ARTICLE IN PRESS

Journal of Urban Management xxx (xxxx) xxx

Contents lists available at ScienceDirect

Journal of Urban Management

journal homepage: www.elsevier.com/locate/jum

Research Article

Identifying functional agglomerations and urban centers using open-source data and machine learning: Framework, applications and planning implications

Pengyu Zhu[a,*], Jianqi Li[b], Zining Wang[c]

[a] Division of Public Policy, The Hong Kong University of Science and Technology, Hong Kong, China
[b] The Hong Kong University of Science and Technology (Guangzhou), Guangzhou, 510000, China
[c] Kyushu University, 744 Motooka Nishi-ku, Fukuoka, 819-0395, Japan

ARTICLE INFO

Keywords:
Agglomeration
Unsupervised machine learning
Point of interest
Urban sub-centers
Multifunction
Clustering analysis

ABSTRACT

Metropolitan polycentricity is a topic that has been extensively studied. Traditional approaches to identifying urban centers have relied on high-resolution survey data, which is limited in many regions. More recent studies utilizing open-access point of interest (POI) data have primarily focused on POI density, often overlooking functional diversity and interactions. This paper introduces a theoretical framework for identifying distinct urban functional agglomerations and centers using open-source POI data. Considering both density and the combination of urban functions, it advances beyond previous density-centered methodologies by revealing multifunction features. Applying k-means clustering to Guangzhou, this study identifies 63 agglomerations and 11 centers, consistent with the city's Urban Development Plan (2018–2035). The results reveal the multifunctional nature of concentrations across multiple scales. The findings offer policy insights for improving land use, facility investment, and transportation planning. The proposed analytical framework can be readily applied to cities or countries with available open-source POI data.

1. Introduction

Cities are central places where diverse social and economic activities happen. An extensive range of literature indicates that polycentricity is one of the most significant characteristics in a metropolis (Garreau, 2011; Giuliano & Small, 1991; Song, 1994; Gordon et al., 1986; McDonald, 1987; Zhu & Guo, 2022). Giuliano and Small (1991) explain that the formation of polycentric patterns is driven by economic forces (e.g., employment), which can produce a dynamic interplay between economic agglomeration and congestion effects. The more economic activities are centralized, the more congestion will be created, acting as a counterforce favoring decentralization. If clustering forces are strong enough, the activities will be centralized in a secondary center. Agarwal et al. (2012) establish that agglomeration economies work at multiple scales in urban areas because of the existence of numerous urban nodes of employment. Giuliano and Small (1999) conclude four forces that shape the formation of an employment sub-center: first, accessibility to the labor force, which ensures that businesses can easily tap into a pool of qualified workers while also enabling feasible commutes; second, economic agglomeration effects, whereby the clustering of firms results in the sharing of resources, knowledge, and infrastructure; third, congestion in the main center, which increasingly prompts firms to locate in less crowded areas and thereby enhance commuting and

* Corresponding author.
E-mail addresses: pengyuzhu@ust.hk (P. Zhu), Jli330@connect.hkust-gz.edu.cn (J. Li), wang.zining.804@s.kyushu-u.ac.jp (Z. Wang).

https://doi.org/10.1016/j.jum.2025.04.004
Received 30 June 2024; Received in revised form 2 April 2025; Accepted 10 April 2025
Available online xxxx

logistical efficiency; and finally, convenient transportation nodes, such as highway intersections or transit stations, that link these emerging sub-centers to the broader metropolitan network. These forces fundamentally give rise to multiple functional agglomerations and urban sub-centers in a metropolis, wherein diverse economic and social activities are clustered in distinct zones to form specialized hubs.

Conventional studies on the identification of urban centers and subcenters have primarily relied on survey data, with employment density serving as the key indicator (Anderson & Bogart, 2001; Bogart & Ferry, 1999; Forstall & Greene, 1997; Giuliano & Small, 1991). These approaches require high-resolution, high-quality geospatial data, often unavailable in many regions. More recent research has shifted to using open-access Point of Interest (POI) data to identify urban centers and agglomerations (Li et al., 2018; Deng et al., 2019; Lou et al., 2019; Lu et al., 2020; Hou & Chen, 2021; Xie, Luan, & Xue, 2021; Zhou et al., 2022; Zhu et al., 2022). This method is timely and effective for capturing contemporary urban dynamics and is accessible in a much more comprehensive range of regions. However, most existing studies focus on analyzing and interpreting the density of POIs. These studies measure how many POIs exist in a given area without considering their functional diversity and interaction.

In this study, we employ clustering analysis to investigate urban centers and agglomerations using POI data. We construct a framework that identifies specific spatial patterns of urban centers and agglomerations, along with their integrated functions. Using k-means to cluster grids based on multiple types of POIs, we capture a more holistic picture of urban functions and spatial structure, revealing how different POI types interact to form urban centers or agglomerations. Rather than focusing solely on POI density, we consider how different types of POIs coexist and complement each other. This approach highlights the interplay and complementarity between various urban functions within a given space. By capturing the diversity and relationships among these functions, the framework provides deeper insights into the specific multifunctional characteristics of urban areas.

The strength of our framework lies in its generalizability and the timeliness of the results it generates. Compared to conventional survey data, POI data—categorized as voluntary geographic information—offers broader spatial coverage, more frequent updates, and greater public accessibility. Unlike official top-down planning documents, POI data reflects actual socioeconomic activities, providing a complementary perspective of urban functionality through the lens of voluntary information. Additionally, the findings presented in this paper offer empirical evidence to enhance the understanding of urban internal structures. These insights are instrumental in informing the design of land-use policies, the development of urban transportation systems, and the formulation of comprehensive structural and growth strategies at the city level. It forms the basis for specific spatial planning and policy recommendations aimed at fostering sustainable and equitable urban development.

The remainder of this paper is organized as follows. Section 2 provides a brief literature review. Section 3 demonstrates the methodology and research framework in this paper. Section 4 discusses the case study and our findings. Discussions and Conclusions are presented in Sections 5 and 6.

2. Literature review

2.1. Studies regarding employment centers and agglomerations based on census data

An urban center is a place where different socioeconomic activities agglomerate (Zhu, 2011). Many existing studies have used employment density as an indicator to identify urban centers. They have developed various methods and criteria to identify employment centers at a metropolitan scale using census data (e.g., employment density). Some use clustering methods based on the minimum cutoff point and indicators of employment density (Anderson & Bogart, 2001; Bogart & Ferry, 1999; Forstall & Greene, 1997; Giuliano & Small, 1991). For instance, Giuliano and Small (1991) define an urban subcenter as a continuous area with more than 1000 employees and a minimum density of 10 employees per acre. Others used nonparametric methods (Lee, 2007; McMillen & McDonald, 1997; Redfearn, 2007). For instance, McMillen and McDonald (1997) apply locally weighted regression estimates of employment density in modeling polycentric cities. Unlike the clustering methods that require the specification of density and total employment cutoffs primarily based on "local knowledge", the nonparametric method does not need any specification.

Most studies on agglomeration use a concentration index to represent agglomerations. Ellison and Glaeser (1997) propose indices that measure the geographical concentration of an industry based on the distribution of employment data. Their index reflects the degree of concentration of an industry at an observed scale. Later research (e.g., Alkay & Hewings, 2012; Lu & Tao, 2009; Rosenthal & Strange, 2001) has utilized this index as a key variable to explore the determinants of industrial agglomeration. Analogous to the Ellison-Glaeser index, Ciccone (2002) proposed a model that measures agglomeration effects using data on value-added, employment, and education. Agarwal et al. (2012) integrate the study of urban centers and agglomerations. They use the clustering method to define urban centers in Los Angeles and explore different industrial concentrations in the downtown area, five largest centers, inside centers, outside centers, and consolidated metropolitan statistical areas. The study reveals that service industry activities tend to located outside the centers and agglomeration-sensitive industries such as manufacturing are more clustered inside the centers.

In summary, a substantial body of literature exists on urban centers and agglomerations based on high-resolution, high-quality survey data. However, such data are not available in many regions, which indicates the generalization of this literature is limited.

2.2. Studies regarding urban centers and urban functional zones based on open-source data

"Big data" has opened up a new era in urban studies. Open-source data like POI data is increasingly exploited in urban studies, especially in developing countries where urban data is deficient and difficult to access. Because POI data contains a large volume of geoinformation that can reflect social and economic behavior patterns, much research in developing countries has been proposed recently to explore the valuable information from POI. The bulk of recent research has used the spatial density of POIs as an indicator of urban centers (Deng et al., 2019; Hou & Chen, 2021; Li et al., 2018; Lou et al., 2019; Lu et al., 2020; Xie, Luan, & Xue, 2021). For example, Deng et al. (2019) use a kernel function and density contour tree based on the density of POIs to detect polycentricity. They identify 37 urban centers in Beijing but do not explore the functions of the urban centers. Studies that distinguish urban functions include Hou and Chen (2021), who extract urban commercial centers and clusters based on commercial-type POIs using the Euclidean distance method. Lu et al. (2020), similarly, apply three methods—location entropy analysis, kernel density estimation, and nearest neighbor index—to different types of POIs to explore the spatial structure of Lanzhou, resulting in a series of maps showing each type of functional agglomeration center independently. However, their research only interprets the density of different POIs separately. This method lacks an integrative approach to understanding how various types of POIs interact and co-exist within urban areas. The functional complexity and the interactions between various types of POIs are often overlooked.

Other scholars have used POI data and other data sources, such as remote sensing data and Open Street Map (OSM) data, to detect urban functional zones (Bao et al., 2020; Hu et al., 2016; Hu & Han, 2019; Liu et al., 2017; Wang et al., 2022; Wu et al., 2021; Yao et al., 2017). They have utilized different types of POIs as an important indicator of urban functions. For example, Hu et al. (2016) classify parcels into different land-use types based on the kernel densities of 10 socioeconomic types of POI data and Landsat images. This type of study focuses on the accuracy of land-use classification at a particular unit level (e.g., transportation zone, parcel, building). It does not explore the intensity of agglomeration effects of urban functions citywide.

To summarize, numerous studies have explored urban centers and functional zones using open-source data. However, many focus exclusively on POI density, without considering their functional roles. Studies that do examine functions often analyze them in isolation, overlooking the interactions and coexistence of diverse urban functions. This leaves several critical questions unanswered: How do different combinations of POIs shape the formation and differentiation of urban centers and agglomerations? How can density and POI interactions be integrated to better understand urban spatial structures? What are the multifunctional characteristics of these centers and agglomerations?

To fill this gap in the literature, we apply clustering analysis to examine urban centers and agglomerations in the inner-city using POI data. Our framework identifies specific and detailed spatial patterns of urban centers and agglomerations and their integrated functions. Unlike previous studies that focus solely on POI density, we consider both the combination and co-occurrence of different urban functions within a given grid. By employing k-means clustering on grids based on multiple POI types, we capture a more comprehensive view of urban functions and spatial structure, revealing how various POI types interact to form urban centers or agglomerations. This offers new insights into the formation and structure of urban centers and agglomerations.

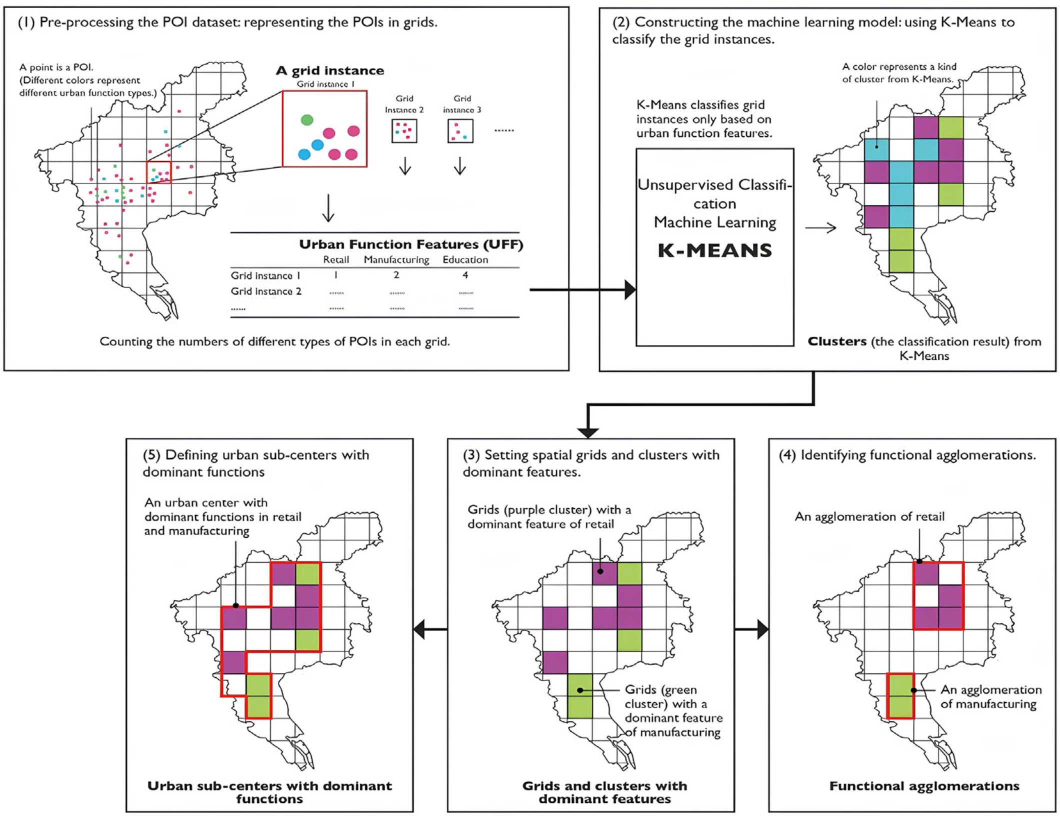

Fig. 1. Core steps of the research framework with a conceptual case.

3. Methodology

3.1. Theoretical framework

In this research, we assume that specific types of economic and social activities tend to agglomerate in a given size area. These areas, whether continuous or not, could share a series of similar activities. We will use open-source POI data and a machine learning approach to identify functional agglomerations and centers with dominant functions. The analytical framework includes five steps: (1) pre-processing the POI dataset; (2) constructing the machine learning model; (3) identifying spatial grids and clusters with dominant functions; (4) identifying functional agglomerations; and (5) defining urban centers with dominant functions. Fig. 1 demonstrates the core steps of our framework with a conceptual case.

3.2. Pre-processing the POI dataset

We mainly utilize Point of Interest (POI) data because of its advantages in reflecting people's daily activities and recognizing different distributions of urban agglomerations (Zhou et al., 2022). The POI is usually a specific point location, indicating diverse social and economic activities. Using different types of POI tags, we can effectively capture spatial characteristics and urban functions in a metropolis. Many map platforms, including Google Maps, OpenStreetMap, and Gaode Map, provide APIs to access the POI dataset for many cities. In this paper, we collect POI data using the Baidu Map API and group them into different urban function types based on the classification tag of the dataset.

We use equally sized grids to represent the POIs and explore the density of each type of POI in different areas. Two types of methods to determine the grid size are considered. One method is based on practical or operational constraints, considering factors beyond the data itself (Tisseyre et al., 2018). Another method is based on the points distribution features. These methods, such as the block kriging and averaging procedure, only consider the data's locational characteristics. Generally, interpolating data on larger grids is helpful for finding differences between spatial zones (Trought & Bramley, 2011). In contrast, interpolating data on smaller grids can capture most of the spatially structured variance but may be susceptible to data noise (Tisseyre et al., 2018). After deciding the grid size, a city can be divided into numerous grids. Each grid contains different numbers of POIs of different urban function types, which are regarded as the grid features. All of these grid features are defined as *Urban Function Features* in our research.

After collecting the urban function features, we clean the training dataset by removing grids without POIs. This accelerates the machine learning process and improves the performance of classification. Next, we proceed to standardize each feature in the training dataset to make all features comparable. The average density of different types of POI in a city can be quite dissimilar, without standardizing features, clustering analysis of agglomeration might result in significant bias. Here, the standardization formula is written as Eq. (1):

$$X' = \frac{X - \mu}{\sigma} \quad (1)$$

Where μ is the mean of the feature value, and σ is the standard deviation of the values of the features. This standardization method is helpful to ensure that each urban function feature has the same weight under the distance-based algorithm *K-means*, which is discussed in the next section.

3.3. Constructing the K-means model

Based on the standardization of the training dataset, this step uses an unsupervised machine learning algorithm (also called clustering algorithm) to identify segmented functional agglomerations. We use the *K-means* algorithm to classify the grids into *k* number of clusters[1] based on their combination of urban function features.

The K-means algorithm has several distinguished advantages in clustering analysis. First, it can quickly classify datasets and achieve convergence in computing. An important step in applying the K-means algorithm is to decide the number of clusters, k. The K-means method initially randomly selects the centroids of these clusters. Each data point is then assigned to the nearest cluster based on its distance to each centroid. This iterative process will not stop until all data points have been assigned and the centroids of the clusters do not change (Géron, 2019, p. 819).

Second, the K-means method performs well in seeking clusters based on multi-dimensional datasets. In this paper, we assume that the clusters of urban function features are globular-shaped agglomerations rather than arbitrarily-shape agglomerations. Fig. 2 is an example of ideal clusters in globular shape (left) and arbitrary shape (right) in three-dimensional space. The left diagram plots five clusters in globular shape in three-dimensional space representing the three urban function features of retail, administration, and education, while the right diagram shows a type of arbitrarily-shaped clusters. Each data point represents a geographical urban grid, plotted according to the value of its retail, administration, and education urban function features. The *K-means* method can also classify multidimensional training datasets, enabling the identification of clusters with similar urban function features based on complex POI data.

[1] "Cluster" in this paper refers to a group of grids with a similar combination of urban function features (a cluster in the urban function features space analyzed by K-means), which is different from the geographical concept of clustering.

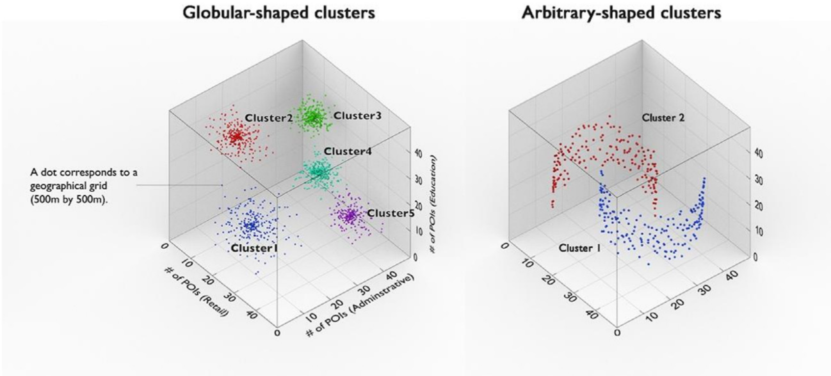

Fig. 2. Examples of ideal clusters in globular shape (left) and arbitrary shape (right) in three-dimensional space.

Before running the *K-means* algorithm, we need to specify the number of clusters, "*k*". Two approaches are used to choose the optimal number of *k*. The first approach is relatively coarse, finding the inflexion point (elbow) on the curve, and plotting the inertia as a function of the number of clusters *k*. Inertia is a performance metric of *K-means*, which is the mean squared distance between each data point and its closest centroid. Another method is more precise and uses the silhouette coefficient (SC) indicator that is calculated for all data points by Eq. (2). In this equation, *a* is the mean distance to other instances in the same cluster, and *b* is the mean nearest-cluster distance. If a SC is close to plus 1 (+1), the data point is well classified in one cluster and far from others. On the other hand, if the SC indicator is close to 0, this indicates that the data point is close to a cluster boundary. Finally, if a coefficient approaches minus 1 (−1), the data point may be wrongly classified. Both the inertia and SC need to be considered together to decide the optimal number of *k*.

$$s = \frac{b - a}{max\ (a, b)} \tag{2}$$

Based on the specified number of *k*, we use the *K-means* clustering analysis to classify the urban grids into different clusters based on urban function features. The clusters are groups of urban grids with a similar combination of urban function features. We use the average value of each urban function feature of all urban grids within a cluster to represent the *urban function features of this cluster*. The urban function features of a cluster are actually the coordinates of the identified cluster centroid in urban function feature space from K-means, which is the average observation of the cluster. And the cluster size is the number of urban grids in the cluster. After classifying urban grids into urban function feature-based clusters, the geographical distribution of each cluster within the urban area can be further analyzed using DBSCAN.

3.4. Identifying the dominant feature(s) in each grid and cluster

As previously indicated, each cluster reflects a particular combination of urban function features. In this paper, we define the dominant feature in a cluster based on whether its urban function feature value surpasses a specified rank threshold among all clusters. For example, if the education and daily life services features of a cluster both exceed this threshold, that cluster is classified as having the dominant features of education and daily life services, indicating a relatively high density of education POIs and daily life services POIs. Each grid within that cluster inherits these dominant features. If a cluster exceeds the threshold for more than one urban function feature, it will be assigned multiple dominant features. Classifying urban grids by their dominant features serves as a first step in analyzing the diverse patterns of spatial agglomeration of different urban functions in the metropolis.

3.5. Identifying agglomerations with different functions

In this paper, an agglomeration with dominant function features refers to a spatial agglomeration of grids with the same dominant function features. As with grids, an agglomeration can have more than one function feature. An agglomeration is defined as containing several grids, characterized by the same dominant features and within a certain threshold of distance. We use the *Density-Based Spatial Clustering of Applications with Noise* (DBSCAN)[2] method to analyze the spatial clustering of grids and specify two parameters to identify agglomerations. One parameter is the minimum number of grids with dominant features required to constitute an agglomeration. The second parameter is the threshold distance, which specifies the maximum distance between grids for them to be considered part of the

[2] DBSCAN is a density-based clustering non-parametric algorithm. Continuous regions with high density of instances are defined as clusters in this algorithm. Consider a set of points in some space to be classified, following is how it works: (1) For every point, DBSCAN counts the numbers of points within a small distance ε from it. This region is called the point's ε-neighborhood. (2) A core point is a point that has at least number of points in its ε-neighborhood. (3) All points in the neighborhood of a core point belong to the same cluster. (4) Any point that is not a core point is regarded as outlier (Géron, 2019).

same agglomeration. When the distance between paired grids exceeds the threshold distance, these two grids cannot be grouped into the same agglomeration. Local circumstances and research aims need to be considered when specifying these two parameters.

This method is a modification of the conventional and influential center definition developed by Giuliano and Small (1991). In their definition, a center consists of a connected group of zones, where each zone surpasses a specified density threshold, D. Collectively, these zones must have a total employment of at least E. Additionally, all zones directly adjacent to the area must have a density lower than D. Zones are considered adjacent if they share at least 0.25 miles of a common boundary. In our study, we applied clusters characterized by dominant urban functions as the criterion for socioeconomic activity density. While Giuliano and Small used total employment as a key measure, we focused on spatial size instead. Therefore, we define an urban center or agglomeration as a spatial cluster consisting of more than a minimum number of grids.

3.6. *Defining urban centers with dominant functions*

An urban center is a place with a high concentration of social and economic activities. Similar to the approach in identifying agglomerations with different functions, we define an urban center as a spatial cluster containing more than a minimum number of grids. However, unlike the approach in identifying agglomerations, the analysis of urban centers looks for spatial clustering of grids with any dominant feature, not just the same dominant features. The dominant features of the urban center are defined as all the dominant features of the grids that comprise it. The dominant functions of a center do not mean the center only has these functions, without any other functions. It only implies the agglomeration effect of the dominant functions is extremely strong in that urban center. Urban centers can be identified via DBSCAN using the same approach used to identify agglomerations.

3.7. *The framework application of Guangzhou*

3.7.1. *The city of Guangzhou*

Guangzhou is one of China's four megacities. It is located in the center of the Pearl River Delta, with an area of 7,400 km^2 and a population of 18 million (2020). It is a polycentric city because originally independent nearby municipalities–Panyu, Huadu, and Zengcheng–were merged into Guangzhou as three administrative districts after 2000. Guangzhou is famous for its new axis planning within the main urban centers. The new axis, which was planned and officially designated in 2000, is where Guangzhou's highest landmark, the Canton Tower, and the Tianhe CBD, are located. Guangzhou possesses diverse industries, a complete set of urban facilities and systems, and a large population, all of which will generate sufficient data and make it easy to uncover the hidden agglomeration pattern of the centers. Guangzhou is a suitable case study for identifying agglomerations and urban centers.

3.7.2. *The POI data collection and pro-processing*

The Gaode Map dataset used in our research consists of 197, 760 POI data points in Guangzhou and covers 14 main types of attributes .[3] As shown in Table 1, we categorize these attributes into eight urban function features. These eight function features will be the supportive indicators to identify the cluster, agglomeration, and urban centers in our discussions. Given that this study aims to detect urban functional areas, we do not investigate residential POI data despite their significant spatial dependency.

As previously described, we standardize the urban function types in the training dataset according to Eq. (1), and drop outlier grids without any POI data from the dataset. This transforms the original 197, 760 POI data points into 5940 grid observations with standardized features.

3.7.3. *Constructing the k-means model*

In Figs. 3 and 4, the inertia and silhouette score methods are plotted for different numbers of k.

As shown in Fig. 5, regardless of the number of clusters specified, each diagram has a primary cluster. Fig. 6 shows the urban function features of the primary cluster for different numbers of clusters, k. Most urban function features (except manufacturing) rank much lower than other clusters, indicating that the POIs in those urban function types are quite sparse. However, this research aims to identify the agglomerations with different functions based on the density of POIs, rather than focusing on the primary cluster. Thus, we exclude the primary cluster and replot the silhouette score curve in Fig. 7.

In all, to achieve an optimal number of k, we must comprehensively consider the inertia, the silhouette score, and the silhouette score excluding the primary cluster. Here, we suggest k be larger than the number of analysis dimensions (eight urban function features). As shown in Figs. 3–7, when k is selected as 10, the silhouette score is relatively higher in the silhouette score curve, diagrams, and the curve excluding the primary class diagram. At the same time, the elbow point of $k = 10$ in the inertia diagram is acceptable at a relatively higher value. Therefore, we determine that the optimal number of k is 10.

[3] The POI dataset for Guangzhou was collected from Baidu Map. The main types of POI data points include transportation facilities, accommodation services, sport and leisure services, public transportation facilities, public facilities, incorporated business, commercial housing, government agency and social organization, life services, education and culture services, shopping services, finance and insurance services, and tourist attraction. Each type is divided into three classifications of major, medium, and minor levels.

Table 1
POI type and represented land use In this study, we divide the administrative boundaries of Guangzhou into 35377 grids of 500m × 500m. There are three rationales for this size of grid. First, the 500m × 500m size is commonly used to divide urban space into grids, due to the high density of POIs in China (Li et al., 2020; Luo et al., 2021; Yang & Diez-Roux, 2012). Second, the 500m × 500m grid is a relatively intuitive distance for understanding the agglomeration patterns and pedestrian behaviors, since 500 m is roughly equivalent to a 5-min walking distance. Third, we compare the 500m × 500m size of grids with other sizes, such as 1000m × 1000m and 250m × 250m. However, the 500 square-meter grid performs better than others in data processing and interpretation.

Original POI Type	Urban Function Type	Number of POIs
Government agencies and social organizations	Public administration	26785
School, educational place	Education	14666
Scientific research institutions	Scientific research	1394
Daily life services	Daily life services	40639
Sport and leisure services	Entertainment	14770
Shopping services, Catering services	Retail & catering	77793
Financial and insurance service	Financial services	12044
Factory	Manufacturing	9669

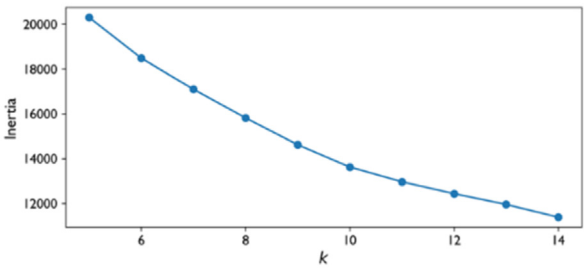

Fig. 3. Inertia.

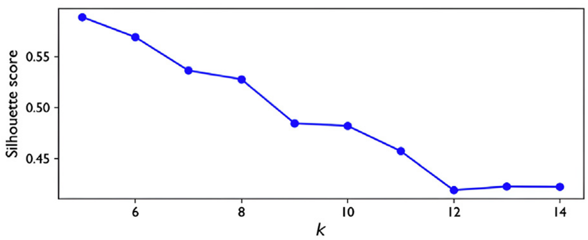

Fig. 4. Silhouette score.

4. Findings

4.1. *Visualizing the results*

Fig. 8 visualizes the 10 clusters of grids with different combinations of urban function features using the *K-means* clustering analysis. In Fig. 9, we visualized the inverted ranking of each urban function feature (the length of the bar) and the feature's rankings among all ten clusters (the text near the end of each bar). The longer the bar is, the higher the urban function feature ranks. For example, the feature of financial services in Cluster 2 has the highest value compared with other features within this cluster. Meanwhile, this feature ranks No. 1 among all 10 clusters. Fig. 9 also reports the number of grids within each cluster and the proportion of grids it accounts for. For instance, *Cluster 2* consists of 32 grids, covering a mere 0.54 percent of the overall study area.

Here, we use the feature's ranking among all clusters rather than using the standardized values directly. One consideration is that the urban function features have been standardized. An alternative consideration is that it would be meaningless to describe the characteristics of each cluster by comparing the numbers of two different featured POIs. For instance, the number of daily life service POIs in a given area can be expected to be much higher than the number of manufacturing POIs (see Table 1). On the other hand, a ranking of each feature allows us to easily interpret the dominant characteristic of each cluster in the urban area.

However, the number of POIs can be useful in comparing one type of urban function feature across clusters. Thus, for further comparison, we scaled back the standardized value of each urban function feature to the original representations and sorted the values from high to low in each urban function type. For instance, *Cluster 2* possesses more than 50 POIs in financial services and ranks No.1 among all clusters (consistent with the result in Fig. 9).

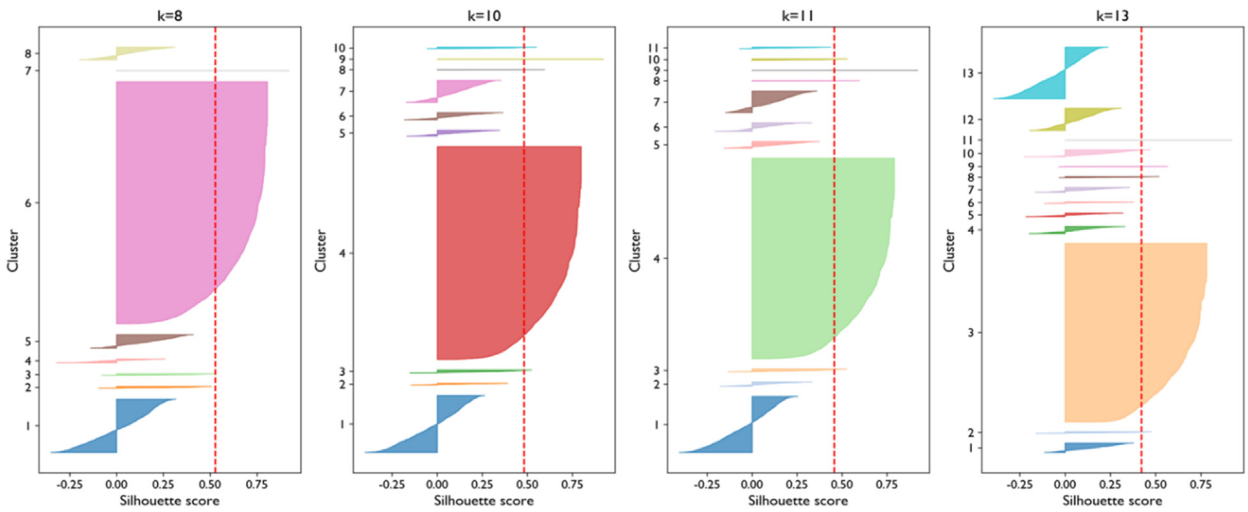

Fig. 5. displays the silhouette coefficients via diagrams in a more informative way. It plots every POI's silhouette coefficient and sorts them by different numbers of clusters. The width of the knife-like shape in the diagram indicates the grid instance's silhouette coefficient in each cluster, while the height denotes the number of grid instances. The dashed line is the mean silhouette coefficient of all grids. If the coefficient of most grid instances in a cluster is shorter than the red dashed line, this indicates that the grid instances in this cluster are close to other clusters and the cluster is bad (Géron, 2019, p. 819).

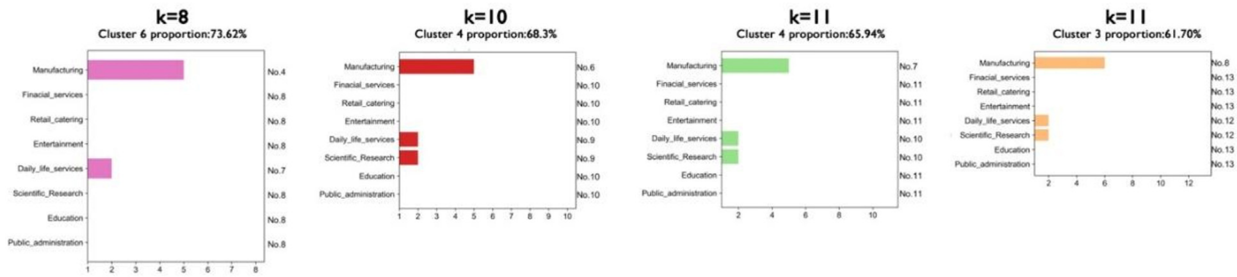

Fig. 6. The centroid value of the largest cluster with different values of k.

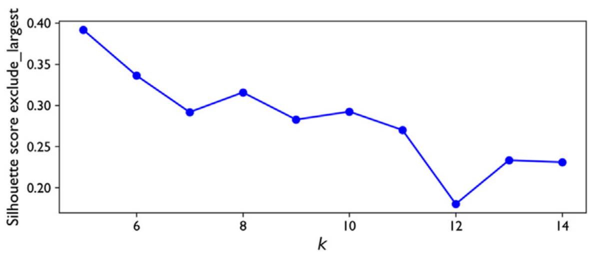

Fig. 7. The silhouette score excluding the primary class.

4.2. *Identifying grids and clusters with a dominant feature*

In this study, we set the rank threshold at second place. Any urban function feature ranking first or second among all clusters is considered a dominant feature. The threshold is set at second place to ensure a sufficient number of samples classified under each dominant feature. Restricting the threshold to only first place would yield too few samples for certain urban functions, limiting our ability to capture meaningful variations across space. By including features ranked up to second place, we achieve a better balance between specificity and representativeness.

In Fig. 9, we can easily recognize that some clusters are characterized by different dominant features. *Cluster 6* is characterized by the dominant feature of public administration. *Cluster 5* stands out in four dominant features—financial services, daily life services, entertainment, and retail catering. Similarly, the dominant features of *Cluster 2* are financial services, entertainment, public

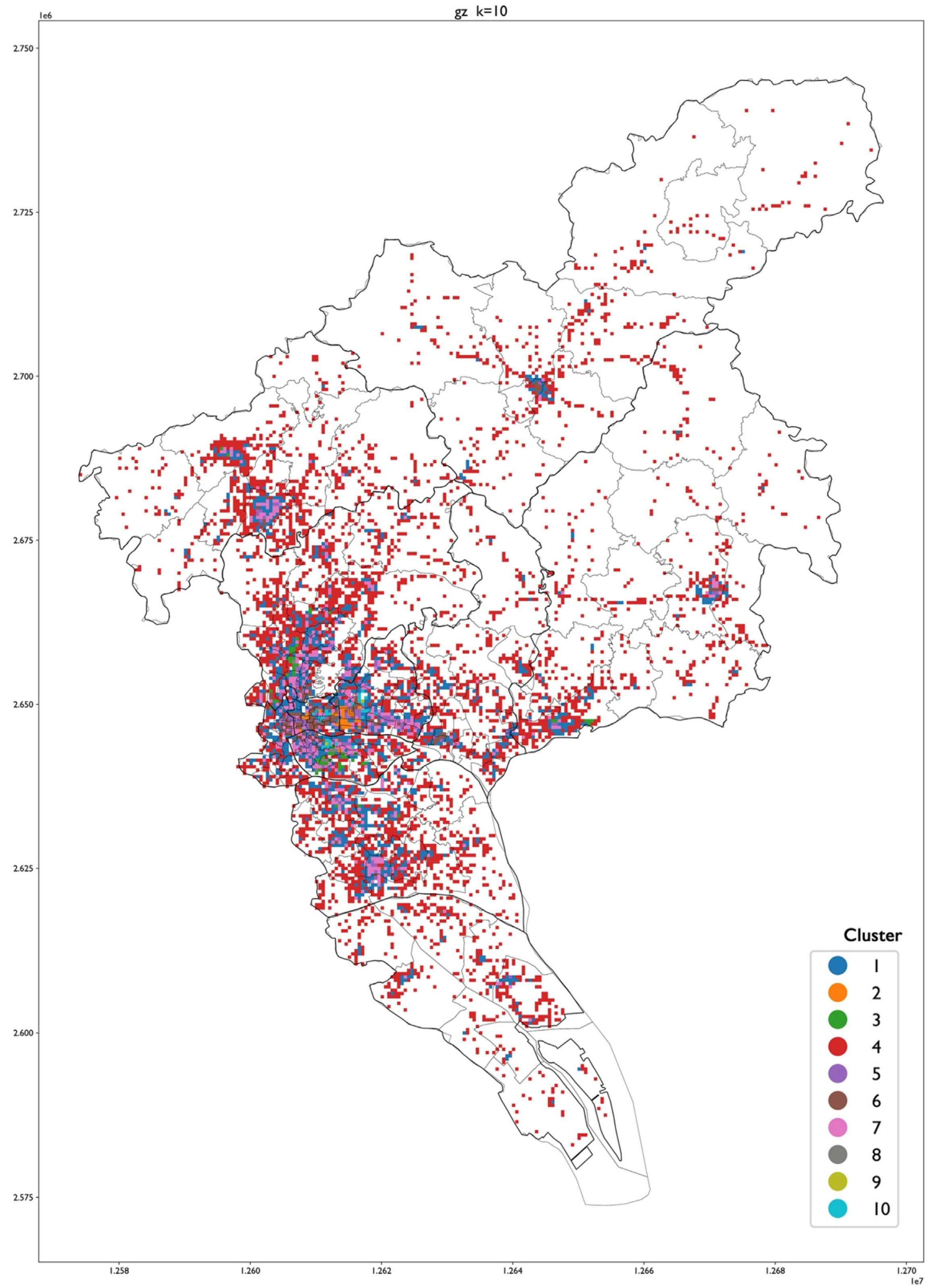

Fig. 8. The geographical distribution of grids of 10 clusters in Guangzhou.

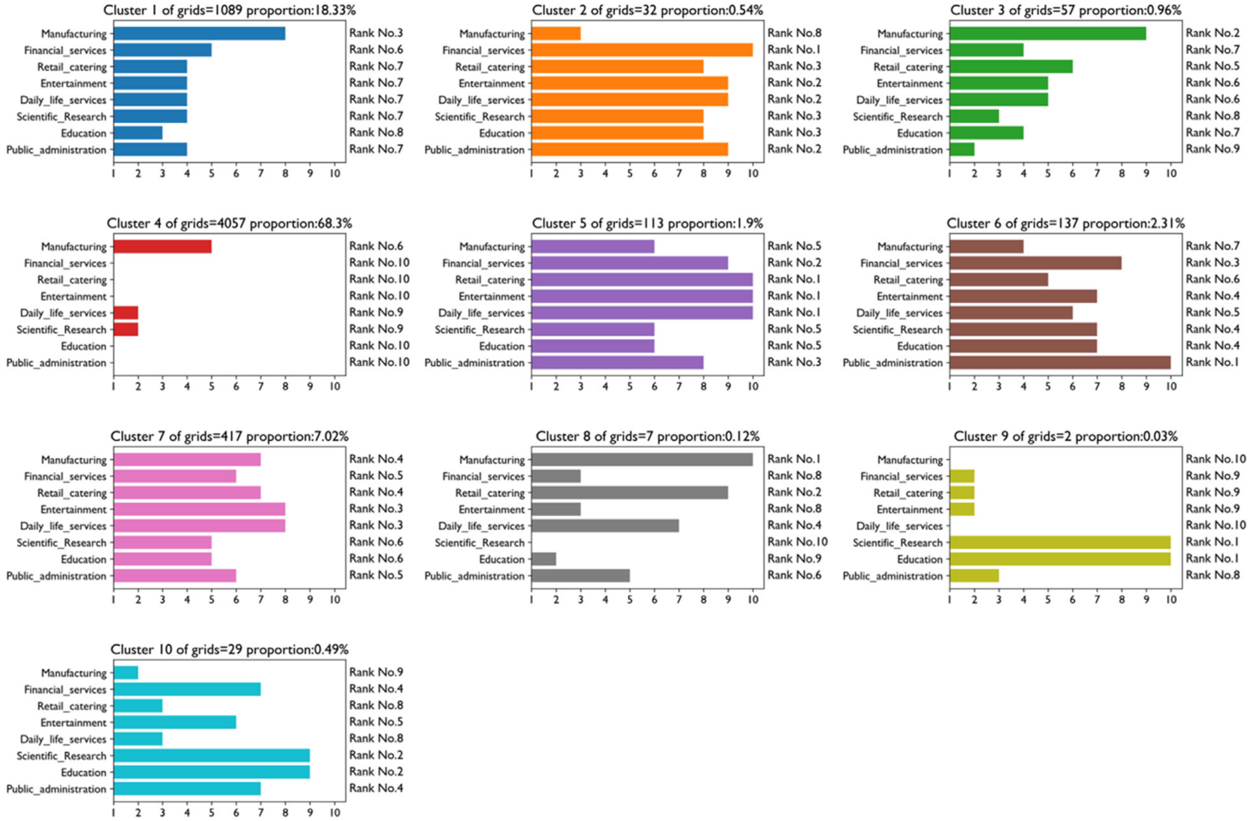

Fig. 9. The rankings of urban function features of ten clusters.

administration, and daily life services. *Cluster 8* is characterized by a high concentration of manufacturing and retail catering, while *Cluster 3* is dominant in manufacturing. *Clusters 10* and *9* are dominated by the function features of education and scientific research. Notably, while each cluster exhibits dominant features, they also reveal multifunctional characteristics, highlighting the variety and relationship of urban functions within space.

Meanwhile, *Clusters 1, 4,* and *7* do not have any dominant features. This indicates that these clustered areas do not present significant agglomerations of urban function features. Hence, we do not focus on these clusters in the coming discussions. The remaining clusters, while not as prominent in urban functions, are nonetheless meaningful. They reflect the spatial interactions of urban functions in other areas and would be valuable in studies focusing on multifunctional features.

4.3. Identifying agglomerations with different functions

Based on the identification of dominantly featured clusters and grids, we proceed to explore their geographical agglomerations. Our results suggest that most of the dominantly featured grids are concentrated in the recognized main urban area of Guangzhou. Fig. 10 plots the grids with dominant features in the recognized main urban area. As shown in Fig. 10, most of the featured grids across different clusters (e.g., public administration, daily life services, entertainment, retail catering, financial services, education, research) are agglomerated in the core, but certain types of grids (e.g., manufacturing) agglomerate in the outer areas.

Furthermore, we quantify the agglomeration of grids with dominant feature(s) using the DBSCAN algorithm. In this study, we define an agglomeration as a set of continuous grids. If grids are not connected at the vertex or boundary, they should belong to different agglomerations. We use the coordinates of each 500m × 500m grid center to conduct the DBSCAN analysis, where the maximum distance (ε) between the center of paired grids is about 710 m and the minimum size of an agglomeration is two grids (0.5 km^2).

As summarized in Table 2, we identified a total of 49 agglomerations citywide. Similar to the dominantly featured grids, most of the agglomerations are concentrated together in the recognized urban area. Fig. 11 plots the agglomerations on the map of the main urban area. Four agglomerations of financial services, entertainment, and daily life services, marked as *FEDP1-FEDP4*, are mainly located in the city core. The largest agglomeration of *FEDP1* is located in the *Tianhe* CBD,[4] consistent with both observed realities and city planning

[4] The Tianhe CBD is one of three national-level CBDs in Guangzhou. It is a government-supported CBD nominated by the State Council in 1990s (Long, 2020).

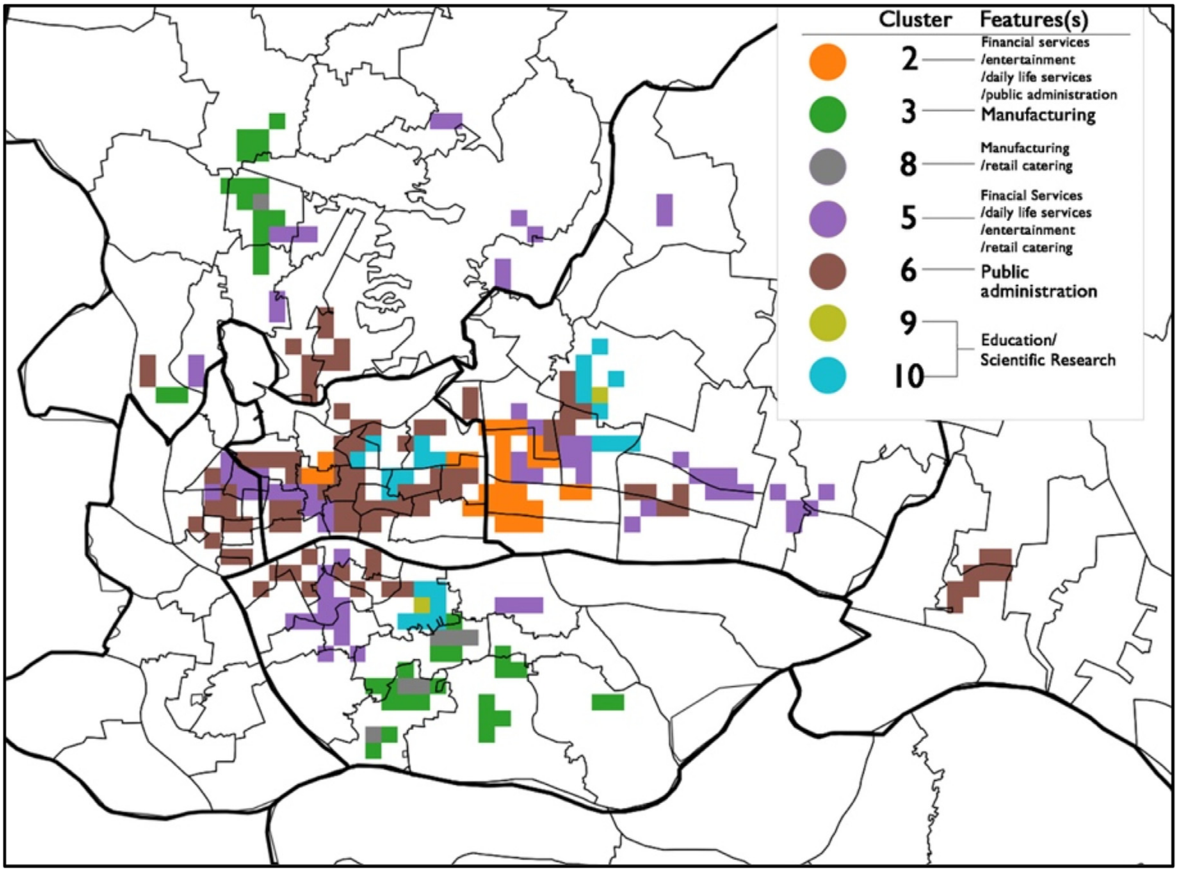

Fig. 10. Functional clusters in the main urban area of Guangzhou.

Table 2
Spatial characteristics of different functional agglomerations of Guangzhou (ε = 710m, minimum size = 2 grids).

Urban Function(s)	Number of agglomerations	Total area (km^2)
Public administration	16	26.25
Education/Scientific research	5	6.25
Financial services/Daily life services/Entertainment/Retail catering/Public administration	24	21.25
Financial services/Daily life services/Entertainment	4	6.75
Manufacturing	12	11
Manufacturing/Retail catering	2	1.25

documentation. This area boasts the most accessible transportation network and the highest urban land value, alongside substantial support from local government initiatives to attract financial industries. For this type of agglomeration, financial services rank highest in prominence among other clusters, followed by entertainment and daily life services. This highlights the multifunctional characteristics of agglomeration within the CBD.

Five agglomerations (*ES1-ES5*) of education and scientific research (e.g., universities, schools, research institutes, etc.) are also located in the city center. *ES1* and *ES3* are present in several universities; however, these agglomerations are absent in the southeastern university town. One explanation is that the educational POIs, which include elementary schools, are more densely concentrated in the urban center. Regarding other urban functions in this type of agglomeration, they rank relatively low, indicating that education and scientific research are the only functions.

Additionally, nine agglomerations of manufacturing (*M1-M9*) are mainly located in the outer areas as shown in Fig. 11. One potential shortcoming of this identification method for manufacturing POI in particular is our method is effectively based on the number of factories rather than their size, which can differ dramatically. Therefore, this identification method may miss important manufacturing agglomerations that consist of a small number of very large factories. However, it still serves to identify agglomerations with a large number of factories effectively. In addition, this kind of agglomeration only has one urban function, which means other urban functions do not tend to coexist with manufacturing.

Relative to other urban function features, agglomerations of public administration (*P1-P13*), financial services, daily life services,

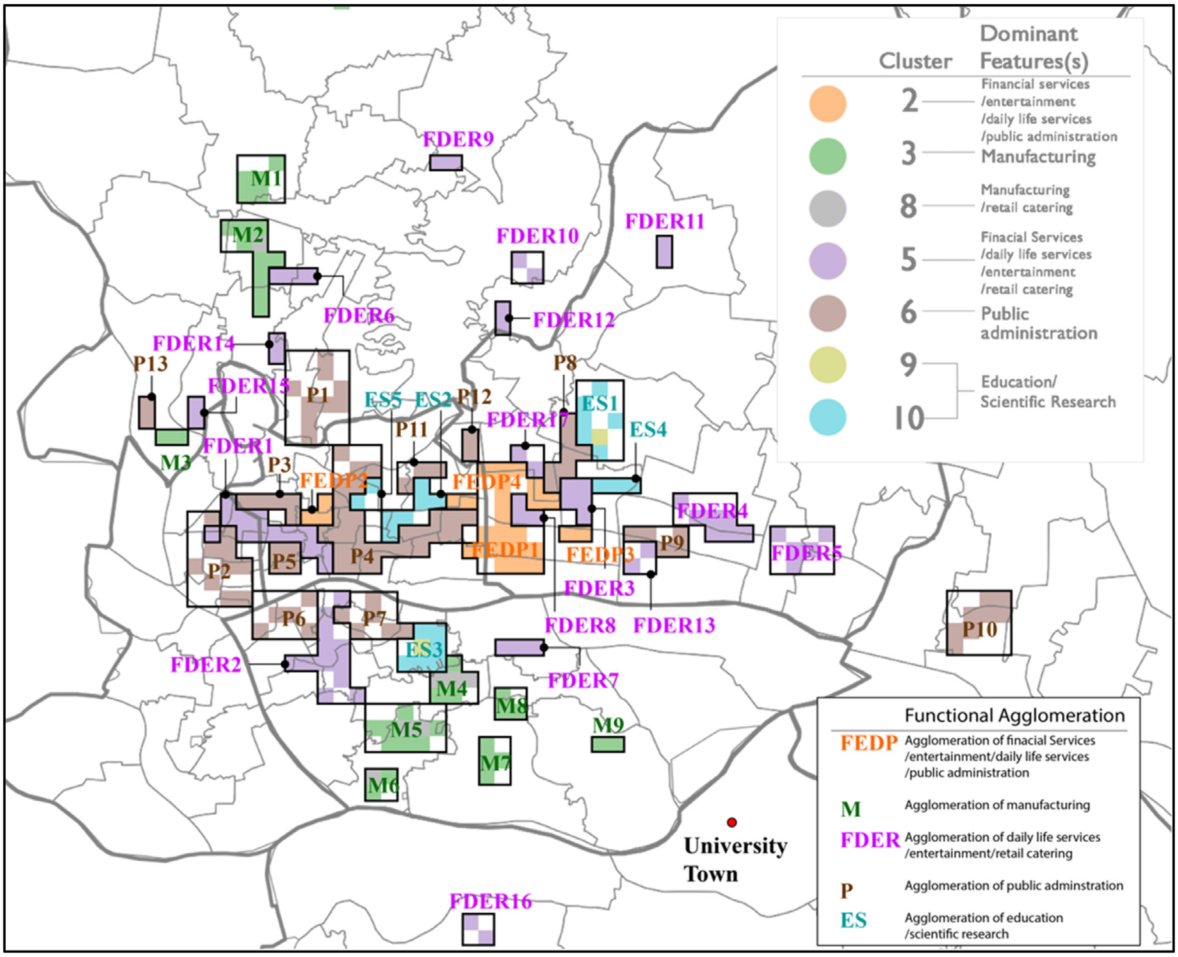

Fig. 11. Agglomerations with different functions in the main urban area of Guangzhou.

entertainment, and retail catering (*FDER1- FDER 17*) are relatively scattered, although most are located in the city center. Guangzhou as the capital of Guangdong province is home to the local and provincial governments and many governmental institutions, and the largest agglomeration of public administration is found in the Yuexiu district, which concentrates most of these institutions, although other agglomerations are dispersed. Meanwhile, agglomerations of services, entertainment, and retail catering are dispersed around the city.

4.4. Identifying urban centers with dominant functions

We proceed to identify urban centers with dominant functions using the DBSCAN method. Considering that centers are usually larger than agglomerations, we reset two DBSCAN parameters. Here, ε equals 1420 m and the minimum size of a center is 1 km^2, consisting of 4 grids. As reported in Table 3, we identify 11 centers within the Guangzhou urban area, including one main center (Center id 1) and 10 sub-centers, each characterized by distinct dominant functions. Fig. 12 compares the identified centers with the latest Guangzhou urban planning document. All the centers we identified align with *the Urban Plan 2018–2035*, except for three newly designated centers in the plan—*the Airport Economic Area, Knowledge Town,* and *Nansha*. Because these centers have only recently been established and are still undergoing development, their socioeconomic activities have not yet clustered sufficiently to be detected through our POI data analysis. As these areas continue to expand, we anticipate that future data will reflect their emerging status more clearly.

In addition, referring to Table 3 and Fig. 12*,* we can further interpret the characteristics of different centers. The main urban center has diverse urban functions, including all defined urban functions. Six of the other centers far from the main urban area are dominated by the functions of financial services, daily life services, entertainment, retail catering, and public administration. The functions of education and scientific research are only featured in the main urban center.

5. Implications

At the city level, the main urban center covers the largest areas (252 km^2) as well as includes all nine defined urban functions, which

Table 3
Characteristics of identified urban centers.

Center id	Urban Functions	Center area (km^2)
1	Education/Scientific research/Public administration/Financial services/Retail Catering/Daily life services/Manufacturing/Retail catering/Entertainment	252
2	Public administration/Financial services/Daily life services/Manufacturing/Retail catering/Entertainment	10
3	Manufacturing	4
4	Financial services/Daily life services/Manufacturing/Retail catering/Entertainment	20
5	Financial services/Daily life services/Manufacturing/Retail catering/Entertainment	4
6	Public administration/Financial services/Daily life services/Retail catering/Entertainment	8
7	Public administration/Financial services/Daily life services/Retail catering/Entertainment	8
8	Public administration/Financial services/Daily life services/Retail catering/Entertainment	10
9	Public administration/Financial services/Daily life services/Retail catering/Entertainment	7
10	Public administration/Financial services/Daily life services/Retail catering/Entertainment	6
11	Public administration/Financial services/Daily life services/Retail catering/Entertainment	4

indicates that this area serves as a multi-functional urban core (see Table 3). The city's land use policies should continue to support such functional diversity, ensuring that high-value urban land is efficiently utilized for numerous services that promote economic growth and social vitality. However, the findings also reveal the other ten sub-centers primarily serve functions related to financial services, daily life services, entertainment, retail, and public administration. Notably, no sub-centers were identified with a focus on education or scientific research. Manufacturing functions even form their own distinct sub-center. There is also one sub-center that is exclusive in manufacturing with an area of 4 km2 in Xintang. To improve the functional balance of these sub-centers, related policies could encourage the introduction of additional functions, particularly education and research institutions, to diversify these areas and balance educational resources.

At the agglomeration level, the manufacturing agglomerations are predominantly located in the outer areas of the city, suggesting a spatial separation of industrial activities from the central urban functions. Given that these agglomerations are singular in function and lack coexistence with other urban services, land use policies should ensure these zones remain well-supported by infrastructure that meets their specific needs, including logistics, utilities, and worker housing. Additionally, the policy could also consider the diversification of the function of these agglomerations, which may contribute to the industrial equipment upgrading.

The research underscores the strong relationship between the accessibility of public transit and the distribution of daily life services, entertainment, and retail. These agglomerations cluster near subway stations, reflecting their reliance on accessibility, especially within walking distance from public transportation. As such, transportation policy should prioritize expanding the public transit network to support the functional development of both the city center and the sub-centers. Improving public transit access in the outer sub-centers could further enhance their attractiveness for both residential and commercial developments, fostering greater connectivity between the city core and peripheral areas (Zhu, Li, Wang, & Huang, 2024). Policies could also explore transit-oriented development (TOD) strategies, ensuring that future transportation hubs are planned alongside diverse urban functions to promote mixed-use environments and reduce the need for long commutes.

6. Conclusions and limitations

In this study, we propose a novel framework to identify specific spatial patterns of urban centers and agglomerations, along with their integrated functions. Using k-means clustering on grids based on multiple types of POIs, we provide a comprehensive view of urban functions and spatial structures, uncovering how different POI types interact to form urban centers or agglomerations. Since POI data is derived from volunteered geographic information, it offers a perspective on urban structure that complements traditional top-down official planning. Unlike other approaches that focus solely on POI density, our framework emphasizes the coexistence and complementarity of different POI types to capture the multifunctional features of urban centers and agglomerations. Furthermore, the framework defines three levels of concentration—clusters with dominant functions, functional agglomerations, and urban centers—offering a multi-scalar perspective on the phenomenon of urban function agglomeration. This nuanced approach provides deeper insights into the spatial and functional dynamics of urban environments.

We apply this framework to Guangzhou and explored nine urban functions: education, scientific research, public administration, financial services, daily life services, manufacturing, retail catering, and entertainment. We first got 10 clusters, seven of which exhibited dominant functions. Next, we further defined 63 agglomerations and 11 centers in Guangzhou at multiple scales. At the city-center level, our findings are largely consistent with the *Guangzhou Urban Development Plan 2018–2035*. All identified centers align with those in the official planning document, except for three newly emerging centers. These discrepancies likely stem from insufficient socioeconomic activity, as indicated by POI data, suggesting that our framework provides a valuable complement to top-down planning. Notably, the primary center, encompassing an area of 252 km^2, supports all major urban functions. In contrast, six peripheral centers are functionally specialized, predominantly oriented around financial services, daily life services, entertainment, retail catering, and public administration.

At the agglomeration level, financial services, entertainment, daily life services, and public administration tend to coexist in the Tianhe CBD. The public administration agglomerations are concentrated in the older Yuexiu district. Educational and scientific research functions, while also clustered in agglomerations within the city core, tend to form separate, more specialized zones, without significant

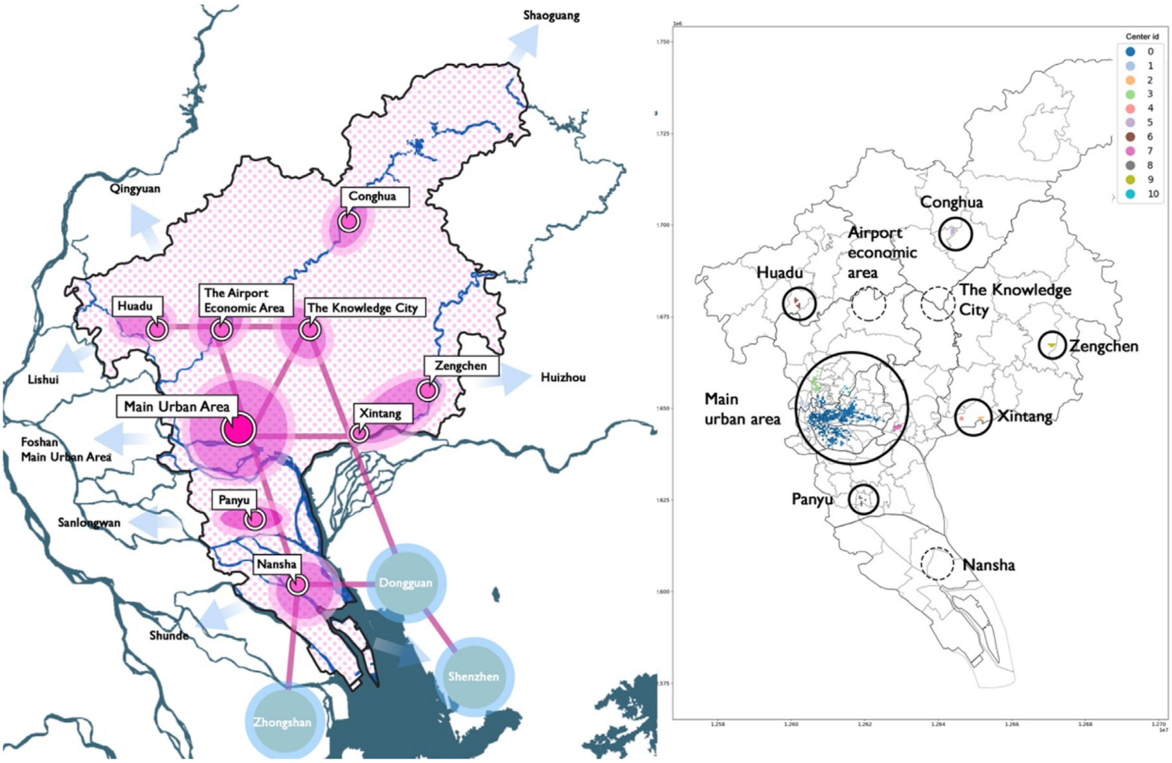

Fig. 12. Comparison between Guangzhou Master Plan (2018–2035) and our estimated urban centers. N*ote.* The Guangzhou Master Plan (left map) is our own drawing based on the original map collected from the Guangzhou Municipal People's Government (2018)

overlap with other urban functions such as financial services or retail. Interestingly, manufacturing agglomerations are mainly located on the outskirts and tend to be functionally singular, with few other urban functions coexisting in these areas. The results of this study suggest several key policy directions for optimizing land use and transportation planning in Guangzhou.

While this study lays out a new framework for analyzing urban polycentricity and provides a case study to demonstrate its application, there is still much need for further extension and testing of our framework. First, this framework can be extended through application to a large number of other cities, including many cities in developing countries where in the past research on urban polycentricity was difficult due to the lack of suitable data. The reliance of our framework on widely available open-source POI data makes it highly generalizable, and future studies can further refine the method by testing it on other cities. Secondly, future research might test the performance of other sizes or shapes of basic spatial units for identifying different types of agglomerations. For instance, by combining our framework with traffic data, it would be possible to use traffic zones instead of grids, which may be more in line with the urban fabric. Thirdly, different parameters for identifying agglomerations and urban sub-centers using DBSCAN can be tested. Different research focuses may demand different specifications for the distance between paired grids and a minimum number of grids. Finally, other methods may be developed to address shortcomings of the POI dataset, due to the fact that the POI data points only record location information (coordinates and function types) but do not include the scale of each POI, such as the useable size of a manufacturing factory or the number of employees. Although this problem has a limited impact on the identification of functional agglomerations and subcenters in our research, future studies may develop supplementary methods to address this shortcoming and to better identify urban agglomerations for urban function features that involve significant variation in the size and characteristics of each POI.

Glossary of Terms and Definitions

- Agglomeration with different functions – an agglomeration with dominant function features refers to a spatial agglomeration of grids with the same dominant function features. An agglomeration is defined as containing several grids, characterized by the same dominant features and within a certain threshold of distance
- Cluster with Dominant feature(s) – we define the dominant feature in a cluster based on whether this urban function feature of the cluster is distinguished from that in other clusters by ranking the feature value. A cluster can contain more than one dominant feature
- Clusters from K-Means – it refers to the cluster in urban function features space but not the geographical cluster
- Dimension of K-Means – it is the dimension space of the urban features. It does not refer to any geographical dimension
  Grid Instance – An instance of K-Means in this research is a grid with urban function features.
- Grids with dominant feature(s) – Each of the grids within a cluster is defined as having the dominant features of the cluster.

- Urban center with dominant functions – the research defines an urban center as gathering grids containing more than a minimum number of grids with any dominant feature within a certain distance. The dominant features of grids the urban center contains are the urban center's dominant functions. The dominant functions of a center do not mean the center only has these functions but without any other functions. It only implies the agglomeration effect of the dominant functions is extremely strong in the center.
- Urban function features (urban function features) –urban function features are the urban function types of POI. The initial urban function feature values are the numbers of POI in urban function types. The values are then standardized before running the K-Means.
- Urban function features of a cluster –We use the average value of each urban function feature of all urban grids within a cluster to represent the urban function features of this cluster. The features of a cluster are actually the coordinates of identified cluster centroid in urban function feature space from K-means.

CRediT authorship contribution statement

Pengyu Zhu: Writing – review & editing, Supervision, Methodology, Conceptualization. **Jianqi Li:** Writing – original draft, Methodology, Formal analysis. **Zining Wang:** Writing – review & editing.

Declaration of compenting interest

All authors declare that they have no conflicts of interest.

Acknowledgement

This work was supported by the Hong Kong Research Grants Council (RGC) Research Fellow Scheme (RFS) [grant number HKUST RFS2425-6H03]; the Hong Kong Public Policy Research (PPR) Funding Scheme [grant number 2024.A7.032.24B]; the Inaugural CLP Research Fellowship Programme Award [grant number RFP2022HKUST]; the Innovation and Technology Support Programme, Hong Kong SAR Government [grant number ITS/003/24SC]; and the Hong Kong Research Grants Council (RGC) Strategic Topics Grant (STG) [grant number STG2/E- 605/23-N].

References

Agarwal, A., Giuliano, G., & Redfearn, C. L. (2012). Strangers in our midst: The usefulness of exploring polycentricity. *The Annals of Regional Science, 48*(2), 433–450.
Alkay, E., & Hewings, G. J. D. (2012). The determinants of agglomeration for the manufacturing sector in the Istanbul metropolitan area. *The Annals of Regional Science, 48*(1), 225–245.
Anderson, N. B., & Bogart, W. T. (2001). The structure of sprawl: Identifying and characterizing employment centers in polycentric metropolitan areas. *The American Journal of Economics and Sociology, 60*(1), 147–169.
Bao, H., Ming, D., Guo, Y., Zhang, K., Zhou, K., & Du, S. (2020). DFCNN-based semantic recognition of urban functional zones by integrating remote sensing data and POI data. *Remote Sensing, 12*(7), 1088.
Bogart, W. T., & Ferry, W. C. (1999). Employment centres in greater Cleveland: Evidence of evolution in a formerly monocentric city. *Urban Studies, 36*(12), 2099–2110.
Ciccone, A. (2002). Agglomeration effects in Europe. *European Economic Review, 46*(2), 213–227.
Deng, Y., Liu, J., Liu, Y., & Luo, A. (2019). Detecting urban polycentric structure from POI data. *ISPRS International Journal of Geo-Information, 8*(6), 283.
Ellison, G., & Glaeser, E. L. (1997). Geographic concentration in US manufacturing industries: A dartboard approach. *Journal of Political Economy, 105*(5), 889–927.
Forstall, R. L., & Greene, R. P. (1997). Defining job concentrations: the Los Angeles case. *Urban Geography, 18*(8), 705–739.
Garreau, J. (2011). *Edge city: Life on the new frontier*. Anchor.
Géron, A. (2019). *Hands-on machine learning with scikit-learn, keras, and TensorFlow : Concepts, tools, and techniques to build intelligent systems*. xxv.
Giuliano, G., & Small, K. A. (1991). Subcenters in the Los Angeles region. *Regional Science and Urban Economics, 21*(2), 163–182.
Giuliano, G., & Small, K. A. (1999). The determinants of growth of employment subcenters. *Journal of Transport Geography, 7*(3), 189–201.
Gordon, P., Richardson, H. W., & Wong, H. L. (1986). The distribution of population and employment in a polycentric city: The case of Los Angeles. *Environment and planning A, 18*(2), 161–173.
Guangzhou Municipal People's Government. (2018). Urban centers structure. *Guangzhou master plan* (pp. 2018–2020).
Hou, G., & Chen, L. (2021). Regional commercial center identification based on POI big data in China. *Arabian Journal of Geosciences, 14*(14), 1–14.
Hu, Y., & Han, Y. (2019). Identification of urban functional areas based on POI data: A case study of the Guangzhou economic and technological development zone. *Sustainability, 11*(5), 1385.
Hu, T., Yang, J., Li, X., & Gong, P. (2016). Mapping urban land use by using landsat images and open social data. *Remote Sensing, 8*(2), 151.
Lee, B. (2007). "Edge" or "edgeless" cities? Urban spatial structure in US metropolitan areas, 1980 to 2000. *Journal of Regional Science, 47*(3), 479–515.
Li, J., Long, Y., & Dang, A. (2018). Live-Work-Play Centers of Chinese cities: Identification and temporal evolution with emerging data. *Computers, Environment and Urban Systems, 71*, 58–66.
Li, S., Zhou, J., Tong, X., Liu, H., Lu, X., & Xiong, H. (2020). Competitive analysis for points of interest. *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining*.
Liu, X., He, J., Yao, Y., Zhang, J., Liang, H., Wang, H., & Hong, Y. (2017). Classifying urban land use by integrating remote sensing and social media data. *International Journal of Geographical Information Science, 31*(8), 1675–1696.
Lou, G., Chen, Q., He, K., Zhou, Y., & Shi, Z. (2019). Using nighttime light data and POI big data to detect the urban centers of Hangzhou. *Remote Sensing, 11*(15), 1821.
Lu, C., Pang, M., Zhang, Y., Li, H., Lu, C., Tang, X., & Cheng, W. (2020). Mapping urban spatial structure based on POI (point of interest) data: A case study of the central city of Lanzhou, China. *ISPRS International Journal of Geo-Information, 9*(2), 92.
Lu, J., & Tao, Z. (2009). Trends and determinants of China's industrial agglomeration. *Journal of Urban Economics, 65*(2), 167–180.
Luo, S., Liu, Y., Du, M., Gao, S., Wang, P., & Liu, X. (2021). The influence of spatial grid division on the layout analysis of urban functional areas. *ISPRS International Journal of Geo-Information, 10*(3), 189.
McDonald, J. F. (1987). The identification of urban employment subcenters. *Journal of Urban Economics, 21*(2), 242–258.
McMillen, D. P., & McDonald, J. F. (1997). A nonparametric analysis of employment density in a polycentric city. *Journal of Regional Science, 37*(4), 591–612.
Redfearn, C. L. (2007). The topography of metropolitan employment: Identifying centers of employment in a polycentric urban area. *Journal of Urban Economics, 61*(3), 519–541.

Rosenthal, S. S., & Strange, W. C. (2001). The determinants of agglomeration. *Journal of Urban Economics, 50*(2), 191–229.

Song, S. (1994). Modelling worker residence distribution in the Los Angeles region. *Urban Studies, 31*(9), 1533–1544.

Tisseyre, B., Leroux, C., Pichon, L., Geraudie, V., & Sari, T. (2018). How to define the optimal grid size to map high resolution spatial data? *Precision Agriculture, 19*(5), 957–971.

Trought, M. C. T., & Bramley, R. G. V. (2011). Vineyard variability in marlborough, New Zealand: Characterising spatial and temporal changes in fruit composition and juice quality in the vineyard. *Australian Journal of Grape and Wine Research, 17*(1), 79–89.

Wang, M., Chen, Z., Rong, H. H., Mu, L., Zhu, P., & Shi, Z. (2022). Ridesharing accessibility from the human eye: Spatial modeling of built environment with street-level images. *Computers, Environment and Urban Systems, 97*, Article 101858.

Wu, R., Wang, J., Zhang, D., & Wang, S. (2021). Identifying different types of urban land use dynamics using Point-of-interest (POI) and Random Forest algorithm: The case of Huizhou, China. *Cities, 114*, Article 103202.

Xie, X., Luan, X., & Xue, Y. (2021). Research on city center identification and optimization strategy based on POI data——taking the four districts of qingdao as an example. *Journal of Physics: Conference Series,1955*, Article 012016.

Yang, Y., & Diez-Roux, A. V. (2012). Walking distance by trip purpose and population subgroups. *American Journal of Preventive Medicine, 43*(1), 11–19.

Yao, Y., Li, X., Liu, X., Liu, P., Liang, Z., Zhang, J., & Mai, K. (2017). Sensing spatial distribution of urban land use by integrating points-of-interest and Google Word2Vec model. *International Journal of Geographical Information Science, 31*(4), 825–848.

Zhou, L., Li, S., Li, C., Shen, G., Yang, H., Zhu, P., Han, H., & Li, B. (2022). Spatial congruency or mismatch? Analyzing the COVID-19 potential infection risk and urban density as businesses reopen. *Cities, 123*, Article 103615.

Zhu, P. (2011). *Telecommuting, travel behavior and residential location choice: Can telecommuting be an effective policy to reduce travel demand? Dissertation*. University of Southern California.

Zhu, P., & Guo, Y. (2022). Telecommuting and trip chaining: Pre-pandemic patterns and implications for the post-pandemic world. *Transportation Research Part D: Transport and Environment, 113*, Article 103524.

Zhu, P., Huang, J., Wang, J., Liu, Y., Li, J., Wang, M., & Qiang, W. (2022). Understanding taxi ridership with spatial spillover effects and temporal dynamics. *Cities, 125*, Article 103637.

Zhu, P., Li, J., Wang, K., & Huang, J. (2024). Exploring spatial heterogeneity in the impact of built environment on taxi ridership using multiscale geographically weighted regression. *Transportation, 51*(5), 1963–1997.