Article

Where Urban Youth Work and Live: A Data-Driven Approach to Identify Urban Functional Areas at a Fine Scale

Yiming Yan [1,2], Yuanyuan Wang [1,3,*], Zhenhong Du [1,2], Feng Zhang [1,2] [ID], Renyi Liu [1,3] and Xinyue Ye [4]

1   School of Earth Sciences, Zhejiang University, 38 Zheda Road, Hangzhou 310027, China;
    yanyiming@zju.edu.cn (Y.Y.); duzhenhong@zju.edu.cn (Z.D.); zfcarnation@zju.edu.cn (F.Z.);
    liurenyi@zju.edu.cn (R.L.)
2   Zhejiang Provincial Key Laboratory of Geographic Information Science, 148 Tianmushan Road,
    Hangzhou 310028, China
3   Ocean Academy, Zhejiang University, 1 Zheda Road, Zhoushan 316021, China
4   Department of Informatics, New Jersey Institute of Technology, Newark, NJ 07102, USA; xye@njit.edu
*   Correspondence: wangyuanyuanxy@zju.edu.cn; Tel.: +86-571-8827-3287

check for
updates

Abstract: As a major labor force of cities, young people provide a huge driving force for urban innovation and development, and contribute to urban industrial upgrading and restructuring. In addition, with the acceleration of urbanization in China, the young floating population has increased rapidly, causing over-urbanization and creating certain social problems. It is important to analyze the demand of urban youth and promote their social integration. With the development of the mobile Internet and the improvement of the city express system, ordering food delivery has become a popular and convenient way to dine, especially in China. Food delivery data have a significant user attribute where the ages of most delivery customers are under 35 years old. In this paper, we introduce food delivery data as a new data source in urban functional zone detection and propose a time-series-based clustering approach to discover the urban hotspot areas of young people. The work and living areas were effectively identified according to the human behavioral characteristics of ordering food delivery. Furthermore, we analyzed the relationship between young people and the industry structure of Hangzhou and discovered that the geographical distribution of the identified work areas was similar to that of the Internet and e-commerce companies. The characteristics of the identified living areas were also analyzed in combination with the distribution of subway lines and residential communities, and it was found that the living areas were mainly distributed along subway lines and that urban villages appeared in the living hotspot regions, indicating that transportation and living cost were two important factors in the choice of residential location for young people. The findings of this paper can help urban industrial and residential planning and young population management.

Keywords: food delivery data; urban youth; urban hotspots; industrial structure; urban village

1. Introduction

The youth population is a group with strong creativity and high consumption power, who provide a strong driving force for urban innovation and development. With the acceleration of urbanization in China, a large number of young people have migrated from villages to cities. The fast-growing young floating population causes over-urbanization, which leads to urban congestion and insufficient infrastructure, and creates certain social problems, especially with respect to housing, employment,

and transportation [1–3]. On the other hand, young people are accustomed to learning new things and can better adapt to the development of emerging industries [4] such as new-generation information and renewable energy technologies, and they contribute to urban industrial upgrading and restructuring. Therefore, it is important to analyze the demand of urban youth and facilitate their social integration to promote sustainable urban development.

Housing and employment are hot research areas for urban issues and are also the focus of urban youth. Many studies use questionnaires or government survey statistics to analyze the urban residential environment and employment situation such as the built environment of residential areas and workplaces, the choice of the commuting mode, and the regional jobs–housing balance [5–7]. However, these survey methods have certain defects in subjective bias, sample size, and timeliness. In recent years, with the widespread use of positioning devices such as civilian GPS (Global Positioning System) technology on mobile terminals and the development and popularization of location-based services and mobile social networks, a large number of trajectory or geotag data are accumulating in daily life and are serving different types of applications [8–10]. These data sources mainly include mobile phones [11–16], taxis [17–21], public or shared bicycles [22–25], transit smart cards [26–29], and so on. They convey human mobility and activity information and provide content and methodological innovation to the research of the spatial-temporal behavior of people in urban areas [8,30].

The identification of urban functional areas such as residential and industrial areas is an important way to understand urban land use structure and population distribution [31,32]. Mao et al. extracted the spatial distribution of urban jobs and housing areas by using a re-clustering algorithm based on the "job–residential factor" index from massive taxi origin-destination (OD) points [17]. Pei et al. identified land use types such as residential, business, commercial, open space and others from mobile phone activity data by using a semi-supervised clustering method [13]. Long et al. analyzed the urban job–housing structure and commuting patterns with the location–time–duration (PTD) data mined from the raw bus smart card data [29]. Zhang et al. used urban bicycle rental records and Point-of-Interest (POI) information to delineate urban functional zones by applying topic modeling techniques where business areas, industrial areas, and residential areas were well detected [25]. However, the above-mentioned data have certain geographical limitations. Taxi GPS data mainly exist in road networks, and the location of the mobile phone data comes from the locations of the base transceiver stations. The smart card data only provide the location of transportation stops and stations, and the origin and destination location of public bicycles are determined by the service stations. These limitations will not be able to effectively identify the complex geographical environment such as communities and colleges where they are outside the road networks or away from the transportation stations. In addition, most studies on the relationship between urban areas with age, gender, and socio-economic factors are based on static survey data. Therefore, there is a lack of a big data-driven approach to detect the geographic distribution of specific age groups.

Due to the wide usage of mobile phones and the convenience of delivery systems, food delivery has become a popular way of eating because it has convenient order services, takes less time, and offers a variety of food options. Food delivery data are trajectory data produced by delivery riders from the store to the customer's place after receiving the order, usually using electric scooters in China. According to the China delivery big data reports issued by the food delivery services (such as *Ele.me, meituan*) in 2017 [33–35], the number of online catering users exceeded 300 million, and the delivery dataset has significant user characteristics as follows: (1) the largest user population are white-collar workers, which account for over 80% of the total delivery orders, and the second largest user population are college students; and (2) the majority of users are under 35 years old. In other words, young people are the major consumer groups of online ordering.

In this paper, food delivery data are introduced as a new data source into urban functional zone detection. Compared with other trajectory or geotag data, the characteristics of delivery data provide a new way to explore the spatially significant areas of the urban youth group. In addition, the electric delivery scooters can enter complex geographical areas such as residential communities,

college campuses, and large industrial parks, and can provide a more accurate dataset for functional area extraction. This paper proposes a time-series-based clustering approach to discover the urban hotspot areas of young people. A case study of Hangzhou, China, which is famous for its Internet economy, was implemented to discuss the relationship between urban youth and the urban industrial structure and the factors that affect the residential location choice of young people.

The remainder of this paper is organized as follows: Section 2 describes the study area and data source, and performs some statistics of the dataset. Section 3 discusses the method for extracting urban functional areas from food delivery data, and the results are verified with office urban planning maps. Section 4 analyzes the spatial distribution of the urban youth population, combined with the industrial structure and the distribution of residential areas in Hangzhou. This paper concludes with a brief summary and discussion in Section 5.

2. Study Area and Data

Hangzhou, capital of East China's Zhejiang Province, is famous for its Internet and digital economy, and it is also home to Chinese Internet and e-commerce giant Alibaba and a number of other world-leading high-tech companies. According to the statistics in 2018, the main business income of Hangzhou's digital economy exceeded 1 trillion yuan, and its added value reached 335.6 billion yuan, accounting for 24.8% of the city's GDP [36]. The study area of this paper is the main city area of Hangzhou, which covers 74 street districts with an area of over 1600 km^2 (Figure 1).
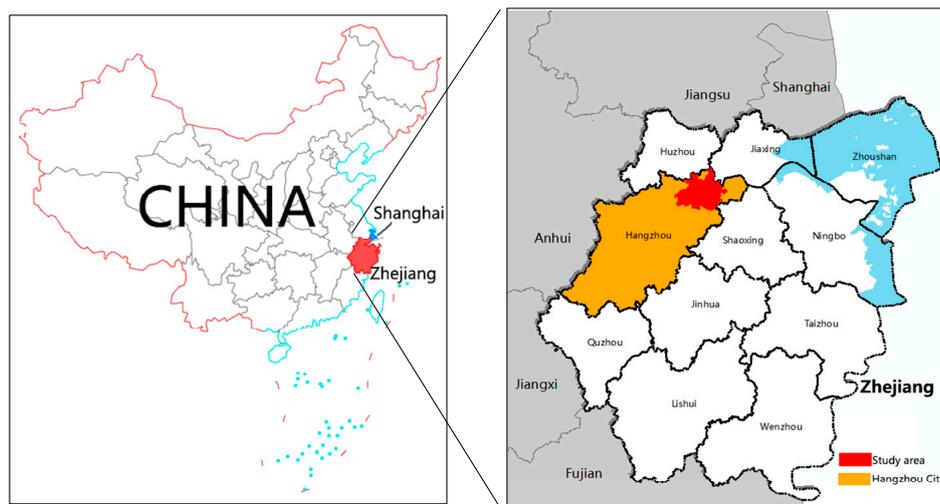

Figure 1. Study area.

The dataset used in this paper was provided by the Chinese logistics company "*Dianwoda*", which recorded food delivery electric scooter trips that occurred in Hangzhou from 28 July 2017 to 26 September 2017, and the user-related information was removed for privacy protection. The original delivery data collected real-time GPS information of the electric scooters during the entire food delivery process. The fields of the dataset mainly include the order ID, rider ID, status, timestamp, latitude, and longitude. The order ID is a unique identifier for a delivery order, while the rider ID is a unique identifier for a delivery rider. Real-time GPS location information is converted to latitude and longitude, accompanied by a time stamp. The field "status" represents different stages of the food delivery process including dispatch, arriving, arrive, leave, delivering, and finish. In this study, the origin and destination (OD) pairs for each delivery order are worthy of our major concern, where the origin for the delivery order refers to the location of the restaurant, and the destination refers to the location of the delivery customer. The record, whose status changes from arrive to leave, was extracted as the origin of one delivery order, and the record, whose status changes from delivering to finish, was

extracted as the destination. In this way, we could obtain the delivery OD dataset by extracting and reorganizing the delivery electric scooter trip data.

In the delivery OD dataset, each record includes the order ID, the departure and arrival time, the origin and destination location, the delivery trip distance, and the delivery time. The delivery OD dataset contained 7,480,402 orders. Table 1 shows a small subset of these data for reference, and Figure 2 shows the location distribution of the food delivery destination points in the study area.

Table 1. A small subset of the delivery origin-destination (OD) dataset. Each row corresponds to an occupied delivery order.

| Order ID	Departure Datetime	Departure lon	Departure lat	Arrival Datetime	Arrival lon	Arrival lat	Duration (sec)	Distance (meter)
166256640	2017-07-29 20:59:48	120.279905	30.305337	2017-07-29 21:19:55	120.28229	30.313441	1206	4187.01
166862850	2017-07-30 13:27:28	120.152728	30.182544	2017-07-30 13:49:26	120.144546	30.161267	1317	4587.01
165838860	2017-07-29 13:57:53	120.32855	30.305576	2017-07-29 14:04:21	120.328559	30.305209	389	1403.50
164347920	2017-07-28 10:18:44	120.382136	30.305377	2017-07-28 10:27:33	120.387719	30.299668	529	1977.62
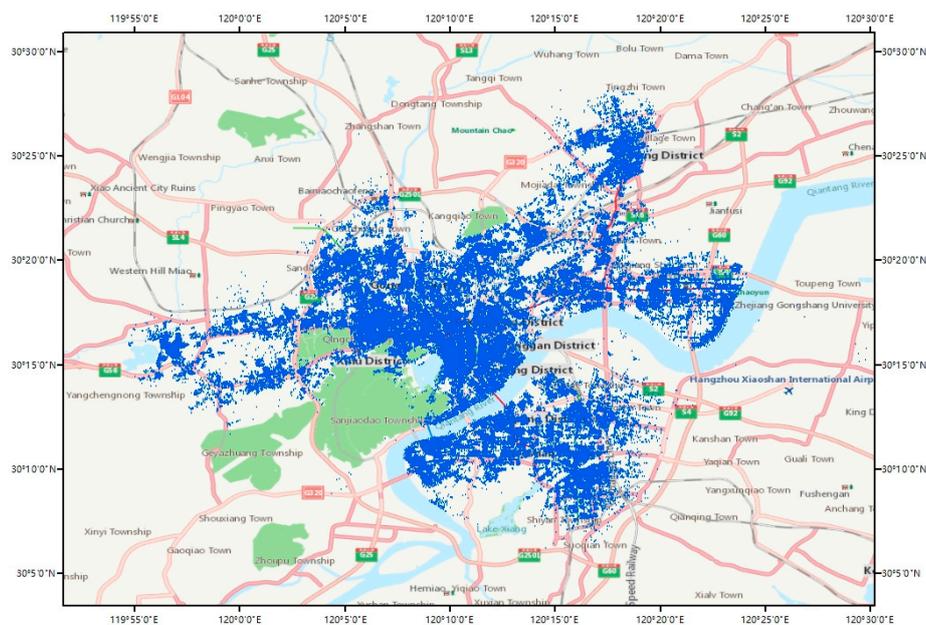
Figure 2. Location distribution of the food delivery destination points in the study area.

As with many large datasets, the delivery OD dataset contains errors. Some errors are easy to identify such as point coordinates out of the research area or order trips with a negative distance or time. Table 2 shows the thresholds used in each filter as well as the quantity of orders that violate the threshold. A trip distance of less than 0 as a negative value was filtered because it is geometrically impossible, and the orders whose trip distance was more than 20 km were also filtered because of their low incidence. The orders with a delivery time between 1 min and 90 min were reserved. Figure 3 shows the delivery duration and trip distance distribution of the delivery OD data after filtering.

Table 2. The error thresholds that are applied to the trip distance and duration of the dataset. The last column shows the percentage of orders that violate each threshold.

| Feature	Lower Threshold	Upper Threshold	Portion of Errors (%)
Trip Distance (meters)	0	20,000.0	0.20
Duration (min)	1	90	0.10
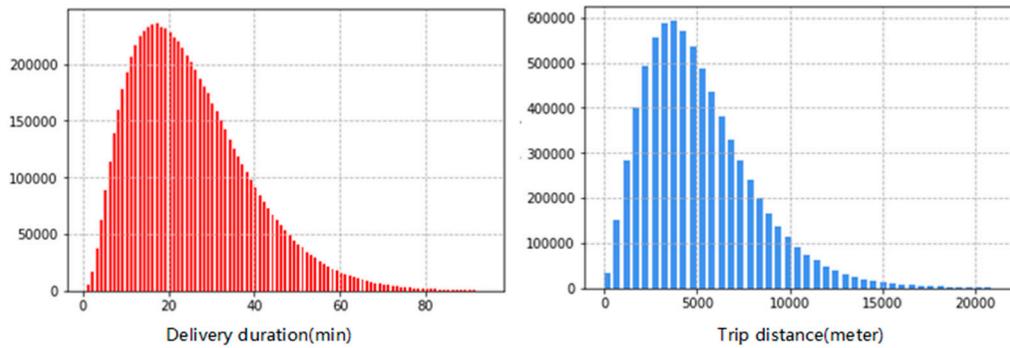
Figure 3. Delivery duration and trip distance distribution after filtering.

We also calculated some statistics to understand the characteristics of the dataset. As shown in Figure 4, we counted the average number of delivery orders of each day in one week and the average number of delivery orders on an hourly basis in one day for both weekdays and weekends. It was observed that people ordered more deliveries on weekends than on weekdays. There were two peaks on each day, one peak from 11:00 to 13:00 and another peak from 17:00 to 20:00, which was similar on both weekdays and weekends.
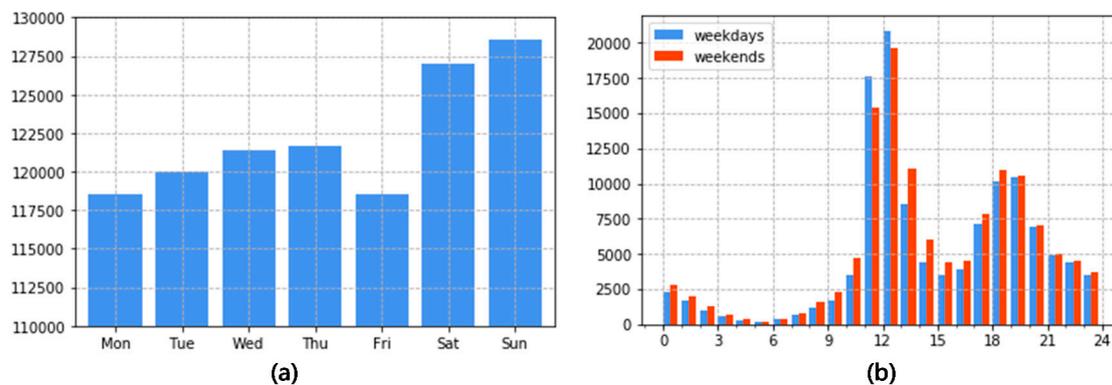

Figure 4. (**a**) Average number of delivery orders each day in a week; (**b**) Average number of delivery orders on an hourly basis on weekdays and weekends.

3. Methods

In this section, we aimed to explore the spatially significant areas of urban youth, and determine the functional types of the extracted areas. The process was divided into the following steps: (1) map the food delivery destination points into a spatiotemporal cube and divide the research area into equal grids; (2) calculate the delivery hot value by using the modified Getis-Ord statistic method, and construct the time series by combining the time pattern with the delivery hot values of each grid; and (3) cluster the time series of the delivery data with the unsupervised k-means method and determine the functional types of clusters based on the behavioral characteristics of ordering food delivery. Finally, the results were verified by office urban planning maps.

3.1. Spatiotemporal Cube Construction

In this paper, the delivery destination points are the main processing target. As shown in Table 1, the destination points have longitude–latitude information and the arrival time, which can be mapped into a three-dimensional Euclidian space. We constructed a spatiotemporal cube model where the cubes had equal space size and time intervals. Each cube was represented as a three-dimensional array $c = (x_c, y_c, t_c)$, where x_c is the number of cubes in the x dimension, y_c is the number of cubes in the y dimension, and t_c is the number of cubes in the time dimension. We set the time size to 1 h and the spatial size to 0.001 degrees (approximately 95 to 115 m). The research area was also divided into equal grids, as shown in Figure 5. After data mapping and aggregation, each spatiotemporal cube stored the count of the destination points in that cube as its attribute.
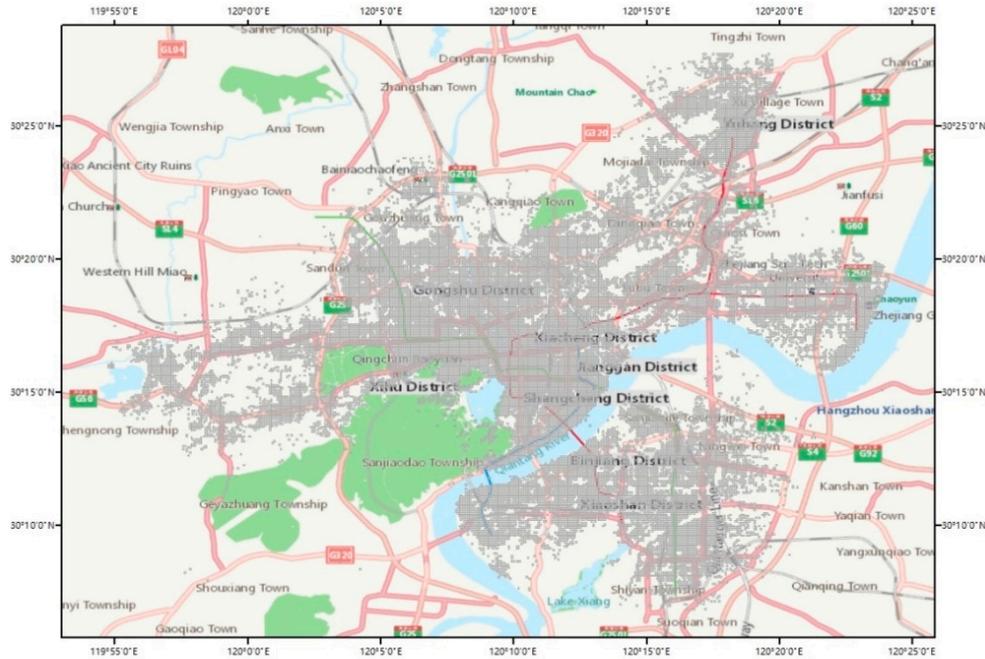
Figure 5. The research area covered by equal grids.

3.2. Time Series Construction

3.2.1. Modified Getis-Ord Statistic Method

The well-known Getis-Ord statistic method was used in this paper to calculate the hot value of each cube. To be a statistically significant hotspot, a feature will have a high value and be surrounded by other features that also have high values. The method provides a z-score that allows users to determine where the features with either high or low values are clustered spatially. Formally, based on the constructed spatiotemporal cube, the Getis-Ord value G_i^* of each cube is defined as [37,38]:

$$G_i^* = \frac{\sum_{j=1}^n w_{i,j} a_j - \bar{a} \sum_{j=1}^n w_{i,j}}{S \sqrt{\frac{n \sum_{j=1}^n w_{i,j}^2 - \left(\sum_{j=1}^n w_{i,j}\right)^2}{n-1}}}, \tag{1}$$

$$\bar{a} = \frac{\sum_{j=1}^n a_j}{n} \text{ and } S = \sqrt{\frac{\sum_{j=1}^n a_j^2}{n} - \bar{a}^2}, \tag{2}$$

where n refers to the total number of cubes, and a_j is the attribute value of cube c_j. $w_{i,j}$ is the spatial weight between cubes i and j, and is defined as follows:

$$w_{i,j} = \begin{cases} 1, & \text{if } c_i \text{ and } c_j \text{ are neighbors} \\ 0, & \text{otherwise} \end{cases}, \tag{3}$$

Here, we define two cubes as neighbors only if the maximum distance between i and j in any of the three coordinates $(x_c, y_c, t_c) \leq 1$. A cube is also its own neighbor. Then, to simplify the computation of the Getis-Ord statistic value, we use $N(c_i)$ to denote the set of neighboring cubes of c_i and $|N(c_i)|$ to denote the number of such neighboring cubes. Therefore, we obtain the following simplified formula for the Getis-Ord statistic:

$$G_i^* = \frac{\sum_{j \in N(c_j)} a_j - \bar{a} \cdot |N(c_i)|}{S \sqrt{\frac{n|N(c_i)| - |N(c_i)|^2}{n-1}}}, \tag{4}$$

3.2.2. The Time Series Construction

It is well known that people usually have different behavior patterns on weekdays and on weekends; thus, we built a 48-h timeline that consisted of 24-h periods on weekdays and weekends. After the Getis-Ord statistic calculation was performed for all the cubes, a group function was used to gather the cubes with the same location and time period, and an aggregate function averaged the Getis-Ord values of the cubes in the same group. Finally, the synthesized time series were obtained as a linear combination of the 48-h timeline and the statistical value of delivery order quantity. Then, each grid would have a time series $TS_k\left(\{g_t^k, t = 1, 2, \ldots, 48; k = 1, 2, \ldots, n\}\right)$, where n is the total number of the grids, and g_t^k is the hot value of grid k at time period t (1–24 for weekdays and 25–48 for weekends).

As the Getis-Ord statistic method is a common method for hotspot analysis, we also obtained 48 delivery hotspot distribution maps that corresponded to 48 time periods. The delivery hotspot distribution during typical time periods (i.e., lunch time, dinner time, and midnight) is visualized in Figure 6.

3.3. Time Series Classification and Identification

The statistical hot values in the time series represent the delivery quality in one region during one period. That is, when the hot value is higher, more delivery orders are placed in this area, and low hot values generally represent less delivery orders. To better identify functional areas, the grids whose hot values were negative in all time periods were removed. To emphasize the characteristics of the delivery data and weaken the impact of regional factors, the hot values of the time series were normalized for each grid.

The clustering method k-means was applied to classify the time series. Unsupervised clustering methods require the number of clusters to be known beforehand. To determine an appropriate value for K, the ratio between the intra-cluster and inter-cluster distances was introduced as a validity index [39], which is defined as:

$$intra - cluster = \frac{1}{N} \sum_{i=1}^{K} \sum_{x \in C_i} (x - z_i)^2, \tag{5}$$

$$inter - cluster = \min_{i \neq j}(z_i - z_j)^2, \tag{6}$$

$$validity = \frac{intra - cluster}{inter - cluster}, \tag{7}$$

where N is the number of instances; K is the number of clusters; and z_i is the centroid of cluster C_i.
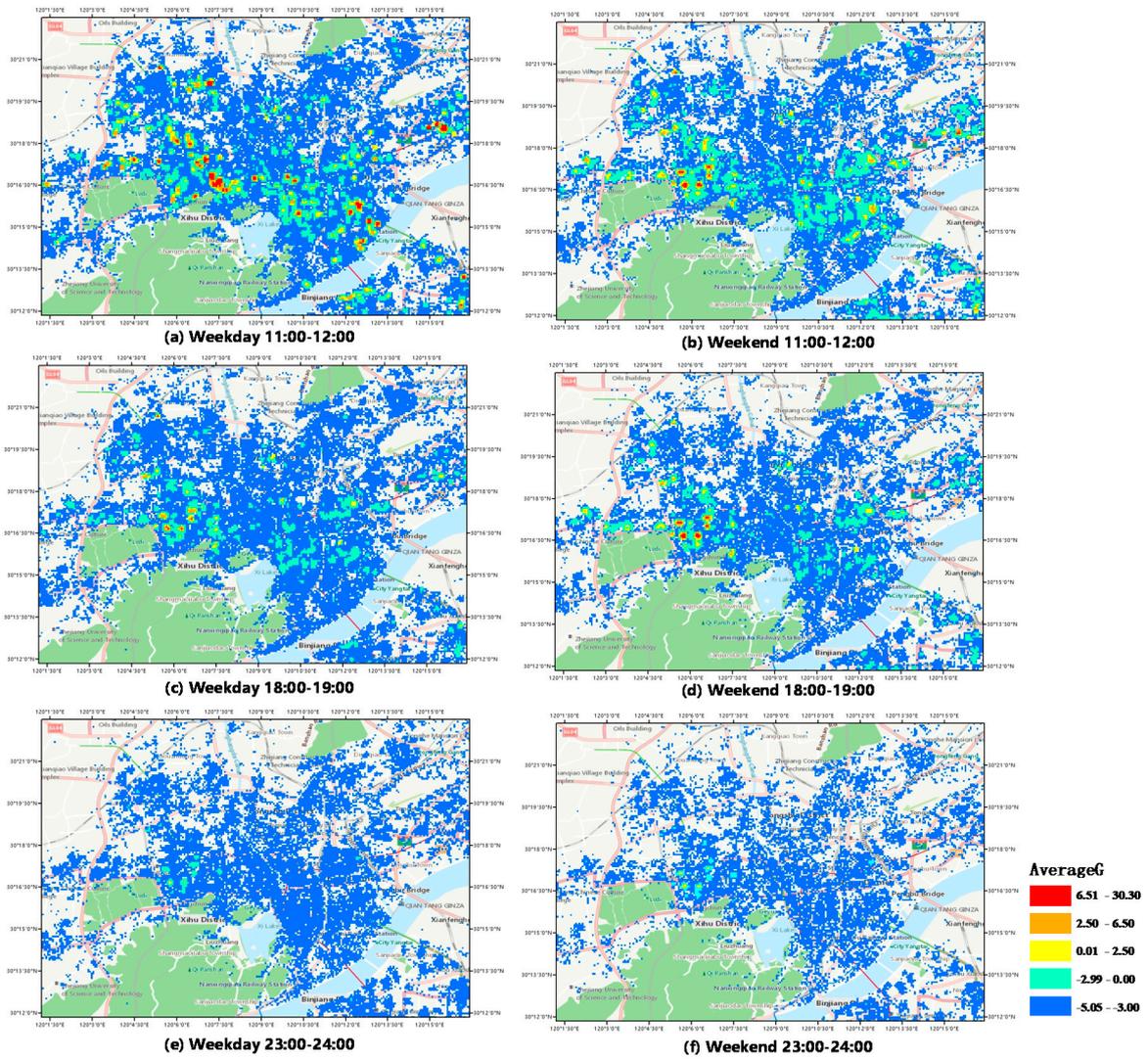
Figure 6. Delivery hotspot distribution during typical time periods.

An ideal partition will minimize the intra-cluster distance and maximize the inter-cluster distance; therefore, the best partition will be the partition that minimizes the validity value. Figure 7 presents the validity values by using the k-means cluster method to the synthesized time series for $K \in \{2, \ldots, 10\}$, and the minimum validity value was obtained when $K = 6$.
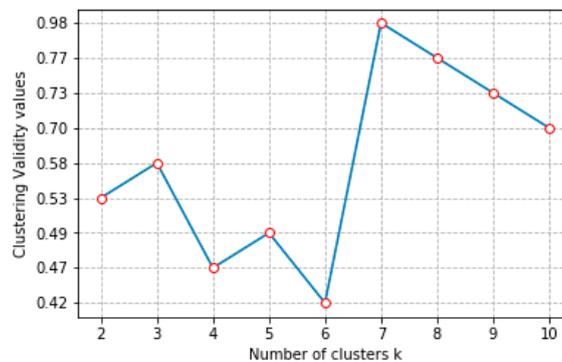

Figure 7. Validity values in different numbers of clusters by using k-means.

Figure 8 represents the cluster results of the time series by using the k-means method with K = 6, where the plotted curves (each consisting of 48 points) represent the cluster centers (red for weekdays and blue for weekends). In addition, the maximum and minimum values of each time period in each class were plotted in a thin color shadow. Figure 9 shows the location distribution of the grids according to the cluster results.
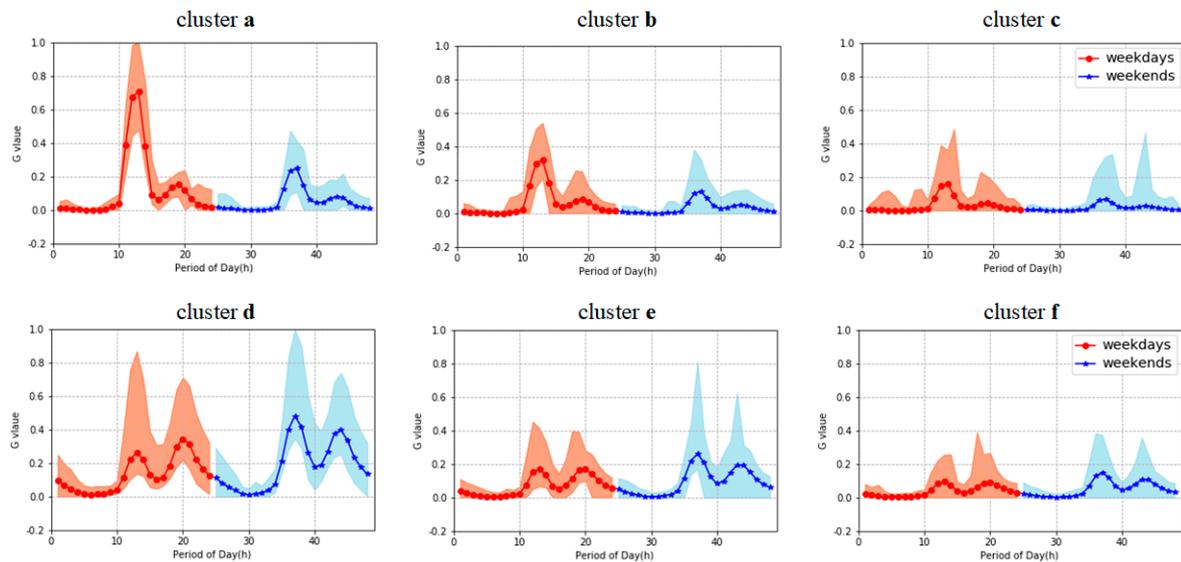
Figure 8. Cluster results of the time series by using the k-means method with $K = 6$.
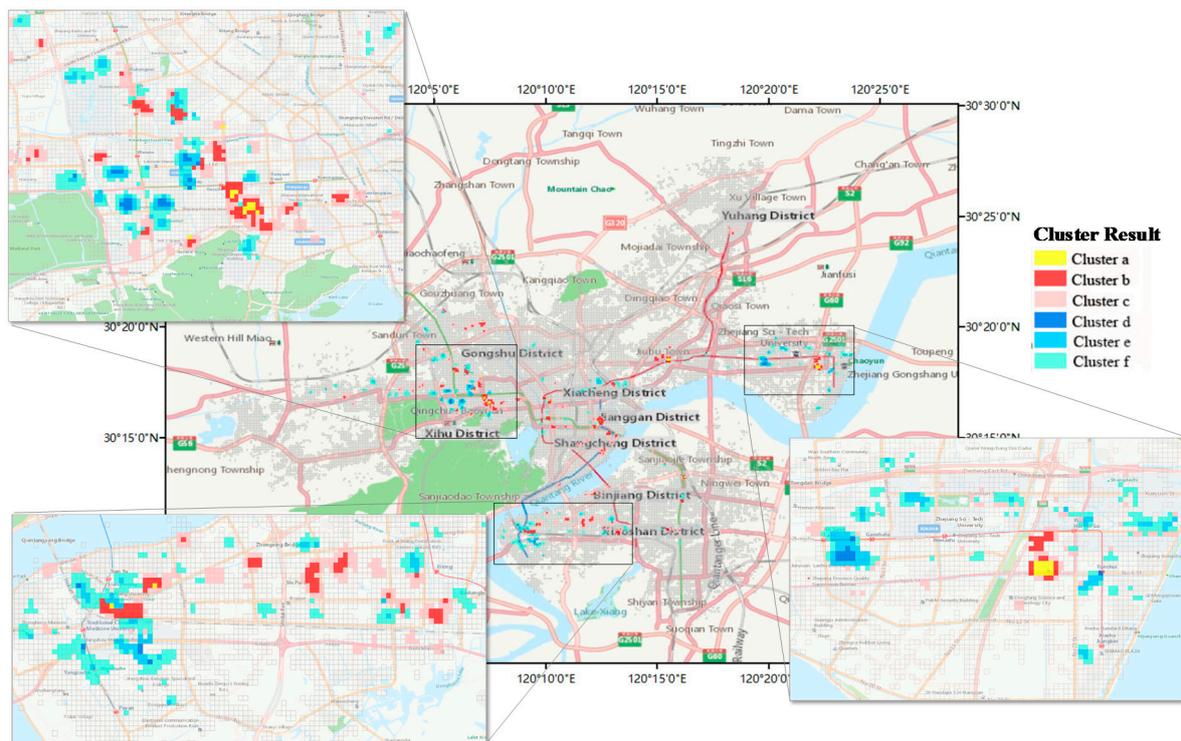
Figure 9. Grid distribution of the cluster results in the research area.

Within different land use areas, people may demonstrate different routine activities. This allows us to determine the social functions of areas based on the customary characteristics of ordering delivery. The peaks of the cluster curves were concentrated in four dining periods (i.e., at lunch time and dinner time on weekdays and weekends). By comparing the curves of the cluster centroids, we found that six

clusters could be divided into two categories based on their similar curve characteristics: one category contains clusters a–c, and another category contains clusters d–f.

Clusters a–c: This category is characterized by the fact that a significant peak appears during lunch time on weekdays. Delivery during weekdays is more than delivery during weekends, and delivery during lunch time is greater than delivery during dinner time both on weekdays and weekends. This category shows a work-related activity characteristic, and the hypothesis is that the areas included in this category are used as industrial parks, office buildings, and/or commercial-related areas.

Clusters d–f: This category representative shows that during weekdays, delivery at dinner time is greater than delivery at lunch time, especially in cluster d. Delivery during weekends is greater than delivery during weekdays. These characteristics imply living areas, where people come back after work on weekdays and stay during weekends, which may include residential communities, apartments, and/or college dormitories.

Clusters in the same category also exhibited different characteristic strength, with a > b > c and d > e > f. The factors that contribute to different cluster characteristic strength may include the following: (1) The number of people in a region varies during different time periods, for example, in industrial parks, the workers are mainly concentrated during the daytime on weekdays; and (2) the dining time is short or limited. The peak time of ordering delivery is related to people's eating habits. However, white-collar workers may be limited by work arrangements, resulting in a dining peak. The strong cluster characteristic also indicates that the social function of this area is significant.

3.4. Cluster Results Validation

To validate our hypotheses, we compared the cluster results against office urban planning maps released by the Hangzhou Planning Bureau, in which the land use types mainly include: residential, commercial, industrial, educational, and green (Figure 10).
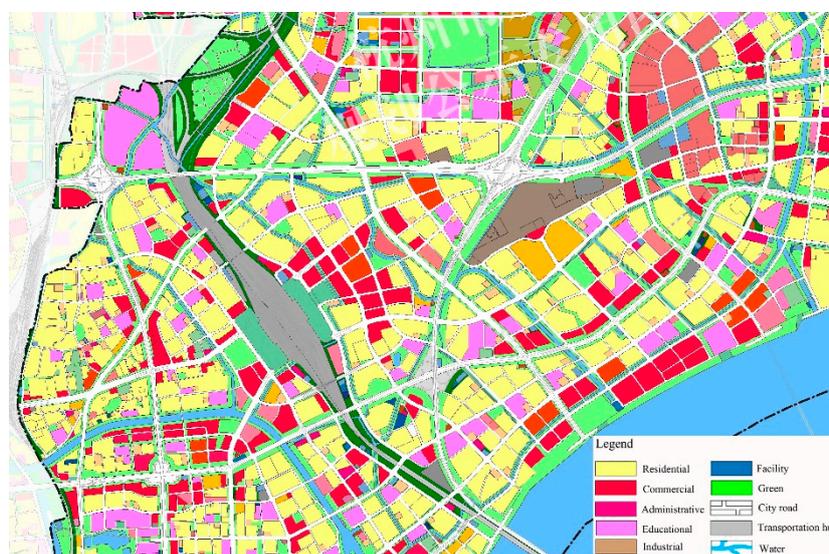

Figure 10. Office urban planning map for Hangzhou (partial).

As shown in Figure 9, the cluster results are displayed in the form of grids on maps. To understand how well the clusters were identified as the work and living areas, we evaluated the percentage of overlap that exists between the grids of the clusters and the official urban planning maps. This way, we would have an understanding of the accuracy of our approach as well as of the difference between the results. It should be noted that this verification method is an approximate measure due to the different granularities of maps or human factors.

Table 3 shows the percentage of overlap between the office planning map (columns) and six cluster results (rows). Here, the office planning map was divided into the six land use types of residential,

commercial, industrial, educational, mixed commercial and residential, and other. Commercial and industrial types refer to work areas on which office buildings and industrial parks are located; residential type areas are obviously living areas; educational areas are mixed areas in which dormitories are living type areas, and research institutes are work type areas. We can overlay the cluster grids with the office planning map to see which land use type the grids belong. Each element in the table represents the percentage of grids that belong to each land use type in one cluster.

According to our assumptions, clusters a–c represent work areas, while clusters d–f represent living areas. The commercial and industrial types are the two land use types that occupy the greatest proportion in clusters a–c, which account for 94.6%, 81.2%, and 66.4% in total, respectively. In cluster a, the industrial type accounts for 64.9% and was more than twice the proportion of the commercial type, which means that large-scale industry parks have a stronger industrial aggregation effect. The residential type obviously occupied the greatest proportion in clusters d–f, which account for 87.2%, 64.7% and 59.0%, respectively. The educational type is the second largest proportion in clusters e and f, perhaps because a college dormitory area, as a type of living area, is a hotspot for delivery due to its large student user group.

Table 3. Percentage of overlap between the official planning maps and the cluster results for Hangzhou.

Cluster Results	Official Planning Map					
	Residential	**Commercial**	**Industrial**	**Educational**	**Mixed Commercial and Residential**	**Other**
a	0	**29.7%**	**64.9%**	0	2.7%	2.7%
b	8.3%	**44.0%**	**37.3%**	1.3%	2.9%	6.2%
c	12.4%	**42.6%**	**23.9%**	4.6%	2.7%	13.8%
d	**87.2%**	2.6%	0	**7.7%**	2.6%	0
e	**64.7%**	3.3%	1.2%	**25.3%**	2.1%	3.3%
f	**59.0%**	10.1%	3.9%	**18.1%**	2.2%	6.6%

Overall, according to the distribution of the land use types in each cluster, the work and living areas extracted by the method of this paper were credible. In addition, it could be discovered that the recognition accuracy of the regional function type was related to the strength of the cluster characteristics. That is, when the cluster characteristic strength is stronger, the accuracy of the regional function type identification is higher.

4. Analysis

4.1. Work Areas

The work areas represent the location distribution of companies, enterprises, and industrial parks. As Hangzhou is famous for its Internet and data economy, we obtained information on Internet and e-commerce companies from a recruitment website called *lagouwang* (https://www.lagou.com) to compare the extracted work areas of this paper. The dataset included the company ID, company name, industry field, company scale, and company address, and all company address text were converted to latitude and longitude coordinates. After filtering out the data that were outside the research area and that contained errors, we finally had 1005 Internet and e-commerce company items. Figure 11 shows the location distribution of the companies.

We used heat maps to visualize the hotspot regions of the work areas, helping to discover the distribution pattern and potential characteristics. The heat maps were made with the kernel density estimation (KDE) method, which calculates the density of the features in a neighborhood around these features. Before making heat maps, three parameters should be set, namely the output cell size, the search radius, and the population field. The cell size was set to 0.0001 degrees, and the search radius was set to 0.005 degrees. The population field value determines the number of times to count the

feature, which could be used to weigh some features more heavily than other features or to allow one point to represent several observations. In this section, the population field was set based on the scale of the companies and the strength of the cluster results' characteristics, as shown in Table 4.
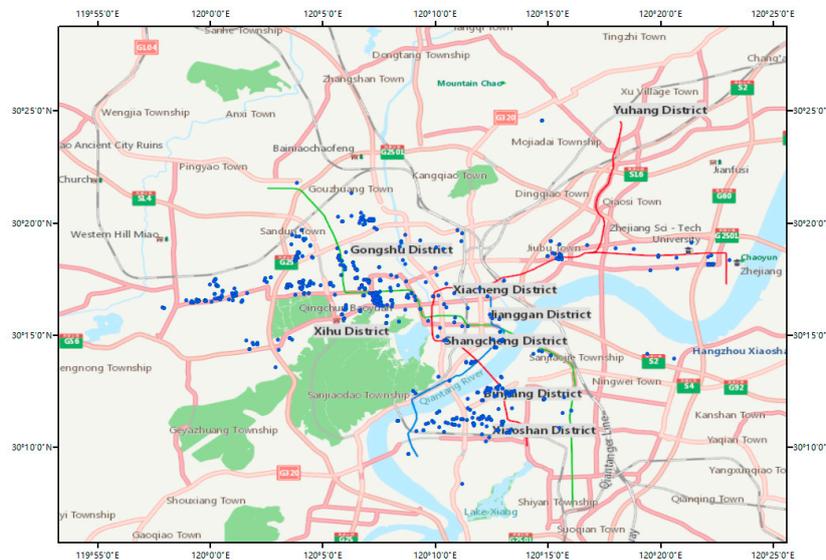

Figure 11. Location distribution of Internet and e-commerce companies from *lagouwang*.

Table 4. The population field set for the kernel density estimation method.

Population Field	1	2	3	4	5	7
Company scale	<15	15–50	50–150	150–500	500–2000	>2000
Cluster result	Cluster c		Cluster b			Cluster a

Figure 12 shows the heat maps made from the Internet and e-commerce companies and the extracted work areas. We discovered that the two heat maps had similar hotspot region distributions. This shows that a large number of young people were working in the Internet-related industries areas. There is a mutual driving relationship between industry agglomeration and talent gathering [40]. Regional industrial structure plays an important role in attracting talent or human capital [41], and the improvement of Internet-related industrial agglomeration level can promote the development of young labor gathering. The enhancement of the young labor gathering level can also promote the strengthening of Internet-related industrial agglomeration in the region. From the phenomenon that the spatial distribution of the delivery customers' work places is similar to the spatial distribution of the Internet and e-commerce companies, we can obtain an inference that Hangzhou's Internet-related industry has a well-planned structure and good development, which can provide sufficient employment and development opportunities for young people.
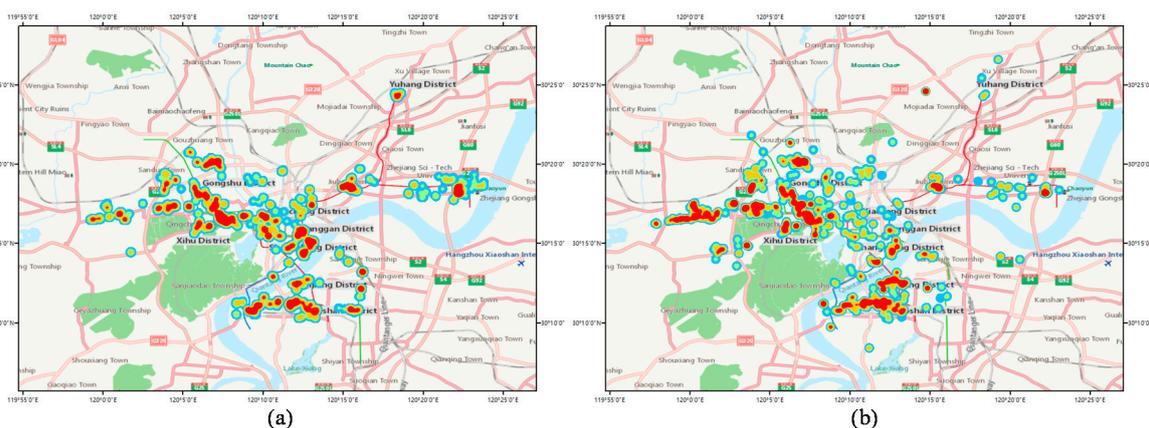
Figure 12. (**a**) Heat map made from the extracted work areas of this paper; (**b**) Heat map made from Internet and e-commerce companies.

The layout of urban industrial planning will affect the distribution of work hotspots. As shown in Figure 13, areas A, B, C, and D are the four high-tech areas planned by the government.

Area A is located in Hangzhou Future Sci-tech City, also called Haichuang Park, which has made great efforts to cultivate industries such as electronic information, new energy and materials, and so on. It has a series of new high-tech enterprises such as the Alibaba Taobao Town and Zhejiang Overseas High-level Talent Innovation Park.

Area B is located in the Xiasha District, which stands for the Hangzhou Economic and Technological Development Area. Xiasha is also a district of colleges and universities, 14 in total, which provide a large number of high-quality young talent. Several high-tech industrial parks are located here including the Hangzhou Singapore Science and Technology Park, Hangzhou Smart Valley Mobile Internet Pioneer Park, and Hangzhou High-Tech Enterprise Incubation Park.

Area C is located in the Hangzhou Binjiang High-Tech Industrial Development Zone, which has followed the leadership of the Scientific Outlook on Development and persisted in the "development of high technology and realization of industrialization". It has many famous Internet companies such as NetEase, Hikvision, Alibaba, and Huawei, and has large industrial parks such as the Hangzhou High-Tech Software Park, Xike Technology Park, and Shangfeng E-Commerce Industrial Park.

Area D is located in Hangzhou Northern New City, whose goal is to become the second largest Internet innovation center in Hangzhou. The large industrial parks located here include Hangzhou North Software Park and Paradise e Valley E-commerce Creative Industry Park.

The establishment of high-tech zones in the city, where locations are generally selected in the peripheral areas of the city, will help the city to gradually change from a single-center form to a multi-center form and drive the development of related industries by sharing resources and overcoming external negative effects to thus effectively promote the formation of industrial clusters.

Area E in Figure 13 is the urban downtown area, and most of the traditional and prosperous business districts are located here such as Wulin, Huanglong, Qianjiang New Town, and QianJiang Century City. There are many office buildings in these business districts, which have good transportation and communication conditions and high-quality supporting facilities, thus attracting a large number of companies to settle here.
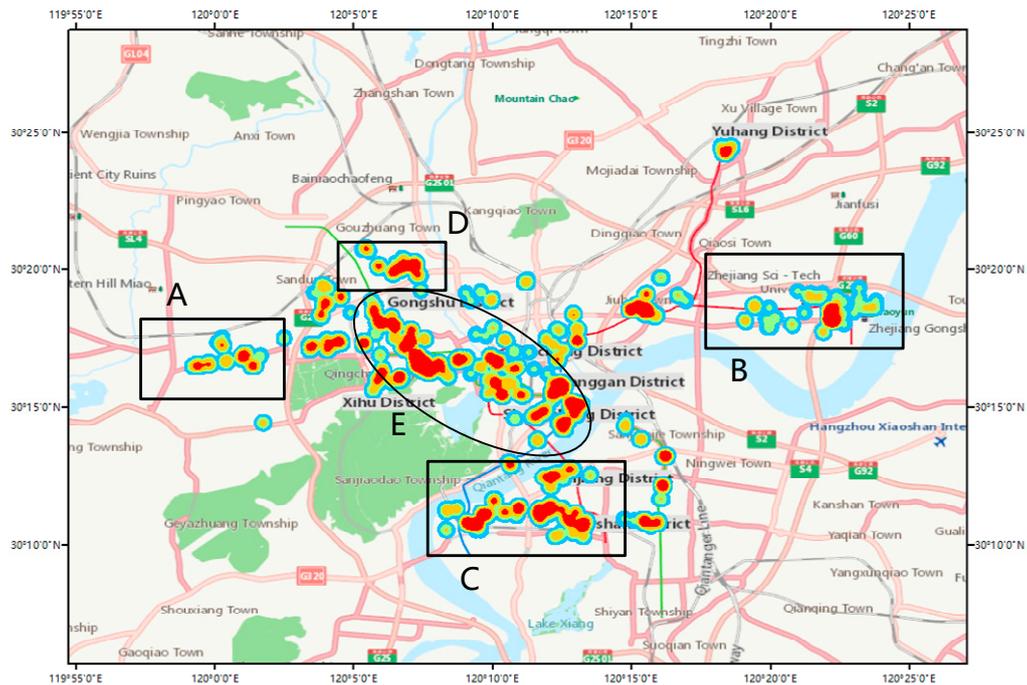
Figure 13. The distribution of high-tech zones and the downtown area in the heat map.

4.2. Living Areas

Like the work areas analysis, we made a heat map to localize the hotspot regions in living areas by using the KDE method. The parameters of the method were set as follows: (1) the cell size was set to 0.0001 degrees; (2) the search radius was set to 0.005 degrees; and (3) for the population field, the clusters d, e, and f were set to 7, 3, and 1, respectively. Figure 14 shows the heat map of the living areas.
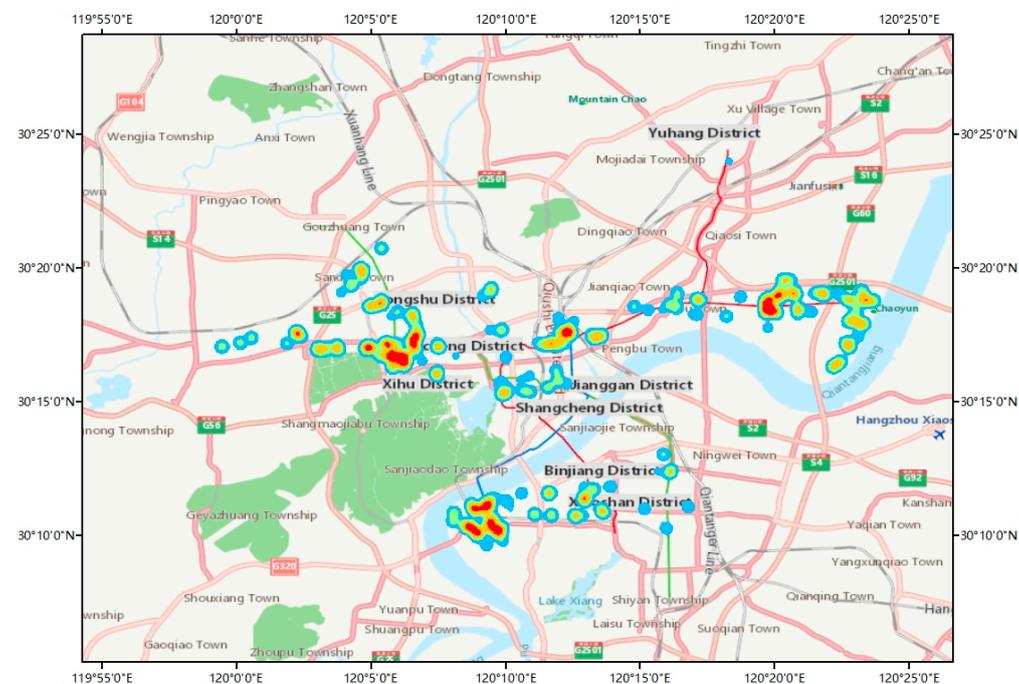
Figure 14. Heat map of the living areas made by the kernel density estimation (KDE) method.

Transportation is one of the important factors that affect the choice of residential location, which is related to commuting time. The subway, as a form of public transportation, has become a major way

for commuting, especially in large cities. Hangzhou has opened subway lines Nos. 1, 2, and 4. We used the subway stations as centers to create buffer zones with a buffer range of 1.5 km and overlay the buffer zones with the heat map of the living areas; the result is shown in Figure 15. From the results of the buffer analysis, it is clear that the living areas are mainly distributed along the subway lines. Statistically, the grids that represent living areas within the buffer zones account for 81.24% of the total.
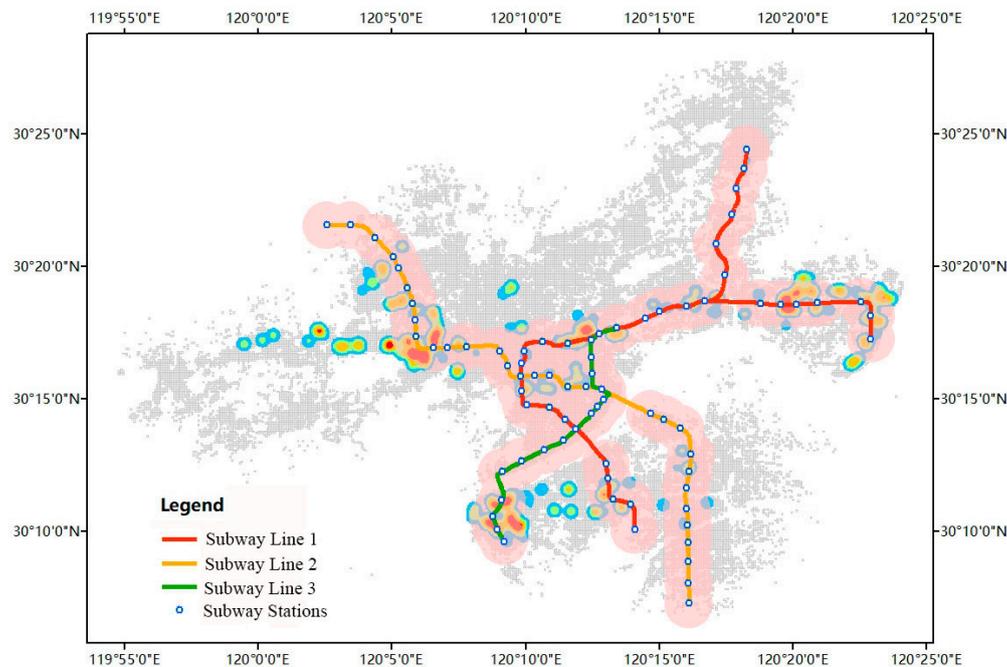

Figure 15. The buffer zones of the Hangzhou subway overlaying the heat map.

It is easy to find that there are four hotspot regions of living areas in the heat map, as shown in Figure 16e. In contrast with online maps such as Google Maps, a large number of residential communities and apartments appear in the hotspot regions, as shown in Figure 16a–d. In addition, we found that urban villages, which are also called *chengzhongcun* in Chinese, also appeared in all hotspot regions. Urban villages are formed when expanding modern urban districts encroach on rural settlements and became transitional neighborhoods under rapid urbanization [1,42].

Hangzhou has a huge job market for Internet-related industries, which has attracted a large number of young talent in recent years and produces a growing demand for housing, especially for low-rent housing. We added the center points of the grids from clusters d and e to the hotspot region A, which were identified as living areas and had more significant functional characteristics. As shown in Figure 17, although residential communities and apartments were densely distributed in area A, the locations of the cluster grids were mostly concentrated in the urban village areas. It shows that the urban villages have become one of the residential choices for many young people, which provides low-rent housing and low-cost living spaces. However, the population of urban villages has strong mobility and instability [43], and there are many private rental housing accommodations in urban villages that are not registered in the local house lease management office, which will make it difficult to monitor the floating population in urban villages. Thus, the method of this paper provides a new way to discover urban villages in cities, which will help to find the hotspot areas of floating populations and promote urban youth population management.

(a) living hotspot region A

(b) living hotspot region B

(c) living hotspot region C

(d) living hotspot region D

(e) the location distribution of living hotspot regions
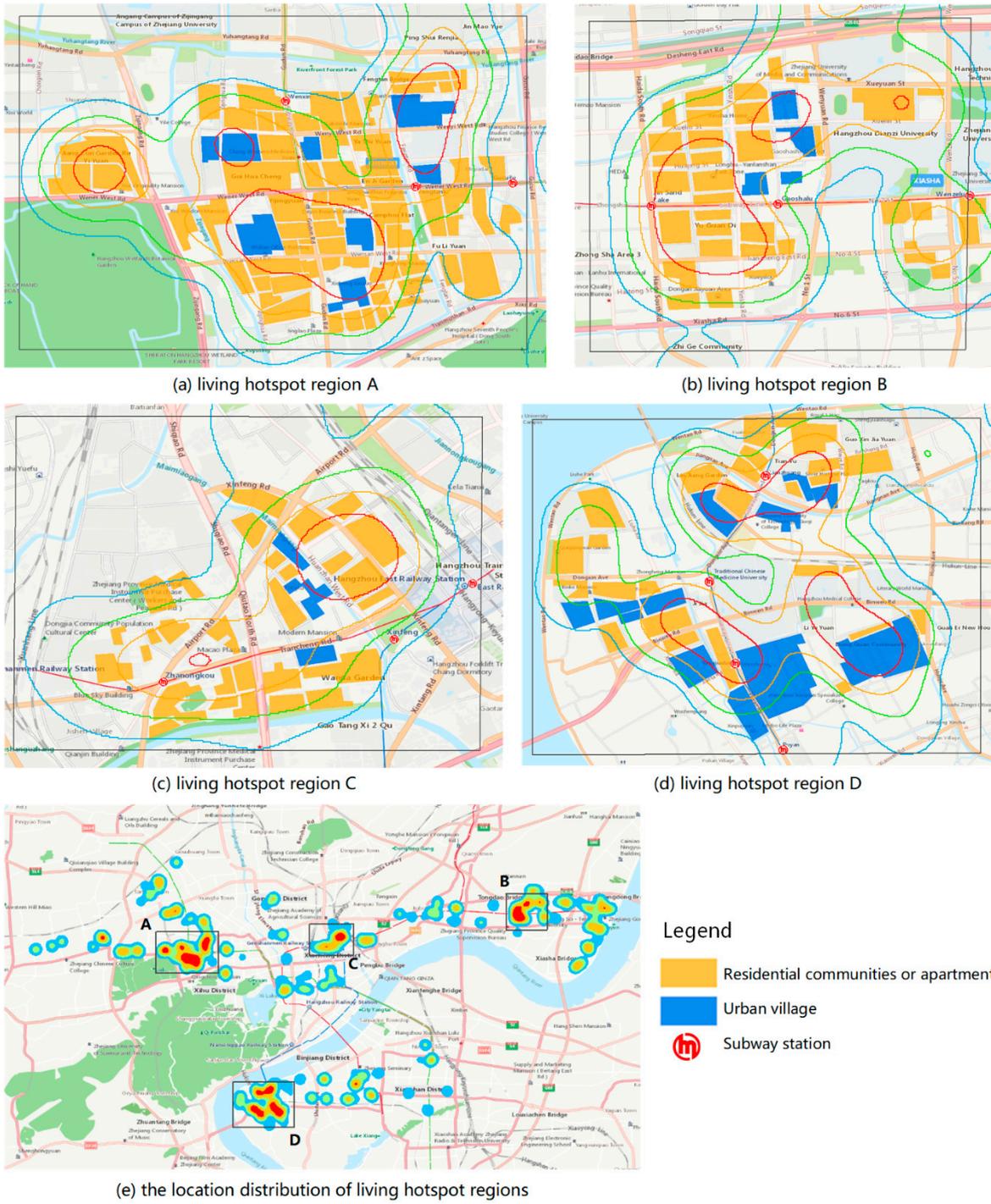
Figure 16. Residential communities, apartments, and urban villages in the hotspot regions of living areas.
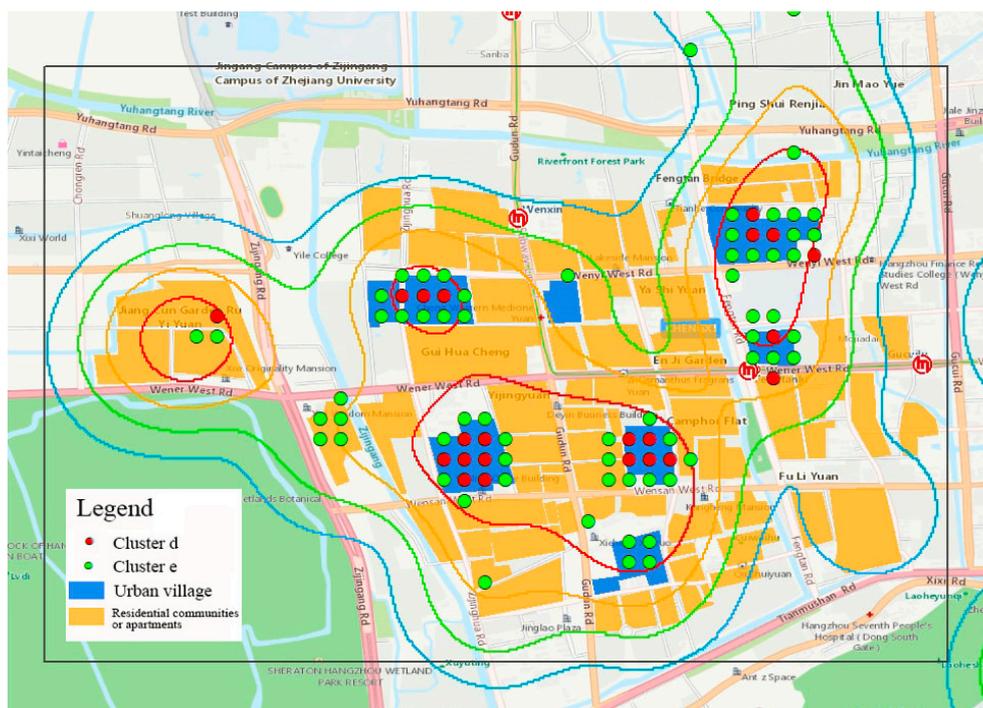
Figure 17. The distribution of clusters d and e in living hotspot region A.

5. Conclusions

In this paper, food delivery data, as a new data source, were introduced into urban computing, and we proposed a time-series-based clustering method to discover the geographical distribution of urban youth. The synthetic time series of food delivery were constructed by combining the weekdays–weekends 48-h timeline with delivery hot values that were calculated by using the modified Getis-Ord statistic method to utilize the characteristics of the delivery data. Then, an unsupervised k-means clustering method was adopted to classify the time series into six classes, and based on the behavioral characteristics of ordering food delivery that differ between job and housing areas, the six classes were identified as two functional types (i.e., work areas and living areas). The identification result was verified by comparing it with the office urban planning map, and the accuracy of the identified work areas was 66.4–94.6%, while the accuracy of the identified living areas was 59–87.2%.

Focusing on the urban youth population, especially white-collar workers, the work and living areas were further analyzed by comparing them with the industrial structure and residential layout in Hangzhou. Heat maps were made with the kernel density estimation method to localize hotspot regions in the work and living areas, with the main findings as follows:

1. The spatial distribution of the delivery customers' work places was similar to the spatial distribution of the Internet and e-commerce companies of Hangzhou. This shows that a large number of young people are working in the Internet-related industries areas planned by the government, and there exists a symbiotic relationship between Hangzhou's Internet-related industrial agglomeration and young labor gathering. The establishment of high-tech zones has effectively attracted young people, and the young labor gathering also promotes the development of the Internet-related industry. There is a mutual driving relationship between Hangzhou's Internet-related industry agglomeration and young labor gathering.

2. Transportation and living costs are the two important factors that affect the choice of residential location for young people. The hotspot living areas of urban youth are mostly located within 1.5 km from the subway stations, and the subway is becoming one of the most important modes

for commuting. Urban villages have become one of the residential choices for many young people, which provide low-rent housing and low-cost living spaces.

In addition, the findings of this paper can not only improve the view of the current state of the city's industrial development, but can also promote floating population management by discovering and monitoring the key areas such as urban villages.

In future studies, machine learning methods can be introduced to improve the classification method for food delivery data, and thus, the identified functional types can be divided into more land use categories. By comparing multi-year food delivery data, we can discover changes in the distribution of work and living areas to track the urbanization process.

References

1. Tan, S.K.; Li, Y.A.; Song, Y.; Luo, X.; Zhou, M.; Zhang, L.; Kuang, B. Influence factors on settlement intention for floating population in urban area: A China study. *Qual. Quant.* **2017**, *51*, 147–176. [CrossRef]
2. Tao, L.; Hui, E.C.; Wong, F.K.; Chen, T. Housing choices of migrant workers in China: Beyond the Hukou perspective. *Habitat Int.* **2015**, *49*, 474–483. [CrossRef]
3. Li, B. Floating population or urban citizens? Status, social provision and circumstances of rural–urban migrants in China. *Soc. Policy Adm.* **2006**, *40*, 174–195. [CrossRef]
4. Frosch, K.H. Workforce age and innovation: A literature survey. *Int. J. Manag. Rev.* **2011**, *13*, 414–430. [CrossRef]
5. Chatman, D.G. How density and mixed uses at the workplace affect personal commercial travel and commute mode choice. *Transp. Res. Rec.* **2003**, *1831*, 193–201. [CrossRef]
6. Peng, Z.-R. The jobs-housing balance and urban commuting. *Urban Stud.* **1997**, *34*, 1215–1235. [CrossRef]
7. Wu, W. Migrant intra-urban residential mobility in urban China. *Hous. Stud.* **2006**, *21*, 745–765. [CrossRef]
8. Liu, Y.; Xiao, Y.; Gao, S.; Kang, C.-G.; Wang, Y.-L. A Review of Human Mobility Research Based on Location Aware Devices. *Geogr. Geo Inf. Sci.* **2011**, *4*, 3.
9. Zheng, Y. Trajectory data mining: An overview. *ACM Trans. Intell. Syst. Technol.* **2015**, *6*, 29. [CrossRef]
10. Feng, Z.; Zhu, Y. A survey on trajectory data mining: Techniques and applications. *IEEE Access* **2016**, *4*, 2056–2067. [CrossRef]
11. Gao, S. Spatio-temporal analytics for exploring human mobility patterns and urban dynamics in the mobile age. *Spat. Cogn. Comput.* **2015**, *15*, 86–114. [CrossRef]
12. Jiang, S.; Ferreira, J.; Gonzalez, M.C. Activity-based human mobility patterns inferred from mobile phone data: A case study of Singapore. *IEEE Trans. Big Data* **2017**, *3*, 208–219. [CrossRef]
13. Pei, T.; Sobolevsky, S.; Ratti, C.; Shaw, S.-L.; Li, T.; Zhou, C. A new insight into land use classification based on aggregated mobile phone data. *Int. J. Geogr. Inf. Sci.* **2014**, *28*, 1988–2007. [CrossRef]
14. Toole, J.L.; Ulm, M.; González, M.C.; Bauer, D. Inferring land use from mobile phone activity. In Proceedings of the ACM SIGKDD International Workshop on Urban Computing, Beijing, China, 12–16 August 2012; pp. 1–8.
15. Yang, X.; Fang, Z.; Yin, L.; Li, J.; Zhou, Y.; Lu, S. Understanding the Spatial Structure of Urban Commuting Using Mobile Phone Location Data: A Case Study of Shenzhen, China. *Sustainability* **2018**, *10*, 1435. [CrossRef]

16. Zhao, Z.; Shaw, S.-L.; Xu, Y.; Lu, F.; Chen, J.; Yin, L. Understanding the bias of call detail records in human mobility research. *Int. J. Geogr. Inf. Sci.* **2016**, *30*, 1738–1762. [CrossRef]

17. Mao, F.; Ji, M.; Liu, T. Mining spatiotemporal patterns of urban dwellers from taxi trajectory data. *Front. Earth Sci.* **2016**, *10*, 205–221. [CrossRef]

18. Tang, J.; Liu, F.; Wang, Y.; Wang, H. Uncovering urban human mobility from large scale taxi GPS data. *Phys. A Stat. Mech. Its Appl.* **2015**, *438*, 140–153. [CrossRef]

19. Zheng, Q.; Zhao, X.; Jin, M. Research on Urban Public Green Space Planning Based on Taxi Data: A Case Study on Three Districts of Shenzhen, China. *Sustainability* **2019**, *11*, 1132. [CrossRef]

20. Pan, G.; Qi, G.; Wu, Z.; Zhang, D.; Li, S. Land-use classification using taxi GPS traces. *IEEE Trans. Intell. Transp. Syst.* **2013**, *14*, 113–123. [CrossRef]

21. Yuan, N.J.; Zheng, Y.; Xie, X.; Wang, Y.Z.; Zheng, K.; Xiong, H. Discovering Urban Functional Zones Using Latent Activity Trajectories. *IEEE Trans. Knowl. Data Eng.* **2015**, *27*, 712–725. [CrossRef]

22. Biehl, A.; Ermagun, A.; Stathopoulos, A. Community mobility MAUP-ing: A socio-spatial investigation of bikeshare demand in Chicago. *J. Transp. Geogr.* **2018**, *66*, 80–90. [CrossRef]

23. Yang, T.; Li, Y.; Zhou, S. System Dynamics Modeling of Dockless Bike-Sharing Program Operations: A Case Study of Mobike in Beijing, China. *Sustainability* **2019**, *11*, 1601. [CrossRef]

24. Zhang, L.; Zhang, J.; Duan, Z.-Y.; Bryde, D. Sustainable bike-sharing systems: Characteristics and commonalities across cases in urban China. *J. Clean. Prod.* **2015**, *97*, 124–133. [CrossRef]

25. Zhang, X.; Li, W.; Zhang, F.; Liu, R.; Du, Z. Identifying Urban Functional Zones Using Public Bicycle Rental Records and Point-of-Interest Data. *ISPRS Int. J. Geo Inf.* **2018**, *7*, 459. [CrossRef]

26. Zhu, Y.; Chen, F.; Li, M.; Wang, Z. Inferring the economic attributes of urban rail transit passengers based on individual mobility using multisource data. *Sustainability* **2018**, *10*, 4178. [CrossRef]

27. Zhou, J.; Murphy, E.; Long, Y. Commuting efficiency in the Beijing metropolitan area: An exploration combining smartcard and travel survey data. *J. Transp. Geogr.* **2014**, *41*, 175–183. [CrossRef]

28. Mohamed, K.; Côme, E.; Oukhellou, L.; Verleysen, M. Clustering smart card data for urban mobility analysis. *IEEE Trans. Intell. Transp. Syst.* **2016**, *18*, 712–728.

29. Long, Y.; Thill, J.-C. Combining smart card data and household travel survey to analyze jobs–housing relationships in Beijing. *Comput. Environ. Urban Syst.* **2015**, *53*, 19–35. [CrossRef]

30. Mazimpaka, J.D.; Timpf, S. Trajectory data mining: A review of methods and applications. *J. Spat. Inf. Sci.* **2016**, *2016*, 61–99. [CrossRef]

31. Hu, Y.; Han, Y. Identification of Urban Functional Areas Based on POI Data: A Case Study of the Guangzhou Economic and Technological Development Zone. *Sustainability* **2019**, *11*, 1385. [CrossRef]

32. Yang, X.; Zhao, Z.; Lu, S. Exploring spatial-temporal patterns of urban human mobility hotspots. *Sustainability* **2016**, *8*, 674. [CrossRef]

33. 2017 China Internet Local Life Services Blue Book. Available online: http://mp.163.com/v2/article/detail/D8KG29RO0518SLLV.html (accessed on 15 December 2018).

34. 2017–2018 China Online Catering Food Market Research Report. Available online: https://www.iimedia.cn/c400/60449.html (accessed on 15 December 2018).

35. 2017 China Delivery Development Research Report. Available online: https://www.sohu.com/a/216309928_800248 (accessed on 15 December 2018).

36. Statistical Communique of Hangzhou on the 2018 National Economic and Social Development. Available online: http://www.hangzhou.gov.cn/col/col805865/index.html (accessed on 15 April 2019).

37. Ord, J.K.; Getis, A. Local Spatial Autocorrelation Statistics: Distributional Issues and an Application. *Geogr. Anal.* **1995**, *27*, 286–306. [CrossRef]

38. Makrai, G. Efficient method for large-scale spatio-temporal hotspot analysis (GIS Cup). In Proceedings of the 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, San Francisco Bay Area, CA, USA, 31 October–3 November 2016; pp. 1–18.

39. Ray, S.; Turi, R.H. Determination of number of clusters in k-means clustering and application in colour image segmentation. In Proceedings of the 4th International Conference on Advances in Pattern Recognition and Digital Techniques, Calcutta, India, 27–29 December 1999; pp. 137–143.

40. Zhang, Y.; Zhu, Z. Research on mutual driving relationship between industrial agglomeration and talent gathering in Jiangsu. In Proceedings of the Second International Conference On Economic and Business Management (FEBM 2017), Shanghai, China, 21–23 October 2017.

41. Florida, R. The economic geography of talent. *Ann. Assoc. Am. Geogr.* **2002**, *92*, 743–755. [CrossRef]

42. Song, Y.; Zenou, Y. Urban villages and housing values in China. *Reg. Sci. Urban Econ.* **2012**, *42*, 495–505. [CrossRef]

43. Wehrhahn, R.; Bercht, A.L.; Krause, C.L.; Azzam, R.; Kluge, F.; Strohschön, R.; Wiethoff, K.; Baier, K. Urban restructuring and social and water-related vulnerability in megacities–the example of the urban village of Xincun, Guangzhou (China). *Die Erde* **2008**, *139*, 227–249.