Article

Urban Analytics and City Science

EPB: Urban Analytics and City Science
0(0) 1–18

DOI: 10.1177/2399808321996705
journals.sagepub.com/home/epb
SAGE

Identifying subcenters with a nonparametric method and ubiquitous point-of-interest data: A case study of 284 Chinese cities

Ying Long
School of Architecture and Hang Lung Center for Real Estate, Key Laboratory of Eco Planning & Green Building, Ministry of Education, Tsinghua University, China

Yimeng Song
Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University; Smart Cities Research Institute, The Hong Kong Polytechnic University, Hong Kong, China

Long Chen
Tsinghua University, China

Abstract

Urban spatial structure, which is primarily defined as the spatial distribution of employment and residences, has been of lasting interest to urban economists, geographers, and planners for good reason. This paper proposes a nonparametric method that combines the Jenks natural break method and the Moran's *I* to identify a city's polycentric structure using point-of-interest density. Specifically, a polycentric city consists of one main center and at least one subcenter. A qualified (sub)center should have a significantly higher density of human activity than its immediate surroundings (locally high) and a relatively higher density than all the other subareas in the city (globally high). Treating Chinese cities as the subject, we ultimately identified 70 cities with polycentric structures from 284 prefecture-level cities in China. In addition, regression analyses were conducted to reveal the predictors of polycentricity among the subjects. The regression results indicate that the total population, GDP, average wage, and urban land area of a city all significantly predict polycentricity. As a whole, this paper provides an alternative and transferrable

Corresponding author:
Yimeng Song, Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hong Kong, China.
Email: yimsong@polyu.edu.hk

method for identifying main centers and subcenters across cities and to reveal common predictors of polycentricity. The proposed method avoids some of the potential problems in the conventional approach, such as the arbitrariness of threshold setting and sensitivity to spatial scales. It can also be replicated rather conveniently, as its input data, such as point-of-interest data, are widely available to the public and the data's validity can be efficiently checked by field trips or other traditional data sources, such as land-use maps or censuses.

Keywords
Urban spatial structure, subcenter, point-of-interest, China

Introduction

Urban spatial structure, which is primarily defined as the spatial distribution of employment and residences, has been of lasting interest to urban economists, geographers, and planners for good reason. On the one hand, a compact urban form (i.e. high concentration of employment and residences) or urban sprawl (i.e. the reverse of a compact urban form) means different benefits and costs to cities. On the other hand, decentralization and polycentric forms, where there are a number of concentrated employment centers and a scattered population distribution, tend to be the norm rather than the exception in the characterization of an increasing number of cities. In the United States, for instance, "edge cities" and even regions "beyond polycentricity" have emerged (Garreau, 1992; Gordon and Richardson, 1996). These new phenomena have profound economic, social, energy, and environmental implications and have also revealed that the underlying dynamics that govern our cities or regions are undergoing great changes. Thus, both academics and practitioners need additional studies to better understand these changes; as such understanding has become a prerequisite for us to better plan and manage the forms and development of cities and regions.

Existing studies have made progress in understanding the changes in urban spatial structure, basically by defining and identifying urban subcenters. However, studies that quantitatively measure polycentricity are dominated by (a) measurements that require local knowledge or multistep computation, such as the selection of density thresholds and the calculation of density functions and residuals; (b) measurements using conventional employment or population data from census and household travel surveys, which are often updated every 5 or 10 years, significantly lagging behind the rapid development and constant changes in our cities and regions; or (c) a focus on only one or a limited number of cities simultaneously, such as Los Angeles (e.g. Giuliano and Small, 1991; Pfister et al., 2000), Chicago (e.g. McDonald, 1987; McMillen and McDonald, 1998), Houston (e.g. Craig and Ng, 2001; McMillen, 2001), Beijing (e.g. Huang et al., 2015; Qin and Han, 2013; Sun et al., 2012; Cai et al., 2017), Hangzhou (e.g. Wen and Tao, 2015; Yue et al., 2010) and Guangzhou (e.g. Wu, 1998). The abovementioned limitations create challenges for accurately depicting the ever-changing urban spatial structure, and the lack of comparable cases also diminishes our ability to explore the factors influencing polycentricity in an urban system.

The concept of new, open and big data (NOBD) has been proposed and developed rapidly over the past decade, and it also brings opportunities for urban studies. Most of the NOBD represents massive data streams that are updated in real time and space and collected by new (sensor) technologies and new social media. For example, Long and Thill

(2015) and Zhou et al. (2017) have turned to emerging NOBD, such as subway smartcard data, to overcome the constraints of conventional data in analyzing intrametropolitan movement. Li et al. (2018) use point-of-interest (POI) data to estimate commuting patterns. The use of NOBD in urban studies has provided decision-makers and analysts with extra sources of user and/or spatial information on every time horizon (Batty, 2012). These features of NOBD can potentially change how we quantify, plan, manage, and/or monitor cities (Batty, 2013). Many examples (e.g. see Alexiou et al., 2016; Batty, 2016; Glaeser et al., 2018; Miller and Tolle, 2016; Song et al., 2018a) have already illustrated that such potential could be realized. However, only a few academics have studied urban spatial structure with NOBD or in combination with the conventional data used in the literature.

Against the above backdrop, this paper aims to develop a generic procedure and method for studying urban spatial structure with NOBD. Specifically, we propose a nonparametric method that combines the Jenks natural breaks method and the local Moran's I to quantitatively identify the main center(s) and subcenters in cities. POIs, rather than conventional population or employment data, are used to measure urban polycentricity in 284 Chinese cities. We find 70 polycentric Chinese cities in 2011. To validate this finding, our results are compared to those based on commonly used data and methods in existing studies. Knowing the number of subcenters in each city, we further explore the determinants of the polycentric structure. Following previous studies (e.g. Li et al., 2016; Liu and Wang, 2016; McMillen and Smith, 2003; Sun and Lv, 2020; Sun et al., 2017), a regression analysis was conducted to reveal the associations between polycentricity and urban characteristics. The regression results indicate that the total population, GDP, average wage, and size of the urban area are all significantly correlated with the number of subcenters in the sample cities.

Literature review

Urban spatial structure, subcenters, and polycentric cities

In the 1980s, researchers proposed the view that modern metropolises are increasingly characterized by the presence of multiple activity nodes rather than a single dominant core (Lee, 2007). These activity nodes or employment and population clusters have been named "suburban downtowns" (Hartshorn and Muller, 1989), "edge cities" (Garreau, 1991), "technopoles" (Scott, 1990), or "employment subcenters" (Giuliano and Small, 1991). Because of these researchers' pioneering work, the concepts of subcenter and polycentric city have been widely discussed. Researchers have proposed and applied various criteria to define subcenters, such as employment size, office and/or retail space, commute flows, the job–housing ratio, and the land-use mix (Cervero, 1989; Giuliano and Small, 1991). Since then, studies have increasingly relied on employment density in defining subcenters in polycentric cities (Craig and Ng, 2001; McMillen, 2001; Pfister et al., 2000). The main attribute of an urban employment subcenter is its significantly higher employment density than the surrounding areas (McDonald, 1987) and its influence on the densities of nearby locations and even the entire region (Giuliano and Small, 1991; Gordon et al., 1986; McMillen, 2001). However, different from the above definition of subcenter, Li et al. (2016) argued that the level of human activities provides the most accurate representation of the city's function, and a city center or subcenter should be a cluster of human activities during some hours of the day. Zhang et al. (2017) also contended that a multicenter structure can be predicted by population, industry, and infrastructure. This new wave of thought calls for a rethinking of subcenter identification.

While studies on urban spatial structure have focused on defining and measuring polycentricity, few have explicitly outlined the importance of doing so. Meijers (2008) was among the first to emphasize the importance of delineating functional urban areas for the measurement of polycentricity in the national urban system in Europe. Later, a study by Zhang and Derudder (2019) found similar results in Chinese urban regions and argued that the polycentricity of an urban region is sensitive to the number of centers identified and included in the calculation. Both of their studies imply that center identification is not an end in itself but has significant implications for understanding the urban and regional system. For example, in the context of polycentric urban regions, studies have tended to analyze spatial organization by examining the functional linkages between urban centers (e.g. Burger et al., 2014; Vasanen, 2012). For studies that examine polycentric urban systems, the polycentricity structure has been used to explain economic productivity (Wang et al., 2019) and the decentralization of poverty (Zhang and Pryce, 2020).

Identifying subcenters with conventional data and NOBD

In the 1980s, scholars started to recognize that the urban spatial structure was changing, which was primarily the result of decreasing transportation costs and the rapid increase in urban sprawl. Efforts have been made to quantify these changes, especially the emergence of subcenters. McDonald (1987) used the positive residuals from a regression of employment density on the distance from the central business district (CBD) in Chicago as a measurement of subcenters. Giuliano and Small (1991) quantitatively defined a subcenter in Los Angeles as a cluster of contiguous tracts that has at least 10 employees per acre in each tract and a minimum of 10,000 employees in the cluster. Since then, employment density has been widely used in identifying subcenters (e.g. Bogart and Ferry, 1999; Craig and Ng, 2001; McDonald and McMillen, 1990; McMillen, 2001; McMillen and McDonald, 1998; Pfister et al., 2000). More recently, the study of subcenter identification remains attractive for scholars, but the data they use have become more diverse. For example, Nasri and Zhang (2018) used fine-grained land use data from Atlanta and Phoenix to identify regional employment subcenters. Liu et al. (2018) employed the 1990 and 2000 Census Transportation Planning Package data and American Community Survey 2006–2010 to explore changes in employment centers in Houston and Dallas.

Among the studies listed above, a number of methods regarding subcenter identification can be generalized. These methods are further classified into two groups according to their density criteria. A table that summarizes the existing methods is provided in the online supplementary materials (see Table S1). One group of methods identifies subcenters based on the absolute density criterion, which means that minimum density thresholds are used to define subcenters. For example, Giuliano and Small (1991) defined a city subcenter as a group of contiguous tracts with at least 10 employees per acre in each tract and a minimum of 10,000 employees in the group. The absolute density criterion has been widely used in existing studies since then, with numerical adjustments to fit different contexts (e.g. see Anderson and Bogart, 2001; Bogart and Ferry, 1999; Huang et al., 2017; Liu and Wang, 2016; Pfister et al., 2000). The key parameter in this method is the minimum density threshold; however, this is also where its primary flaw and corresponding technical difficulties lie when being applied. The minimum density is somewhat arbitrary, and its justification relies heavily on individual researchers' knowledge and observations of the study area (Huang, 2015; Lee, 2007). This makes it difficult if not impossible to expand the study coverage across time and space or to duplicate one study in another context.

The other group is based on the relative density criterion, including the use of various employment density functions, either parametric or nonparametric (Table S1). McDonald (1987) first proposed a more objective method, defining a subcenter as a cluster of significantly positive residuals from a simple regression of employment density on the distance from the CBD. McDonald and McMillen (1990) subsequently used the same method to identify employment subcenters in Chicago, and their results were verified by local knowledge. Similarly, Craig and Ng (2001) and McMillen (2001) have all identified subcenters in sample U.S. cities based on an employment density function. Later, Sun et al. (2012) and Huang et al. (2015) also conducted density regressions in their studies identifying subcenters in Beijing.

In the era of information, the development of technology has generated massive data that are available from multiple sources, which in turn encourages and triggers rapid development and innovations in multiple disciplines. In urban studies, NOBD have the ability to fill the gaps in conventional research methods and data and to build the path towards bottom-up planning and design (Li et al., 2016). Some of the most widely used NOBD in urban studies today include POIs, cellular signaling data, public transit smartcard transactions, social media posts, and location-based service (LBS) data from smartphone apps. NOBD share some common attributes, including finer resolution, a wider scope of coverage, and a higher update frequency, all of which make them excellent data for analyzing our city systems. Studies on urban spatial structure have also embraced NOBD. POI or similar LBS data (such as Baidu Heatmap and Didi taxi records) have been used in quantifying urban polycentricity (Duan et al., 2018; Guo et al., 2015; Li et al., 2016), delineating city boundaries (Chen et al., 2020; Hollenstein and Purves, 2010; Song et al., 2020; Xu and Gao, 2016), and identifying population movements (Becker et al., 2011; Song et al., 2019). However, compared to conventional studies on urban spatial structure, emerging studies using NOBD are still quite limited.

Factors influencing urban polycentricity

Acknowledging the fact that polycentric cities are emerging worldwide, scholars have started to explore the mechanisms behind this evolution. A growing but limited body of literature has examined the association between urban polycentricity, which is usually represented by the number of subcenters, and physical and socioeconomic urban characteristics. In doing so, these studies aim to generate preliminary but important hypotheses regarding the formation of polycentric cities. For example, in a Poisson regression with 62 large American urban areas, McMillen and Smith (2003) found that the key explanatory variables for the number of subcenters are population and commuting costs. Liu and Wang (2016) revealed that higher degrees of polycentricity are associated with cities in fragmented landscapes, and conditioning on topographic characteristics and total land area, GDP per capita is positively associated with high polycentricity in Eastern China. Similar analyses were also carried out by Li et al. (2016), Sun et al. (2017), and Sun and Lv (2020), although different predictors were used in different urban areas. These studies have enriched the theoretical underpinnings of urban spatial structure evolution; however, the empirical evidence remains insufficient.

To summarize, our literature review indicates that center and subcenter identification is still the mainstream in existing studies on urban spatial structure. Most of the existing studies still use conventional demographic data (which do not accurately reflect the spatial structure, nor do it allow for diachronic comparative studies) and conventional methods to identify subcenters for a limited number of cities, which has greatly limited the significance of the findings. Although NOBD has been utilized in related research, there is still room for improvement. In addition, the discussion of the factors influencing polycentricity is insufficient (Lan et al., 2019). Aiming to fill these gaps, this paper attempts to expand the literature by employing POIs and a new nonparametric method for center/subcenter identification, taking 284 Chinese cities as the study subject. This enables us to simultaneously identify the main center and subcenters (if any) of hundreds of cities and to investigate the possible determinants of polycentricity across these cities.

Study area and data

Study area and unit of analysis

In this paper, we focus on prefecture cities ("地级市" in Chinese) and above in the Chinese context. There are 284 such cities in total.

It is noteworthy that the definition of "city" has continued to change since the earliest concentration of human settlements. A widely accepted definition of city is an area delimited by administrative boundaries, which are demarcated by governments for administrative purposes, such as taxes, governance, and censuses. A common attribute of cities defined by administrative boundaries is that they often cover some nonurban areas, and this attribute has attracted urban planners' and researchers' attention in recent years. They have argued that administrative boundaries could pose challenges to urban studies, as most human activities occur only in urbanized areas. How we define "city" actually has an impact on how we understand and solve urban issues. Emerging studies have been undertaken to redefine "city." Long (2016), for instance, applied percolation theory in light of newly emerging big and open data and proposed an alternative definition of city based on road junctions. Song et al. (2018b) used POI data and road networks to redefine Chinese cities by their functional areas.

We are aware of the above progress, but we still adopt administrative boundaries to define "city" in this study for a number of reasons. First, using administrative boundaries to define a city or region is not unique to China. Similar well-known concepts are the metropolitan statistical area (MSA) and county in the United States, both of which include a large portion of rural areas around urbanized areas. Second, a majority of the previous studies on urban spatial structure and subcenter identification employed administrative cities, e.g. Gordon et al. (1986), McDonald (1987), Pfister et al. (2000), Dai et al. (2014), Yu and Wu (2016) and Li et al. (2016). Our results can be easily compared with these studies if we are using the same definition of a city. Third, the administrative boundaries of a city are relatively stable, which greatly facilitates historical comparisons and longitudinal analyses, while the lack of NOBD for the past simply makes it impossible to define older cities using NOBD techniques. Fourth, administrative boundaries matter when we must formulate plans or policies, which are often valid for areas or stakeholders within certain administrative boundaries.

Data

Our literature review indicates that the city's main center and subcenters have historically been defined and measured as clusters of employment or residences above a certain threshold. However, we argue that a center or subcenter should be able to provide versatile urban functions and to support various human activities, including not only residences and employment but also recreation, education, and social interactions. As one of the emerging

NOBD, POIs record the locational information and attributes of different places in cities. Such places are highly correlated with people's daily lives, such as restaurants, shopping malls, office buildings, bus stations, banks, etc., which make indicators based on POI data good proxies for diverse urban functions and human activities. To quantitatively identify and analyze the urban spatial structure, a total of 5,281,382 POIs in 2011 for mainland China were collected from the Baidu Map application programming interface (http://lbsyun.baidu.com/).

Methodologies

As identified in the literature review, both the absolute density and the relative density criteria in subcenter identification share a common principle: a subcenter should exhibit a local peak in density within the metropolitan area. In this paper, we follow the same principle and define a city's main center as a spatial cluster area (one or more subdistricts) with the highest level of human activities, while a subcenter is a subarea with a relatively higher level of human activity but is geographically separated from the main center. In other words, a qualified subcenter should have a significantly higher human activity level than its surroundings (locally high) and relatively higher density among all the subdistricts in the city (globally high). In addition, we think a good identification method should be able to produce reasonable results even if researchers are unfamiliar with the study area, which means that the method should be nonparametric and transferrable across time and space. Moreover, the method should be relatively easy to operate and be automated to support the analysis of subcenters for many cities simultaneously. Last but not least, the method should take advantage of NOBD, which not only have the merits identified previously but have also become increasingly available at a low cost.

In light of the above, we employ POI density, a typical NOBD, in a nonparametric approach to main center and subcenter identification, and the computational flowchart of this method can be found in the online supplementary materials (Figure S1). In our method, the POI density of each subdistrict in a city is calculated first. Either the Jenks natural breaks or the Anselin local Moran's I is then used to classify and identify the main center and subcenters, depending on the number of subdistricts in each city. The local Moran's I is an inferential spatial pattern analysis tool grounded in probability theory, and the application of the tool requires a decent sample size to meet the requirements of the law of large numbers (LLN) and the central limit theorem (CLT), while a small sample size may lead to errors of inference (Bivand et al., 2019). A sample size of 30 is a widely recognized and accepted empirical value in the academic community to meet the minimum requirements of both the LLN and the CLT (Ott and Longnecker, 2015) and is thus used as a threshold for model selection in this study.

For cities with fewer than 30 subdistricts, Jenks natural breaks is used to divide all the subdistricts into several classes based on the distribution of POI density, and a goodness of variance fit greater than 0.8 is used as the general threshold to accept the classification result (Jenks, 1967). Next, the subdistricts that belong to the top two classes from the natural break result are identified as the main center and subcenter candidates (as shown in Figure 1a). We then define the candidate subdistrict with the highest POI density as the main center of the city. For all the other candidates, if one subdistrict is geographically contiguous with the main center, it will also be identified as part of the main center, while a subdistrict that is geographically separated from the main center will be defined as a subcenter (as shown in Figure 1a).

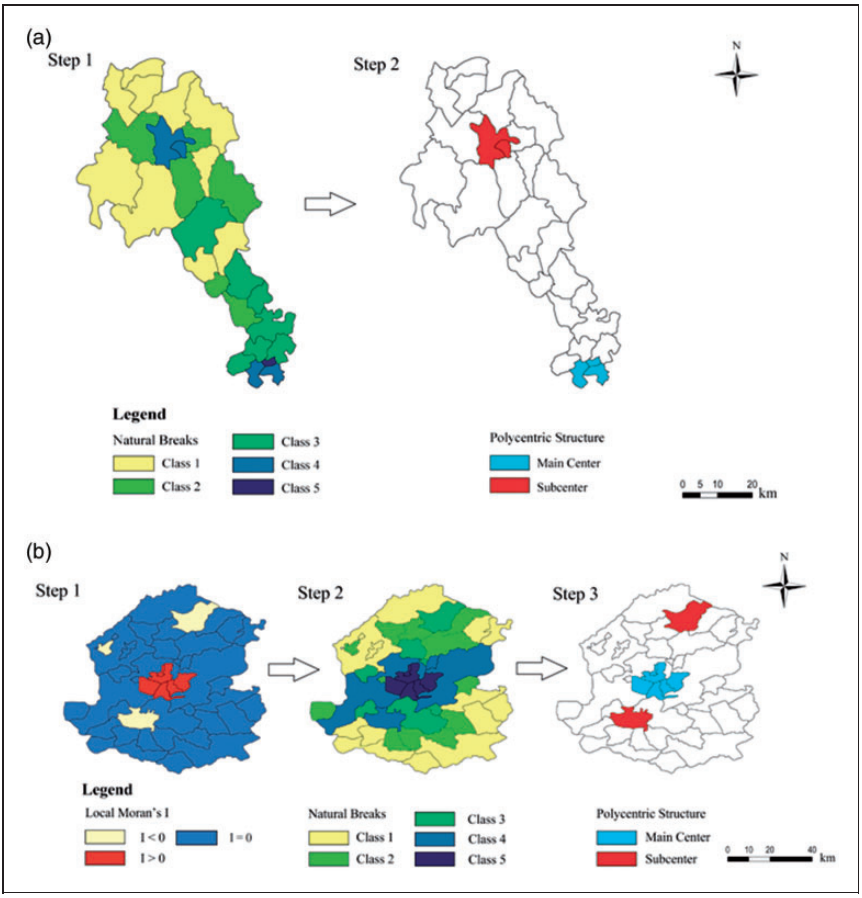

Figure 1. The urban spatial structure identification process: (a) uses Huangshan (less than 30 subdistricts) as an example to demonstrate the process using Jenks natural breaks; (b) uses Shenyang (more than 30 subdistricts) as an example to demonstrate the process of combining Moran's I and Jenks natural breaks.

For cities that have more than 30 subdistricts, the local Moran's I (Anselin, 1995) is calculated before selecting the subdistricts with higher POI density using the above method (as shown in Figure 1b). The local Moran's I is a widely used spatial statistics tool for cluster and outlier analysis in economics, resource management, biogeography, political geography, and demographics.

The local Moran's I statistic for spatial associations is given as:

$$I_i = \frac{x_i - \bar{X}}{S_i^2} \sum_{j=1, j \neq i}^{n} w_{i,j}(x_j - \bar{X}) \tag{1}$$

$$S_i^2 = \frac{\sum_{j=1, j \neq i}^{n} (x_j - \bar{X})^2}{n-1} - \bar{X}^2 \tag{2}$$

where x_i and x_j are the attributes of features i and j, respectively, $\bar{X}$ is the mean of the corresponding attribute, $\omega_{i,j}$ is the spatial weight between features i and j, S_i^2 is the global sample variance, and n is the total number of features. In this case, x_i refers to the density of POIs in subdistrict i.

When the local Moran's I results in a positive value, it means that a subdistrict and its surrounding subdistricts form a high-density cluster or a low-density cluster. A negative result indicates that a subdistrict with relatively higher POI density is surrounded by subdistricts with lower POI densities, or vice versa. In this work, if the local Moran's I value of a subdistrict is greater than 0 (for a 95% confidence interval), this subdistrict and the subdistricts adjacent to it that belong to the first class of the Jenks natural breaks result are defined as the main center of the city. If the subdistrict has a negative I value but its POI density is classified in the top three, it is then defined as a subcenter, as shown in Figure 1a.

This method we propose not only considers the density attributes of different subdistricts when identifying local peaks but also takes their spatial relationships into account. Moreover, this nonparametric method requires only the density of POIs as input but does not require local knowledge of the study area and therefore can be reproduced across different contexts.

Results and validation

Subcenters in Chinese cities

We found 70 polycentric cities in mainland China, most of which lie in east-central China (Figure 2). The population in these polycentric cities ranges from 850,000 (Tongchuan city) to 33 million (Chongqing city), with a mean of 5 million and a standard deviation of 4.5 million, which indicates an extremely large degree of deviation among the cities. Of the 70 polycentric cities, 31 cities had more than 5 million people in 2011, which implies that polycentricity exists not only in large cities but also in small cities.

Among the 70 polycentric cities, the number of subcenters ranges from 1 to 12, and 57% of these cities have only one subcenter identified, while only 10 cities had more than 2

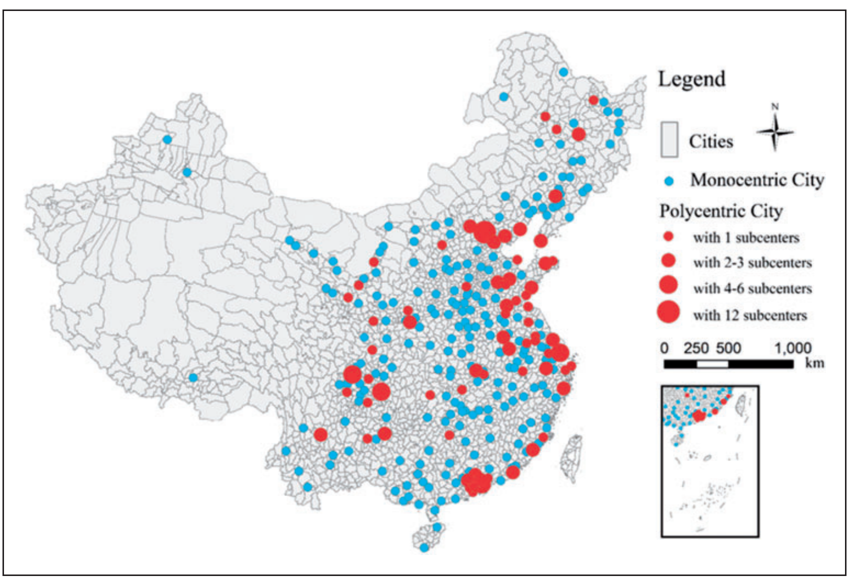

Figure 2. The distribution of polycentric and monocentric cities in China.

subcenters in 2011. The spatial distributions of the main centers and subcenters in the 70 polycentric cities are shown in the online supplementary materials (see Figure S2).

The spatial pattern of subcenters

The number of subcenters and the average distance between the main center and the subcenters in a city provide a straightforward way to depict its spatial structure. Such distances vary from city to city and are determined by multiple factors, such as terrain characteristics, the land area of the city, and the layout of the road network. In this study, we define the centroids of the main center and the subcenter polygons in each city as origins and destinations and then calculate the mean linear distance (MD) between them. We also employ the ratio of the MD and the square root of the land area as an indicator to quantify the spatial distribution characteristics (SDCs) of the main center and the subcenter(s) (see more details in Table S2 in the online supplementary materials).

Among the 70 cities considered, the longest MD is found in Lanzhou city at 93.69 km, while the shortest is 5.72 km in Weihai city, and the MDs of most cities (62.9%) are between 15 and 30 km. The SDC takes the geometric area of a city into account and reflects the geographic scope of the subcenters, and a larger SDC normally indicates that the subcenters are located near the periphery of the city boundaries, while a smaller SDC implies a more compact distribution of the subcenters around the main center.

Result validation

As we employ new data and methods in center and subcenter identification, it is necessary for us to verify that our results are valid and robust. To perform the validation, we first compare the results (see Section 'Subcenters in Chinese cities') with those based on the

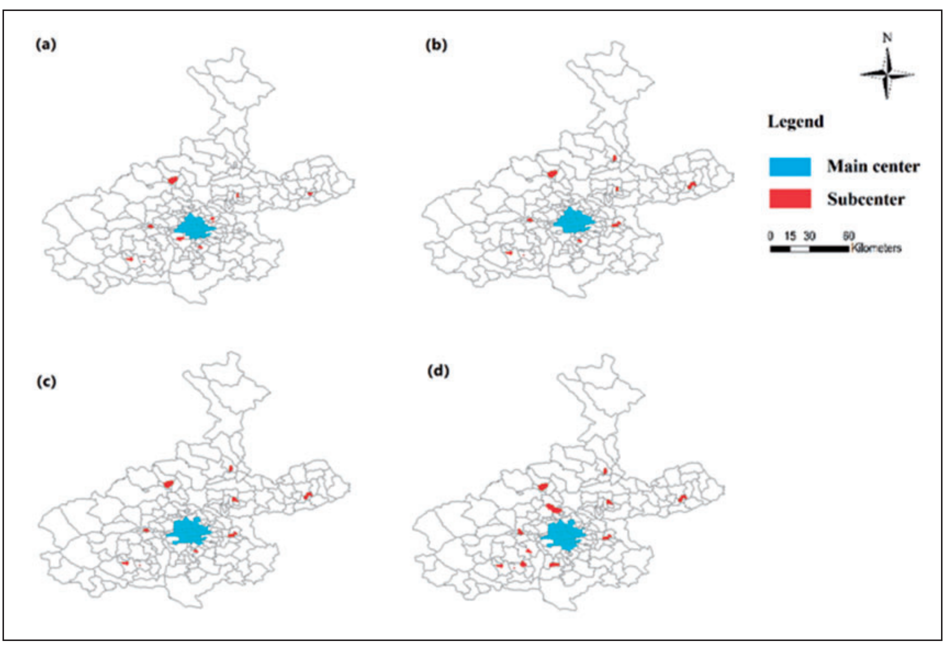

Figure 3. The main center and subcenters identified in Beijing: (a) using job density data from 2000; (b) using job density data from 2005; (c) using job density data from 2009; (d) using POI density data from 2011.

existing methods and traditional/census data. Let us take Beijing as an example. In Beijing, we were fortunate to have employment data at the township level (subdistrict) for three different years: 2000, 2005, and 2009. By using the existing method highlighted in Section ‘Methodologies’, we obtain the job density-based polycentric structures of Beijing in these years (see Figure 3a to c). By contrast, Figure 3d shows the results of this study.

Figure 3a-d share a common subset of subcenters, which implies that these subcenters might remain constant as time progresses. Compared to Figure 3a-c, notably, Figure 3d identifies several new subcenters in the northwest and southeast. Our local knowledge and site audits tell us that the conventional employment data alone could fail to capture high-density residential communities with diverse open spaces, public facilities, and commercial stores but few office buildings. Baidu heatmaps, which have been widely used as a proxy for the concentration of human activities, can also corroborate this. When compared with the findings of Li et al. (2016), who used a Baidu heatmap to capture a wide spectrum of human activities, Figure 3d’s additional subcenters relative to Figure 3a-c are defendable.

In addition to the above cross-checking, we compare the results of this study (Figure 4b&d) with those based on the conventional/existing method proposed by McMillen (2001) (Figure 4a and c). Figure 4a&c are adapted from Sun et al. (2012) and Sun and Wei (2014). The comparisons indicate that (a) our results identify more subcenters than those of other scholars—the reasons for this have been highlighted above—and (b) a subset of the subcenters in our results are identical to those of other scholars. The above shows that this study’s method can at least identify subcenters comparable to those of the conventional/existing methods. The former might also identify subcenters that the latter cannot.

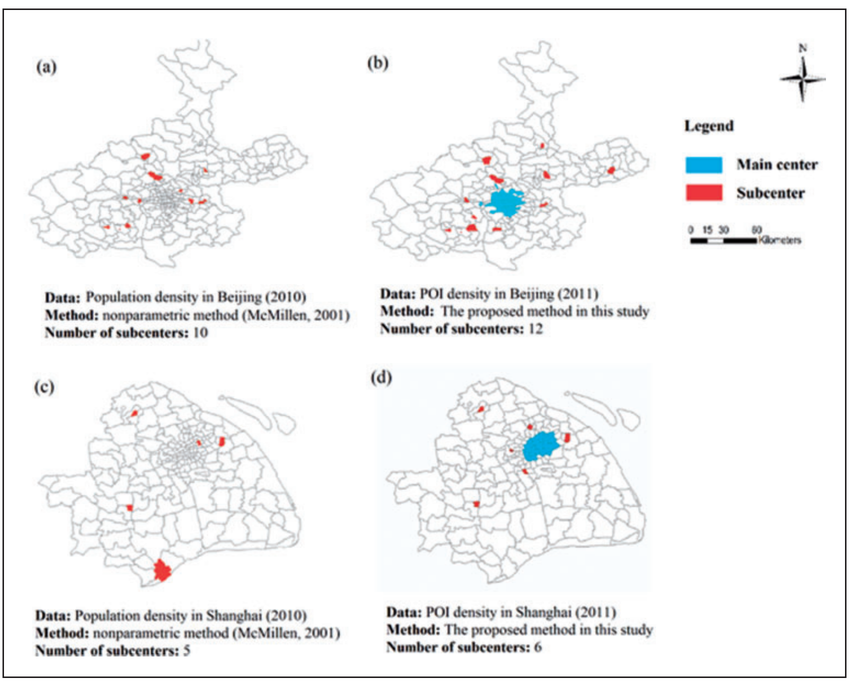

Figure 4. Subcenters identified by existing studies and by this study.
POI: point-of-interest.

Table 1. OLS regression results.

Variable	Coefficient	t	p
Constant	–0.13	–0.975	0.330
Population in the year 2010 (millions)	–0.257	–4.727	0.000
GDP (millions of yuan)	0.557	12.695	0.000
Average wages of staff and workers	0.080	2.457	0.015
Area of construction land (sq.km.)	0.052	0.884	0.378
Urban land area (sq.km.)	0.141	2.729	0.007
Density of road intersections	–0.003	–0.079	0.937
R^2	0.468		
Adjusted R^2	0.456		

Note: This sample includes 284 cities in China. The dependent variable for the regressions is the number of subcenters.
GDP: gross domestic product.

Regression analysis of the relationship between polycentricity and urban characteristics

When scholars find a way to identify subcenters, the factors influencing urban polycentricity become of interest. Growing but limited number of studies have regressed urban polycentricity on urban characteristics to examine their correlations. In this study, 70 out of 284 Chinese cities are identified with polycentric structures, which constitutes a decent sample size for analyzing the association between urban polycentricity and its factors of influence. Specifically, a multivariate regression model was created to examine the correlation between the number of subcenters and the cities' geographical, socioeconomic, and demographic characteristics as potential factors of influence. Given that there is no consensus concerning the selection of the independent variables in existing studies, this paper chose the factors of influence in accordance with the most widely used variables in the literature (cf. Li et al., 2016; Liu and Wang, 2016; McMillen and Smith, 2003; Sun and Lv, 2020; Sun et al., 2017) and took full consideration of data availability as well. Ultimately, the total population, the GDP, the average wage of staff and workers, the area of construction land, the total urban land area, and the density of road intersections were used as predictors in the regression analysis.

Each independent variable is normalized to eliminate differences among their dimensions using min–max normalization. The ordinary least squares (OLS) regression results are given in Table 1. At the 0.05 level of significance, the total population shows a significant and negative correlation with the number of subcenters, which means that cities with a larger population have fewer subcenters on average when other factors are kept constant. This result is surprising at first glance since we noticed that the cities with more subcenters are populous metropolises such as Beijing and Shanghai. Then, we further examined the other less populous cities and found that the majority of midwestern cities have a relatively large population but only one subcenter. Such cities are still in the early stages of urban polycentralization, and the concentration of the population is the driving force behind the formation of their first subcenter (Zhao, Dang and Wang, 2009). Total urban land area is found to be positively correlated with the number of subcenters, while the area of construction land is not significant. It is not hard to anticipate this result since larger cities potentially provide enough space for subcenters to emerge and grow. Both GDP and the average wage of employees reflect the economic development level of a city, and both are significantly correlated with the number of subcenters. The positive coefficients on GDP and average wage verify that subcenters can also generate economies of scale through agglomeration and thus contribute to the economy. Although a new approach and new data were used in this study to identify urban subcenters, we had some findings regarding the correlations between polycentricity and urban characteristics that are consistent with those of previous studies, such as the negative correlation with total population (Zhao, Dang and Wang, 2009), and the positive impacts from urban land area (Liu and Wang, 2016) and GDP (Lip et al., 2018; Liu and Wang, 2016; Sun et al., 2017).

Discussion and conclusions

This study examines the feasibility and reliability of using a newly proposed nonparametric method to identify city subcenters with widely available POI data at the subdistrict level. By implementing the approach for 284 prefecture-level Chinese cities, 70 cities were identified with at least one subcenter in 2011. To validate the robustness of the proposed method and data, the results are compared with existing studies using conventional methods and data. The findings indicate that both the new method and the POI data can deliver validated results for subcenter identification. To further elaborate on our results, a regression model was created to analyze the relationships between the number of subcenters and certain selected demographic, geographic, and socioeconomic characteristics of the city. The regression results indicate that the number of subcenters in a city is positively associated with land area and economic performance but negatively correlated with population size, particularly for midwestern cities in China. Such findings may guide local governments and planning officials to identify their city's stage in polycentralization and to be prepared for the related changes in urban spatial structure.

Overall, we think this paper makes two main contributions to the literature on urban spatial structure (especially urban polycentricity). First, the introduction of POI data in subcenter identification provides us with an alternative mindset for urban spatial structure, which accounts for versatile urban functions and supports various human activities. By contrast, most if not all of the existing studies, define subcenters as clusters of employment or population. Cities have continued to evolve since those studies were conducted. Human activities, for instance, have become much more diverse and complex in recent years given the advent of the Internet of Things, the sharing economy, crowdsourcing, and smartphones. Therefore, main centers and subcenters in a city support a much wider variety of human activities, not only for work but also for lifestyles, leisure, (random) encounters, exhibitions, meetings, and entertainment (cf. Li et al., 2016). Thanks to the rapid development and vast availability of NOBD, we can now redefine and identify main centers and subcenters based on information that is richer, more continuous, and more current (even real-time) than ever before (cf. Batty, 2016). In this study, we illustrate the potential of POI data in identifying main centers and subcenters, and such attributes could enable urban planners and researchers to understand our cities more profoundly, efficiently, and reliably. They could also in theory support more timely and proactive policy- and decision-making in our cities.

Second, we proposed a nonparametric method using Jenks natural breaks and Moran's I to identify and locate main centers and subcenters in cities. The validity and robustness of the method have been verified by comparing the corresponding results with those in existing studies. Our method has the advantages of avoiding some of the potential problems in the existing ones, such as arbitrariness in setting thresholds for population and/or employment density and sensitivity to the spatial scale and the complexity of calculating the density surface. Moreover, the wide availability of POI data and the replicable method we propose in this study make it possible for researchers to reproduce their studies in new contexts.

Undoubtedly, there is still room for improvement in future research. First, the use of POIs as point data has limitations in describing the coverage and scale of the subjects. For instance, one cannot tell from the point data whether a commercial POI is a local grocery store or a chain supermarket. Unlike conventional employment or population data that capture the exact number of individuals, POI data provide a proxy for urban functionality and some description of the aggregated level of human activities. Second, we attempt to explore the factors that may be associated with urban structure after we identify polycentric cities in China. Admittedly, our sample size for the regression analysis is highly constrained by the total number of prefecture cities in China. Although the regression results are statistically significant, their practical significance could benefit from an expansion of the sample size or from conducting a panel analysis with multiple years of data. Finally, the main purpose of this study is to illustrate the feasibility and strength of the nonparametric method and NOBD in studying urban spatial structure; therefore, it used only data from 2011 and failed to conduct a longitudinal study of the main centers and subcenters. However, we know that there is no single best approach in center and subcenter identification, and all the merits and drawbacks should be discussed in terms of specific contexts. For example, this paper proposes a refined nonparametric approach with ubiquitous POI data, which has most of its advantages in large-scale analyses and comparisons. In our future studies, we could also combine multiple sources of NOBD, such as POI data and LBS data, including but not limited to car GPS, smartwatches, security cameras, and food delivery, as well as conventional census and survey data across multiple geographic and temporal scales. This combination would enable us to gain more knowledge about and insights into urban spatial structures, answering questions such as the following: What kind of POI composition would allow a subcenter to attract a constant flow of people? How long does a visitor usually stay at a subcenter? How does the exchange of people influence the attractiveness of different subcenters?

Acknowledgements

The authors would like to thank Dr. Zhou Jiangping of the University of Hong Kong for his generous support in this research.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The research was supported by the National Natural Science Foundation of China (Grant No. 42001385).

ORCID iDs

Yimeng Song https://orcid.org/0000-0001-9558-1220
Long Chen https://orcid.org/0000-0002-0972-449X

Supplemental material

Supplemental material for this article is available online.

References

Alexiou A, Singleton A and Longley PA (2016) A classification of multidimensional open data for urban morphology. *Built Environment* 42(3): 382–395.

Anderson NB and Bogart WT (2001) The structure of sprawl: Identifying and characterizing employment centers in polycentric metropolitan areas. *American Journal of Economics and Sociology* 60(1): 147–169.
Anselin L (1995) Local indicators of spatial association-LISA. *Geographical Analysis* 27(2): 93–115.
Batty M (2012) Smart cities, big data. *Environment and Planning B: Planning and Design* 39(2): 191–193.
Batty M (2013) Big data, smart cities and city planning. *Dialogues in Human Geography* 3(3): 274–279.
Batty M (2016) Big data and the city. *Built Environment* 42(3): 321–337.
Becker RA, Caceres R, Hanson K, et al. (2011) A tale of one city: Using cellular network data for urban planning. *IEEE Pervasive Computing* 10(4): 18–26.
Bivand R, Muller W and Reder M (2009) Power calculation for global and local Moran's I. *Computational Statistics & Data Analysis* 58(8): 2859–2872.
Bogart WT and Ferry WC (1999) Employment centres in greater Cleveland: Evidence of evolution in a formerly monocentric city. *Urban Studies* 36(12): 2099–2110.
Burger MJ, Van Der Knaap B and Wall RS (2014) Polycentricity and the multiplexity of urban networks. *European Planning Studies* 22(4): 816–840.
Cai J, Huang B and Song Y (2017) Using multi-source geospatial big data to identify the structure of polycentric cities. *Remote Sensing of Environment* 202: 210–221.
Cervero R (1989) *America's Suburban Centers: The Land Use-Transportation Link*. Boston: Unwin Hyman.
Chen B, Song Y, Huang B, et al. (2020) A novel method to extract urban human settlements by integrating remote sensing and mobile phone locations. *Science of Remote Sensing* 1: 100003.
Craig SG and Ng PT (2001) Using quantile smoothing splines to identify employment subcenters in a multicentric urban area. *Journal of Urban Economics* 49(1): 100–120.
Duan Y, Liu Y, Liu X, et al. (2018) Identification of polycentric urban structure of Central Chongqing using points of interest big data. *Journal of Natural Resources* 33(5): 788–800.
Garreau J (1991) *Edge City*. New York: Doubleday.
Garreau J (1992) *Edge City: Life on the New Frontier (1st Anchor Books ed.)*. New York: Anchor Books.
Giuliano G and Small KA (1991) Subcenters in the Los Angeles region. *Regional Science and Urban Economics* 21(2): 163–182.
Glaeser EL, Kominers SD, Luca M, et al. (2018) Big data and big cities: The promises and limitations of improved measures of urban life. *Economic Inquiry* 56(1): 114–137.
Gordon P and Richardson HW (1996) Beyond polycentricity: The dispersed metropolis, Los Angeles, 1970-1990. *Journal of the American Planning Association* 62(3): 289–295.
Gordon P, Richardson HW and Wong HL (1986) The distribution of population and employment in a polycentric city: The case of Los Angeles. *Environment and Planning A: Economy and Space* 18(2): 161–173.
Guo J, Lü Y and Shen T (2015) Urban spatial structure based on point pattern analysis—Taking Beijing metropolitan area as a case. *Economic Geography* 35(8): 68–74.
Hartshorn TA and Muller PO (1989) Suburban downtowns and the transformation of metropolitan Atlanta's business landscape. *Urban Geography* 10(4): 375–395.
Hollenstein L and Purves R (2010) Exploring place through user-generated content: Using Flickr tags to describe city cores. *Journal of Spatial Information Science* 2010(1): 21–48.
Huang D, Liu Z and Zhao X (2015) Monocentric or polycentric? The urban spatial structure of employment in Beijing. *Sustainability* 7(9): 11632–11656.
Huang D, Liu Z, Zhao X, et al. (2017) Emerging polycentric megacity in China: An examination of employment subcenters and their influence on population distribution in Beijing. *Cities* 69: 36–45.
Jenks GF (1967) The data model concept in statistical mapping. *International Yearbook of Cartography* 7: 186–190.
Lan F, Da H, Wen H, et al. (2019) Spatial structure evolution of urban agglomerations and its driving factors in Mainland China: From the monocentric to the polycentric dimension. *Sustainability* 11(3): 610.

Lee B (2007) “Edge” or “Edgeless” cities? Urban spatial structure in U.S. metropolitan areas, 1980 to 2000. *Journal of Regional Science* 47(3): 479–515.

Li J, Li M, Long Y, et al. (2016) China polycentric cities based on *Baidu Heatmap*. *Shanghai Urban Planning Review* 3: 30–36.

Li M, Kwan MP, Wang F, et al. (2018) Using points-of-interest data to estimate commuting patterns in Central Shanghai, China. *Journal of Transport Geography* 72: 201–210.

Liu X, Pan Q, King L, et al. (2018) Analysing the changes of employment subcentres: A comparison study of Houston and Dallas. *Urban Studies*. Epub ahead of print 9 October 2018. DOI: 10.1177/0042098018789554.

Liu X and Wang M (2016) How polycentric is urban China and why? A case study of 318 cities. *Landscape and Urban Planning* 151: 10–20.

Long Y and Thill JC (2015) Combining smart card data and household travel survey to analyze jobs–housing relationships in Beijing. *Computers, Environment and Urban Systems* 53: 19–35.

Long Y (2016) Redefining Chinese city system with emerging new data. *Applied Geography* 75: 36–48.

McDonald JF (1987) The identification of urban employment subcenters. *Journal of Urban Economics* 21(2): 242–258.

McDonald JF and McMillen DP (1990) Employment subcenters and land values in a polycentric urban area: The case of Chicago. *Environment and Planning A: Economy and Space* 22(12): 1561–1574.

McMillen DP (2001) Nonparametric employment subcenter identification. *Journal of Urban Economics* 50(3): 448–473.

McMillen DP and McDonald JF (1998) Suburban subcenters and employment density in metropolitan Chicago. *Journal of Urban Economics* 43(2): 157–180.

McMillen DP and Smith SC (2003) The number of subcenters in large urban areas. *Journal of Urban Economics* 53(3): 321–338.

Meijers E (2008) Measuring polycentricity and its promises. *European Planning Studies* 16(9): 1313–1323.

Miller HJ and Tolle K (2016) Big data for healthy cities: Using location-aware technologies, open data and 3D urban models to design healthier built environments. *Built Environment* 42(3): 441–456.

Nasri A and Zhang L (2018) Employment subcenters, polycentricity, and travel behavior: The tale of two cities in the U.S. *International Conference on Transportation and Development 2018*. pp. 94–104.

Ott RL and Longnecker MT (2015) *An Introduction to Statistical Methods and Data Analysis*. Nelson Education.

Pfister N, Freestone R and Murphy P (2000) Polycentricity or dispersion?: Changes in center employment in metropolitan Sydney, 1981 to 1996. *Urban Geography* 21(5): 428–442.

Qin B and Han S (2013) Emerging polycentricity in Beijing: Evidence from housing price variations (2001-2005). *Urban Studies* 50(10): 2006–2023.

Scott AJ (1990) The technopoles of Southern California. *Environment and Planning A: Economy and Space* 22(12): 1575–1605.

Song Y, Huang B, Cai J, et al. (2018a) Dynamic assessments of population exposure to urban greenspace using multi-source big data. *Science of the Total Environment* 634: 1315–1325.

Song Y, Long Y, Wu P, et al. (2018b) Are all cities with similar urban form or not? Redefining cities with ubiquitous points of interest and evaluating them with indicators at city and block levels in China. *International Journal of Geographical Information Science* 32(12): 2447–2476.

Song Y, Huang B, He Q, et al. (2019) Dynamic assessment of PM2. 5 exposure and health risk using remote sensing and geo-spatial big data. *Environmental Pollution* 253: 288–296.

Song Y, Chen B and Kwan MP (2020) How does urban expansion impact people’s exposure to green environments? A comparative study of 290 Chinese cities. *Journal of Cleaner Production* 246: 119018

Sun B, Hua J, Li W, et al. (2017) Spatial structure change and influencing factors of city clusters in China: From monocentric to polycentric based on population distribution. *Progress in Geography* 36(10): 1294–1303.
Sun B and Wei X (2014) Spatial distribution and structure evolution of employment and population in Shanghai Metropolitan Area. *Acta Geographica* 69: 6.
Sun T and Lv Y (2020) Employment centers and polycentric spatial development in Chinese cities: A multi-scale analysis. *Cities* 99: 102617.
Sun TS, Wang LL and Li GP (2012) Distributions of population and employment and evolution of spatial structures in the Beijing metropolitan area. *Acta Geographica Sinica* 67(6): 829–840.
Vasanen A (2012) Functional polycentricity: Examining metropolitan spatial structure through the connectivity of urban sub-centres. *Urban Studies* 49(16): 3627–3644.
Wang M, Derudder B and Liu X (2019) Polycentric urban development and economic productivity in China: A multiscalar analysis. *Environment and Planning A: Economy and Space* 51(8): 1622–1643.
Wen H and Tao Y (2015) Polycentric urban structure and housing price in the transitional China: Evidence from Hangzhou. *Habitat International* 46: 138–146.
Wu F (1998) The new structure of building provision and the transformation of the urban landscape in metropolitan Guangzhou, China. *Urban Studies* 35(2): 259–283.
Xu Z and Gao X (2016) A novel method for identifying the boundary of urban built-up areas with POI data. *Acta Geographica Sinica* 71(6): 928–939.
Yue W, Liu Y and Fan P (2010) Polycentric urban development: The case of Hangzhou. *Environment and Planning A: Economy and Space* 42(3): 563–577.
Yu T and Wu W (2016) Monocentric or Polycentric? A Study on Urban Employment Sub-centers in Beijing. *Urban Planning Forum* 229(3): 21–29.
Zhao J, Dang X and Wang X (2009) The evolution of spatial structure of urban agglomerations: Empirical evidence from western China. *Economic Review* 4: 27–34.
Zhang L, Yue W and Liu Y (2017) Multidimensional analysis of the polycentric urban spatial structure: A case study of Hangzhou. *Economic Geography* 37(6): 67–75.
Zhang ML and Pryce G (2020) The dynamics of poverty, employment and access to amenities in polycentric cities: Measuring the decentralisation of poverty and its impacts in England and Wales. *Urban Studies* 57(10): 2015–2030.
Zhang W and Derudder B (2019) How sensitive are measures of polycentricity to the choice of 'centres'? A methodological and empirical exploration. *Urban Studies* 56(16): 3339–3357.
Zhou J, Wang M and Long Y (2017) Big data for intrametropolitan human movement studies: A case study of bus commuters based on smart card data. *International Review for Spatial Planning and Sustainable Development* 5(3): 100–115.

Ying Long is now an Associate Professor in School of Architecture, Tsinghua University, China. His research focuses on urban planning, quantitative urban studies, and applied urban modeling. Before he joined Tsinghua University, he has been worked for Beijing Institute of City Planning as a senior planner for 11 years. Familiar with planning practices in China and versed in the international literature, Dr. Long's academic studies creatively integrates international methods and experiences with local planning practices. Dr. Long is also the founder of Beijing City Lab (BCL, www.beijingcitylab.org), an open research network for quantitative urban studies. More information is available at http://www.beijingcitylab.com/longy.

Yimeng Song is currently a research assistant professor in the Department of Land Surveying and Geo-Informatics at The Hong Kong Polytechnic University. He received his BEng from China University of Geosciences (Beijing) in 2012, BEcon from Peking

University in 2012, MSc from The Chinese University of Hong Kong in 2014, and PhD from The Chinese University of Hong Kong in 2019. Before joining The Hong Kong Polytechnic University, Dr Song served as a postdoctoral researcher in the Department of Urban Planning and Design at the University of Hong Kong. His research mainly focuses on the application of spatial big data analytics in the field of urban environment, human mobility, exposure assessment, and environmental health.

Long Chen is now a Post-doc Research Associate in School of Architecture, Tsinghua University, China. He got his PhD degree at Florida State University, focusing on the topic of sustainable urban form. His research interests include quantitative urban studies, sustainable development and big data.