Article

Portraying Citizens' Occupations and Assessing Urban Occupation Mixture with Mobile Phone Data: A Novel Spatiotemporal Analytical Framework

Xiaoming Zhang [1,2], Feng Gao [1,2,3,*], Shunyi Liao [1,2], Fan Zhou [1,2], Guanfang Cai [1,2] and Shaoying Li [3]

[1]   Guangzhou Urban Planning & Design Survey Research Institute, Guangzhou 510060, China; zhangxiaoming@gzpi.com.cn (X.Z.); liaoshunyi@gzpi.com.cn (S.L.); zhoufan@snnu.edu.cn (F.Z.); gc603@rutgers.edu (G.C.)
[2]   Guangdong Enterprise Key Laboratory for Urban Sensing, Monitoring and Early Warning, Guangzhou 510060, China
[3]   School of Geography and Remote Sensing, Guangzhou University, Guangzhou 510006, China; lsy@gzhu.edu.cn
[*]   Correspondence: 2111801048@e.gzhu.edu.cn

Abstract: Mobile phone data is a typical type of big data with great potential to explore human mobility and individual portrait identification. Previous studies in population classifications with mobile phone data only focused on spatiotemporal mobility patterns and their clusters. In this study, a novel spatiotemporal analytical framework with an integration of spatial mobility patterns and non-spatial behavior, through smart phone APP (applications) usage preference, was proposed to portray citizens' occupations in Guangzhou center through mobile phone data. An occupation mixture index (OMI) was proposed to assess the spatial patterns of occupation diversity. The results showed that (1) six types of typical urban occupations were identified: financial practitioners, wholesalers and sole traders, IT (information technology) practitioners, express staff, teachers, and medical staff. (2) Tianhe and Yuexiu district accounted for most employed population. Wholesalers and sole traders were found to be highly dependent on location with the most obvious industrial cluster. (3) Two centers of high OMI were identified: Zhujiang New Town CBD and Tianhe Smart City (High-Tech Development Zone). It was noted that CBD has a more profound effect on local as well as nearby OMI, while the scope of influence Tianhe Smart City has on OMI is limited and isolated. This study firstly integrated both spatial mobility and non-spatial behavior into individual portrait identification with mobile phone data, which provides new perspectives and methods for the management and development of smart city in the era of big data.

Keywords: mobile phone data; employed population; citizens' occupations; spatiotemporal mobility; APPs

1. Introduction

With the development of communication technology and big data, the trajectory of big data with both temporal and spatial information has become the focus of urban and spatial research, especially such topics as mobile phone data [1–7], social platform user density data [8,9], ridesourcing trajectory data [10,11], Baidu mobility data [12], and bike sharing riding record data [13–15] as well as rail transit ridership data [16,17]. Among them, mobile phone data is the one that can capture not only users' spatiotemporal mobility patterns but also users' non-spatial attribute information such as age, gender, and so on. Hence, mobile phone data is of great potential to mine citizens' mobility patterns and to classify them.

Related works on cellphone big data can be divided into two categories: non-spatial studies and spatiotemporal studies. Firstly, non-spatial studies mainly focus on demographics from cellphone data such as gender, age, and so on. For instance, Tu et al. (2021)

found that the change of long-term usage of APPs was associated with transition in demographic attributes including civil status, family size, and economic status [18]. Peltonen et al. (2018) found significant differences in APP usage across 44 countries with APP user data and user survey data, and further revealed the geographic boundaries [19]. Kumar et al. (2020) explored influencing factors on APP usage among the youth of eastern Bhutan considering the mobile operating system, APP type, APP category, APP features, and advertisements [20]. Malmi et al. (2016) improved the effectiveness of targeted advertisements by studying the predictability of user demographics (age, race, and income) based on the list of a user's apps, and found that the most predictable attribute was gender while the hardest to predict was income [21]. Ernsting et al. (2017) tried to change health-related behaviors and manage chronic conditions with health APPs and population-based surveys, and found that the associations of APP use and characteristics with actual health behaviors was significant [22].

Second, previous spatiotemporal studies were mainly conducted on job–housing balance, human mobility patterns, the classification of urban functional area, and the identification of population types and so on. For example, Zhou et al. (2020) explored the modifiable areal unit problems in jobs–housing balance and employment self-containment of Shenzhen with mobile positioning data [23]. Zhou et al. (2017) analyzed the spatial variation of self-containment of employment and jobs–housing balance with location and housing prices based on cellphone data in Shenzhen [24]. Qian et al. (2021) used mobile phone data to identify visitors and analyze the spatial correlations between tourism places in Shanghai [25]. Guo et al. (2020) identified the population's temporal exposure to PM2.5 pollution with mobile phone data in Shenzhen [26]. Lee et al. (2018) explored the urban activity and mobility patterns and compared the spatial dispersion of residences and other activity locations with daily and hourly mobile phone records of 10 cities in South Korea [27]. Louail et al. (2015) developed an origin–destination matrix method to categorize residents' mobility flow and classify the city based on the flow patterns with a long time series of mobile phone data in Spain [28]. Willberg et al. (2021) explored the mobility between urban and rural and the influence of multi-local living spaces on population dynamics in Finland during the COVID-19 crisis in 2020 with mobile phone datasets, and found that a population decline in urban centers and an increase in rural areas were observed, which is strongly correlated to secondary housing [3]. Gong et al. (2020) identified the spatial distribution of three types of residents' activity at different scales and uncovered the relationship between the built environment and activity space with mobile phone data [29]. Gao et al. (2015) identified the human mobility patterns and intra-urban communication dynamics in China with mobile phone data [30]. Yin et al. (2021) proposed a method to mine human activity chains from large-scale mobile phone data by integrating both the spatial and temporal features of daily activities with varying weights [31]. Cao et al. (2021) proposed an approach for exploring urban mobility networks based on mobile phone tracking data in Shenzhen [32].

In terms of population classification with mobile phone data, several studies have attempted to categorize population types based on various methods and aspects. Jiang et al. (2012) categorized population into eight and seven representative groups according to the spatiotemporal characteristics of their activities during weekdays and weekends including students, regular workers, early bird workers, afternoon workers, the stay-at-home, the morning adventurers, the afternoon adventurers, and the overnight adventurers [33]. Ding et al. (2019) categorized population into permanent and floating populations based on the variation of activity spatiotemporal characteristics [34]. Li et al. (2021) first explored the individual mobility pattern by clustering the GPS (Global Positioning System) locations with DBSCAN (Density-Based Spatial Clustering of Applications with Noise) algorithm and then spatialized the clusters with Google Maps [1]. It finally classified the population into two-point-one-line patterns and the situation of a single cluster, which reflected white-collar workers working in enterprises or institutions, and college students and home freelancers, respectively [1]. In sum, cases above attempted to cluster population according

to their spatiotemporal characteristics and mobility. It is noted that only spatial behavior was focused on in previous studies, while non-spatial behavior and attributes were ignored due to available data such as users' age, gender, and APP usage preferences. However, even though classifying population based on spatial behavior is effective and accurate, it is a fact that there are originally non-spatial differences between people in ages, genders, and especially occupations.

Therefore, several gaps need to be filled in population classification with mobile phone data. First, the classification criteria in previous studies were based on citizens' daily mobility, that is, inferring citizens' types by the fine scale information about their spatial and temporal locations and trajectories, rather than classifying population based on their essential differences, such as ages, gender, or occupations. Second, previous studies only focused on the spatiotemporal mobility patterns with a long and fine series of mobile phone data, while the non-spatial behaviors or attributes are also important. Third, the integration of non-spatial demographics attributes and spatiotemporal dynamic information was rare in existing studies.

The aim of this study is to fill the research gaps mentioned above by proposing a novel spatiotemporal analytical framework with the integration of both spatial mobility patterns and non-spatial behavior, as identified through smart phone APP usage preference, to portray citizens' occupations. An occupation mixture index was proposed in this study to further assess the spatial distribution of urban occupation diversity in Guangzhou center with mobile phone data. This research firstly combines both spatial mobility patterns and non-spatial behavior (APP usage preference) to classify population into different occupations, which helps fill the gap in individual portrait identification with mobile phone data, and provides a novel framework to make full use of mobile phone data. It contributes to the management and development of smart cities in the era of big data.

The study consists of four parts. Section 2 presents the data source, study area, and methodology. Section 3 presents the results and discussions of the study including identification of the employed population, portraying citizens' occupations, and assessing urban occupation mixture. Section 4 concludes the study results and presents the prospect of future research. The framework of this study is shown as Figure 1.
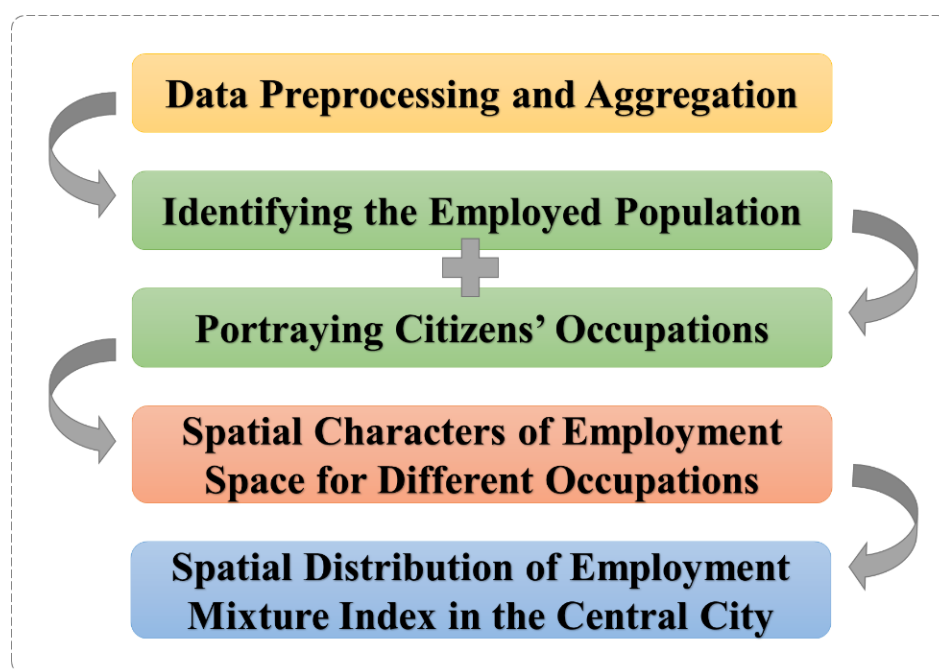
Figure 1. The framework of the study.

2. Data and Methods

2.1. Study Area

The study area was located in the center of Guangzhou, one of the largest cities in southern China, consisting of four districts: Tianhe, Yuexiu, Haizhu, and Liwan district (Figure 2). Although the center area of 325.4 km² accounts for only 4.49% of the city, the population of the city center is 5.7344 million, accounting for 37.5% of the city's population in 2020. Guangzhou, as one of the largest cities in China, is also located in the forefront of China's reform and opening up. With rapid economic development and more employment opportunities, it has attracted a large number of foreign populations. For a long time, migrant workers have become local citizens, which has a great impact on the social and economic development of Guangzhou.
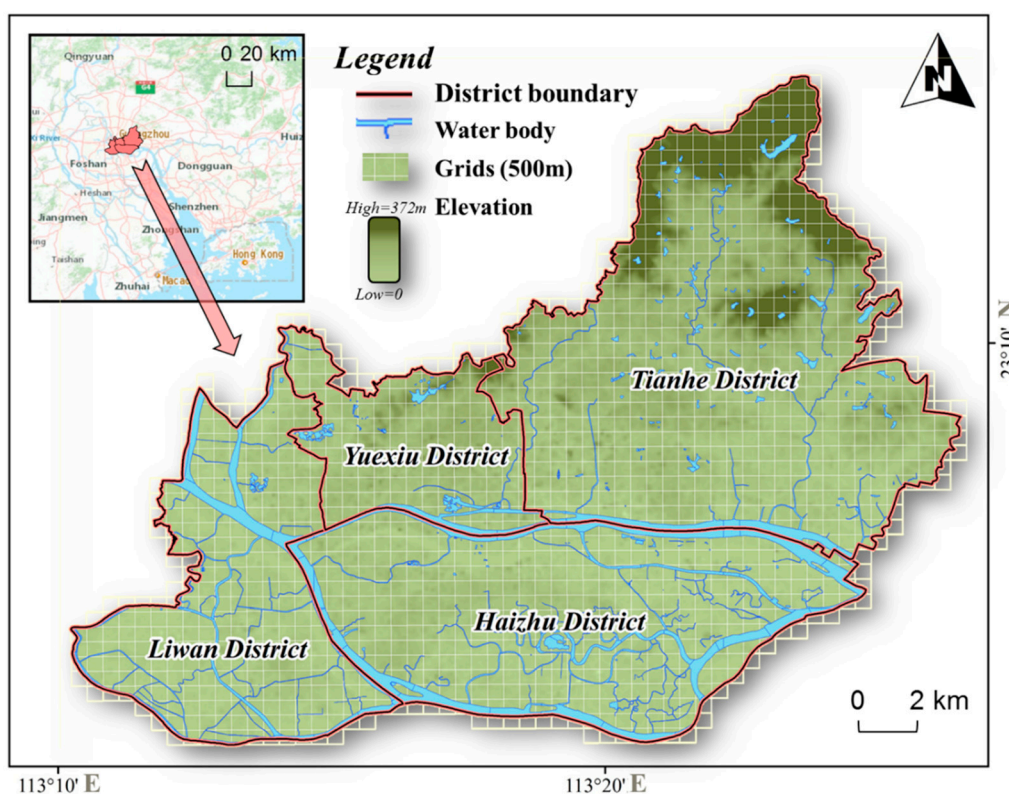

Figure 2. Study area.

2.2. Data

Mobile phone data used in this study were acquired from the local major communications company in Guangzhou center in November 2020, during which the COVID-19 blockade measures were lifted, and the country resumed normal production and life. The data are anonymous, and each mobile phone data contains user ID, time stamp, base station location number, event type (such as receiving and calling, receiving and sending SMS, location update, APP usage time statistics), etc. The data used in this paper recorded 5.8–5.9 million different mobile phone identification numbers in the center of Guangzhou, with a total of 200–300 million pieces of data per day. There are about 80,000 base stations in the city, and the distance between them is about 50–100 m. It recorded the information about users' spatiotemporal behavior and APP visiting behavior, including users' base station location of every 30 min and daily statistics of smart phone APP use duration in November. By cooperating with the communication company, the cellphone data applied in this study recorded the frequency of APP usage, a list that records each APP's daily use duration for each user. Therefore, each APP's daily use duration was recorded and sorted, and each category of APP daily use duration of each user was also recorded. To avoid

deviation, social and recreational APPs in China (such as WeChat, QQ, Weibo, and Douyin, etc.) were removed from APP use statistics since this type of APP is unhelpful to portray users' occupations.

Mobile phone data has the characteristics of dynamic, continuous, almost full coverage of urban and rural space and high holding rate, which can better reflect people's overall spatiotemporal behavior. Mobile phone data can be categorized into two types: location data of base stations, and the CSV files recording users' locations and APP visiting duration from the SQL database.

POI (point of interest) data was obtained from Baidu Map API (http://lbsyun.baidu.com/ (accessed on 22 November 2020)) to classify employment space based on different types of occupational facilities' POI in the process of portraying citizens' occupations.

2.3. Pre-Processing of Mobile Phone Data

The pre-processing of mobile phone data in this study can be divided into two parts: non-spatial pre-processing and spatial pre-processing.

Non-spatial pre-processing of mobile phone data includes creating the database and data cleaning. First, one month of mobile phone data were obtained as text files through the Hive-SQL platform of the communications company with the access rights of non-private data. That is, private data such as phone number and other personal information of users were encrypted to protect privacy. Second, a file geodatabase was created in ESRI ArcMap 10.7 software. Third, text files were converted into shape files according to the coordinates of base stations and imported into a geodatabase. Then, data cleaning was conducted including removing invalid fields and GPS drift points [1]. After comparing the time intervals of raw data acquisition, it was noted that the raw data acquisition frequency ranged from 7 s to 1 day. The data obtained within 7 s intervals accounted for 1.2% while 92.4% of data were obtained within 30 min. Therefore, the data with an interval within 30 min were filtered as valid data to form the trajectory chains of users.

Spatial pre-processing of mobile phone data mainly includes the spatial aggregation of station-based mobile phone data. Since the spatial information of mobile phone data is based on stations, data aggregation is necessary for further analysis. The density method and grid aggregation were two major ways to process the raw mobile phone data. The density method was capable in visualization but fell short of identifying the population of specific blocks. Therefore, to better portray citizens' occupations and assess the mixture of urban occupations, a grid was used to aggregate statistics of mobile phone data. Although the study area is located in downtown Guangzhou with high density distribution of base stations, suitable grid size should not be too large or too small. Oversized grids might ignore some details of spatial heterogeneity, while too finely sized grids might lead to low coverage of base stations. Therefore, we compared ten candidate grid sizes' coverage of base stations (Figure 3) and found out that a 500 m grid might be suitable with station coverage of 85%, which was also the turning point among ten sizes. The coverages were relatively low and growth rates were high when sizes were smaller than 500 m and growth rates started to slow down and be stable with grids larger than 500 m.

The main purpose of this study is to identify the typical occupation types of the employed population. The employed population (or workplaces of cellphone users) were identified from the cellphone big data based on the analytical framework in previous studies [35–37]. That is, if the cellphone users stayed at one place of activity for the duration not less than 6 h during typical working time (i.e., from 9 a.m. to 6 p.m. for most workers) at least three weekdays a week in a month, then those cellphone users were identified as the employed population, and that activity place was identified as the workplace (Figure 4). The time slot selection of the processing was determined by residents' journey-to-work trip patterns from the travel survey data in Guangzhou (Guangzhou Urban Planning and Design Survey Research Institute, 2020). In this way, the employed population were identified based on the 8-h work rule in China, though some were excluded such as night shift workers, who were in the minority. Moreover, based on previous studies, those with

residences (from 0 a.m. to 7 a.m.) and workplaces (from 9 a.m. to 6 p.m.) with the same base station or within 100 m were also omitted, since that they might be housewives or retirees [35–37]. In this manner, around 5.83 million employed population were identified.
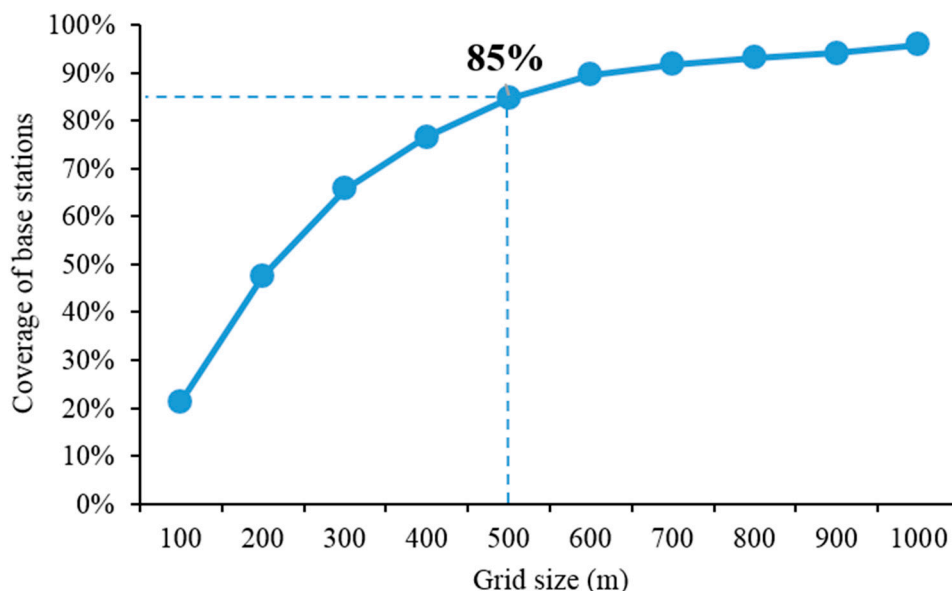

Figure 3. Coverage of base stations at different grid sizes.

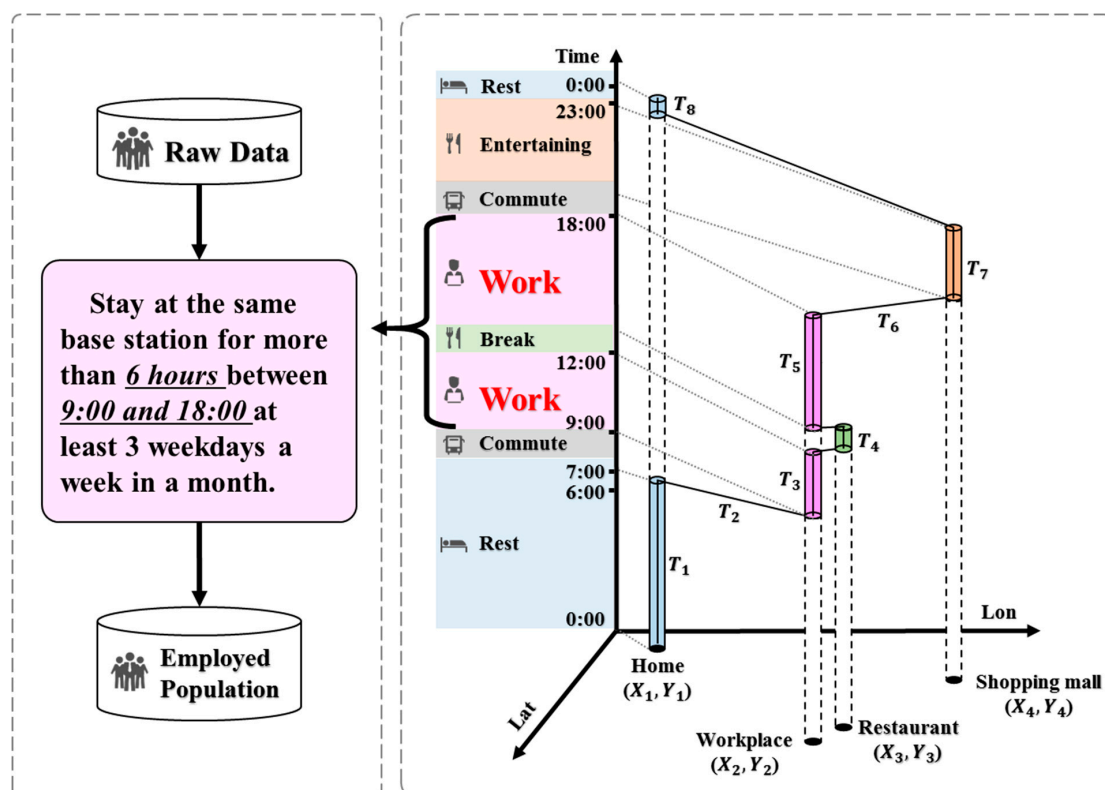*2.4. Processing of Extracting Employed Population*

Figure 4. Workflow of extracting the employed population.

2.5. Processing of Portraying Citizens' Occupations

　　The main purpose and novelty of the study is to propose the framework to portray citizens' occupations based on users' spatiotemporal dynamics behavior and APP visiting

behavior (Figure 5). It should be noted that urban citizens are from all walks of life, and hence portraying citizens' occupations with mobile phone data is difficult to cover all kinds of jobs. Based on the employed population extracted, this study combined daily mobility patterns and APP visiting duration to try to portray some typical occupation types with mobile phone data. The workflow of portraying citizens' occupations is as follows (Figure 5), including two parts: APP visiting frequency analysis and spatiotemporal mobility patterns analysis.
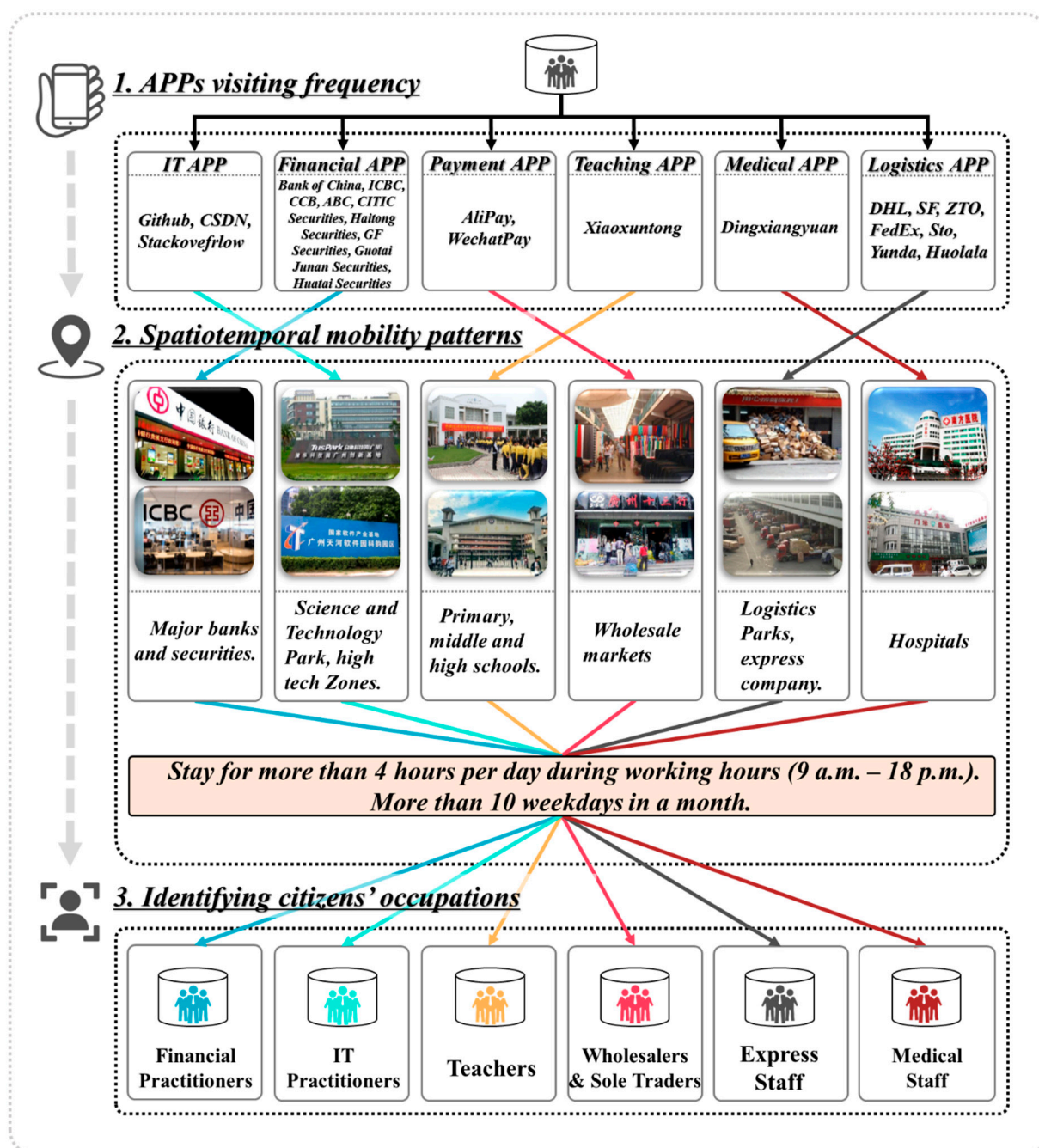

Figure 5. Workflow of portraying citizens' occupations.

Considering the results of the preliminary investigation and data experiment, six types of occupations were set as the goals of this study: financial practitioners, IT practitioners, teachers, wholesalers and sole traders, express staff, and medical staff. Moreover, both APPs and employment space were classified into the same six types. Employment space

was identified with POI data including banks and securities, science parks and high tech zones, schools, wholesale markets, express companies and logistics parks, and hospitals. Each type of occupation was identified according to the rules as follows.

- Financial practitioners:

In terms of APP visiting frequency, in addition to the common instant messaging and entertainment APPs, financial APPs were the most frequently used type of APPs in a month, such as China's major banks and securities APPs (Bank of China, Industrial and Commercial Bank of China, Agricultural Bank of China, China Construction Bank, CITIC Securities, GF Securities, Haitong Securities, and so on).

In terms of spatiotemporal mobility patterns, users were served by base stations around banks and securities companies for more than 4 h a day, 10 days a month.

- IT practitioners:

In terms of APP visiting frequency, IT APPs were the most frequently used APP types in a month, such as GitHub, CSDN, and Stackoverflow, which are popular APPs among programmers or IT practitioners in China.

In terms of spatiotemporal mobility patterns, users were served by base stations around science and technology park and high-tech zones for more than 4 h a day, 10 days a month.

- Teachers:

In terms of APP visiting frequency, Xiaoxuntong was the most frequently used APP in a month. It is the most popular APP used by teaching staff for managing students, issuing notices and contacting parents among primary, middle, and high schools in Guangzhou.

In terms of spatiotemporal mobility patterns, users were served by base stations around primary, middle, and high schools for more than 4 h a day, 10 days a month.

- Wholesalers and sole traders:

In terms of APP visiting frequency, payment APPs were the most frequently used APP types in a month, such as AliPay and WeChat Pay, which are popular instant payment APPs in China. Because of the wide popularity of this type of APPs in China, we further compared the use frequency: the average frequency of payment APPs of all samples was 7 times a day. Therefore, here we only selected those users with payment APPs use frequency higher than twice the overall average frequency.

In terms of spatiotemporal mobility patterns, users were served by base stations around wholesale markets for more than 4 h a day, 10 days a month.

- Express staff:

In terms of APP visiting frequency, express APPs were the most frequently used APP types in a month, such as SF express, FedEx express, DHL express, STO express, and so on, which were official APPs of major express companies in China.

In terms of spatiotemporal mobility patterns, users were served by base stations around logistics parks and express companies for more than 4 h a day, 10 days a month.

- Medical staff:

In terms of APP visiting frequency, Dingxiangyuan was the most frequently used APP in a month. It is the largest APP and social platform for medical staff in China.

In terms of spatiotemporal mobility patterns, users were served by base stations around hospitals for more than 4 h a day, 10 days a month.

2.6. Assessing the Urban Occupation Mixture Index

The main framework of this study mentioned above has firstly extracted the employed population, and then portrayed citizens' occupation based on APP visiting behavior and spatiotemporal mobility patterns with mobile phone data. We further assessed the mixture

of occupations in the study area by proposing an urban occupation mixture index (OMI) based on Shannon entropy as follows [38].

$$OMI = -\sum_{i=1}^{N} p_i \log_n p_i \qquad (1)$$

where *OMI* represents the value of occupation mixture index; p_i represents the percentage of the ith type of occupation; n represents the number of occupation types.

2.7. Spatial Analysis Methods

Spatial analysis methods were also applied in this study to further explore the spatial characteristic of the distribution of the employed population and population of different occupations portrayed in this study. For example, kernel density estimation method was applied to compare the employments core of the study area, and Getis-Ord Gi* analysis was used to identify the hot spot and cold spot of each occupation portrayed in the study; it was also applied to identify hot spot areas of the occupation mixture index. Finally, Moran's I was used to explore the spatial agglomeration of different types of occupations.

3. Results and Discussion

3.1. Identification of Employed Population

The identification results of the employed population are shown in Figure 6. To test the accuracy of the employed population identified in this study, the study area of 79 census blocks were taken as the spatial unit, and the employed population of each block was counted. Linear correlation analysis was conducted between the employed population and the total population of the latest census data. The results showed that there was a significantly positive correlation between them with a correlation coefficient of 0.81, which was a very strong correlation. The employed population identified by this method can effectively reflect the real population distribution.
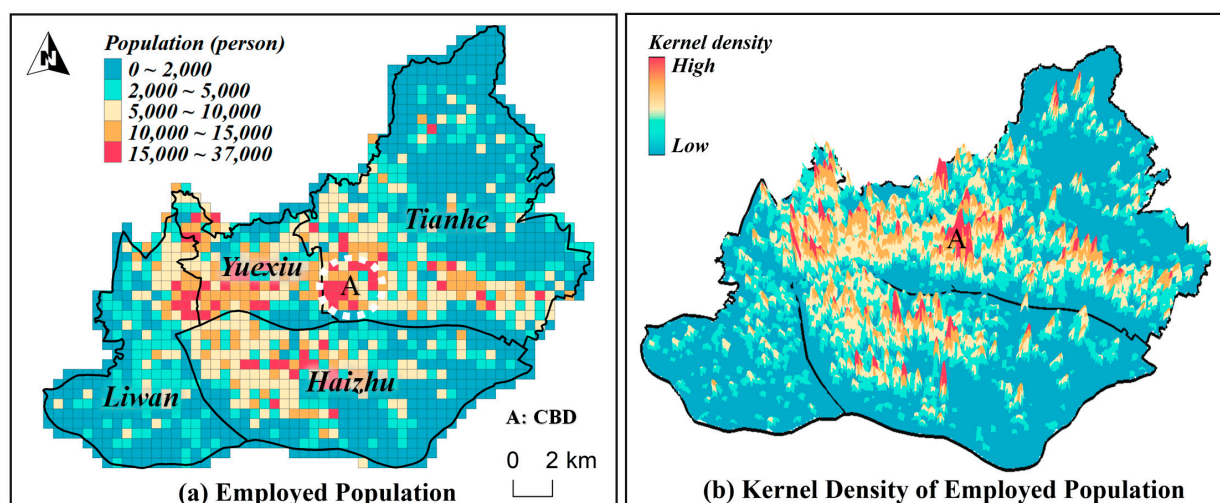

Figure 6. Spatial distribution (**a**) and kernel density (**b**) of employed population.

The grid-based distribution map (Figure 6a) shows that the employed population was mainly distributed in the center of study area: southwestern of Tianhe district, middle of Haizhu district, most of Yuexiu district, and northeast of Liwan district. To be specific, areas with high employed population were compared in a 3D kernel density map (Figure 6b), and it turns out that area A was the core of a highly concentrated employed population, where Zhujiang New Town (CBD) is located. The spatial distribution of population presents a multi-center pattern with CBD as the core, high value in the core area and low value in the edge area.

3.2. Portraying Citizens' Occupations

Based on the results of the employed population identified in the last chapter, citizens' occupations were portrayed according to smart phone APP visiting frequencies and their spatiotemporal mobility patterns (Figure 7); the hot spot analysis results are shown as Figure 8.
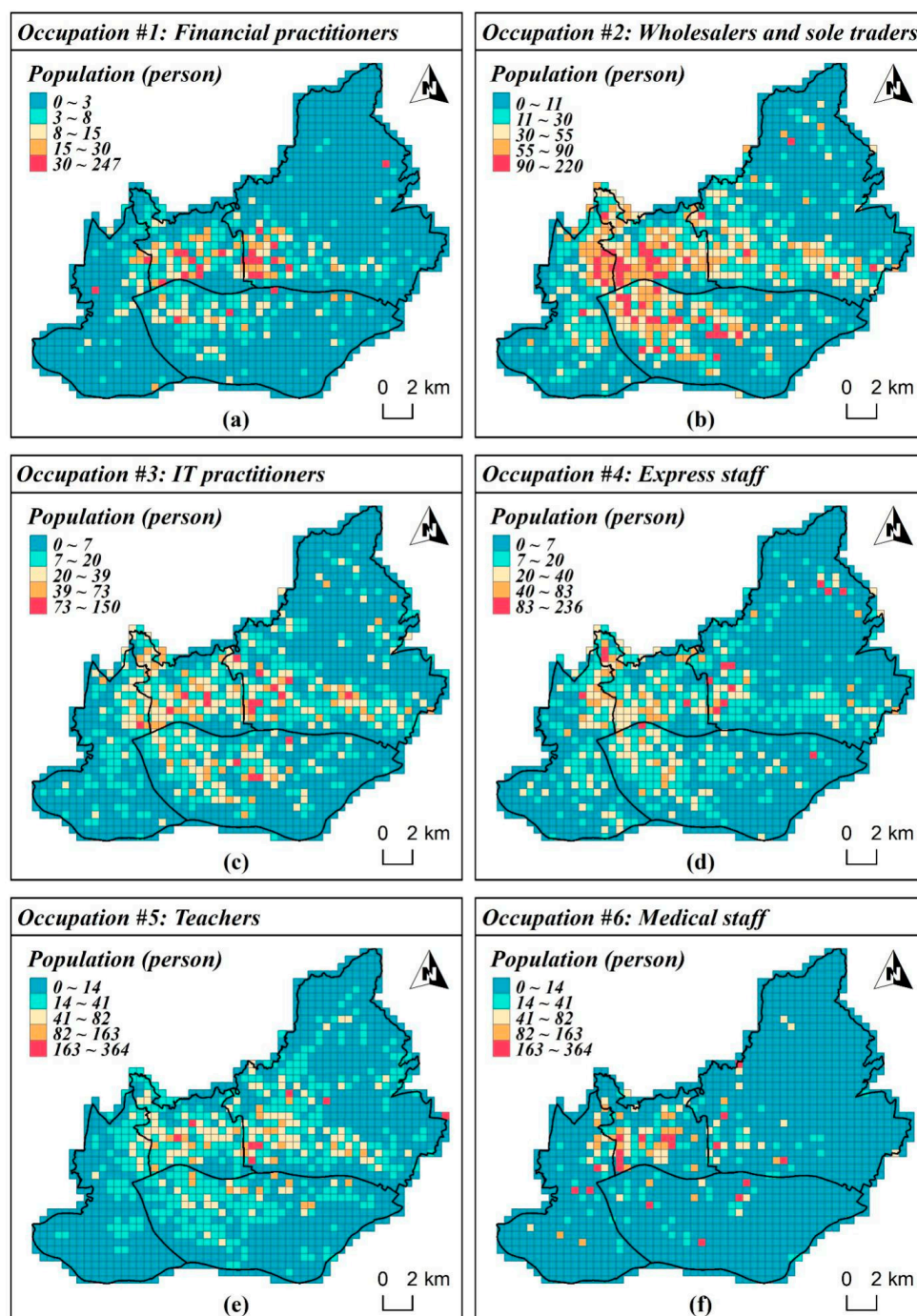
Figure 7. Employment space of financial practitioners (**a**), wholesalers and sole traders (**b**), IT practitioners (**c**), express staff (**d**), teachers (**e**), and medical staff (**f**).
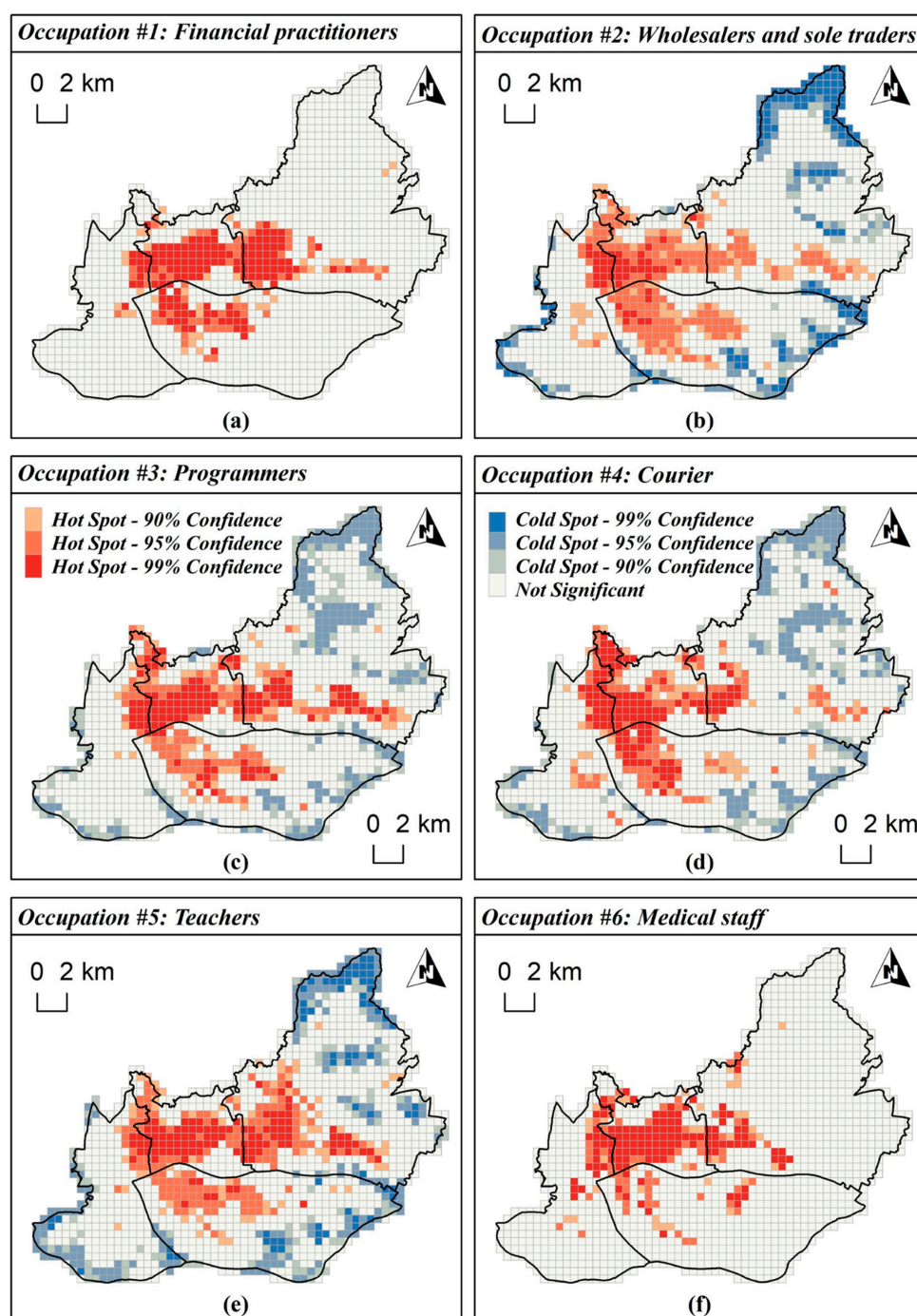
Figure 8. Hotspot analysis of employment space of financial practitioners (**a**), wholesalers and sole traders (**b**), IT practitioners (**c**), express staff (**d**), teachers (**e**), and medical staff (**f**).

- **Financial practitioners:**

Tianhe district has the largest number of financial practitioners (11 high value grids), followed by Yuexiu district (8 high value grids), Haizhu district (3 high value grids), and Liwan district (2 high value grids). A hot spot area with 99% confidence accounted for 10.2% of the total area. Most of them were located in Yuexiu district and southwest of Tianhe district. The distribution center of its hot spots was located in the middle of Yuexiu district and southwest of Tianhe district.

- **Wholesalers and sole traders:**

Haizhu district has the largest number of wholesalers and sole traders (21 high value grids), followed by Yuexiu district (17 high value grids), Liwan district (7 high value grids), and Tianhe district (5 high value grids). The hot spot area of wholesalers and sole traders with 99% confidence accounted for 3.2% of the total area. Most of them were located in Yuexiu, Liwan, and Haizhu districts. The distribution center of its hot spots was located around the junction of Yuexiu District and Liwan District.

- **IT practitioners:**

Tianhe district has the largest number of IT practitioners (9 high value grids), followed by Yuexiu and Haizhu district (4 high value grids), and Liwan district (1 high value grid). The hot spot area of IT practitioners with 99% confidence accounted for 8.6% of the total area. Most of them were located in Yuexiu and Tianhe district. Unlike other occupations, the spatial distribution of IT practitioners' hot spot area with 99% was significantly poly-centric. The largest continuous distribution was located in Yuexiu district, and the junction of Yuexiu and Liwan district; the secondary continuous distribution was distributed in southern Tianhe district and the middle of Haizhu district.

- **Express staff:**

Tianhe district has the largest number of express staff (9 high value grids), followed by Yuexiu district (3 high value grids), Haizhu district (2 high value grids), and Liwan district (1 high value grid). The hot spot area of express staff with 99% confidence accounted for 8.5% of the total area. Most of them were located in Yuexiu and Haizhu districts. It is noted that the spatial distribution of express staff's hot spot area with 99% confidence was the most continuous among six occupations, with Yuexiu district as the center covering the surrounding areas including Tianhe, Haizhu, and Liwan districts.

- **Teachers:**

Tianhe district has the largest number of teachers (6 high value grids), followed by Yuexiu district (2 high value grids), and Haizhu and Liwan districts (0 high value grids). The hot spot area of teachers with 99% confidence accounted for 7.7% of the total area. Most of them were located in Yuexiu and Tianhe districts. The distribution of hot spot areas was similar to that of IT practitioners, which also had a polycentric pattern.
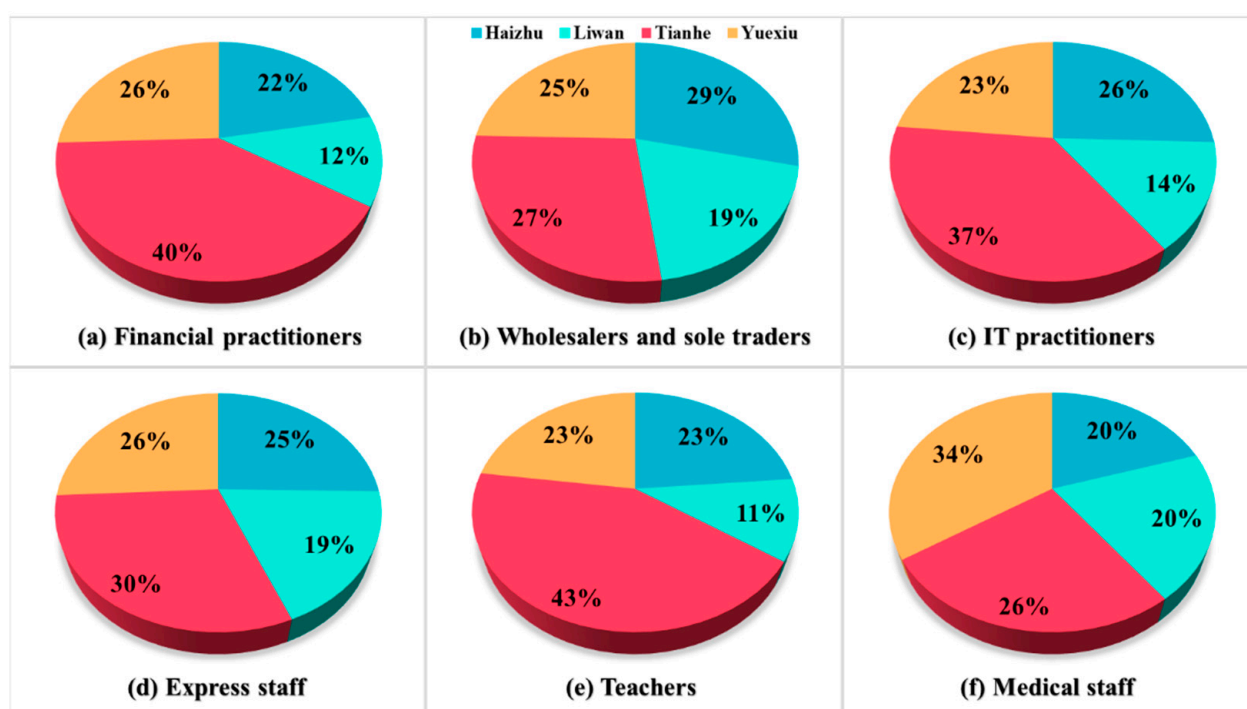
- **Medical staff:**

Yuexiu district has the largest number of medical staff (8 high value grids), followed by Tianhe and Haizhu districts (4 high value grids), and Liwan district (3 high value grids). The hot spot area of medical staff with 99% confidence accounted for 8.9% of the total area. Most of them were located in Yuexiu district. The distribution center of its hot spots was located in Yuexiu district, and areas of Tianhe and Liwan districts around the junction with Yuexiu district.

Due to the defect of incomplete coverage of mobile data samples, direct comparative study on the number of practitioners in various occupations might easily lead to biased results. Therefore, it is more appropriate and scientific to perform spatial analysis on the above, and compare them independently but spatially. Figure 9 shows the proportions of each district of six occupations. It is noted that Tianhe district is the main population distribution area of these six occupations with the average proportion of 33.8%, followed by Yuexiu district (26.2%), Haizhu district (24.2%), and Liwan district (15.8%). These four districts are the center of Guangzhou, playing different roles in the process of urbanization and rapid economic development. In the late 1980s and early 1990s, with the gradual development of China's reform and opening up, Yuexiu District became a central business district with commercial agglomeration, headquarters gathering, and convenient transportation. At the beginning of the 21st century, Tianhe District gradually became the focus of urban economic development planning. Until around 2010, the CBD of Tianhe District, with Zhujiang New Town as the core, has gradually flourished. This area includes a financial and trade area, business area, commercial shopping area, administrative office area, high-rise residential area, cultural activity area, etc., which integrates multiple urban

functions and promotes the further economic development of Guangzhou. Since then, Tianhe CBD has become the center of Guangzhou's economic development, while Yuexiu old CBD has become a mature business district around CBD because of its geographical proximity with Tianhe. Liwan and Haizhu districts are the peripheral areas in the center of Guangzhou, and their economic functions are inferior to Tianhe and Yuexiu districts. Therefore, Tianhe and Yuexiu district accounted for the majority of employed population portrayed in this study (Figure 9).


Figure 9. Proportion of each district of financial practitioners (**a**), wholesalers and sole traders (**b**), IT practitioners (**c**), express staff (**d**), teachers (**e**), and medical staff (**f**).

To further compare the spatial agglomeration characteristics of different occupations, global Moran's I was applied in this study (Figure 10). Moran's I for six occupations were significantly positive, which indicated that spatial distribution of six occupations presents a strong spatial agglomeration feature. Among the six types of occupations portrayed in this study, Moran's I of wholesalers and sole traders was the highest with a value higher than 0.5, followed by IT practitioners, teachers, express staff, medical staff, and financial practitioners. It is noted that the Moran's I of wholesalers and sole traders was much higher than others, which indicated that as a type of typical traditional occupation, wholesalers and sole traders in Guangzhou were highly dependent on location, and showed an obvious feature of industrial clusters around comprehensive wholesale markets and professional wholesale markets in Liwan and Yuexiu districts.

### 3.3. Assessing Urban Occupation Mixture

Although the types of occupations portrayed were insufficient, they comprehensively covered tertiary industry (financial practitioners and wholesalers and sole traders), high-tech industry (IT practitioners) and public service industry (teachers, express staff, and medical staff). Therefore, it is necessary to assess the mixture of urban occupation with these results. To assess the urban occupation mixture, the OMI proposed in this study and Getis-Ord Gi* were applied (Figure 11). The higher the OMI, the more diverse the types of occupations in the region.
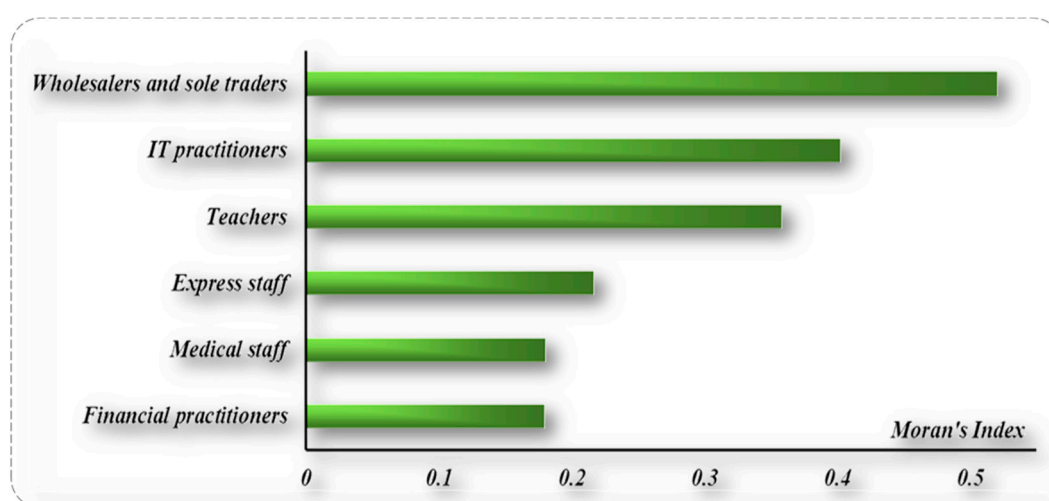
Figure 10. Moran's I of different occupations.

Figure 11 shows that most hot spot areas of OMI with 99% confidence were distributed in Tianhe district, while cold spot areas of OMI were located in the peripheral areas, especially to the northeast and east of the study area. The hot spot areas of OMI with 99% confidence presented a double center distribution: the main center was located in Zhujiang New Town in Tianhe district, and the sub-center was located in Tianhe Smart City in the east of Tianhe district. It was observed that there was obvious difference in spatial distribution between these two hot spot centers. First, the area of hot spot in Zhujiang New Town (A) was much larger than that in Tianhe Smart City (B). Then, the hot spot area A was widely spread to the north and south, with the Zhujiang New Town CBD as the center (Figure 11c). On the contrary, the hot spot area B was distributed in an island pattern (Figure 11d). It indicated that the influence scope of CBD on occupation mixture (Figure 11e) was much greater than Tianhe Smart City (Figure 11f). Usually, an area with high levels of occupation mixture means that firstly it is rich in supporting resources and facilities; secondly, there are great attractions to people from all walks of life; and finally, a high level of occupation mixture leads to more comprehensive city functions serving the locals.

On the one hand, the results reflect that as a CBD with nearly 20 years of development, Zhujiang New Town has the highest level of occupation mixture with the greatest scope of influence (Figure 11e), which might be attributed to the following aspects. First, at the beginning of the planning and construction period, the government's supporting policies and planning were implemented. Second, the economic foundation brought by long term development was also important. Then, CBD was advantageous with a solid foundation in infrastructure and traffic accessibility since it was located in the urban geometric center. These factors are conducive to the development, transformation, and upgrading of the local economy, and the regional industry gradually changed from single to multiple. Therefore, a high level of occupation mixture was observed not only inside the CBD core area, but also the around the adjacent area. On the other hand, Tianhe Smart City, located in the northeast of Tianhe, is the key planning and construction area of Tianhe district in the past five years. It is a High-Tech Development Zone invested by the government in the suburbs of the city center. The development of Tianhe smart city relies more on the government's investment and planning, rather than a new city center formed spontaneously. Tianhe smart city has a short development time, a long distance from the city center, and the infrastructure needs to be improved. As a result, the OMI of this area was high, but the influence scope was much smaller than that of CBD, and therefore the impact on the surrounding areas was relatively weak.
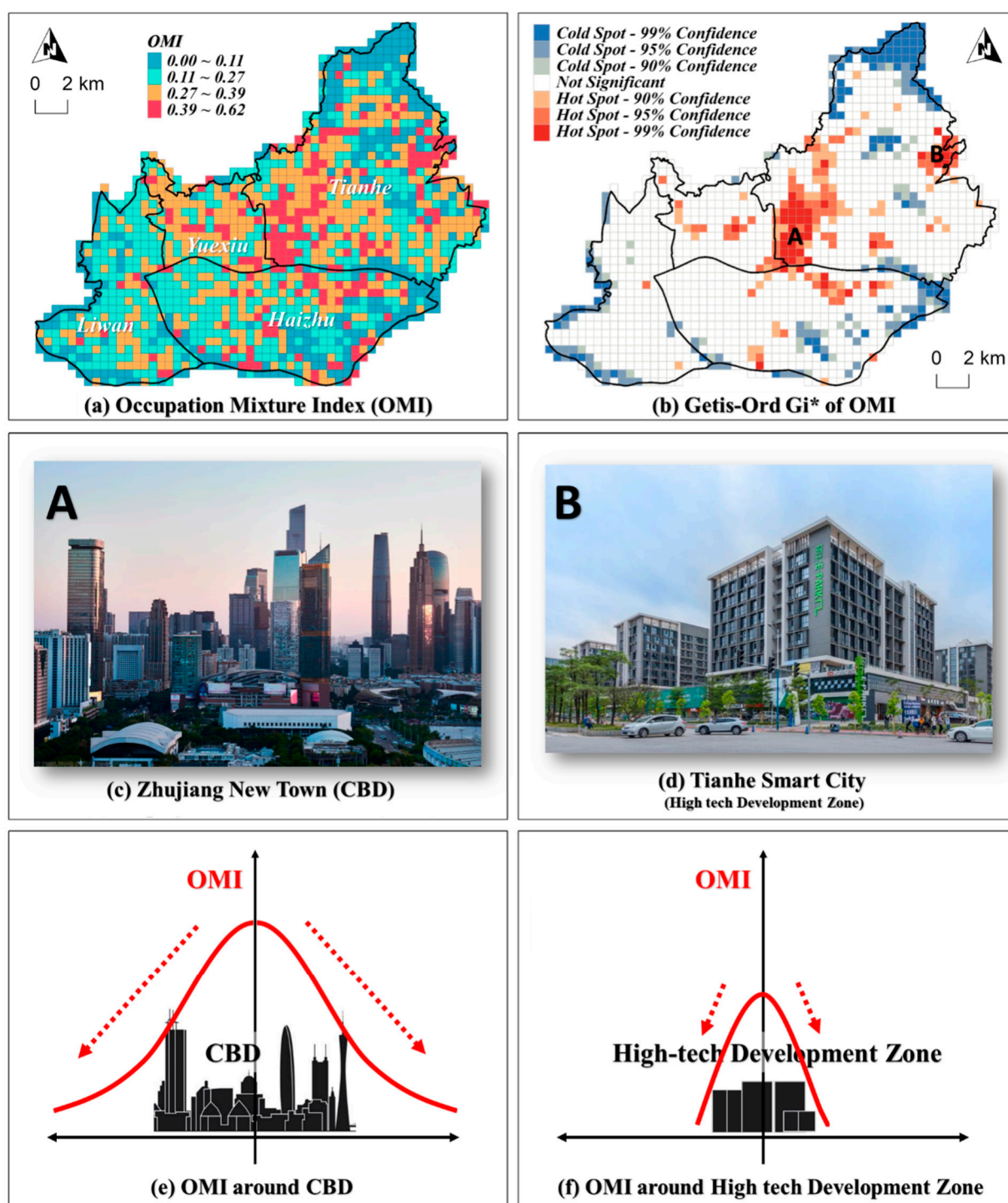
Figure 11. Spatial distribution (**a**) and Getis-Ord Gi* (**b**) of OMI. Area A: Zhujiang New Town CBD (**c**). Area B: Tianhe Smart City (**d**). OMI around CBD (**e**). OMI around High-Tech Development Zone (**f**).

This study identified two centers of high occupation mixture: CBD and High-Tech Development Zone with mobile phone data. The results indicated that CBD has a profound effect on local as well as nearby occupation diversity. Oppositely, the scope of influence of the High-tech Development Zone in the suburbs on occupation mixture was relatively isolated. It hints that the newly-built high-tech development zones, satellite cities, and sub-centers in suburban areas have levels of OMI as high as CBD. When the planning and construction of these areas are improved, such as traffic conditions and infrastructure, their

level of OMI will be improved, and the driving effect of these areas on the surrounding areas will also be improved accordingly.

4. Conclusions

This study firstly proposed a spatiotemporal analytical framework to portray citizens' occupations and assess urban occupation mixture with mobile phone data. In this study, employed population were first extracted from the full samples based on users' active time and location. Then, six types of typical urban occupations were identified based on users' APPs visiting frequency and spatiotemporal mobility behavior. Finally, an urban occupation mixture index (OMI) was proposed to assess the spatial distribution of local occupation mixture in the center of Guangzhou.

In conclusion, in the first part of the study, employed population were extracted according to users' activities during working hours (9:00–18:00), which was similar to the previous studies' process. Correlation between employed population extracted and latest census data was tested with correlation coefficient of 0.81, proving that the accuracy was acceptable. Second, six types of typical urban occupations were identified: financial practitioners, wholesalers and sole traders, IT practitioners, express staff, teachers, and medical staff. Tianhe and Yuexiu district accounted for the majority of employed population of these occupations in this study. Wholesalers and sole traders in Guangzhou were highly dependent on location being the most obvious characteristic of the industrial cluster. Finally, most hot spot areas of the occupation mixture index (OMI) were observed in Tianhe district, while cold spot areas were located in the peripheral areas. Two centers of high occupation mixture were identified: Zhujiang New Town CBD and Tianhe Smart City (High-Tech Development Zone), which presented a different spatial distribution pattern that CBD that had a more profound effect on local as well as nearby occupational diversity. However, the influence scope of Tianhe Smart City on occupation mixture was limited and isolated.

This study provides a novel spatiotemporal analytical framework with mobile phone data, and it was the first time to integrate users' APP use behavior and spatiotemporal mobility pattern to portray users' occupation, and attain further access to the urban occupation mixture of the metropolitan center. Mobile phone data is a type of geospatial big data of high research value. While most studies solely focus on the users' spatial mobility behavior, this study proved that the combination of users' spatial and non-spatial behavior was more capable and effective in urban studies with mobile phone data.

Despite the merits of this study, there are some limitations that need to be addressed in the future. First, though the employed population was tested to be significantly correlated with the total population of census data, the reliability of the results of the occupational portrait identification was not directly examined. It is similar with previous studies with mobile phone data in difficulties of direct validation [1,23,24,35]. The occupations of the employed population inferred from mobile phone data were difficult to validate for two reasons. On the one hand, the mobile phone data used in this study comply with the principles and laws of privacy protection. In other words, the key fields related to personal identity information (such as name, gender, age, phone number, etc.) were removed and confidential. Hence, it is difficult to perform sampling validation by identifying authors themselves or people known. On the other hand, at present, the statistical data of practitioners of each specific occupation are not enough to support further validation. Therefore, if possible, it still needs to be verified with other data sources such as comprehensive survey data [35,36] to further improve the reliability of research conclusions. Second, in fact, both types and work shifts of occupations are complex and diverse. This study portrayed six kinds of typical occupational practitioners, who comply with the eight-hour (nine-to-five or six) working day rule. Therefore, biases were inevitably imported into the results. For instance, those working in night shift were excluded in this study, and though the most-used APPs were considered in occupation identification together with users' spatiotemporal mobility, the problem that a person's most-used APP happens to be related to an interest or hobby rather than occupation remains to be addressed and

validated with other data in the future. Third, both employed population and resident population should be considered in the future to analyze the jobs–housing balance and commuting behavior. Finally, the problem of sample coverage still exists due to the market share of the mobile operator, and the elders who seldom use mobile phones. Therefore, the sampling expansion method should be applied to solve the problem in the future.

References

1. Li, C.; Hu, J.; Dai, Z.; Fan, Z.; Wu, Z. Understanding Individual Mobility Pattern and Portrait Depiction Based on Mobile Phone Data. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 666. [CrossRef]
2. Gan, T.; Li, W.; He, L.; Li, J. Intracity Pandemic Risk Evaluation Using Mobile Phone Data: The Case of Shanghai during COVID-19. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 715. [CrossRef]
3. Willberg, E.; Järv, O.; Väisänen, T.; Toivonen, T. Escaping from Cities during the COVID-19 Crisis: Using Mobile Phone Data to Trace Mobility in Finland. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 103. [CrossRef]
4. Chen, Q.; Hu, Z.; Su, H.; Tang, X.; Yu, K. Understanding travel patterns of tourists from mobile phone data: A case study in Hainan. In Proceedings of the 2018 IEEE International Conference on Big Data and Smart Computing (BigComp), Shanghai, China, 15–17 January 2018; Volume 1, pp. 45–51.
5. Doyle, J.; Huang, P.; Farrell, R.; McLoone, S. Population mobility dynamics estimated from mobile telephony data. *J. Urban Technol.* **2014**, *21*, 109–132. [CrossRef]
6. Manfredini, F.; Pucci, P.; Tagliolato, P. Toward a systemic use of manifold cell phone network data for urban analysis and planning. *J. Urban Technol.* **2014**, *21*, 39–59. [CrossRef]
7. Niu, X.; Yue, Y.; Zhou, X.; Zhang, X. How Urban Factors Affect the Spatiotemporal Distribution of Infectious Diseases in Addition to Intercity Population Movement in China. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 615. [CrossRef]
8. Chen, Y.; Liu, X.; Li, X.; Liu, X.; Yao, Y.; Hu, G.; Xu, X.; Pei, F. Delineating urban functional areas with building-level social media data: A dynamic time warping (DTW) distance based k-medoids method. *Landsc. Urban Plan.* **2017**, *160*, 48–60. [CrossRef]
9. Cao, Z.; Wu, Z.; Li, S.; Guo, G.; Song, S.; Deng, Y.; Ma, W.; Sun, H.; Guan, W. Explicit Spatializing Heat-Exposure Risk and Local Associated Factors by Coupling Social Media Data and Automatic Meteorological Station Data. *Environ. Res.* **2020**, *188*, 109813. [CrossRef]
10. Brown, A.; LaValle, W. Hailing a change: Comparing taxi and ridehail service quality in Los Angeles. *Transportation* **2021**, *48*, 1007–1031. [CrossRef]
11. Yang, H.; Liang, Y.; Yang, L. Equitable? Exploring ridesourcing waiting time and its determinants. *Transp. Res. Part D Transp. Environ.* **2021**, *93*, 102774. [CrossRef]
12. Li, C.; Wu, Z.; Zhu, L.; Liu, L.; Zhang, C. Changes of Spatiotemporal Pattern and Network Characteristic in Population Flow under COVID-19 Epidemic. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 145. [CrossRef]
13. Gao, F.; Li, S.; Tan, Z.; Zhang, X.; Lai, Z.; Tan, Z. How Is Urban Greenness Spatially Associated with Dockless Bike Sharing Usage on Weekdays, Weekends, and Holidays? *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 238. [CrossRef]
14. Gao, F.; Li, S.; Tan, Z.; Wu, Z.; Zhang, X.; Huang, G.; Huang, Z. Understanding the modifiable areal unit problem in dockless bike sharing usage and exploring the interactive effects of built environment factors. *Int. J. Geogr. Inf. Sci.* **2021**, *35*, 1–21. [CrossRef]
15. Li, S.; Zhuang, C.; Tan, Z.; Gao, F.; Lai, Z.; Wu, Z. Inferring the trip purposes and uncovering spatio-temporal activity patterns from dockless shared bike dataset in Shenzhen, China. *J. Transp. Geogr.* **2021**, *91*, 102974. [CrossRef]
16. Li, S.; Lyu, D.; Liu, X.; Tan, Z.; Gao, F.; Huang, G.; Wu, Z. The varying patterns of rail transit ridership and their relationships with fine-scale built environment factors: Big data analytics from Guangzhou. *Cities* **2020**, *99*, 102580. [CrossRef]
17. Li, S.; Lyu, D.; Huang, G.; Zhang, X.; Gao, F.; Chen, Y.; Liu, X. Spatially varying impacts of built environment factors on rail transit ridership at station level: A case study in Guangzhou, China. *J. Transp. Geogr.* **2020**, *82*, 102631. [CrossRef]
18. Tu, Z.; Cao, H.; Lagerspetz, E.; Fan, Y.; Flores, H.; Tarkoma, S.; Nurmi, P.; Li, Y. Demographics of mobile app usage: Long-term analysis of mobile app usage. *CCF Trans. Pervasive Comput. Interact.* **2021**, *5*. [CrossRef]

19. Peltonen, E.; Lagerspetz, E.; Hamberg, J.; Mehrotra, A.; Musolesi, M.; Nurmi, P.; Tarkoma, S. The hidden image of mobile apps: Geographic, demographic, and cultural factors in mobile usage. In Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services, ACM, MobileHCI '18, New York, NY, USA, 3–6 September 2018. [CrossRef]

20. Kumar, P.; Gowlikar, D. *Analysis of Mobile App Usage Preferences of University Students in Eastern Bhutan*; Elsevier: Amsterdam, The Netherlands, 2021. [CrossRef]

21. Malmi, E.; Weber, I. You Are What Apps You Use: Demographic Prediction Based on User's Apps. In Proceedings of the International AAAI Conference on Web and Social Media, ICWSM, Cologne, Germany, 17–20 May 2016.

22. Ernsting, C.; Dombrowski, S.; Šteko, M.; O'Sullivan, J.; Kanzler, M.; Kuhlmey, A.; Gellert, P. Using Smartphones and Health Apps to Change and Manage Health Behaviors: A Population-Based Survey. *J. Med. Internet Res.* **2017**, *19*, e101. [CrossRef] [PubMed]

23. Zhou, X.; Yeh, A.G.O. Understanding the modifiable areal unit problem and identifying appropriate spatial unit in jobs–housing balance and employment self-containment using big data. *Transportation* **2020**. [CrossRef]

24. Zhou, X.; Yeh, A.G.O.; Yue, Y. Spatial variation of self-containment and jobs-housing balance in Shenzhen using cellphone big data. *J. Transp. Geogr.* **2018**, *68*, 102–108. [CrossRef]

25. Qian, C.; Li, W.; Duan, Z.; Yang, D.; Ran, B. Using mobile phone data to determine spatial correlations between tourism facilities. *J. Transp. Geogr.* **2021**, *92*, 103018. [CrossRef]

26. Guo, H.; Li, W.; Yao, F.; Wu, J.; Zhou, X.; Yue, Y.; Yeh, A.G.O. Who are more exposed to PM2.5 pollution: A mobile phone data approach. *Environ. Int.* **2020**, *143*, 105821. [CrossRef] [PubMed]

27. Lee, K.-S.; You, S.Y.; Eom, J.K.; Song, J.; Min, J.H. Urban spatiotemporal analysis using mobile phone data: Case study of medium- and large-sized Korean cities. *Habitat Int.* **2018**, *73*, 6–15. [CrossRef]

28. Louail, T.; Lenormand, M.; Picornell, M.; Cantú, O.G.; Herranz, R.; Frias-Martinez, E.; Ramasco, J.J.; Barthelemy, M. Uncovering the spatial structure of mobility networks. *Nat. Commun.* **2015**, *6*, 6007. [CrossRef] [PubMed]

29. Gong, L.; Jin, M.; Liu, Q.; Gong, Y.; Liu, Y. Identifying Urban Residents' Activity Space at Multiple Geographic Scales Using Mobile Phone Data. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 241. [CrossRef]

30. Gao, S. Spatio-Temporal Analytics for Exploring Human Mobility Patterns and Urban Dynamics in the Mobile Age. *Spat. Cogn. Comput. Interdiscip. J.* **2015**, *15*, 86–114. [CrossRef]

31. Yin, L.; Lin, N.; Zhao, Z. Mining Daily Activity Chains from Large-Scale Mobile Phone Location Data. *Cities* **2021**, *109*, 103013. [CrossRef]

32. Cao, J.; Li, Q.; Tu, W.; Gao, Q.; Cao, R.; Zhong, C. Resolving urban mobility networks from individual travel graphs using massive-scale mobile phone tracking data. *Cities* **2021**, *110*, 103077. [CrossRef]

33. Jiang, S.; Ferreira, J.; González, M. Clustering daily patterns of human activities in the city. *Data Min. Knowl. Discov.* **2012**, *25*, 478–510. [CrossRef]

34. Ding, D.; Mao, H.; Lu, Z. Research on population type recognition based on mobile signaling data. In Proceedings of the 14th China Intelligent Transportation Conference, Qingdao, China, 31 October 2019; pp. 417–428.

35. Tu, W.; Cao, J.; Yue, Y.; Shaw, S.; Zhou, M.; Wang, Z.; Chang, X.; Xu, Y.; Li, Q. Coupling mobile phone and social media data: A new approach to understanding urban functions and diurnal patterns. *Int. J. Geogr. Inf. Sci.* **2017**, *31*, 2331–2358. [CrossRef]

36. Liu, K.; Murayama, Y.; Ichinose, T. Exploring the relationship between functional urban polycentricity and the regional characteristics of human mobility: A multi-view analysis in the Tokyo metropolitan area. *Cities* **2021**, *111*, 103109. [CrossRef]

37. Zhou, X.; Yeh, A.; Li, W.; Yue, Y. A commuting spectrum analysis of the jobs–housing balance and self-containment of employment with mobile phone location big data. *Environ. Plan. B* **2018**. [CrossRef]

38. Shannon, C.E. A mathematical theory of communication. *Bell Syst. Tech. J.* **1948**. [CrossRef]