Citation: SONG Cheng, CHEN Jiachao, LI Caixia, AI Guantao. Urban Vitality Zone and Central District Identification Based on Big Data: A Case Study in Guangzhou City [J]. Urban Transport of China, 2020 (04): 71–78.

Urban Vitality Zone and Central District Identification Based on Big Data: A Case Study in Guangzhou City

SONG Cheng, CHEN Jiachao, LI Caixia, AI Guantao

Guangzhou Transport Planning Research Institute, Guangzhou 510030, Guangdong Province, China

Abstract: The urban vitality zone and the central district can offer important basic information to understand and research the city, which can also serve as the premise of making urban management and control polices. It is difficult to define the boundary of an urban vitality zone and a central district accurately depending on traditional survey data, which is usually the qualitative analysis. The emergence of big data provides conditions for boundary identification of urban functional areas. Based on cellular signaling, Internet location data, Point of Interest (POI) and other multi-source data, this paper develops two static indicators (comprehensive population density and POI density) as well as two dynamic indicators (urban center accessibility and point density of business and leisure activities) so as to comprehensively identify the urban vitality zone and central district boundaries. The combination of dynamic and static indicators overcomes the limitations of boundary identification with single index or only static indexes. Taking Guangzhou as an example, the paper demonstrates the practicability of the method and concludes that the scope of Guangzhou's vitality zone and central district is equivalent to that of world-class international cities. **DOI:** 10.13813/j.cn11-5141/u.2020.0028-en

Keywords: cellular signaling data; POI; vitality zone; central district; boundary identification; kernel density; Guangzhou

1 Introduction

Urban spatial structure reflects the spatial distribution and combination of urban elements. It is the spatial reflection of urban economic and social structures and the spatial form of the existence and development of society and economy. Zhou and Ye [1] divided the research on China's urban spatial structure into three stages: the initial stage in which western theories were introduced to China which then started empirical research; the accumulation stage in which China accumulated empirical researches; the diversified stage in which China's urban spatial patterns were summarized and new urban patterns were studied, which is China's current stage. Wang, Wu and Liu [2] conducted a quantitative analysis of the four elements that defined urban space: the spatial element, the temporal element, the flow element and the gravitational element. They used economic potential and traffic weights to determine the spatial scope of a metropolitan area and proposed a method to determine the radius and structure of the metropolitan area by integrating the four elements. Besides, they conducted an empirical analysis on Chengdu Metropolitan Area, and finally obtained the radius and structure of Chengdu Metropolitan Area. When defining the space of Lanzhou Metropolitan Area, Guo and Feng [3] integrated the economic distance and the quality of life of surrounding cities into the calculation of the field strength and gravity model, which makes their definition method more useful as a reference. Using the Point of Interest (POI) data provided by Baidu, Inc. as the basis, Liu, Chen and Kong [4] proposed a method to identify the urban spatial structure with big data technology. Based on more than 400,000 POI samples collected from the major urban area of Chongqing, Duan et al. [5] used kernel density estimation, natural breaks method, proximity analysis and other methods to identify the structure and influence area of the city and its multiple centers with different functions, considering the spatial distribution and aggregation of the overall and different types of POI data. Taking Panjin City in Liaoning Province as an example, Ming [6] used POI data and kernel density estimation to analyze the spatial distribution pattern of various facilities in the city and identified the urban layout. With the assistance of massive spatio-temporal data obtained from Baidu Maps, Chen [7] took the main urban area of Shanghai as an example, focused on service accessibility and element completeness, and introduced accessibility analysis, isochron analysis, and statistical analysis that is based on POI classification. Chen [7] then quantitatively examined the development of multi-center space from three aspects: coverage, promptness and

Received: 2019-03-26
Supported by: Reserve Talent Training Program in the Guangzhou "Lingnan Talent Project"
First author: SONG Cheng (1983–), male, from Zigui of Hubei Province, master's degree, is mainly engaged in the research on traffic modeling and simulation. E-mail: 510659684@qq.com

capacity of service to explore new methods for evaluating the spatial development of cities with multiple centers.

In summary, most of previous studies on urban space identification adopted a single indicator. Although some adopted multiple indicators, they only focused on static indicators without considering dynamic indicators. Using one single indicator can only reflect the characteristics of one aspect and ignore others, whereas the formation of the urban spatial structure is affected by various factors and one single indicator fails to truly reflect the urban spatial patterns. In addition, using static only indicators cannot reflect the influence of a city's mobility on its spatial pattern.

2 Factors and indicators for identifying urban development boundary

An urban vitality zone is defined as an area with the most concentrated public activities in a city, such as political, economic and cultural activities. It is an area with a high-level concentration of urban functions, which is a major component in a system of urban public activities. It provides facilities and space for economic, political, cultural and other activities in a city and the region where the city is located, and it is spatially different from other areas of the city. The central district is an area with the most mature development in a city. It is mainly featured as multi-functional and comprehensive, which is a main area that gathers urban production and service. It is also the center of political and administrative control, transportation, information and technology as well as a talent-intensive area.

The urban vitality zone and the central district are areas with the most developed urban transportation network, which are the areas with the most intensive travel activities. Boundary identification of these areas must consider the impact of both static and dynamic indicators. Therefore, this paper constructs a comprehensive identification system that includes two static indicators and two dynamic indicators. The static indicators are comprehensive population density and POI density, which represent population agglomeration and infrastructure configuration. The dynamic indicators are urban center accessibility and point density of business and leisure activities, which reflect the attractiveness to passenger flow by regional transportation connections and transportation convenience in the existing transportation system.

1) Comprehensive population density

Comprehensive population density pe_i is different from the residential population density, which reflects the normal population size per unit area. It is a comprehensive indicator that considers both residential employment population and reflects the status of residence and employment. In general, areas with higher comprehensive population density are more developed.

$$pe_i = \frac{emp_i + pop_i * \alpha}{area_i}, \quad (1)$$

where pe_i is comprehensive population density of area i (person·km^{-2}); emp_i is the number of jobs in area i; pop_i is the number of residents in area i; $area_i$ is the area of area i (km^2); α is the population adjustment coefficient, which is used to reflect the influence of active population. The value of α can be determined as the ratio of permanent residents to employees in a city.

2) POI Density

POI includes hospitals, schools, restaurants, hotels, supermarkets, shopping malls, major office buildings, major transportation hubs, parks, gas stations, banks, recreation facilities, government agencies and other fundamental facilities for basic urban services. POI density reflects the density of public service facilities in an area and its development to a certain extent.

$$poi_i = \frac{\sum_{j=1}^{n} poi_{ij}}{area_i}, \quad (2)$$

where poi_i is POI density of area i (number of POIs per square kilometer); poi_{ij} is the number of type j POIs in area i .

3) Urban center accessibility

An urban center could be the location of the city government or a landmark, and urban center accessibility reflects the convenience to travel from an area to the urban center. The higher urban center accessibility of an area indicates a more developed transportation system in the area, which can more truly reflect the development of the area.

$$t_i = \min\left(t_{cari}, t_{busi}, t_{\mathrm{metro}i}\right) \quad (3)$$

where t_i is the minimum travel time from area i to the urban center /(min); t_{cari}, t_{busi}, and t_{metroi} indicate the travel time from area i to the urban center by car, bus, and rail transit, respectively/(min).

4) Point density of business and leisure activities

The density of business and leisure activities reflects not only the normal travel intensity of the people who live or work in an area but the travel intensity of visitors. Therefore, it represents the area's overall attractiveness to passenger flow, which indicates the concentration of traffic flow and maturity of important economic, political and cultural activities in developed urban areas. An area could have very high comprehensive population density, such as an industrial area where a large number of workers, but low point density of business and leisure activities, and such an area would not represent the most mature and active area of the city.

$$d_i = \frac{\sum_{j=1}^{n} d_{ij}}{area_i}, \quad (4)$$

where d_i is point density of business and leisure activities of area i (number of activity points per square kilometer), and d_{ij} is the number of type j activity points in area i.

3 Identification methods of urban development boundaries

3.1 Obtaining indicators based on big data

It is difficult to obtain the four major indicators defined above in the first 10 years of the 21st century. Comprehensive population density can be obtained through statistics; however, the statistics of permanent residents and employees cannot reflect the actual population in large cities, especially in metropolises and megacities. On the other hand, data granularity is coarse (the smallest granularity is at the neighborhood committee level), which cannot achieve the data granularity requirements needed for identifying urban development boundaries. Before the rise of big data, POI density cannot be obtained by traditional methods. Although urban center accessibility can be obtained with the help of traffic modeling, it is difficult to guarantee the accuracy of modeling results. Point density of business and leisure activities cannot be obtained by traditional methods, either. In recent years, the rise of big data has provided an effective way to obtain these indicators, and data granularity can well support the identification of urban development boundaries. The methods to obtain each indicator are listed as follows:

1) Comprehensive population density can be obtained with the help of information technology, such as long-term cellular signaling data. The data can be processed with certain algorithms to gain the accurate spatial distribution of residents and employees (Live-Work), which can be further processed to obtain comprehensive population density of an area. The relevant algorithm is shown in Fig. 1.

2) POI density can be obtained mainly through the POI data provided by Baidu, Inc., which is counted in small areas.

3) Urban center accessibility can better reflect the connection between an area and the city center, so the accuracy of this indicator is essential to the identification of urban space. Internet-based location data, such as navigation data from AutoNavi and Tencent, is an ideal source for this indicator. It can be used to obtain the fine-grained travel time matrix with small intervals to show the actual travel time among different traffic zones, which can support to accurately identify the travel time from a location to any area. The travel time matrix based on the Internet is more accurate. It reflects the actual travel time of travelers instead of the modeled travel time based on the traffic assignment results of a certain period from a traffic model. Generally, the shorter travel time to the city center indicates stronger accessibility and a more developed area.

4) Point density of business and leisure activities must be obtained through cellular signaling data which can provide the distribution of activity points and thus activity point density. Different from traditional algorithms, this paper proposes a new method to identify activity points that can fit user's travel characteristics. This method adopts cellular signaling data and Internet-based location data, which are triggered by high-frequency signals and temporally and spatially correlated. The data are based on the Live-Work correspondence and spatio-temporal kernel clustering analysis. The detailed process of this method is shown in Fig. 2.

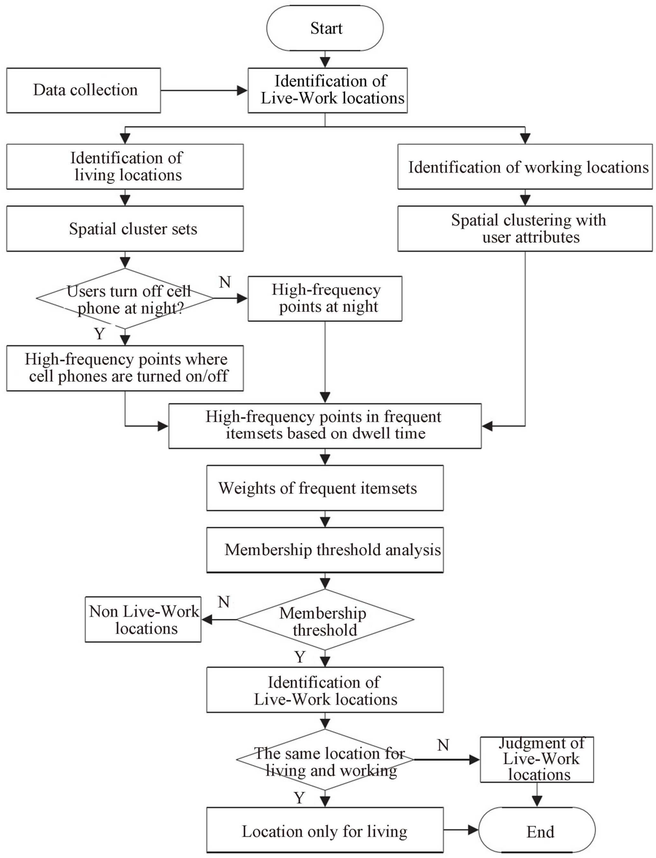


Fig. 1 Identification of Live-Work location based on cellular signaling data

3.2 Calculation of comprehensive indicators

The four indicators reflect the development of an area from different aspects. Therefore, it is necessary to weight the four indicators to obtain a comprehensive one reflecting multiple factors, which will be used to identify urban development boundaries. Expert opinions have been collected to determine the weights of the four indicators. Through a survey from 20 experts in the field of urban planning, the final weights of the four indicators are determined to be 0.22, 0.18, 0.33, and 0.27 respectively.

After the weights of the four indicators are determined, it is necessary to normalize the indicators to ensure that their magnitudes are comparable. This paper normalizes each indicator based on their ratios to the threshold which is determined by dividing a city into multiple traffic zones. Smaller traffic zones are preferred. In general, there should be more than 3,000 traffic zones for metropolises and megacities and 1,500 for large cities. Each indicator is calculated for each traffic zone and the distribution curve is then plotted by sorting all indicator values in an ascending order. The 80% quantile value of the distribution curve is selected as the

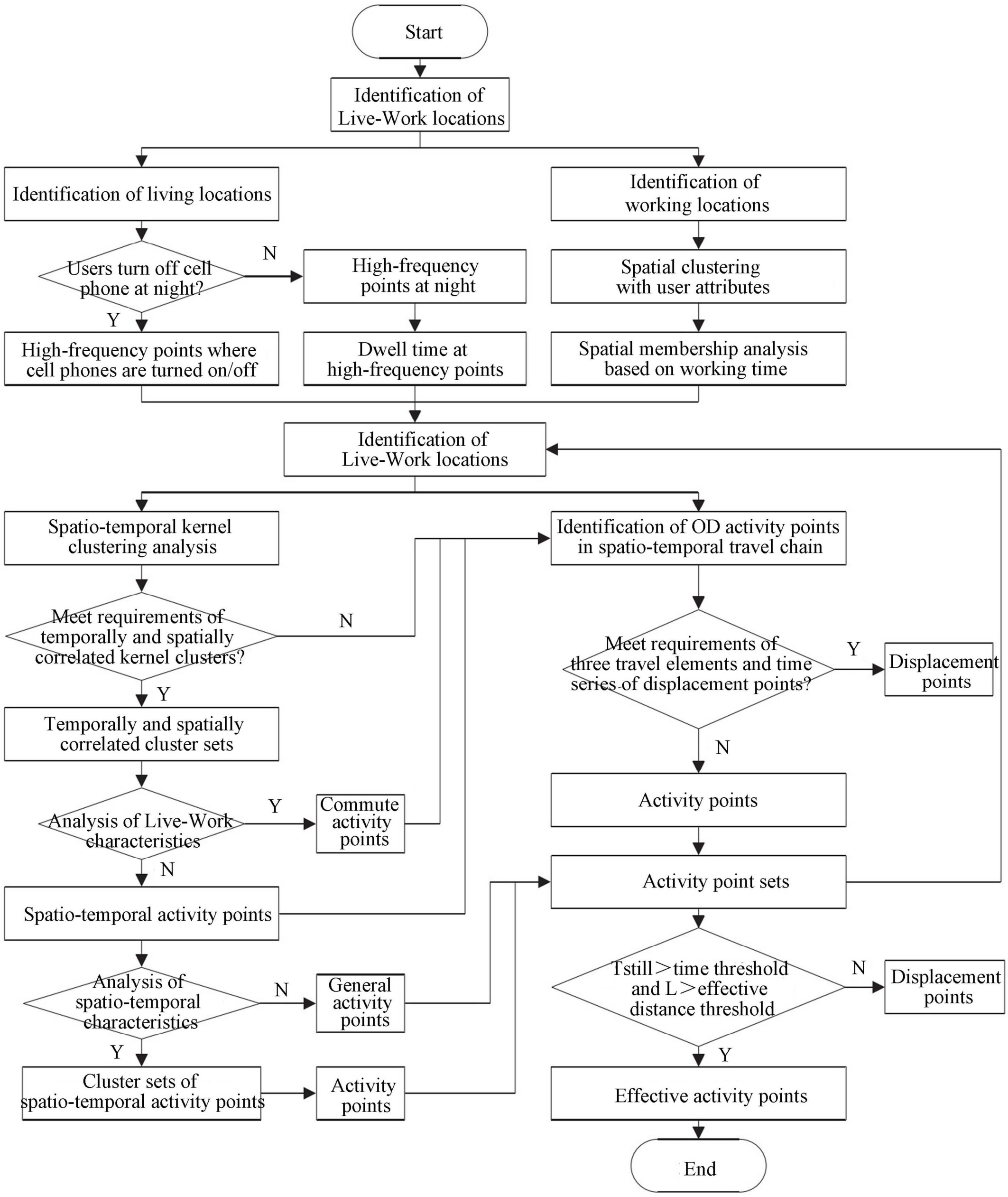


Fig. 2 Identification of activity points based on cellular signaling data

threshold of the indicator. The normalized value of the indicator in a traffic zone is then calculated as the ratio of the original indicator value to the threshold of the indicator. The comprehensive indicator of each traffic zone can be calculated as follows:

$$\overline{A}_{ij} = A_{ij}/\theta_j, \tag{5}$$

$$\lambda_i = \sum_{j=1}^{n} \overline{A}_{ij} * \beta_j, \tag{6}$$

where $\overline{A}_{ij}$ is the normalized value of indicator j in traffic zone i; A_{ij} is the calculated value of indicator j in traffic zone i; θ_j is the threshold of indicator j ; λ_i is the comprehensive indicator of traffic zone i; β_j is the weight coefficient of indicator j .

During normalization, comprehensive population density, POI density, and point density of business and leisure activities are all positive factors, while urban center accessibility is a negative factor. Therefore, the normalization of urban center accessibility should be conducted on the reciprocal of its originally calculated value.

3.3 Identification of urban development boundaries based on kernel density estimation

In recent years, kernel density estimation has been widely used in the exploration of urban hot spots [8–9]. This method can be used to calculate the density of spatial elements, such as points and lines, in the proximity of an area and to continuously simulate density distribution in order to reflect the distribution of spatial elements with the kernel density of each grid in an image [5]. It is suggested to estimate the density of proximity based on the kernel density of the comprehensive indicator in each traffic zone and compare the results of kernel density analysis under different search radii in order to select the optimal search radius that is suitable for this research. The kernel density function is calculated as follows:

$$f(x)=\sum_{i=1}^{n}\frac{1}{\pi r^2}\varphi\left(\frac{d_{ix}}{r}\right) \tag{7}$$

where $f(x)$ is the kernel density value of the comprehensive indicator per unit area; r is the search radius (km); n is the total number of samples; d_{ix} is the distance between traffic zone i and x (km); φ is the weight for distance. This paper uses the comprehensive indicator and kernel density estimation to identify urban development boundaries.

4 Identification of urban space of Guangzhou City

4.1 Four major indicators and their thresholds

4.1.1 Data source

For boundary identification of the vitality zone and the central district of Guangzhou, this paper mainly employs cellular signaling data from China Unicom, location data from Tencent and POI data from Baidu. How the data are used is shown in Tab. 1, and the number of POIs is shown in Tab. 2.

4.1.2 Calculation of indicators for each traffic zone

Based on various information and data, the four main indicators for each of the 3,989 traffic zones in Guangzhou are calculated. The spatial distribution of each indicator is shown in Fig. 3.

Tab. 2 Distribution of POI data in Guangzhou

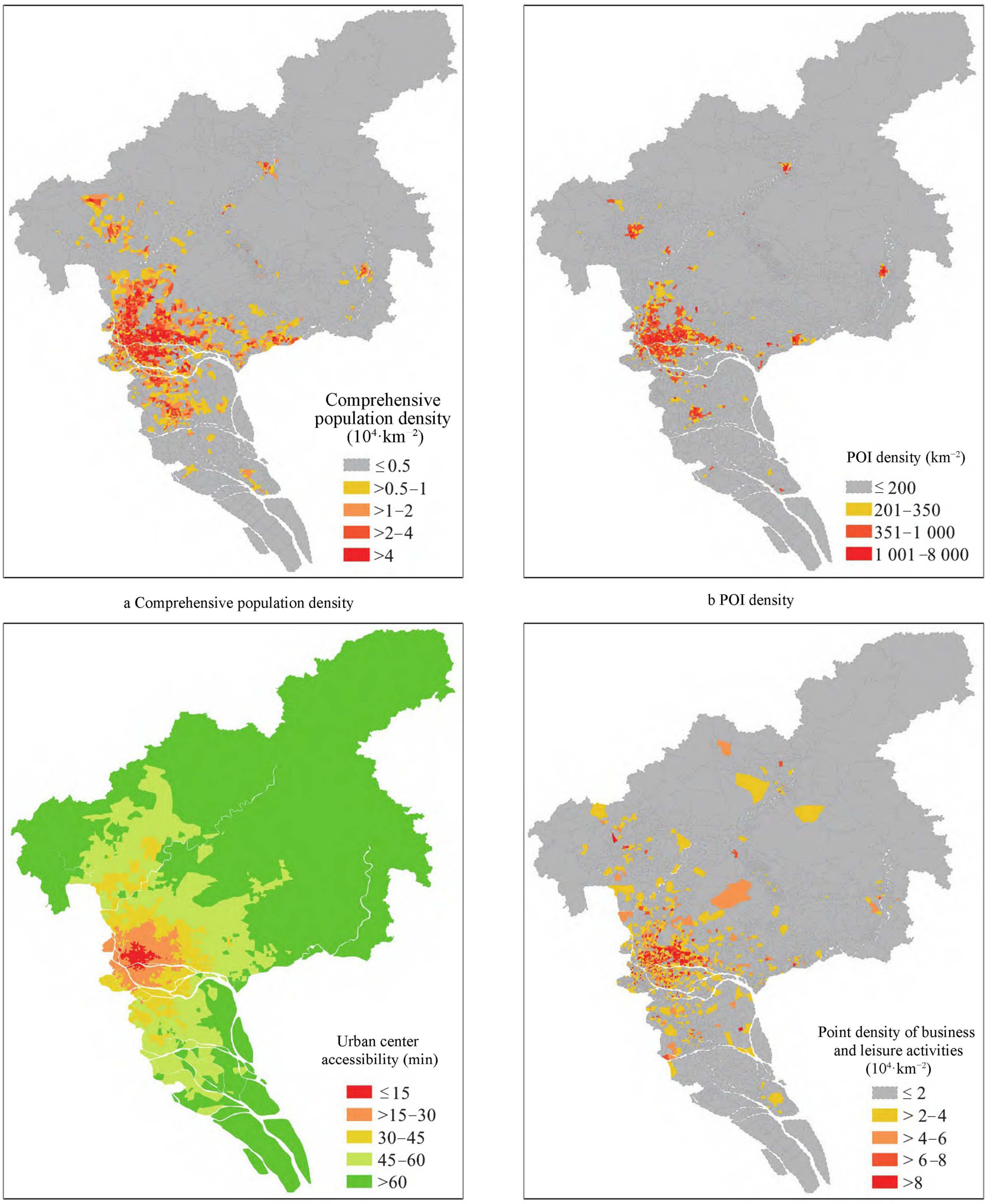
Category	Number of POIs	Percentage/%
Shopping mall	51 615	29.2
School	4 742	2.7
Hospital and health care facility	4 164	2.4
Leisure and entertainment facility	3 726	2.1
Hotel	3 389	1.9
Bank	2 649	1.5
Sport and fitness center	1 317	0.7
Cultural venue	543	0.3
Restaurant	21 336	12.1
Daily life service facility	17 811	10.1
Corporate	27 456	15.6
Institution	9 193	5.2
Infrastructure	12 695	7.2
Tourist service	248	0.1
Others	15 581	8.8

4.1.3 Determination of thresholds and normalization of indicators

The indicator values for all traffic zones are sorted in an ascending order to determine the threshold of an indicator, and the 80% quantile value is selected as the threshold of the indicator (see Fig. 4). The thresholds of comprehensive population density, POI density, urban center accessibility and point density of business and leisure activities are $25{,}000\cdot km^{-2}$, $350\cdot km^{-2}$, 62 min and $45{,}000\cdot km^{-2}$, respectively. After the indicators are normalized and weighted based on weight coefficients, the comprehensive indicator of each traffic zone is calculated, which is shown in Fig. 5.

Tab. 1 Specific acquisition methods of each indicator

Indicator	Data source	Calculation method	Parameter
Comprehensive population density	Cellular signaling data from China Unicom and location data from Tencent	Distribution of Live-Work locations is identified based on long-term data, and 3,989 traffic zones are determined through clustering. The density values are calculated by Formula (1) and then normalized by Formula (5).	α = the number of employees/the number of residents in Guangzhou = 862/1,450 = 0.59
POI density	POI data from Baidu	The calculation is based on the number of POIs and Formula (2).	
Urban center accessibility	Location data from Tencent	Tencent's location data is used to obtain the actual travel time matrices of cars, buses and rail transit for the 3,989 traffic zones, and the accessibility is calculated by Formula (3).	
Point density of business and leisure activities	Cellular signaling data from China Unicom and location data from Tencent	Cellular signaling data from China Unicom is used to obtain business and leisure activity points. Tencent data is used to obtain the market share of China Unicom, which is used to expand the samples. The density is calculated by Formula (4).	The market share of China Unicom is around 18%.



a Comprehensive population density

b POI density

c Urban center accessibility

d Point density of business and leisure activities

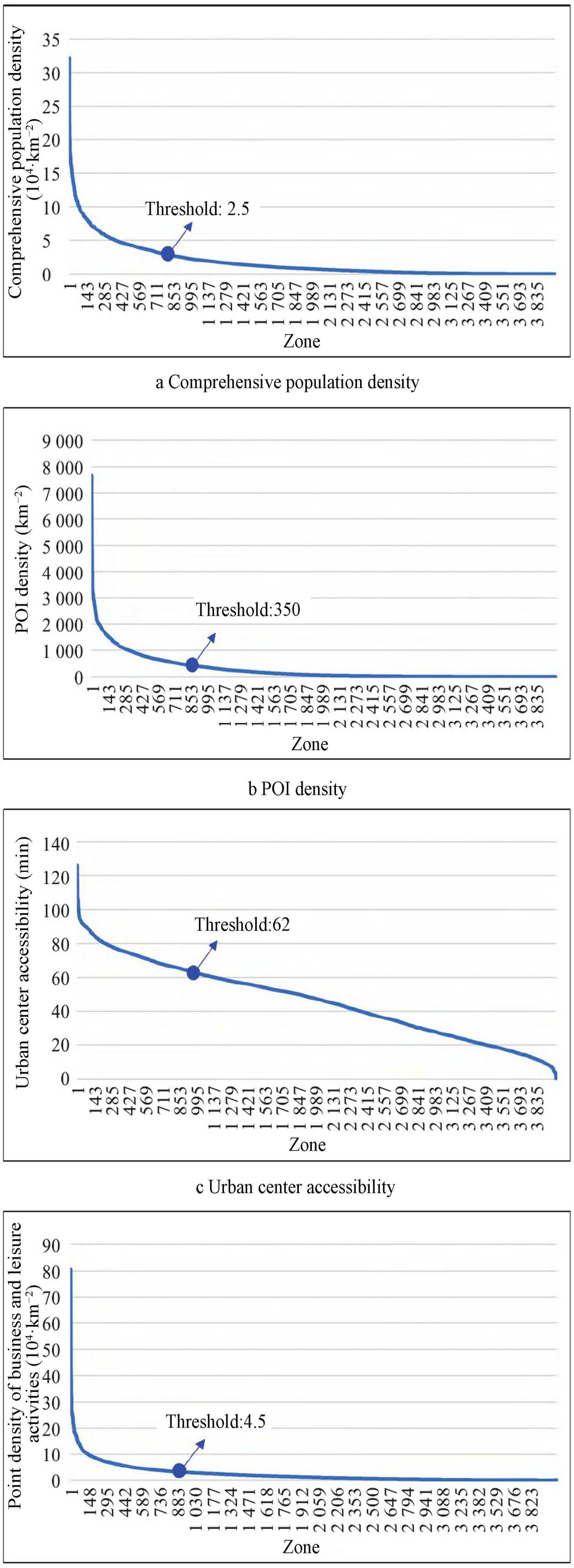
Fig. 3 Spatial distribution of four indicators

4.2 Boundary identification of Guangzhou's vitality zone and central district

The kernel density of the comprehensive indicator is calculated by performing kernel density estimation, and the results are shown in Fig. 6. Based on the results of kernel density and the consideration of the city's current roadway network, Guangzhou's vitality zone and central district are identified and their areas are around 93 km^2 and 600 km^2, respectively. This result can guide the formulation of the city's traffic policies and provide foundation for the city's space planning.

4.3 Comparison with similar cities

Compared with four first-class international cities (i.e., New York, London, Tokyo and Paris), the area of Guangzhou's vitality zone identified by this research ranges between Central Six Districts of Tokyo and Little Paris District of Paris; the area of Guangzhou's central district is close to that of Beijing, Tokyo and Paris (see Tab. 3). From the perspective of the areas of the vitality zone and the central district, Guangzhou's urban development has reached the level of first-class international cities.



a Comprehensive population density

b POI density

c Urban center accessibility

d Point density of business and leisure activities

Fig. 4 Distribution sorting curve of four indicators

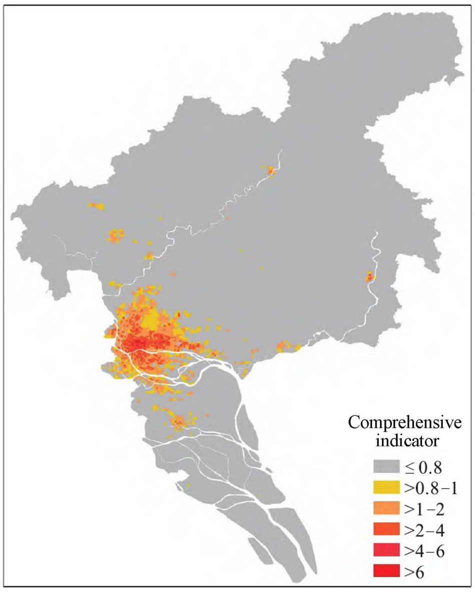


Fig. 5 Spatial distribution of comprehensive indicator

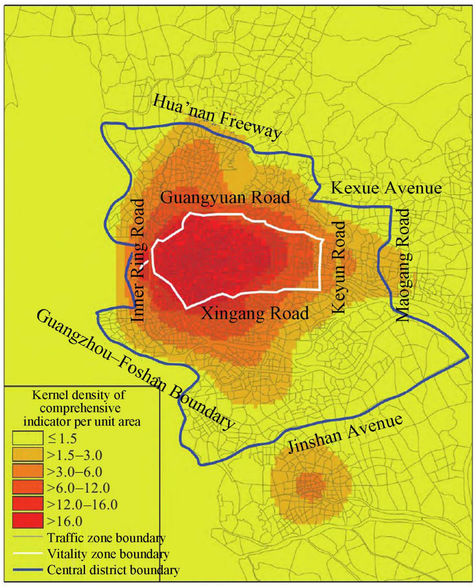


Fig. 6 Boundaries of Guangzhou's vitality zone and central district identified by kernel density estimation

Tab. 3 Comparison of the identified vitality zone and central district in Guangzhou and those of other compatible cities

Cities	Area of vitality zone (km^2)	Area of central district (km^2)
Guangzhou	93	600
Beijing	93	667
New York	59.5	
London	27	320
Tokyo	71.8	617
Paris	105	657

5 Conclusion

This paper proposes a quantitative method to identify urban development boundaries depending on comprehensive analysis with multiple indicators. It is an extension of the identification method that uses a single indicator. In addition to static indicators, namely population aggregation and infrastructure configuration, the paper also introduces dynamic indicators, i.e., accessibility and attractiveness to passenger flow, to reflect the influence of urban mobility on spatial patterns. This method makes the identification of urban development boundaries more realistic and provides more accurate support for the implementation of urban control policies. However, this paper only identifies urban development boundaries from the perspective of transportation, namely the identification of physical boundaries. In fact, in addition to physical boundaries, urban development boundaries also include economic ties, social cultural bonds and sense of identity. The research in these aspects needs to be further improved.

References

[1] Zhou Chunshan, Ye Changdong. Progress on studies of urban spatial structure in China [J]. 地理科学进展, 2013, 32 (7): 1030–1038 (in Chinese).

[2] Wang Jianwei, Wu Yongping, Liu Xin. Method of Circle Region Classification of Metropolitan Region [J]. Journal of Architecture and Civil Engineering, 2007 (2): 91–94 (in Chinese).

[3] Guo Aijun, Feng Qiyuan. 兰州都市圈空间界定方法研究 [J]. 甘肃社会科学, 2009 (6): 137–140 (in Chinese).

[4] Liu Pengcheng, Chen Chao, Kong Bin, et al. 基于 POI 数据的镇江中心城区空间结构识别与特征解析 [J]. 江苏城市规划, 2017 (9): 7–12 (in Chinese).

[5] Duan Yaming, Liu Yong, Liu Xiuhua, et al. Identification of Polycentric Urban Structure of Central Chongqing Using Points of Interest Big Data [J]. Journal of Natural Resources, 2018, 33 (5): 788–800.

[6] Ming Yingnan. 基于网络开放数据的城市空间布局识别研究: 以辽宁省盘锦市为例[C]//中国城市规划学会. 共享与品质: 2018 中国城市规划年会论文集. Beijing: China Architecture Publishing, 2018: 1073–1082 (in Chinese).

[7] Chen Yingxue. Study on Polycentric Urban Spatial Development Evaluation Approaches Based on the Webmap: A Case of Shanghai Main Urban Area [J]. Shanghai Urban Planning Review, 2017 (6): 96–102.

[8] Chu H J, Liau C J, Lin C H, et al. Integration of Fuzzy Cluster Analysis and Kernel Density Estimation for Tracking Typhoon Trajectories in the Taiwan Region [J]. Expert Systems with Applications, 2012, 39 (10): 9451–9457.

[9] Xie Z, Yan J. Kernel Density Estimation of Traffic Accidents in a Network Space [J]. Computers Environment & Urban Systems, 2008, 32 (5): 396–406.

