have a certain outdoor feature. Therefore, based on previous research and taking into account the behavioral characteristics of local residents, seven types of activities were selected as the main types of daily activities (Table 1), based on which the subsequent step of text classification was carried out.

**Table 1.** The identification of activity categories.

Activity Category	Feature Words
Social	Party, dates, gathering, friends, visit, etc.
Eating	Eat, drink, restaurant, barbecue, hot pot, etc.
Entertainment	Movies, parks, theaters, KTV, games, Happy Valley, zoos, bars, etc.
Shopping	Shop, supermarkets, Carrefour, Walmart, Uniqlo, etc.
Studying	Study, school, read, class, library, teacher, etc.
Sports	Sports, running, balls, exercise, fitness, hiking, cycling, swimming, etc.
Working	Business trip, meeting, interview, job, internship, duty, overtime, etc.

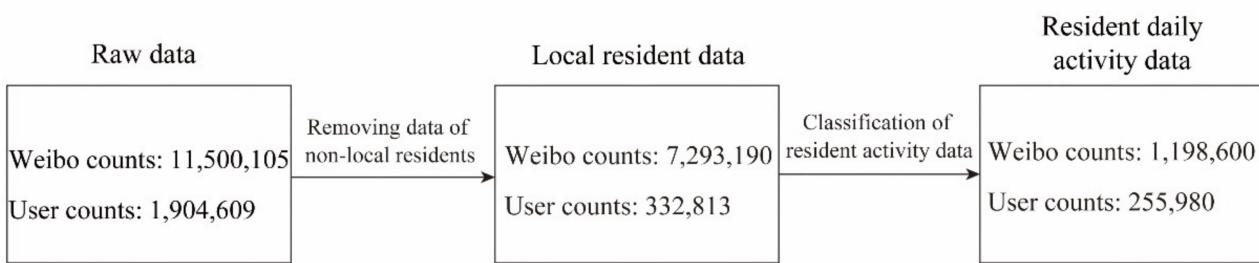
#### 4.2. Classification of Residents' Daily Activities Based on BERT

Since the research object is the daily activities of Beijing residents, the original datasets need to be screened to exclude the user data of non-local residents. Therefore, according to the specific filtering rules (the time and frequency of user posts), 7,293,190 pieces of Weibo data of local residents were ultimately obtained. At the same time, as this research only focused on the Weibo items about the residents' daily activities, the original Weibo datasets were filtered to select the related items using the Bidirectional Encoder Representations from Transformers (BERT). It is one kind of language encoder, released by Google in 2018, able to translate the input sentences or paragraphs into corresponding semantic features, which has performed amazingly well and become an important recent advancement in NLP [55].

In this research, a text classification model was constructed based on the BERT to perform multiple classifications on the crawled Weibo texts. Specifically, first, 70,000 items were randomly selected as the training samples. For each item, if it was related to different residents' activities, it was labeled as 1–7 manually, otherwise labeled as 0 (Table 2) [64,65]. Second, using machine learning and the original BERT model to pretrain 70,000 training data and verify the classification accuracy, by adjusting the corresponding parameters and the number of iterations for several times under the experiment, the trained text multiclassification model was obtained (the overall accuracy exceeded 87%). Third, based on the derived classifier, all the Weibo items were input to BERT and the items relating to the various residents' activities were classified. Then, 1,198,600 pieces of data of the daily activities of Beijing residents in 2019 were identified. Finally, a further 5000 randomly selected social media posts from each of the seven categories of activity data were manually validated and an average accuracy of 94.12% was achieved, thereby verifying the excellent classification effect of the model. The detailed data screening process is presented in Figure 2.

**Table 2.** Manual labeling example.

Weibo Items	Label
It's so boring!	0 (Irrelevant)
An extraordinarily enjoyable team building~	1 (Social)
This hot pot is really delicious	2 (Eating)
Take a stroll around the Forbidden City	3 (Entertainment)
Come out to shop!	4 (Shopping)
There are many things to learn, trying to learn	5 (Studying)
Five kilometers completed	6 (Sports)
I'm still struggling in the office at this hour	7 (Working)

**Figure 2.** The data filtering process.

#### 4.3. The Identification of Residential Activity Clusters

In order to reveal the clustering patterns of various residents' activities in the geospatial range and their distribution combinations, this study uses the activity density and type ratio methods with reference to the study of function, mixing the degree of spatial unit [66,67] to identify the clustering patterns and activity combinations of daily residents' activities in Beijing.

First, the activity density method was employed to calculate the proportions of different types of resident activity in each grid to the total number of corresponding types of resident activity in the study area, which were calculated as follows:

$$A_{ij} = \frac{P_{ij}}{\sum_{j=1}^n P_{ij}}, \quad (1)$$

where  $A_{ij}$  is the number of type  $i$  activities in grid  $j$  as a proportion of the total number of type  $i$  activities in all spatial units in the study area,  $P_{ij}$  indicates the number of type  $i$  activities in grid  $j$ , and  $n$  indicates the number of grids in the study area.

Next, the relative proportions of the densities of different types of residential activity in each grid were calculated to reflect their type-ratio characteristics. These proportions were calculated as follows:

$$AC_{ij} = \frac{A_{ij}}{\sum_{i=1}^n A_{ij}}, \quad (2)$$

where  $AC_{ij}$  is the ratio of type  $i$  residential activities in spatial unit  $j$ , and  $A_{ij}$  has the same meaning as in Equation (1).

In the identification of residential activity clusters, if the proportion of one type of activity in the grid is  $\geq 50\%$ , the spatial unit is dominated by a single activity; if the proportion of all types of activity in the grid is below 50%, the unit is considered to be a co-located cluster space of multiple activities. Particularly, in this study, the co-located clusters were subdivided and the activities with a proportion of  $\geq 25\%$  of the type of activity in the grid were considered as the dominant activity type within the spatial unit. If the proportion of all types is less than 25%, this indicates that the frequency density of

different types of public services is relatively evenly distributed and there are no clearly dominant activities.

Furthermore, a vibrant urban space must maintain sufficient diversity to meet the diversity of people's needs [68]. To calculate the mix of daily activities of residents within different types of activity clusters, this study draws on the concept of measuring the land-use mix, which is commonly used in urban research [69], to construct a characteristic indicator for the measurement of activity diversity while taking into account the number and types of activities. The formula is as follows:

$$M_j = \frac{-\sum_{i=1}^k q_i \ln q_i}{\ln(k)} \quad (3)$$

where  $M_j$  denotes the activity diversity index of grid  $j$ ,  $q_i$  denotes the ratio of the number of type  $i$  activities in grid  $j$  to the total number of activities in the spatial unit  $j$ , and  $k$  denotes the number of activity types in spatial unit  $j$ . The activity diversity index  $M_j$  has a value range from 0 to 1, and its size reflects the degree of mixing of different activities; a larger value indicates a more balanced distribution of various types of activities in the spatial unit and a higher activity diversity, while a smaller value indicates a more homogeneous distribution of activity types and a lower diversity.

#### 4.4. Analysis of Influencing Factors

The Geodetector was used to explore the causes of the spatial and temporal heterogeneity of the residents' daily activities. It consists of four components: risk detection, factor detection, ecological detection and interaction detection, to detect geo-spatial heterogeneity and reveal the driving forces behind it [70,71]. The method is good at analyzing typological quantities, detecting both numerical and qualitative data, and is unique in its ability to investigate the interaction between two explanatory variables to a response variable. As the decision-making process of the residents' activities is the result of a combination of factors, the spatial differentiation mechanism and diversity of residents' daily activities cannot be separated from the systematic analysis of multiple influencing factors. According to urban diversity theory [72], the formation of diversity is closely related to factors such as population, land, and transportation. Based on this, the Geodetector was employed to analyze the influences of various factors, including the socioeconomic attributes, facility configuration, and location conditions of the grid, on the formation of diversity in residential activity clusters. Specifically, the dependent variable was considered to be the resident activity pattern in the grid, while the explanatory variables included the population density, land price, traffic accessibility, and others (see Table 3). In short, factor detection and interaction detection were used to reveal the influencing mechanisms of the clustering pattern and diversity of residents' daily activities.

**Table 3.** The index system of the influencing factors of residents' daily activity cluster patterns.

Influencing Factor	Explanatory Variables
Population distribution	Population density (X1)
Land use	Land price (X2)
Activity facilities	POI density(X3)
Traffic accessibility	Distance to the nearest subway station(X4)
Location conditions	Distance to the city center (X5)
Functional positioning	Urban planning positioning (X6)

## 5. Results

### 5.1. Semantic Characteristics of Residents' Daily Activities

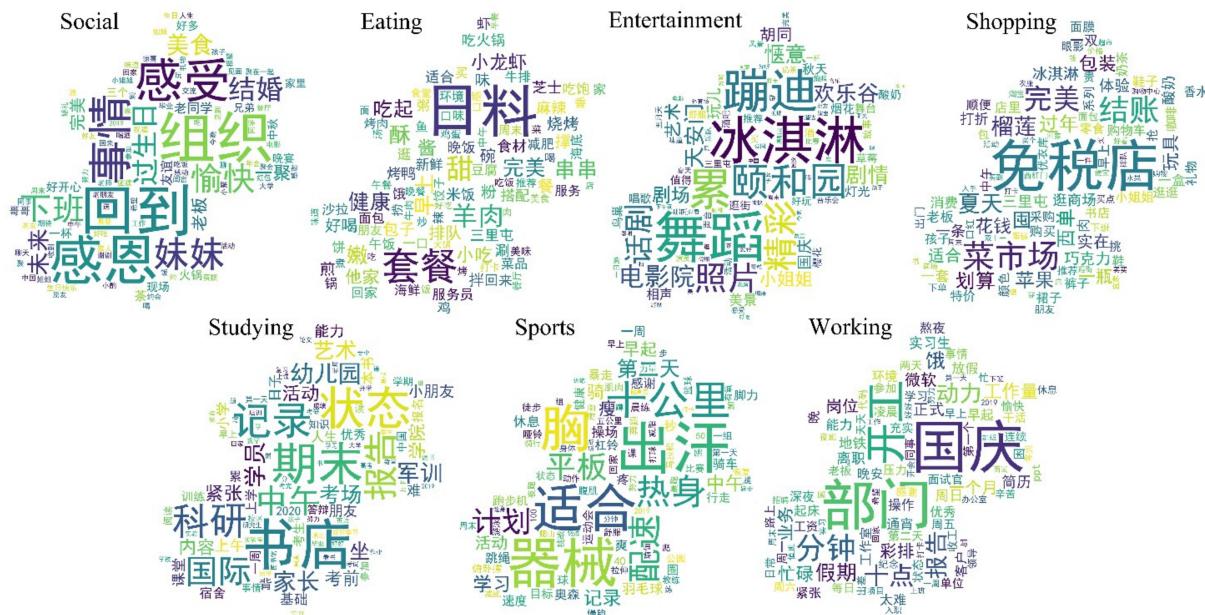
According to the results of text classification processing, 1,198,600 pieces of data for seven types of activities were identified, and the percentages of the numbers of various

activities and the top 10 highest-frequency words are reported in Table 4. Entertainment, eating, and studying activities were found to account for more than 80% of the total. It indicates that residents are more inclined to share these activities on online communities as opposed to work and social activities, with entertainment and study spaces becoming the main physical spaces corresponding to virtual online spaces.

**Table 4.** Numbers of activity categories and high-frequency words.

Activity Category	Number	Proportion	High-Frequency Words
Social	45,879	3.83%	friends, party, eat, potluck, gift, attend, get together, receive, thanks, small gathering
Eating	329,134	27.46%	eat, delicious, taste, restaurant, breakfast, hot pot, meat, rice, restaurant, dish
Entertainment	466,386	38.91%	eat, check-in, film, weekend, play, drink, go, live, holiday, shop
Shopping	37,007	3.09%	buy, eat, stroll, supermarket, clothes, store, delicious, shopping, bought, good-looking
Studying	186,233	15.54%	study, exam, teacher, write, classes, graduation, class, homework, library, graduate entrance exam
Sports	92,764	7.74%	run, check-in, exercise, fitness, training, workout, gym, minutes, walk, practice
Working	41,197	3.44%	overtime, work, at work, interview, after work, first day, effort, weekend, check-in, writing

Based on the classified activity data, the top 100 highest-frequency keywords for each type of activity were extracted by word frequency statistics and arranged in reverse order (i.e., for the top 100 words in each type of activity, the No. 1 word was swapped with the No. 100 word, the No. 2 word was swapped with the No. 99 word, etc.), and were displayed in word clouds (Figure 3). For example, social activities often imply positive emotions, eating activities reflect the dishes and taste preferences of the residents' daily diets and some important places and restaurants, and healthy eating is also highly recognized. Entertainment, shopping, and sports reflect the corresponding specific types of activities and main venues. Studying highlights an intense state and atmosphere of learning, in addition to recording the activities themselves, such as exams and assignments. Moreover, working reflects more complex emotions about work itself (e.g., effort, motivated, nervous, too hard, etc.).

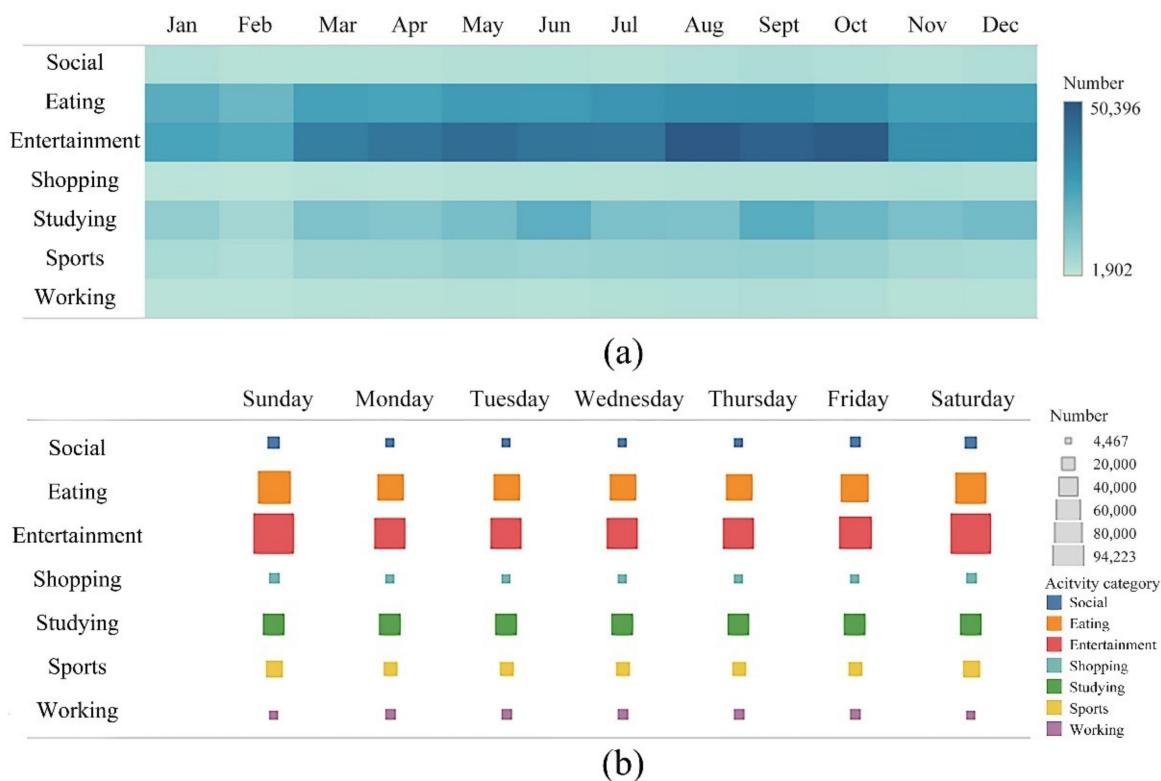


**Figure 3.** Word clouds for various types of resident activities.

## 5.2. Temporal Distribution of Residents' Daily Activities

### 5.2.1. Activity Dynamics on a Long Timescale

The Weibo data of residents' daily activities in Beijing in 2019 were counted separately by months and days. First, from the results of the month-to-month statistics (Figure 4a), the activity intensities in terms of months and the types of resident activities were found to have large variations. Specifically, eating and entertainment were found to have the highest intensities and to be closely related to the temporal distribution of holidays, with obvious temporal clustering characteristics (May, August–October). Studying was found to have a strong correlation with China's own education system. The start of the term (March and September) and the end of the term (June and December) were found to have significantly more Weibo posts related to studying than the other periods. The temporal distribution of sports was found to be more seasonally correlated, with the fewest data related to sports in the winter and the most in the summer. In contrast, social, shopping, and working activities were found to be less intense and very evenly distributed between months. Secondly, after residents' daily activities in 2019 were counted by days (Figure 4b), it was found that the intensity of activities was significantly higher on weekends than on weekdays. Particularly, the intensities of entertainment, eating, social, and shopping activities were found to have more pronounced increases. However, working was found to be significantly less intense on weekends than on weekdays. In addition, the intensities of activities on Mondays and Fridays were also found to be more prominent due to the influence of weekend activities.

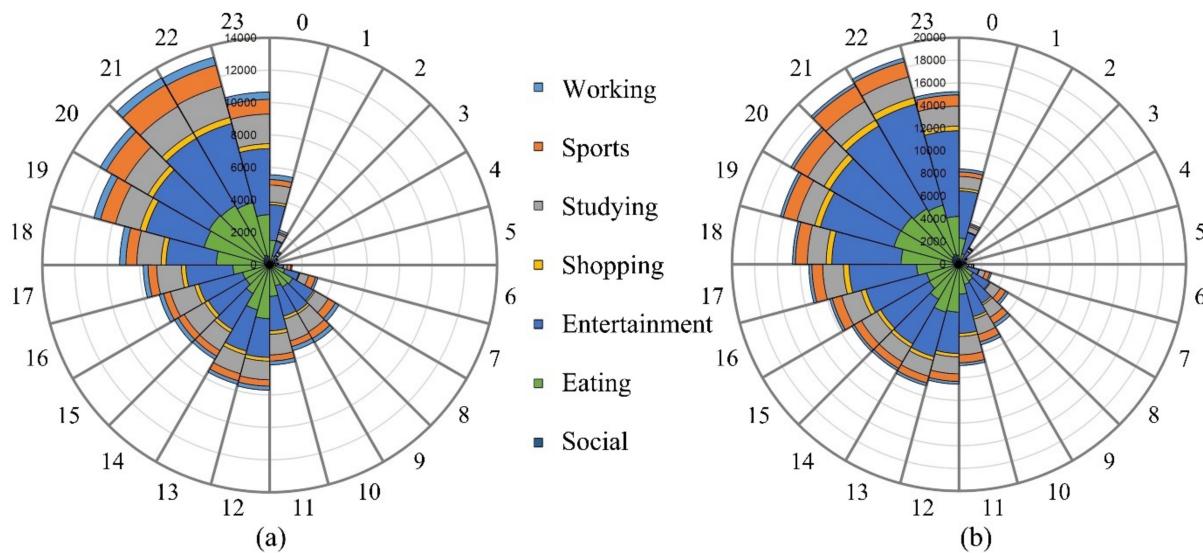


**Figure 4.** The (a) monthly and (b) daily temporal distribution patterns of residents' daily activities.

### 5.2.2. Activity Dynamics on a Short Timescale: The Activity Rhythm on Weekends Is Delayed by One Hour as Compared to That on Weekdays

The intraday distribution characteristics of residents' daily activities on weekdays and weekends were further examined. To reduce the impact of weekend activities on weekday activities, the activity data for Monday and Friday were excluded (see Figure 5). It was found that, overall, there is a clear temporal pattern for the seven types of activities, i.e., the intensity of residential activity is lowest before dawn, it fluctuates while increasing during

the day, and it peaks at night. However, as compared to that on weekdays, the intensity of activity on weekends was found to be higher and longer, and the activity rhythm was found to be delayed by one hour, i.e., the nighttime sleep period (the six hours in the day with the lowest activity intensity) was found to be between 2:00 and 7:00 on weekends, compared to 1:00 and 6:00 on weekdays. Moreover, the peak of lunchtime activity was found to be 13:00 on weekends, as compared to 12:00 on weekdays. The intensity of activity at night continued to increase until 22:00 hours on weekends as compared to 21:00 hours on weekdays. This result is very similar to the pattern of mood changes observed by Golder et al. (2011) for Twitter users, i.e., people are happier on weekends, but the morning peak of positive affect is delayed by two hours [73]. This indicates that late bedtimes and late starts are becoming the norm on weekends, due to the increase in discretionary time and activities, resulting in a higher intensity and delayed pace of activity on weekends. However, compared to other countries, the daily activities of Chinese residents are more strongly influenced by the traditional routine, and the relative delay in activities on weekends is shorter.

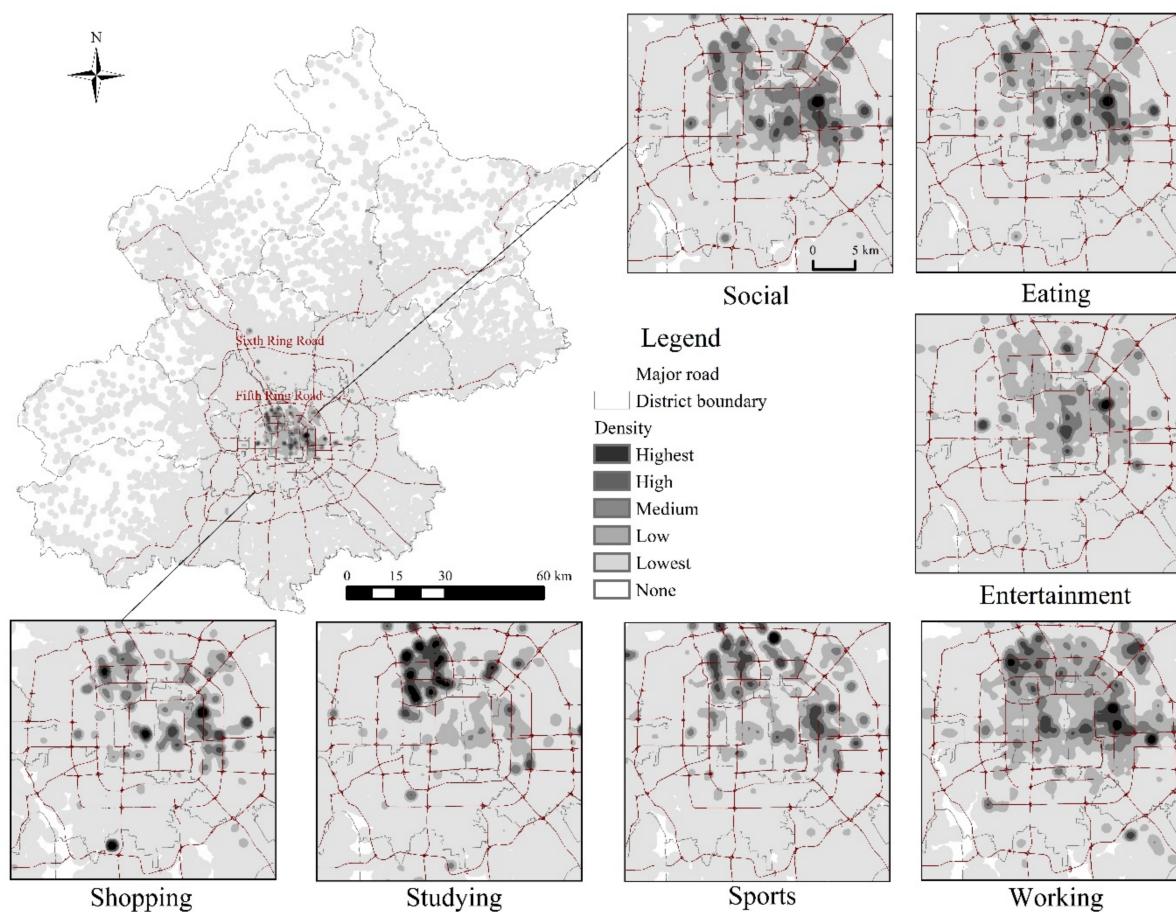


**Figure 5.** The hourly temporal distribution patterns of residents' daily activities on (a) weekdays and (b) weekends.

### 5.3. Spatial Characteristics of Residents' Daily Activities

#### 5.3.1. Distribution of Hotspot Areas for Residents' Daily Activities

Kernel density estimation (KDE) is a non-parametric estimation method for the analysis of the density of geographic elements in the surrounding area [74]. KDE was used to carve out hotspot areas for the daily activities of Beijing residents, and the results are presented in Figure 6. Overall, the spatial distribution reveals that the hotspots of the residents' daily activities are mainly concentrated in the central city within the Fifth Ring Road. Moreover, the spatial distribution of residents' daily activities within the Fifth Ring Road is characterized by significant differences between the north and south, with the hotspots of various activities mainly located in the northern areas of the Fifth Ring Road. This reveals that the spatial structure of Beijing's city is still dominated by the traditional northern city model from the perspective of resident activity. In addition, due to the characteristics of various activities and the differences in the distributions of different types of activity facilities, the spatial characteristics of various activities were also found to be highly variable, as shown in Table 5. In general, the hotspot distributions of various activities and activity facilities exhibit spatial co-location patterns, especially some of the comprehensive activity facility clusters, which often become mixed clustering areas for multiple types of activities.



**Figure 6.** The kernel density distribution of overall activity and various daily activities.

**Table 5.** The spatial distribution characteristics of various activities and hotspot areas.

Activity Category	Spatial Distribution Characteristics	Hotspot Areas and Their Properties
Social	One center and multi-points	Comprehensive business district: Workers' Stadium-Sanlitun, Xidan, Wangfujing Street, Zhongguancun, Wudaokou, etc.
Eating	One center and multi-point	Comprehensive business district: Workers' Stadium-Sanlitun, Xidan, Wangfujing Street, Zhongguancun, Wudaokou, etc.
Entertainment	Central symmetry	Comprehensive business district and main attractions: Workers' Stadium-Sanlitun, Tiananmen Square, Wukesong, National Olympic Sports Center, etc.
Shopping	Polycentric	Grand shopping mall: Sanlitun, Xidan, Zhongguancun, LIVAT Centre, etc.
Studying	Multi-point clustering	Locations of universities: Tsinghua University, Peking University, Renmin University of China, Beijing Institute of Technology, etc.
Sports	One center and multi-points	Locations of parks and universities: Olympic Forest Park, sites of some universities
Working	Multi-point clustering	Financial and economic-technological development areas: Pan-CBD area, Financial Street, Zhongguancun, Yizhuang

The nearest neighbor index (NNI) is an indicator that characterizes the proximity and mutual relationship between point-like geographic elements in a particular region [74]. The NNI was used to quantitatively measure the degree of agglomeration of various daily activities. The results indicate that the NNI for each type of activity is much less than 1 (Table 6), and there is significant spatial agglomeration for all types of activities. However, there is some variability in the degree of agglomeration of the various types of activity,

with eating, entertainment, and studying being more spatially agglomerated, followed by shopping and sports, and social and working activities being relatively less agglomerated.

**Table 6.** The NNIs of various resident activities.

Activity Category	Number	Z-Score	p-Value	Nearest Neighbor Index
Social	45,879	−327.0001	0.000 *	0.2020
Eating	329,134	−961.5882	0.000	0.1239
Entertainment	466,382	−1138.3747	0.000	0.1287
Shopping	37,007	−299.2567	0.000	0.1868
Studying	186,232	−717.7203	0.000	0.1306
Sports	92,764	−488.7824	0.000	0.1611
Working	41,197	−313.0133	0.000	0.1939

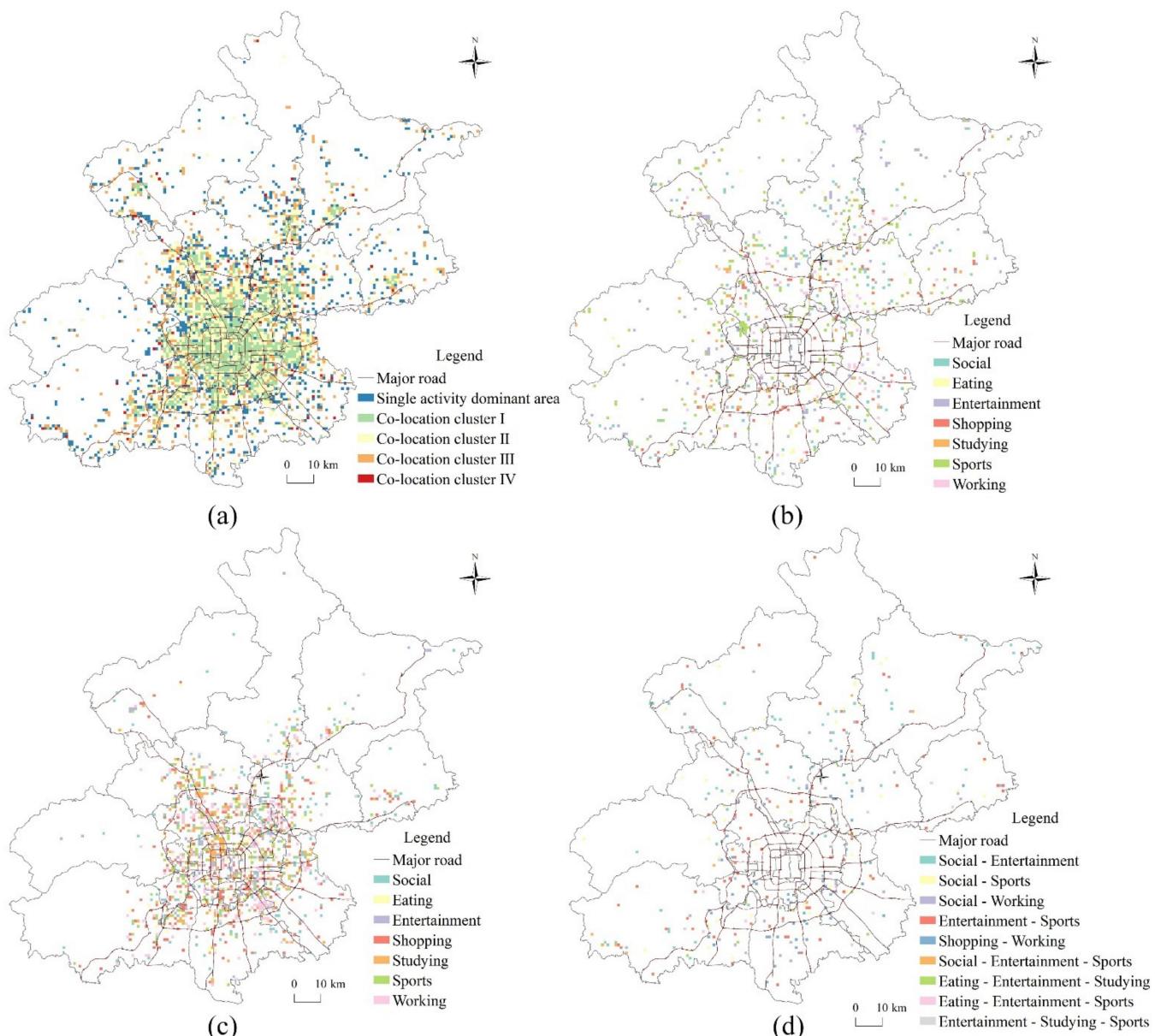
0.000 \* indicates that the result is significant at the 99.9% level.

### 5.3.2. Identifying Residential Activity Clusters

To understand the common agglomeration characteristics and activity combination distribution patterns of different types of activities, the activity density and type ratio methods were utilized to identify the resident activity clusters, which were divided into two major categories: single-activity-dominant areas and multiple activity co-location clusters. Among them, the co-location clusters can be subdivided into four subcategories, namely co-location clusters I to IV, according to the distribution of activities in the grid. Ultimately, the daily activity clusters of Beijing residents are divided into five patterns. The specific definitions and characteristics of the activity patterns are reported in Table 7, and the spatial distributions of activity clusters are presented in Figure 7.

**Table 7.** The definitions of resident activity patterns and spatial distribution characteristics.

Residential Activity Cluster	Definition	Major Characteristics	Proportion of Distribution Range	Distribution Areas
Single-activity-dominant areas	The proportion of one activity in the grid is more than 50%	Uneven distribution of activities within the grids with only one dominant activity	22.59%	Areas outside the Fifth Ring Road
Co-location cluster I	The proportion of all activities in the grid is less than 25%.	The distributions of various activities within the grids are balanced, with no dominant activity	22.33%	Central city within the Sixth Ring Road and the urban district in distant suburbs
Co-location cluster II	The proportion of only one activity in the grid is above 25% and below 50%	The distributions of various activities within the grids are relatively balanced, with only one activity being relatively dominant	33.99%	Central city within the Sixth Ring Road and the urban district in distant suburbs
Co-location cluster III	The proportions of two activities in the grid are above 25% and below 50%	The distributions of various activities within the grids are relatively moderate, with two activities being relatively dominant	19.16%	Mainly outside the Fifth Ring Road, with a scattered distribution within the Fifth Ring Road
Co-location cluster IV	The proportions of three activities in the grid are above 25% and below 50%	The distributions of various activities within the grids are relatively uneven, with three activities being relatively dominant	1.93%	Areas outside the Fifth Ring Road

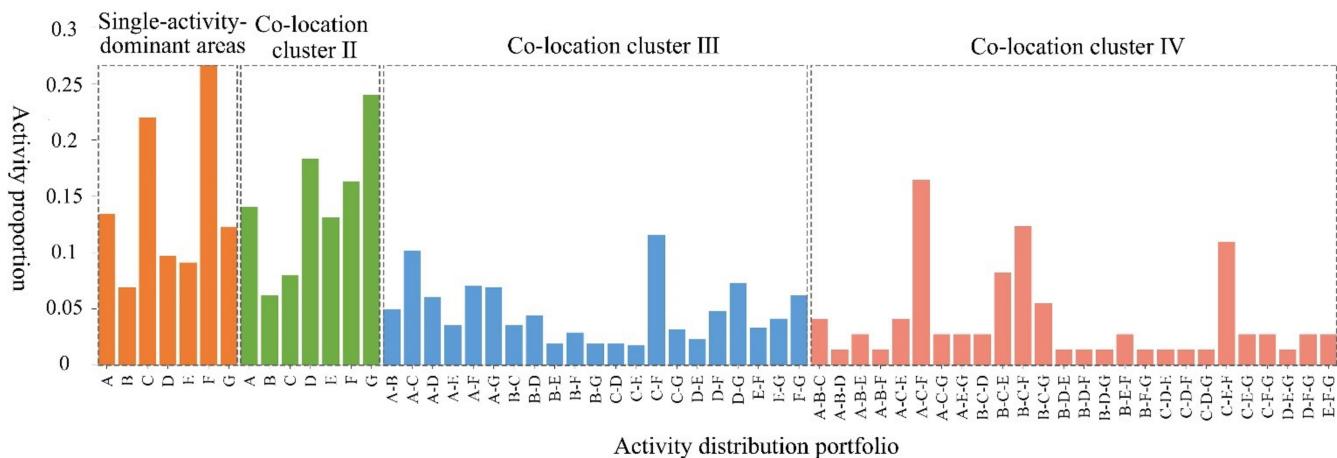


**Figure 7.** The spatial distributions of residential activity clusters and activity combination patterns. **(a)** Division of residential activity clusters; **(b)** single-activity-dominant areas; **(c)** co-location cluster II; **(d)** co-location clusters III and IV.

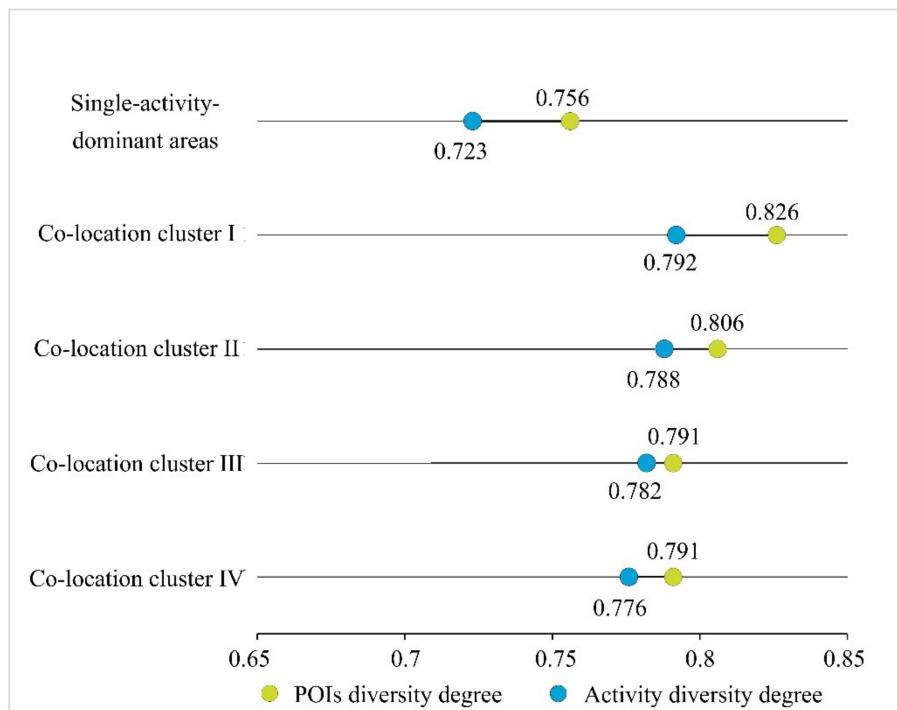
Overall, the activity areas in the central urban district within the Fifth Ring Road and the urban district in the outer suburbs and counties were dominated by co-location clusters I and II. It reflects the relatively rich and balanced distribution of activities in the central city and the urban district in the outer suburbs and counties as places where various high-density socioeconomic activities take place. However, the internal spatial distributions of the different residential daily activity clusters were found to vary considerably, with the exception of co-location cluster I, for which the activity combination could not be broken down. The specific activity patterns within the four different types of activity clusters are exhibited in Figure 8.

In addition, the information entropy was used to calculate the diversity of residents' daily activities to explore the balance of the activity distributions within different clusters. Moreover, for comparative analysis, the same method was employed to calculate the POI diversity to characterize the balance of activity facilities within the activity clusters (Figure 9). It was found that the mean values of activity diversity in various activity clusters exhibited the following decreasing pattern: co-location cluster I > co-location cluster II > co-location

cluster III > co-location cluster IV > single-activity-dominant areas. Moreover, the distributions of POI and activity diversities were found to have a high degree of matching. It indicates that the balanced co-location clusters have high activity diversity, which corresponds to the high accessibility of activity facilities and a variety of facility types. The monolithic co-location clusters were found to have the second-highest activity diversity, and single-activity-dominant areas were found to have low activity diversity, which corresponded to the low accessibility of facilities and a relatively homogeneous composition of activity types.



**Figure 8.** The activity combination portfolio and proportions within different co-location clusters (A: social; B: eating; C: entertainment; D: shopping; E: studying; F: sports; G: working).



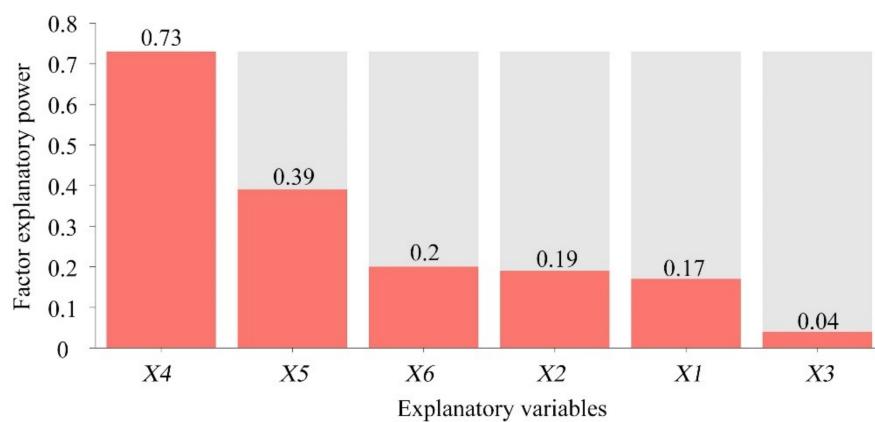
**Figure 9.** The diversity degrees of POIs and resident activity within different clusters.

#### 5.4. Analysis of Influencing Factors

##### 5.4.1. Main Influencing Factors and Explanatory Power

Factor detection was used to measure the factor explanatory power of various variables for the residential activity pattern. The results show that all kinds of explanatory variables

have passed the 0.001 level significance test, indicating that these explanatory variables are important factors influencing the co-location clustering of residential activities in the Beijing. Figure 10 shows the explanatory power of specific influencing factors. The order of explanatory power of the factors is “ $X_4$  (distance to the nearest subway station) >  $X_5$  (distance to the city center) >  $X_6$  (urban planning positioning) >  $X_2$  (land price) >  $X_1$  (population density) >  $X_3$  (POI density)”. The distance to the nearest subway station has the largest explanatory power, reaching 0.73, indicating that the residential activity pattern of Beijing is most strongly influenced by the accessibility of transport. The more convenient the transport conditions, the more significant the co-location cluster of the residents' daily activities. The distance to the city center is the next most important factor, indicating that macro-location conditions play an important role in the resident activities pattern. At the same time, the urban planning positioning in the area of the research unit also has an important influence in the formation of its activity clustering pattern. In addition, the factor explanatory power of land price and population density is also greater than 0.1, which also has a certain influence on the activity. However, the POI density in the research unit has a weaker influence on the type of activity cluster pattern, and its factor explanatory power is only 0.04, reflecting that the facilities configuration in the research area is not highly correlated with resident activity.



**Figure 10.** The explanatory powers of various influencing factors.

#### 5.4.2. Analysis of Factor Interactions

Interaction detection was used to analyze whether the interaction of two different influencing factors enhances or weakens their explanatory power for the dependent variable, and can effectively reveal the impact of the joint action of two types of explanatory variables on the resident activity pattern (Table 8). The results demonstrate that the explanatory power of any two influencing factors tends to increase after a two-by-two interaction, which indicates that the resident activity pattern is jointly constrained by the sub-factors of each dimension. Specifically, the type of factor interaction is “Enhance bi-”, i.e., the explanatory power of the factors after interacting is significantly stronger than that of a single factor, but not higher than the sum of the explanatory powers of two factors acting independently. Overall, the order of the top five interaction results was found to be as follows:  $X_4 \cap X_5$ ,  $X_2 \cap X_4$ ,  $X_1 \cap X_4$ ,  $X_4 \cap X_6$ ,  $X_3 \cap X_4$ . These results indicate that the resident activity pattern is most significantly affected by the combination of a micro-location condition, represented by the distance to the nearest subway station, and a macro-location condition, represented by the distance to the city center. Moreover, although the influences of land price and population density are minor when they act independently, they remain important basic factors that cannot be ignored.

**Table 8.** Interaction results of different influencing factors.

Interaction Factors	q-Value	The Type of Factor Interaction
X4 ∩ X5	0.80	Enhance bi-
X4 ∩ X2	0.75	Enhance bi-
X4 ∩ X6	0.74	Enhance bi-
X4 ∩ X1	0.74	Enhance bi-
X4 ∩ X3	0.73	Enhance bi-
X5 ∩ X6	0.54	Enhance bi-
X5 ∩ X1	0.42	Enhance bi-
X5 ∩ X3	0.39	Enhance bi-
X5 ∩ X2	0.39	Enhance bi-
X2 ∩ X6	0.39	Enhance bi-
X2 ∩ X1	0.28	Enhance bi-
X6 ∩ X1	0.28	Enhance bi-
X6 ∩ X3	0.25	Enhance bi-
X2 ∩ X3	0.20	Enhance bi-
X1 ∩ X3	0.19	Enhance bi-

## 6. Discussion and Conclusions

User-initiated social media data, based on a social network platform, contain a wealth of information on resident behavior dynamics, which is of great significance for the understanding of the spatiotemporal patterns and dynamic laws of resident activities in the information age [60]. Nevertheless, existing related research remains limited in terms of mining the resident activity information in the social media data, while various spatiotemporal clustering methods and their variants are usually used to cluster digital footprints from social media platforms, and while keyword extraction or topic model clustering is performed to identify various resident daily activities. Most of these methods only consider the location information of resident activity while ignoring the geographic background, which leads to certain problems in the identification of activity types [48]. In this regard, some scholars have pointed out that social media big data contain not only spatial information (e.g., locations, place names, etc.), but, more importantly, rich contextual and semantic information. Via NLP technology, spatiotemporal information can be extracted from text, and the semantics of the places, resident activities, and emotional experiences behind the text can also be mined [17].

The present study was based on a comprehensive integrated method that combines NLP technology and spatiotemporal analysis to achieve the organic integration of textual and spatiotemporal information. The results revealed that the BERT-based text classification model achieved excellent results in identifying residents' daily activities with an accuracy of more than 90%, which can effectively solve the current problems of the low utilization of social media text data and the poor integration of spatiotemporal and semantic information. Meanwhile, it also provides a solid data foundation for the full exploration of the spatiotemporal patterns and laws of human behavior activities hidden behind social media data, and provides a new research framework for the study of residents' daily activities in the mobile information era. Furthermore, based on the perspective of residents' daily activities, this study treated residents' daily activities and their spatiotemporal information as a complete system, and comprehensively explored the diversified and heterogeneous characteristics of resident activities, thereby solving the existing problem of the segmentation of the residents' activity types and providing a useful exploration and scientific guidance for the comprehensive and systematic revelation of urban dynamics, urban rhythms, and urban spatial structures from the perspective of the residents' daily activities.

The findings of this research can be summarized as follows. First, residents are more inclined to share their entertainment, eating, and studying activities in online communities as opposed to their work and social activities, with entertainment and study spaces becoming the main physical spaces corresponding to virtual online spaces. Second, the

distribution differences and types of resident activities are closely related to the characteristics of the activities and holiday arrangements. However, compared to that on weekdays, the intensity of activity on weekends was found to be higher and longer, and the activity rhythm was found to be delayed by one hour. Third, there was significant spatial clustering of resident daily activities, with the main hotspot areas concentrated in the central city within the Fifth Ring Road and exhibiting differentiation characteristics between the north and south, with more activities in the north. The cluster patterns of resident daily activities can be divided into five modes, namely single-activity-dominant areas and multiple activity co-location clusters (co-location clusters I-IV). There are certain differences between the spatial distributions and activity combination types of various cluster patterns. In general, while the co-location cluster pattern has taken shape in Beijing, the proportion of balanced co-location cluster areas remains low. These cluster areas are mainly concentrated in the central urban district within the Sixth Ring Road and some urban districts in the outer suburbs and counties. Finally, the results indicate that the location conditions, especially the micro-location condition (distance to the nearest subway station), are the main factors that affect the resident activity cluster pattern. However, land price and population density, despite their limited influence when acting independently, remain fundamental influencing factors of the resident activity cluster patterns that cannot be ignored.

However, user data from various social media platforms are often affected by spurious correlation problems and their spatial and temporal dynamics may be partially linked to accidental events [75,76]. Therefore, the relevant data should be cleaned and filtered before being fed into the model for subsequent operations. Meanwhile, due to the biased nature of various social media data in terms of user groups, the relevant conclusions and patterns obtained by using social media data should be limited in scope (e.g., the relevant conclusions in this paper mainly reflect the activity patterns of relatively young groups). In order to improve the quality and reliability of the conclusions, other data from different crowd-sourcing platforms can be combined to corroborate them in a subsequent study [77]. In addition, the in-depth mining of social text data should continue to be strengthened, and multi-label technology should be employed to identify text information types and hidden content more efficiently. Moreover, spatiotemporal correlation data should be combined with information about other urban elements to effectively connect the users' places of residence, work, and activities, construct the daily life chains of residents, and improve the overall perception of residents' daily life space and the dynamic understanding of the urban spatial structure. Such research will provide more effective technical, theoretical support, and a governance basis with which to solve the practical problems of residents' daily activities and construct an efficient, convenient, and livable living pattern.

**Author Contributions:** Conceptualization, methodology, data curation, writing—review and editing: Jian Liu, Bin Meng and Juan Wang; investigation, software, visualization, writing—original and draft preparation: Jian Liu, Siyu Chen, Bin Tian and Guoqing Zhi; funding acquisition and project administration: Bin Meng and Juan Wang. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the National Key Research and Development Program of China (2017 YFB 0503605), the National Natural Science Foundation of China (41671165) and the Academic Research Projects of Beijing Union University (ZK40202001).

**Data Availability Statement:** The data is available from the authors upon reasonable request.

**Acknowledgments:** We would like to thank the anonymous reviewers for their insightful comments and substantial help in improving this article. We also thank Dongsheng Zhan for providing the valuable data and technical support.

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

## References

1. Salomon, I. Telecommunications and travel relationships: A review. *Transp. Res. A Gen.* **1986**, *20*, 223–238. [[CrossRef](#)]
2. Schwanen, T.; Kwan, M.P. The Internet, mobile phone and space-time constraints. *Geoforum* **2008**, *39*, 1362–1377. [[CrossRef](#)]

3. Kestens, Y.; Lebel, A.; Daniel, M.; Theriault, M.; Pampalon, R. Using experienced activity spaces to measure foodscape exposure. *Health Place* **2010**, *16*, 1094–1103. [[CrossRef](#)] [[PubMed](#)]
4. Vallée, J.; Cadot, E.; Roustit, C.; Parizot, I.; Chauvin, P. The role of daily mobility in mental health inequalities: The interactive influence of activity space and neighbourhood of residence on depression. *Soc. Sci. Med.* **2011**, *73*, 1133–1144. [[CrossRef](#)]
5. Widener, M.J.; Farber, S.; Neutens, T.; Horner, M. Spatiotemporal accessibility to supermarkets using public transit: An interaction potential approach in Cincinnati, Ohio. *J. Transp. Geogr.* **2015**, *42*, 72–83. [[CrossRef](#)]
6. Brockmann, D.; Hufnagel, L.; Geisel, T. The scaling laws of human travel. *Nature* **2006**, *439*, 462–465. [[CrossRef](#)] [[PubMed](#)]
7. González, M.C.; Hidalgo, C.A.; Barabasi, A.L. Understanding individual human mobility patterns. *Nature* **2008**, *453*, 779–782. [[CrossRef](#)] [[PubMed](#)]
8. Batty, M.; Axhausen, K.W.; Giannotti, F.; Pozdnoukhov, A.; Bazzani, A.; Wachowicz, M.; Ouzounis, G.; Portugali, Y. Smart cities of the future. *Eur. Phys. J. Spec. Top.* **2012**, *214*, 481–518. [[CrossRef](#)]
9. Osorio-Arjona, J.; García-Palomares, J.C. Social media and urban mobility: Using twitter to calculate home-work travel matrices. *Cities* **2019**, *89*, 268–280. [[CrossRef](#)]
10. Gong, Y.; Lin, Y.; Duan, Z. Exploring the spatiotemporal structure of dynamic urban space using metro smart card records. *Comput. Environ. Urban Syst.* **2017**, *64*, 169–183. [[CrossRef](#)]
11. Xu, Y.; Belyi, A.; Bojic, I.; Ratti, C. Human mobility and socioeconomic status: Analysis of Singapore and Boston. *Comput. Environ. Urban Syst.* **2018**, *72*, 51–67. [[CrossRef](#)]
12. Martí, 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](#)]
13. Lazer, D.; Pentland, A.; Adamic, L.; Aral, S.; Barabasi, A.L.; Brewer, D.; Christakis, N.; Contractor, N.; Fowler, J.; Gutmann, M.; et al. Computational Social Science. *Science* **2009**, *323*, 721–723. [[CrossRef](#)]
14. Song, C.; Qu, Z.; Blumm, N.; Barabasi, A.L. Limits of predictability in human mobility. *Science* **2010**, *327*, 1018–1021. [[CrossRef](#)]
15. Hssan, S.; Zhan, X.Y.; Ukkusuri, S.V. Understanding urban human activity and mobility patterns using large-scale location-based data from online social media. In Proceedings of the 2nd ACM SIGKDD International Workshop on Urban Computing, Washington, DC, USA, 13–17 January 2013; pp. 1–8. [[CrossRef](#)]
16. Fu, C.; McKenzie, G.; Frias-Martinez, V.; Stewart, K. Identifying spatiotemporal urban activities through linguistic signatures. *Comput. Environ. Urban Syst.* **2018**, *72*, 25–37. [[CrossRef](#)]
17. Liu, Y.; Yuan, Y.H.; Zhang, F. Mining urban perceptions from social media data. *J. Spat. Int. Sci.* **2020**, *20*, 51–55. [[CrossRef](#)]
18. Hägerstrand, T. What about people in Regional Science? In *Papers of the Regional Science Association*; Springer: New York, NY, USA, 1970; Volume 24, pp. 6–21. [[CrossRef](#)]
19. Parkes, D.N.; Thrift, N. Timing Space and Spacing Time. *Environ. Plan. A* **1975**, *7*, 651–670. [[CrossRef](#)]
20. Miller, H.J. Modelling accessibility using space-time prism concepts within geographical information systems. *Int. J. Geogr. Inf. Syst.* **1991**, *5*, 287–301. [[CrossRef](#)]
21. Kwan, M.P. Gender, the Home-Work Link, and Space-Time Patterns of Nonemployment Activities. *Econ. Geogr.* **1999**, *75*, 370–394. [[CrossRef](#)]
22. Chen, J.; Shaw, S.-L.; Yu, H.; Lu, F.; Chai, Y.; Jia, Q. Exploratory data analysis of activity diary data: A space–time GIS approach. *J. Transp. Geogr.* **2011**, *19*, 394–404. [[CrossRef](#)]
23. Axhausen, K.W.; Zimmermann, A.; Schönenfelder, S.; Rindsfüser, G.; Haupt, T. Observing the rhythms of daily life: A six-week travel diary. *Transportation* **2002**, *29*, 95–124. [[CrossRef](#)]
24. Schönenfelder, S.; Axhausen, K.W. Activity spaces: Measures of social exclusion? *Transp. Policy* **2003**, *10*, 273–286. [[CrossRef](#)]
25. Ettema, D.; van der Lippe, T. Weekly rhythms in task and time allocation of households. *Transportation* **2009**, *36*, 113–129. [[CrossRef](#)]
26. Vilhelmson, B. Daily mobility and the use of time for different activities. The case of Sweden. *GeoJournal* **1999**, *48*, 177–185. [[CrossRef](#)]
27. Ewing, R.; Robert, C. Travel and the Built Environment: A Synthesis. *Transp. Res. Rec.* **2001**, *1780*, 87–114. [[CrossRef](#)]
28. Maat, K.; van Wee, B.; Stead, D. Land Use and Travel Behaviour: Expected Effects from the Perspective of Utility Theory and Activity-Based Theories. *Environ. Plan. B Plan. Des.* **2016**, *32*, 33–46. [[CrossRef](#)]
29. Schlich, R.; Axhausen, K.W. Habitual travel behaviour: Evidence from a six-week travel diary. *Transportation* **2003**, *30*, 13–36. [[CrossRef](#)]
30. Huang, Q.; Wong, D.W.S. Modeling and Visualizing Regular Human Mobility Patterns with Uncertainty: An Example Using Twitter Data. *Ann. Assoc. Am. Geogr.* **2015**, *105*, 1179–1197. [[CrossRef](#)]
31. Ríos, S.A.; Muñoz, R. Land Use detection with cell phone data using topic models: Case Santiago, Chile. *Comput. Environ. Urban Syst.* **2017**, *61*, 39–48. [[CrossRef](#)]
32. Mayer-Schönberger, V.; Cukier, K. *Big Data: A Revolution That Will Transform How We Live, Work, and Think*; Houghton Mifflin Harcourt: Boston, MA, USA, 2013.
33. Rhee, I.; Shin, M.; Hong, S.; Lee, K.; Kim, S.J.; Chong, S. On the Levy-walk nature of human mobility. *IEEE/ACM Trans. Netw.* **2011**, *19*, 630–643. [[CrossRef](#)]
34. Sun, L.; Axhausen, K.W.; Lee, D.H.; Huang, X. Understanding metropolitan patterns of daily encounters. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 13774–13779. [[CrossRef](#)] [[PubMed](#)]

35. Lara, O.D.; Labrador, M.A. A Survey on Human Activity Recognition using Wearable Sensors. *IEEE Commun. Surv. Tutor.* **2013**, *15*, 1192–1209. [[CrossRef](#)]
36. Ahas, R.; Aasa, A.; Yuan, Y.; Raubal, M.; Smoreda, Z.; Liu, Y.; Ziemlicki, C.; Tiru, M.; Zook, M. Everyday space-time geographies: Using mobile phone-based sensor data to monitor urban activity in Harbin, Paris, and Tallinn. *Int. J. Geogr. Inf. Sci.* **2015**, *29*, 2017–2039. [[CrossRef](#)]
37. Kandt, J.; Leak, A. Examining inclusive mobility through smartcard data: What shall we make of senior citizens' declining bus patronage in the West Midlands? *J. Transp. Geogr.* **2019**, *79*, 102474. [[CrossRef](#)]
38. Csáji, B.C.; Browet, A.; Traag, V.A.; Delvenne, J.C.; Huens, E.; Van Dooren, P.; Smoreda, Z.; Blondel, V.D. Exploring the mobility of mobile phone users. *Phys. A* **2013**, *392*, 1459–1473. [[CrossRef](#)]
39. Calabrese, F.; Diao, M.; Lorenzo, G.D.; Ferreira, J., Jr.; Ratti, C. Understanding individual mobility patterns from urban sensing data: A mobile phone trace example. *Transp. Res. C Emerg. Technol.* **2013**, *26*, 301–313. [[CrossRef](#)]
40. Liu, Y.; Liu, X.; Gao, S.; Gong, L.; Kang, C.; Zhi, Y.; Chi, G.; Shi, L. Social Sensing: A New Approach to Understanding Our Socioeconomic Environments. *Ann. Assoc. Am. Geogr.* **2015**, *105*, 512–530. [[CrossRef](#)]
41. Watts, D.J. A twenty-first century science. *Nature* **2007**, *445*, 489. [[CrossRef](#)] [[PubMed](#)]
42. Cao, G.; Wang, S.; Hwang, M.; Padmanabhan, A.; Zhang, Z.; Soltani, K. A scalable framework for spatiotemporal analysis of location-based social media data. *Comput. Environ. Urban Syst.* **2015**, *51*, 70–82. [[CrossRef](#)]
43. Luo, F.; Cao, G.; Mulligan, K.; Li, X. Explore spatiotemporal and demographic characteristics of human mobility via Twitter: A case study of Chicago. *Appl. Geogr.* **2016**, *70*, 11–25. [[CrossRef](#)]
44. Hawelka, B.; Sitko, I.; Beinat, E.; Sobolevsky, S.; Kazakopoulos, P.; Ratti, C. Geo-located Twitter as proxy for global mobility patterns. *Cartogr. Geogr. Inf. Sci.* **2014**, *41*, 260–271. [[CrossRef](#)]
45. Bao, Y.; Huang, Z.; Li, L.; Wang, Y.; Liu, Y. A BiLSTM-CNN model for predicting users' next locations based on geotagged social media. *Int. J. Geogr. Inf. Sci.* **2020**, *35*, 639–660. [[CrossRef](#)]
46. Crooks, A.; Pfoser, D.; Jenkins, A.; Croitoru, A.; Stefanidis, A.; Smith, D.; Karagiorgou, S.; Efentakis, A.; Lamprianidis, G. Crowdsourcing urban form and function. *Int. J. Geogr. Inf. Sci.* **2015**, *29*, 720–741. [[CrossRef](#)]
47. Huang, Q.Y.; Wong, D.W.S. Activity patterns, socioeconomic status and urban spatial structure: What can social media tell us? *Int. J. Geogr. Inf. Sci.* **2016**, *30*, 1873–1898. [[CrossRef](#)]
48. Liu, X.; Huang, Q.; Gao, S.; Xia, J. Activity knowledge discovery: Detecting collective and individual activities with digital footprints and open source geographic data. *Comput. Environ. Urban Syst.* **2021**, *85*, 101551. [[CrossRef](#)]
49. Lansley, G.; Longley, P.A. The geography of Twitter topics in London. *Comput. Environ. Urban Syst.* **2016**, *58*, 85–96. [[CrossRef](#)]
50. Jendryke, M.; Balz, T.; McClure, S.C.; Liao, M. Putting people in the picture: Combining big location-based social media data and remote sensing imagery for enhanced contextual urban information in Shanghai. *Comput. Environ. Urban Syst.* **2017**, *62*, 99–112. [[CrossRef](#)]
51. Ye, C.; Zhang, F.; Mu, L.; Gao, Y.; Liu, Y. Urban function recognition by integrating social media and street-level imagery. *Environ. Plan. B* **2020**, *1*–15. [[CrossRef](#)]
52. Hasan, S.; Ukkusuri, S.V. Urban activity pattern classification using topic models from online geo-location data. *Transp. Res. C Emerg. Technol.* **2014**, *44*, 363–381. [[CrossRef](#)]
53. Tsou, M.-H.; Yang, J.-A.; Lusher, D.; Han, S.; Spitzberg, B.; Gawron, J.M.; Gupta, D.; An, L. Mapping social activities and concepts with social media (Twitter) and web search engines (Yahoo and Bing): A case study in 2012 US Presidential Election. *Cartogr. Geogr. Inf. Sci.* **2013**, *40*, 337–348. [[CrossRef](#)]
54. Yang, J.-A.; Tsou, M.-H.; Jung, C.-T.; Allen, C.; Spitzberg, B.H.; Gawron, J.M.; Han, S.-Y. Social media analytics and research testbed (SMART): Exploring spatiotemporal patterns of human dynamics with geo-targeted social media messages. *Big Data Soc.* **2016**, *3*, 1–19. [[CrossRef](#)]
55. Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics, Minneapolis, MN, USA, 2–7 June 2019; pp. 4171–4186.
56. Alvares, L.O.; Bogorny, V.; Kuijpers, B.; Moelans, B.; Fern, J.A.; Macedo, E.D.; Palma, A.T. Towards semantic trajectory knowledge discovery. In *Data Mining and Knowledge Discovery*; Hasselt University: Limbourg, Belgium, 2007; Volume 12.
57. Aggarwal, C.C.; Wang, H. Text mining in social networks. In *Social Network Data Analytics*; Aggarwal, C.C., Ed.; Springer: Cham, Switzerland, 2011; Chapter 13; pp. 353–378. [[CrossRef](#)]
58. Beijing Statistical Bulletin on National Economic and Social Development 2019. Available online: [http://www.beijing.gov.cn/gongkai/shuju/tjgb/202003/t20200302\\_1838196.html](http://www.beijing.gov.cn/gongkai/shuju/tjgb/202003/t20200302_1838196.html) (accessed on 12 March 2021).
59. Cai, J.X.; Huang, B.; Song, Y.M. Using multi-source geospatial big data to identify the structure of polycentric cities. *Remote Sens. Environ.* **2017**, *202*, 210–221. [[CrossRef](#)]
60. Owuor, I.; Hochmair, H.H. An Overview of Social Media Apps and their Potential Role in Geospatial Research. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 526. [[CrossRef](#)]
61. Batty, M. The pulse of the city. *Environ. Plan. B* **2010**, *37*, 575–577. [[CrossRef](#)]
62. Axhausen, K.W.; Tommy, G. Activity-based approaches to travel analysis: Conceptual frameworks, models, and research problems. *Transp. Rev.* **1992**, *12*, 323–341. [[CrossRef](#)]
63. Harvey, A.S. Guidelines for time use data collection. *Soc. Indic. Res.* **1993**, *30*, 197–228. [[CrossRef](#)]

64. Doherty, S.T. Should we abandon activity type analysis? Redefining activities by their salient attributes. *Transportation* **2006**, *33*, 517–536. [[CrossRef](#)]
65. Wang, J.; Meng, B.; Pei, T.; Du, Y.Y.; Zhang, J.Q.; Chen, S.Y.; Tian, B.; Zhi, G.Q. Mapping the exposure and sensitivity to heat wave events in China's megacities. *Sci. Total Environ.* **2021**, *755*, 142734. [[CrossRef](#)] [[PubMed](#)]
66. Zhan, D.S.; Xie, C.X.; Zhang, W.Z.; Ding, L.; Xu, J.X.; Zhen, M.C. Identifying mixed functions of urban public service facilities in Beijing by cumulative opportunity accessibility method. *J. Geo-Inf. Sci.* **2020**, *22*, 1320–1329. (In Chinese) [[CrossRef](#)]
67. Liu, L.; Chen, H.; Liu, T. Study on urban spatial function mixture and individual activity space from the perspectives of resident activity. *IEEE Access* **2020**, *8*, 184137–184150. [[CrossRef](#)]
68. Wong, K.; Domroes, M. Users' perception of Kowloon Park, Hong Kong: Visiting patterns and scenic aspects. *Chin. Geogr. Sci.* **2004**, *14*, 269–275. [[CrossRef](#)]
69. Maoh, H.; Tang, Z. Determinants of normal and extreme commute distance in a sprawled midsize Canadian city: Evidence from Windsor, Canada. *J. Transp. Geogr.* **2012**, *25*, 50–57. [[CrossRef](#)]
70. Wang, J.F.; Li, X.H.; Christakos, G.; Liao, Y.L.; Zhang, T.; Gu, X.; Zheng, X.Y. Geographical Detectors-Based Health Risk Assessment and its Application in the Neural Tube Defects Study of the Heshun Region, China. *Int. J. Geogr. Inf. Sci.* **2010**, *24*, 107–127. [[CrossRef](#)]
71. Wang, J.F.; Xu, C.D. Geodetector: Principle and prospective. *Acta Geogr. Sin.* **2017**, *72*, 116–134. (In Chinese) [[CrossRef](#)]
72. Jacobs, J. *The Death and Life of Great American Cities*; Vintage: New York, NY, USA, 1961.
73. Golder, S.A.; Macy, M.W. Diurnal and seasonal mood vary with work, sleep, and daylength across diverse cultures. *Science* **2011**, *333*, 1878–1881. [[CrossRef](#)] [[PubMed](#)]
74. Wang, J.F.; Liao, Y.L.; Liu, X. *Analysis on Spatial Data*; Science Press: Beijing, China, 2014. (In Chinese)
75. Braaksma, B.; Zelemburg, K. “Re-make/Re-model”: Should big data change the modelling paradigm in official statistics? *Stat. J. IAOS* **2015**, *31*, 193–202. [[CrossRef](#)]
76. Janssens, A.C.J.W.; Kraft, P. Research Conducted Using Data Obtained through Online Communities: Ethical Implications of Methodological Limitations. *PLoS Med.* **2012**, *9*, e1001328. [[CrossRef](#)] [[PubMed](#)]
77. Kovacs-Györi, A.; Ristea, A.; Havas, C.; Mehaffy, M.; Hochmair, H.H.; Resch, B.; Juhasz, L.; Lehner, A.; Ramasubramanian, L.; Blaschke, T. Opportunities and Challenges of Geospatial Analysis for Promoting Urban Livability in the Era of Big Data and Machine Learning. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 752. [[CrossRef](#)]