Highlights

- We demonstrate the feasibility of applying carpooling big data in metropolitan studies.
- We propose a data-driven three-step method to characterize the metropolitan polycentricity in-depth and comprehensively
- Beijing Metropolitan Region has a hierarchical polycentric structure and an influence sphere beyond the administrative boundary.
- The heterogeneity of human activity performance and role for each regional center is remarkable.

Title Page (with Author Details)

# Characterizing the Polycentric Spatial Structure of Beijing Metropolitan Region Using Carpooling Big Data

Xiaobing LIU[1], Xuedong YAN[1*], Helena TITHERIDGE[2], Wei WANG[1], Rui WANG[3], Yang LIU[4]

1 MOT Key Laboratory of Transport Industry of Big Data Application Technologies for Comprehensive Transport, School of Traffic and Transportation, Beijing Jiaotong University, Beijing 100044, PR China

2 Centre for Transport Studies, University College London, Gower Street, London, WC1E 6BT, UK

3 School of Civil Engineering, Beijing Jiaotong University, Beijing, 100044, P. R. China

4 The Belt and Road Initiative Construction Promotion Center, National Development and Reform Commission, Beijing, 100824, P. R. China

* Correspondence: xdyan@bjtu.edu.cn

# Characterizing the Polycentric Spatial Structure of Beijing Metropolitan Region Using Carpooling Big Data

**Abstract**

Polycentric metropolitan regions are a high-level urbanization form characterized with dynamic layout, fuzzy boundary and various human mobility performances. Owing to the complexity of polycentricity, it can be difficult to understand their spatial structure characteristics merely based on conventional survey data and method. This poses a challenge for authorities wishing to make effective urban land use and transport policies. Fortunately, the presence and availability of big data provides an opportunity for scholars to explore the complex metropolitan spatial structures, but there are still some research limitations in terms of data use and processing, unit scale, and method. To address these limitations, we proposed a three-step method to apply the carpooling big data in metropolitan analysis including: first, locating the metropolitan sub-centers; second, delimiting the metropolitan sphere; third, measuring the performance of polycentric structure. The developed method was tested in Beijing Metropolitan Region and the results show that the polycentric metropolitan region represents a hierarchical regional center system: one primary center interacting with seven surrounding secondary centers. These metropolitan centers have a strong attraction, which results in the continuous expansion beyond the administrative boundary to radiate more adjacent jurisdictions. Furthermore, the heterogeneity of human activity performance and role for each regional center is remarkable. It is necessary to consider the specific role of each sub-center when making metropolitan transport and land use policies. Compared with previous studies, the proposed method has the advantages of being more reliable, accurate and comprehensive in characterizing the polycentric spatial structure. The application of carpooling big data and the proposed method would provide a novel perspective for research on the other metropolitan regions.

***Key words*:** Polycentric spatial structure, functional boundary, carpooling, commuting, Beijing Metropolitan Region

## 1. Introduction

In recent decades, the urban sprawl and job decentralization have given rise to metropolitan regions (MRs) that extend geographically beyond the boundaries of single urban cores to multiple interconnected centers (Meijers &Burger, 2010). Urban planners have realized that the development of multiple centers with mixed use has become a necessary choice for megacities to overcome typical urban diseases around the central business district (CBD), such as traffic congestion, environmental pollution, and the heat island effect (Liu et al., 2020). Although it is still arguable about which urban form is the most efficient and sustainable, the polycentric development is considered as a normative planning strategy to reach important objectives in terms of enhancing regional economic competitiveness, environmental sustainability and social cohesion (Davoudi, 2003). The characterization of metropolitan polycentricity, more generally, urban spatial structure, has become an important research topic (Schleith et al., 2016; Lin et al., 2015, Zhen et al., 2017).

The metropolitan polycentric spatial structures are often characterized with dynamic layout and fuzzy boundary, as well as various human activity performances of multiple regional centers (Veneri 2013; Fang and Yu, 2017; Hu et al., 2018; Liu et al., 2020). Traditionally, regions and their structures have been measured based on survey data (Wong and Huang, 2017), most of which are static, limited by survey cycle time, are either expensive, or gathered for administrative purposes (Elwood et al., 2012). Owing to the complexity of polycentricity, it can be difficult to understand their spatial structure characteristics merely based on conventional survey data and method. This poses a challenge to implement this planning strategy in practice, such as designing sustainable land use and transport policies that are effective across planning areas with multiple municipalities. Fortunately, the presence and availability of big data provides us an opportunity to address this challenge. Some scholars have attempted to investigate the urban polycentric structure based on diverse big data (Wong and Huang, 2017; Zhang et al., 2017; Zhen et al., 2017; Wan et al., 2018), but there are still some research limitations in terms of data use and processing, unit scale, and method.

First, a more competent dataset and the innovations about data application need to be emphasized. The current higher level of information communication technology (ICT) and associated device usage record a large amount of activity data from nearly all residents (Lynch, 2008; Allam and Newman, 2018). Scholars and planners have used massive night light data from satellite images (Gao et al., 2015; Zhang and Su, 2016), geo-web data from mobile applications (Sobolevsky et al., 2013; Wong and Huang, 2017), and taxi GPS data (Liu et al., 2015; Zhang et al., 2018) to better depict spatial performance of human activities. Some features of these data sources, however, limit their usage for exploring the metropolitan structure in practice. For example, the saturation effects of night light data make it difficult to reflect the intensity and spatial distribution of human activities exactly, especially in developed regions (Liu et al., 2012), while the Geo-referenced data from mobile applications have the disadvantages of positional uncertainty and representation vagueness (Li et al., 2013; Longley et al., 2015). On the other hand, seldom studies have focused on this topic on a delicate scale of data application, such as a grid level. One reason is that obtaining a high-quality dataset is difficult in traditional approaches. Another reason is the unsolved issues in data consistency, especially for different data resources, inconsistent scales and diverse formats (Liu et al., 2020).

Second, most previous studies directly used the administrative divisions in the topic of metropolitan polycentricity. Limited to data sources or just for convenience, without exception, most of the studies on BMR (Long et al., 2013; Zhou et al., 2014) or metropolitan regions in other countries (Angel and Blei, 2016; Burger et al., 2011; Veneri 2013) ignored the territory problem. Simply using the administrative divisions as the geographical divisions would hinder the sophisticated investigation into the regional development (Shi and Cao, 2020) and cause the unpredictable regional bias due to inconsistency in size (Liu et al., 2020). Moreover, it is more likely to encounter a modifiable areal unit problem when applying these local administrative units in comparative analyses across countries (Veneri 2013).

Third, the heterogeneity of human activity performance and role for each regional center is largely overlooked, which may be also due to the insufficient data source. Owing to different geographical and social environment, the regional centralization versus decentralization and clustering versus dispersion performance can be various not only from one country (or region) to another (Veneri 2013; Hu et al., 2018), but also among different centers within a same metropolitan region. A comprehensive investigation on this regional centers' heterogeneity is necessary to determine the priority of public resource assignment and make more targeted land use and transport policies. The combination of morphological and functional approach on characterizing the polycentric structure is a good choice for contemporary complex MRs (Riguelle et al., 2007). However, most studies involved the performance of sub-centers merely consider one specific facet, such as the job density and share (Angel and Blei, 2016), job-housing relationship (Lin et al., 2015), commuting duration (Hu et al., 2018). Furthermore, works on the regional centers' roles in metropolitan regions receive much less attention. Only Giuliano and Small (1991) conducted a cluster analysis using 32 centers as observations and eight industry shares as variables. They found that the more service-oriented centers tend to be at higher densities and somewhat closer to the core area.

To address these limitations in previous studies, we first use the carpooling big data under a grid-based Geographic Information System (GIS) environment and propose a three-step method: first, identifying the metropolitan CBD and sub-centers by a grid-based clustering algorithm; second, delimiting the metropolitan sphere of influence based on a three-fold judgment criterion; third, measuring the human activity performance and role of each center using two set of morphological and functional indexes. The emerging carpooling big data can help put this three-step task into practice.

The objective of this paper is to characterize the metropolitan polycentric spatial structure in-depth and comprehensively with the advantage of big data. More specifically, first, we need to demonstrate the feasibility of carpooling data in metropolitan studies and find the way to use these data. Furthermore, we need to determine the advanced clustering algorithm, delimiting approach and measurement system based on the carpooling data and literature review, to realize the proposed three-step method. Last, applying our data and method in the Beijing Metropolitan Area, we hope the associated results and findings can provide valuable insights for metropolitan land use and transport planning.

The rest of this paper is organized as follows. Section 2 presents a review of the relevant literature. Section 3 details the methodology used to measure the metropolitan spatial structure. The proposed method is tested in the case of Beijing Metropolitan Region and the results are analyzed and compared with similar studies in Section 4. Section 5 conducts a comparison with other works and provides some policy suggestions based on the results. Finally, Section 6 summarizes our major conclusions and some points for future research.

## 2. Literature review

### 2.1 The spatial structure of metropolitan regions

The design of urban transport and land use policies are frequently on the basis of people's perceptions of the current spatial structure of cities or regions (Angel and Blei, 2016). These perceptions inform decision-makers of what can and should be done — in terms of public plans and investments as well as regulatory reforms of land use — to improve urban land use and transportation systems in the coming years. Therefore, scholars in related fields have been working on defining regions and their spatial structure, especially on the functional regions with complex structures, e.g., the Metropolitan Regions. A metropolitan region can be thought of a multi-functional region consisting of a densely populated urban core and its less-populated surrounding territories, sharing industry, infrastructure, and housing (Squires, 2002). From the perspective of spatial scope, a metropolitan region is similar with a large metropolitan area belt defined by Fang and Yu (2017), which usually comprises multiple mega-cities and tens of millions of populations.

In the abstract, the term metropolitan spatial structure can be regarded the discernible patterns in the distribution of human activity in cities (Anas et al., 1998), especially the discernible patterns in the distribution of residences and workplaces and the commuting flows that connect them to each other (Angel and Blei, 2016). The latter study argued there can be five types of spatial structures in cities: the Maximum Disorder model, the Mosaic of Live-Work Communities model, the Monocentric City model, the Polycentric City model, and the Constrained Dispersal model. Among them, the Polycentric City model was defined as that workers commute to a discrete set of identifiable employment sub-centers—including but not restricted to the CBD—located throughout the metropolitan region.

In recent decades, worldwide metropolitan spatial structure has experienced great changes along with population decentralization or regional integration. The classic monocentric model has gradually lost its power to explain these evolutions (Clark, 2000). In western cities, the polycentric model has been widely involved in metropolitan structure studies (Burger et al., 2011; Veneri, 2013), while currently the disperse model has also been proposed in some large western metropolises (Dong, 2013; Angel and Blei, 2016). As a contrast, the evolutions of metropolitan structure in developing countries are at a slow pace; most studies focus on the transformation of metropolitan regions from monocentric to polycentric (Fernandez-Maldonado et al., 2014; Hashem and Mehdi, 2017). In China, under the influences of both the market force and government interventions, many large urban areas, such as Beijing, Shanghai, Guangzhou and Shenzhen, also present polycentric structure (Liu et al., 2015; Huang et al., 2017; Lv et al., 2017), although the number and the size of employment sub-centers tend to be limited. Exploring the polycentric spatial structure can provide a wider knowledge of metropolitan spatial organization, which is significant to make scientific spatial planning policies and public resource assignments.

### 2.2 The characterization of metropolitan polycentric spatial structure

The previous studies on the characterization of metropolitan polycentricity frequently focused on one or more of these three broad issues: a) the identification on the regional sub-centers; b) the delineation of the metropolitan spatial extension; c) the measurement on the human activity performance (especially the employment performance).

A necessary first step in the characterization of polycentric MA concerns the identification of metropolitan sub-centers (Anas, Arnott, & Small, 1998). The identification of sub-centers can provide a wider understanding of metropolitan spatial organization, which is necessary for any spatial planning policy (Veneri, 2013). Numerous studies have examined the location of sub-centers and their boundaries by identifying centers (Veneri, 2013, Fernandez-Maldonado et al., 2014, Huang et al., 2017; Hu et al., 2018). Although various practical approaches have been proposed for identifying layout of sub-centers, the employment density-based indexes are most widely applied (Zhou et al., 2001; Angel and Blei, 2016; Guzman et al., 2017). Zhou et al. (2001), for instance, measured the centrality of a city using urban employment data for five industries in China. Considering the work-commuting flows do not represent all the movements that take place in a metropolitan region, we may neglect the urban nodes that can indeed be central for activities related to consumption, study and leisure in their way. As a consequence, it is necessary to distinguish the concept of employment sub-center from the wider one of urban sub-center. Veneri (2013) indicated that a metropolitan sub-center must have a minimum degree of productive variety and can supply a wide range of urban functions. The point density of origins and destinations (OD) of resident trips based on GPS trajectory data, involving a variety of human activities, can help us investigate which area has higher agglomeration capacity and productive variety in an urban system (Yue et al., 2012; Liu et al., 2015), which can be a rational centrality index for locating the CBD and other general sub-centers.

As a complex, dynamic and huge systems, metropolitan spatial structure are typically characterized by fuzzy boundaries. Defining the spatial boundaries of MRs from a variety of aspects is one of the traditional tasks in urban geography and planning (Ouředníček et al., 2018). A major reason behind the need to delineate the metropolitan regions is that official information at that scale are frequently based on administrative or legally-defined regions (Moreno-Monroy et al., 2020), while the latter cannot adapt timely to rapid changes in spatial population and economic activities, causing a persistent misalignment between legal and functional boundaries. Metropolitan regions are frequently delimited by functional approaches, relying on commuting ties between local units and regional centers (Bosker et al., 2019). In practice, for example, Japan set the standard of its metropolitan regions with the number of commuting population and the proportion of the population commuting to the central area of the metropolis in the 1960s (Fang and Yu, 2017). Since then, commuting density index has become a universally accepted determinant of the metropolitan circles' boundaries (Schleith et al. 2018; Ouředníček et al., 2018). Such methods are likely to be accurate to delineate metropolitan regions, but the lack of commuting data in many countries limit a global and consistent delineation (Moreno-Monroy et al., 2020). Another method frequently used in looking at the potential region scope is the accessibility measures. A trade-off between economies and diseconomies of commuting to metropolitan sub-centers can determine the growth boundary of MRs to some degree. One of the classic accessibility measures applied is the time-threshold based contour measure, also be called isochrone measure (Geurs & van Wee, 2004; Sánchez-Mateos et al., 2014). The isochrone measure provides evidence of the spatial scale expansion of urban regions by the increasing number of municipalities, people and jobs that can be reached within a certain time budget. Although this indicator is considered straightforward for implementation and interpretation, it has some theoretical shortcomings. First, the wide variety of travel time budgets used in literature means the difficulty of establishing a unique value of the time threshold, which greatly varies from country to country (Reggiani et al., 2011). Second, it does not take into account a distance-decay function to weight the opportunities (Sánchez-Mateos et al., 2014). Hence the area delimited by a travel time budget value should only be considered as a potential interaction metropolitan sphere.

There are also plenty of scholars focusing on the specific human activity performance of metropolitan polycentric structure, especially the employment performance, such as the regional job-housing relationship, interaction intensity between centers, and commuting efficiency. Two main approaches have been used to measure these performances– morphological and functional (Veneri, 2013; Sánchez-Mateos et al., 2014). The morphological approach is based on identifying nodes (centers) and characterizing them in terms of size and complementarities to other nodes (Giuliano and Small, 1993). A growing body of literature attempts to measure spatial structure by investigating the job-housing relationship for cities or regions (Wan et al. 2018; Zhang et al., 2017), while Lee & Gordon (2011) and Angel & Blei (2016) used the share of jobs in sub-centers (and CBD) to explore the whether a metropolitan structure has polycentric structure. The functional approach is based on characterizing centers by their interconnecting flows (Sánchez-Mateos et al., 2014). In previous studies, scholars mainly measured the spatial flows patterns in metropolitan regions from two perspectives. The first concerns the flow intensity. The flows of people and freight are key ties that connect the discrete physical resources of a city into an integrated system, and flow intensities can represent the spatial-interaction strengths between places. Based on the measurement of flow intensity to centers, a series of indexes were proposed to reveal spatial structure of cities or regions, such as the network dominance index (Limtanakool et al., 2007), the flow centrality (Veneri, 2013), the connection intensity (Zhen et al. 2017). The second focus is on the flow cost. (or travel cost), e.g. passenger travel time (or distance). Some scholars have studied the impact of polycentric structure on commuting time (Lin et al., 2015; Zhao et al., 2011) and others explored complex metropolitan structures by using a travel-time based accessibility index to show the interplay between the transport network and land use (Li et al., 2018; Sánchez-Mateos, et al., 2014). Furthermore, a number of scholars (Zhen et al. 2017; Chen et al., 2014) have suggested that a multi-criteria approach needs to be adopted to better understand the human activity performance of complex polycentric structure.

Furthermore, some scholars have recognized it is more rigorous and accurate to measure the performance of spatial structure on the basis of valid center layout and functional boundary in a given metropolitan region (Zhen et al., 2017; Sun and Lv, 2020). However, limited by data or just for convenient, most studies on metropolitan performance paid less attention on these two steps, but directly use directly took the lower-level administrative divisions as the regional centers and took the boundary of higher-level administrative division as the scope of whole study area.

### 2.3 The potential of carpooling big data in metropolitan studies

With the advent of the sharing economy era, on-demand carpooling services have become popular in many countries by their benefits of reducing travel costs, total fuel consumption, and carbon emissions compared to driving in single-occupancy vehicles. Carpooling trip data have two key advantages compared with conventional taxi trip data in metropolitan studies. First, smartphone-based carpooling mainly caters for commuting trips; commuting flows can be used to effectively uncover the spatial structure of an urban system (Angel and Blei, 2016). In general, non-professional carpooling drivers have their own jobs, so commuting is their primary travel purpose. Yongqi et al. (2018) conducted an empirical study on internet based ride-sharing travel patterns and demonstrated that carpooling primarily serves commuters from the perspective of data visualization and mathematical method. Second, the service scope of carpooling trips can spread over the whole metropolitan area. Carpooling can be a feeder for public transit to support commuting, and other travel activities, between suburban and urban areas, central and satellite cities. Some research has also implicitly viewed the application scope of carpooling as the metropolitan area (Xing et al., 2009; Najmi et al., 2017). Due to its commuting function and broader service scope, carpooling big data has huge advantages for exploring metropolitan spatial structures, which have not been utilized for metropolitan study to date.

## 3. Methodology

### 3.1 Identifying the study area

Beijing is located on the North China Plain and covers an area of 16,400 km$^2$. It includes 16 urban, suburban, and rural districts, with 21.71 million permanent residents in 2017 (BMBS, 2018). According to the new “Beijing General City Planning (2016-2030)”[1], the administrative region of Beijing has four different functional areas based on the layout of its urban space: a) the central city area (six inner districts including Xicheng district, Dongcheng district, Haidian district, Chaoyang district, Shijingshan district and Fengtai district); b) the city sub-center (i.e. Tongzhou district); c) the new city on the plain, including four suburban districts – Daxing district, Fangshan district, Changping district, Shunyi district, and one planned community – Yizhuang economic development zone, located within Daxing district; d) the eco-conservation area (the mountainous area, comprising of the five remaining districts). The locations of these four areas are shown in Fig. 1 (right). Based on the conceptual definition of Metropolitan regions, the Beijing Metropolitan Region (BMR) can be said to comprise of the highly-populated central city area and its surrounding close-connected territories. Most of previous works focusing on the BMR, simply took the Beijing administrative region as the study area (Long et al., 2013; Tian et al., 2010). Given the continual sprawl of this metropolitan region, however, we cannot determine intuitively whether Beijing's administrative boundary is identical to the functional boundary of BMR or not. In general, the size of the BMR ought to be smaller than Beijing-Tianjin-Hebei Urban Agglomeration (BTH-UA), i.e., the broad region covering Beijing, Tianjin and 11 prefectural cities of the neighboring Hebei Province, also shown in Fig. 1 (left). Therefore, we take the wider BTH-UA as our initial study area before delineating the BMR.

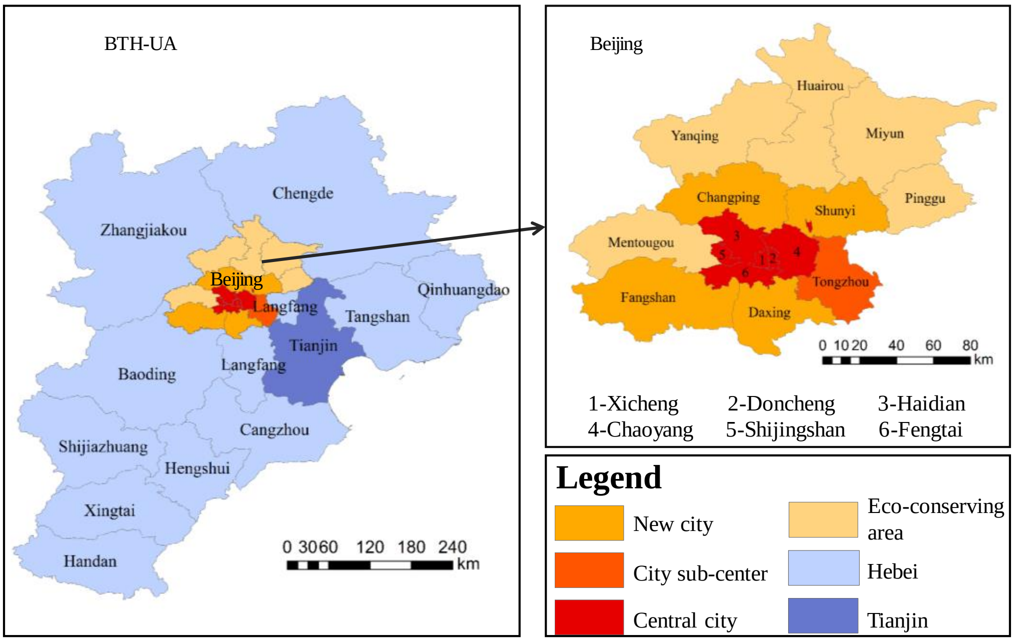



**Fig. 1**. The Beijing-Tianjin-Hebei Urban Agglomeration and the fourfold functional components of Beijing

Although we cannot ascertain, at this stage, the specific sphere of the BMR, we do know the urban area of Beijing is frequently regarded as the core of the BMR and even

[1] http://www.bjghw.gov.cn/web/ztgh/ztgh000.html

of BTH-UA. A preliminary visualized analysis of the spatial structure of Beijing was thus conducted using the density distribution of the OD points of the carpooling trips, as presented in Fig. 2. Most of the carpooling trips took place within the 6th-ring-road of Beijing, aggregating to be some highly-populated centers, while few people travel by carpooling in the outer suburbs. There are a large number of carpooling trips to/from railway stations and the airport, as well as to/from the traditional Central Business District (CBD). Intuitively, the BMR doesn't seem to have a uniform polycentric structure, but has one continuous large-scale settlement within the 5th-ring-road and some small-scale settlements scattered around the 6th-ring-road. In other words, the BMR has a hierarchical polycentric structure. In reality, this form of metropolitan structure is common globally, especially in developing countries (Lin et al., 2015).

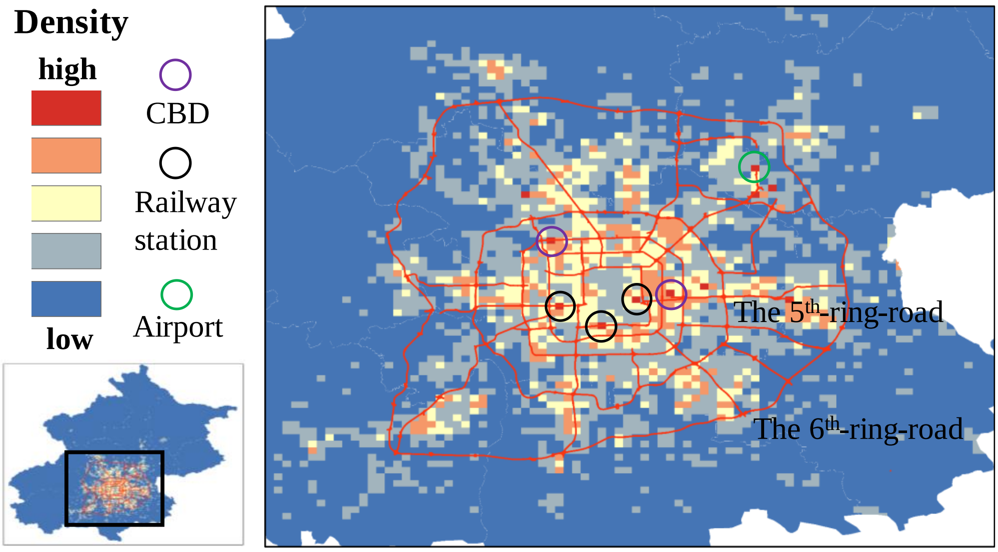



**Fig. 2**. The spatial distribution of carpooling trips in Beijing

3.2 Dataset and preliminary analysis

The dataset used here contains 15 million randomly sampled records of carpooling trips that occurred in BTH-UA between October 2017 and December 2017 (92 days in total). These carpooling trips were provided by an application-based system named DiDi Hitch, which was developed by the DiDi transportation company. DiDi is the largest ride-hailing service company in China and one of the largest on-demand ride sourcing service platforms in the world (DiDi, 2018). There are 922,021 carpooling drivers and 4,074,158 passengers included in the dataset. Each trip record includes a unique identifier for each driver and passenger, passengers' pick-up/drop-off locations (longitude and latitude) and the associated time stamp, as well as the actual distance travelled. Abnormal data where distance travelled was less than 1km or travel time was less than 5 minutes was removed from database, removing only 94,550 trips in total. To investigate the characteristics of the Beijing's carpooling big data, we conducted statistical analysis on the temporal and spatial distribution of the carpooling trips as shown in Fig. 3.

From the temporal perspective, the morning peak (7:00-9:00) and evening peak

(17:00-19:00) are obvious on workdays (Monday to Friday); up to 35% of daily trips are made during these times, while the same period on non-workdays only accounts for 26% of daily trips. In contrast, only 20% of conventional taxi trips are made within peak hours (Yongqi et al., 2018). This suggests a higher proportion of carpooling trips are made by commuters compared with taxi trips; this accords with the commuting function of carpooling trips demonstrated in previous works (Liu et al., 2019; Yongqi et al. 2018). For this dataset, we assumed that most carpoolers departing between 6:00 to 9:00 on workdays would be commuting for three reasons. Firstly, commuting trips in Beijing are generally concentrated within peak hours of workdays (BTI, 2018). There is no reason to suspect carpooling trips would be an exception. Secondly, people living in outer suburbs, especially out of Beijing, are likely to need more time to travel to their inter-city workplaces and thus may set off earlier. Taking Beijing as destination, for example, the percentage of inter-city carpooling trips departing to total trips from 6:00 to 7:00 on workdays is higher than the percentage departing during other hours; the former accounts for 12%, while later hours less than 4% on average. Thirdly, the evening peak is likely to include a higher proportion of leisure travel, with a proportion of commuters travelling to entertainment venues rather than going straight home (Yongqi et al., 2018). The inflection point of hourly carpooling trips at 19:00-20:00 shown in Fig. 3(top) may result from some people going home from entertainment venues.

From the spatial perspective, not only are there intra-city commuting carpooling trips, but some commuters travel from their residential cities to another one, shown in Fig. 3 (bottom). The average distance of morning commuting carpooling trips is 23.1km, which is much higher than the average distance travelled by other passenger transportation modes in Beijing, which are, for example, 9.9 km and 13.3 km for taxi trips and urban rail transit trips respectively (BTI, 2018). Moreover, the inter-city carpooling trips have a longer average travel distance (83.4km) compared to intra-city carpooling trips. This implies that the service scope of carpooling can exceed the administrative boundary of Beijing and the may spread throughout the BMR. Moreover, the influence sphere of BMR seems not accordance with the administrative boundary of Beijing. This analysis supports our premise that carpooling data can be used to represent commuting flows of the metropolitan region and characterize the metropolitan structure.

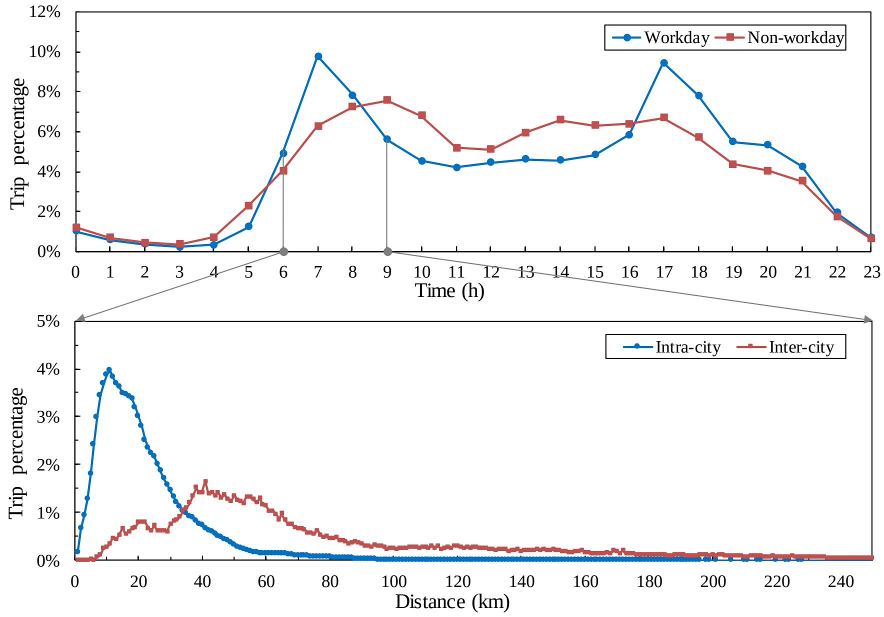



**Fig. 3**. Workday and non-workday carpooling orders number distribution by hour (top) and intra-city and inter-city commuting carpooling trips distribution by traveling distance (bottom)

Furthermore, we tested whether carpooling trips data could substitute for household travel surveys to describe the commuting demand of all residents. To do this we collected data on the size of the employed population for all cities in the BTH-UA to represent the real commuting demand, and explored its correlation with the distribution of carpooling trips. Considering Beijing's employment population and trip numbers have different orders of magnitude from the other cities, we took the logarithm for both variables, as shown in Fig. 4. With the R-squared and elasticity coefficients equal to 0.66 and 1.49 respectively, there is a relatively high positive log-linear correlation between commuting carpooling trips and commuting population. This suggests using carpooling trips made within morning peak hours to represent the commuting flows of residents in the BMR is a reasonable assumption.

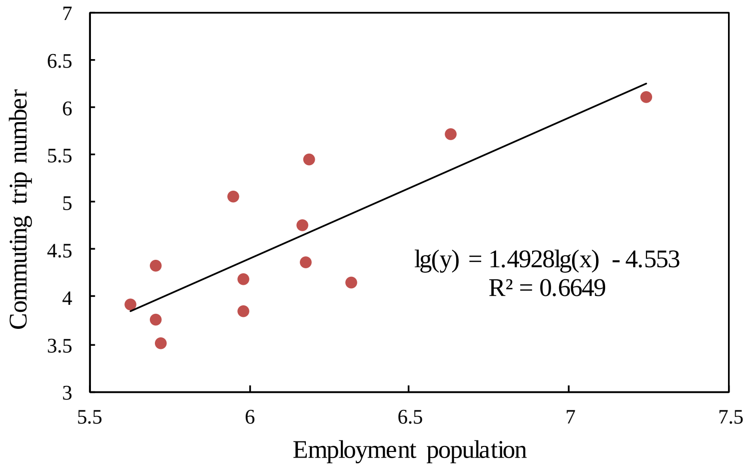



**Fig. 4**. Log-linear fitting for commuting carpooling trips and employment population

### 3.3 Methods and tools of data analysis

#### 3.3.1 Research framework

Given our preliminary identification of the BMR and analysis on the carpooling trips data, Fig. 5 outlines the three-step method used to measure the polycentric metropolitan structure. Firstly, we developed a grid-based clustering algorithm to identify the CBD and sub-centers of the metropolitan region. Secondly, we delineated the specific metropolitan functional sphere based on the regional commuting intensity and commuting accessibility to centers. Lastly, combining the morphological approach and functional approach, we developed two sets of indexes to measure human activity performance and investigate the possible role of each center, visualized by the last two concept maps, respectively. The multi-criteria quantitative indexes, including three density-based indexes and three flow-based indexes, estimated by the carpooling trip data within the defined metropolitan sphere. We would introduce the specific method and define the index system in more details in the subsequent sections.

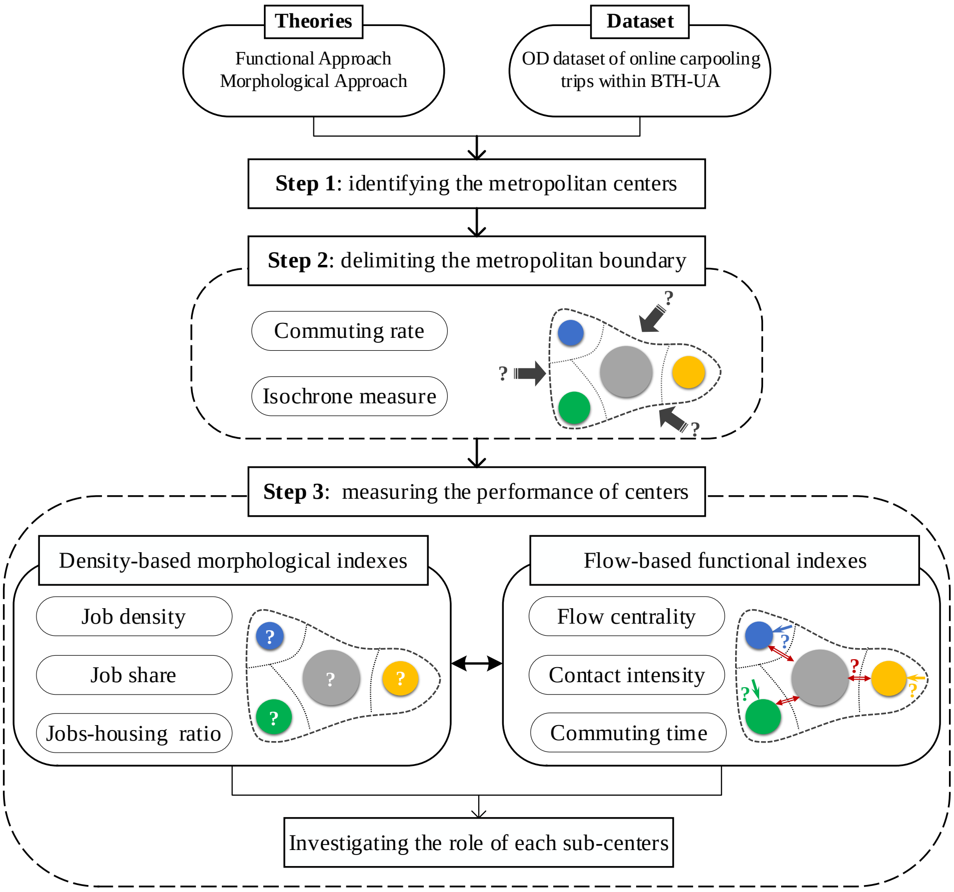



**Fig. 5**. Method framework of this study based on carpooling big data

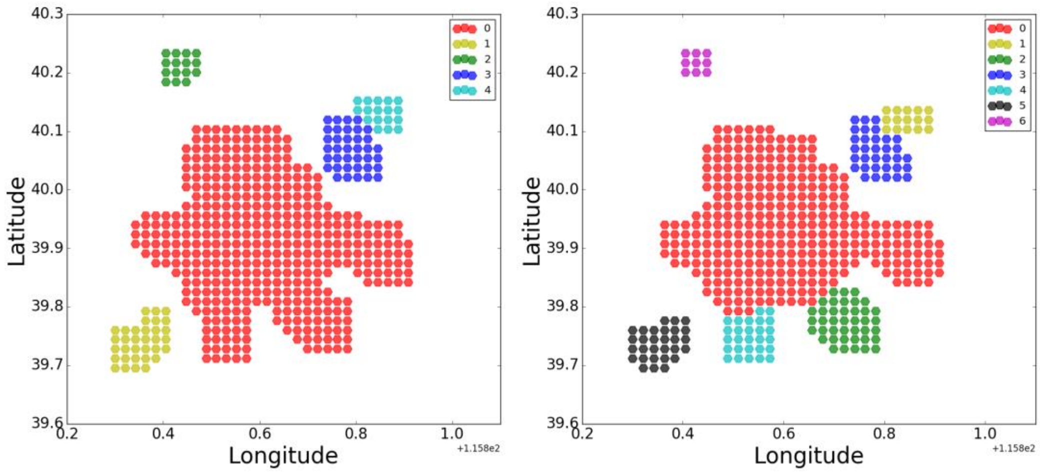
#### 3.3.2 Algorithm on identifying the regional centers

The Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm is widely used to form clustering in large scale data due to its simple calculation structure and low computing cost (Tang et al., 2015; Ester et al., 1996). Taking clusters of origin and destination points as metropolitan centers can transcend the limitation of administrative units.

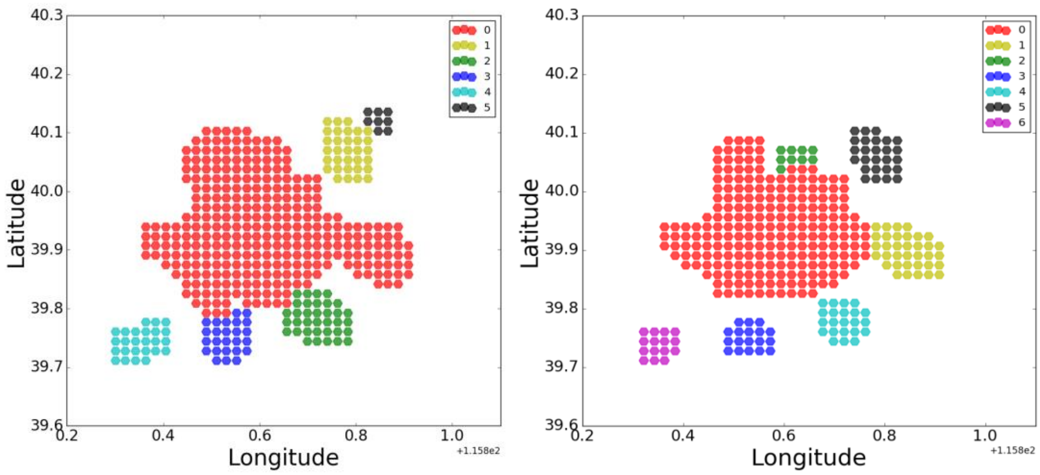
In carpooling trip dataset, although we know the position where a customer is picked up or dropped off, the exact place or building that the customer comes from or goes to is unknown. Given that a small spatial unit usually has a single land use, we can reasonably aggregate trips to obtain spatial interactions between these small spatial units. These small units could be traffic analysis zones (TAZs), grids, or parcels segmented by major roads. Due to a lack of TAZ data, we take grids as the basic unit of density clustering.

There are two parameters we need to set before conducting this grid-based clustering method (Liu et al., 2017). We set the parameter ε (search radius) as the smallest 2-cell neighborhood to guarantee the basic search scope only covers one adjacent unit in each direction and obtain accurate clustering results. As for the MinPts (the minimum number of OD points within the 2-cell search scope to form a cluster), we need to choose a rational value based on the local situation as follows.

Focusing on the region surrounded by the 6th-ring-road, i.e. the central city areas and inner suburbs of Beijing (see Fig. 2), we partitioned this area into 1,050 (30 lines×35 rows) cells with a unit area of 1.8km×1.8km; these are a similar size to the latest (2010) Traffic Analysis Areas (TAZs) for this area. We obtain preliminary cluster results based on four values of MinPts using the grid-based DBSCAN algorithm, shown in Fig. 6. Obviously, as MinPts rises, the total number grid cells within clusters reduces, but the separation between the central cluster and outer clusters increases. Compared with Fig. 6 (b) and (c), when the density threshold is 200,000, there are less clusters (only five) in Fig. 6 (a) and its central cluster (the red cells) is so dominant that it consumes some outer clusters. When the value of MinPts reaches 300,000 in Fig. 6 (d), the separation between the central cluster and the outer clusters is more evident at the cost of outer clusters vanishing. We take the cluster results with parameter MinPts=230,000 in Fig. 6 (b) as the final sub-center system; this captures more outer clusters whilst matching the five new cities on the plain, introduced in Beijing city plans (as shown in Fig. 1).



**(a)** **(b)**




**(c)** **(d)**

**Fig. 6**. The clustering results when the density threshold $minPts$ takes (a) 200000, (b) 230000, (c) 250000 and (d) 300000, respectively.

3.3.3 Method on delimiting the metropolitan influence sphere

We determine whether a certain region belongs to a metropolitan region based on a threefold judgment criterion: a) regional commuting population number; b) regional commuting intensity to the metropolitan sub-centers; c) regional commuting accessibility to the metropolitan sub-centers. We disperse the study area as grids under the GIS environment; a grid can be regarded as a part of the metropolitan region, if it has the certain commuting population, higher commuting interaction with metropolitan centers and is reachable within a rational time threshold. This grid-based boundary is dynamic and fully independent from local jurisdictions boundaries with cross-country comparability.

For the first judgment criterion a), therefore, we can exclude the grids generating less commuting trips than a preset lower threshold to extract the grids (regions) with sufficient commuting populations. For the judgment criterion b), we use the carpooling-based commuting rate (CR) as a measurement of the commuting interaction to the metropolitan centers. Based on the regional unit of grid, $CR_k$ here is the ratio between the sum of commuting carpooling trips $\sum_{i=1}^{m} N_{ki}^{o}$ from a certain grid $k$ to every sub-center $i$ and the total number of commuting trips $N_k^o$ from grid $k$, shown in Eq. 1. $\Omega = \{1, 2, \cdots m\}$ is the set of sub-centers and $i \in \Omega$; $\Phi = \{1, 2, \cdots n\}$ is the set of grids $k \in \Phi$. Note that the set of sub-centers is the subset of the set of grids, i.e. $\Omega \in \Phi$. A contour map of all grids' CR was used to visualize the distribution of sub-centers' influence; this was produced using the interpolation algorithm embedded in the ArcGIS software.

$$CR_k = \sum_{i=1}^{m} N_{ki}^{o} \Big/ N_k^{o} \qquad (1)$$

For the third judgment criterion c), the isochrone or contour measure can be used to define catchment areas by determining their limits within certain travel times to the metropolitan centers, assessing the number of accessible job opportunities within each time threshold. This isochrone measure is formulated in Eq.2 as an expression of accessibility index $AI$ depending on a Boolean function $x_k^t$ and on the sum of job opportunities to all centers $\sum_{i=1}^{m} N_{ki}^{o}$ from grid $k$. The Boolean function $x_k^t = 1$ if the commuting times of major carpoolers from grid $k$ to centers less than predetermined time threshold $t$ and $x_k^t = 0$, otherwise. Accessibility index $AI$ is the sum of commuting trips from all the associated grids.

$$AI = \sum_{k=1}^{n} \sum_{i=1}^{m} x_k^t N_{ki}^{o} \qquad (2)$$

To avoid the theoretical shortcomings mentioned in literature, in this paper, (1) we pick out the cells (grids) with sufficient commuting population and commuting intensity and visualize their spatial distribution as initial metropolitan sphere; (2) we depict a sequence of isochrone maps with different commuting time thresholds and select a isochrone map approximate to the former spatial distribution; (3) we delimit the metropolitan boundary based on the overlapping content of the former initial sphere and the latter isochrone map.

#### 3.3.4 Measurement on the performance of polycentric structure

We measure the performance of a metropolitan region based on two sets of indexes: three density based indexes including the job density (JD), job share (JS) and job-housing ratio (JHR); three flow-based indexes including the flow-centrality ratio (FCR), connection intensity (CI) and time-threshold based cumulative trip ratio (CTR). These indexes are calculated based on the information of carpooling trips within above delimited metropolitan sphere.

To investigate the morphological patterns of sub-centers, we used the employment aggregation performance of each sub-center as measurement indexes. 1) Job density (JD) is the number of jobs to each sub-center per unit area. Since the number of jobs for each area was not available, we used a proxy based on commuting carpooling trips; so it is in following indexes. 2) Job share ($JS_i$) is the percent of a sub-center's job number accounting for the total jobs within the metropolitan region, shown in Eq.3. The $N_i^d$ is the commuting carpooling trips to the sub-center $i$ and the $N_k^d$ is the commuting trips to the grid $k$. 3) Job-housing ratio ($JHR_i$) is the ratio of total employment number to local employed residents number within each sub-center, shown in Eq.4. $N_i^d$ and $N_i^o$ is the commuting carpooling trips taking sub-center $i$ as destination and origin, respectively.

$$JS_i = N_i^d \Big/ \sum_{k=1}^{n} N_k^d \qquad (3)$$

$$JHR_i = N_i^d \big/ N_i^o \tag{4}$$

To explore the functional performance of the sub-center system, three measurement indexes are proposed based on the carpooling trip flows between sub-centers, from the two perspectives of flow intensity and flow cost.

Flow-centrality ratio is another form of human activity based regional centrality index, besides the OD density. In this paper, flow-centrality ratio is the ratio of regional in-degree index to the associated out-degree index. The former represents the number of flows that directly enter each sub-center, while the latter is the number of flows that directly exit each sub-center. Hence the flow-centrality ratio ( $FCR_i$ ) for sub-center $i$ is computed based on the formula Eq.5, where the in-degree indicator $I_{ik}$ is the number of carpooling trips (or commuting carpooling trips) towards the sub-center $i$ from the grid $k$ and the out-degree indicator $O_{ki}$ is the number of carpooling trips (or commuting carpooling trips) from the sub-center $i$ towards the grid $k$. Note that any of the grid $k$ is not in the associated sub-center $i$. We can compare this functional centrality index with the trip density index we used in identifying the sub-center to examine the regional central role in a metropolitan network.

$$FCR_i = \sum\nolimits_{k=1}^{n} I_{ik} \Big/ \sum\nolimits_{k=1}^{n} O_{ki} \tag{5}$$

Connection intensity is another essential index to analyze the potential function of each sub-center. For the sub-center $l$, its connection intensity $CI_{lj}$ with sub-center $j$ is the percentage of carpooling trips towards sub-center $j$ from sub-center $l$ accounting for all carpooling trips from the sub-center $l$, where $l, j \in \Omega$ and $l \neq j$, shown in Eq.6. A higher value of $CI_{lj}$ means sub-center $l$ has a closer connection with sub-center $j$.

$$CI_{lj} = O_{jl} \Big/ \sum\nolimits_{i=1}^{m} O_{il} \tag{6}$$

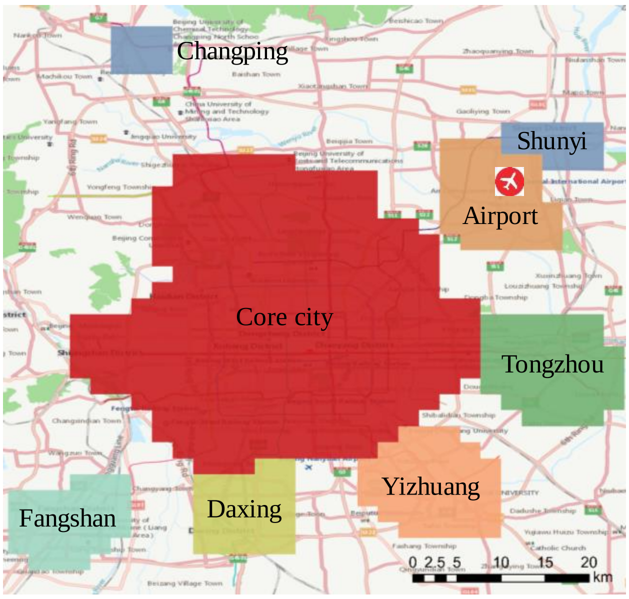
The commuting time distribution of passenger flows to each center can help us explore the level of flow cost and traffic performance in a given metropolitan network. Taking sub-center $i$ as a destination, the time-threshold based cumulative trip ratio ($CTR_i^t$) is the ratio between the sum of commuting carpooling trips $I_{ik}^t$ from grid $k$ that can reach the sub-center $i$ within a certain time threshold $t$ and all commuting trips $I_{ik}$ from grid $k$ to this sub-center, shown in Eq. 7. For example, a $CTR_i^{30\min}$ value of 0.75 indicates that 75% of all jobs (commuting carpooling trips) in metropolitan region can reach sub-center $i$ within a particular time threshold of 30 minutes. The use of a relative value eliminates ill effects due to the large variations of population scale between higher-order centers and lower-order centers.

$$CTR_i^t = \sum_{k=1}^{n} I_{ik}^t \Big/ \sum_{k=1}^{n} I_{ik} \tag{7}$$

# 4. Results

## 4.1 Clustering the layout of metropolitan centers

Mapping the clustering results onto the Beijing road network, we replaced the cluster codes with the name of corresponding administrative districts or planned districts that locate each cluster (Fig.7). These clusters identify the built-up areas of the inner urban and suburban areas. Note that we separated the Tongzhou cluster from the largest central cluster (red cells) considering its relatively isolated topologies and independent administrative attribution[2]. As expected, the current BMR is a hierarchical polycentric sub-center system. The majority of the area covered by the six inner districts of Beijing constitutes the core city (the largest cluster), i.e. the primary center or the higher-order sub-center. The built-up areas of the Tongzhou district and five new cities, as well as the new settlement around airport can be regard as secondary centers (or lower-order sub-centers). Some basic information on the sub-centers of the BMR is given in Table 1. Both the trip numbers and the area of the core city are larger than the sum of all other seven secondary centers together. The core city also has the highest OD point density; this further illustrates the core city's dominant role within the BMR. Of the secondary centers, the Tongzhou cluster is the largest in each value. With the smallest area among all the secondary centers, the Changping cluster has the second highest density; this may be due to its more intensive build-up area. Overall, 56% of carpooling trips are from or to these sub-centers and 78% of these trips associated with sub-centers pick up or drop off in the core city.




[2] Tongzhou district was declared as Beijing's administrative sub-center by local authorities in 2015.

**Fig. 7**. Spatial location of central cities in Beijing Metropolitan region

**Table. 1**. Basic statistical analysis of sub-centers

| Hierarchy | Primary | Secondary | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Centers | Core city | Tongzhou | Yizhuang | Airport | Fangshan | Daxing | Shunyi | Changping |
| Carpooling trip number ($10^3$) | 6872 | 1107 | 827 | 762 | 532 | 513 | 301 | 219 |
| Area ($km^2$) | 888 | 139 | 130 | 107 | 94 | 87 | 49 | 29 |
| OD Density ($10^3/km^2$) | 10.91 | 8.55 | 7.31 | 7.49 | 6.10 | 6.16 | 6.51 | 7.69 |

4.2 Delimiting the metropolitan boundary

Before defining the boundary of this metropolitan region, the broader area of BTH-UA was divided into 7000 grid cells (70 lines×100 rows), each with an area of 7km×7km. The larger grid cells (than the grid used in sub-centers identification) are to ensure the sufficient carpooling trips and commuting population in each grid cell that the proposed positive correlation between commuting carpooling trips and commuting population applies.

For the first judgment criteria of regional commuting population constraint, we preset a filter threshold of 65 trips per grid cell and remove cells with less origin points for commuting rate estimation. Sixty-five trips could guarantee there is at least one trip every workday on average during the three months covered by the sample data. There are 657 grids cell left, less than 10% of total grid cells.

For the second judgment criteria of commuting intensity, the commuting rate of each grid cell to sub-centers is estimated by the Eq.1. Then we used the Kriging interpolation method to smooth the commuting rate spatial distribution and produce a contour map of the commuting rate, shown in Fig. 8. Note we take the commuting rate of 5% as the lower commuting intensity threshold and we only include and depict the grid cells with commuting rate beyond this threshold. The region comprised by all of these qualified grids is defined as the metropolitan commuting sphere (MCS).

Remarkably, the metropolitan commuting sphere of the sub-centers is beyond Beijing administrative district, gradually decaying from inside to outside the BMR. For the continuous settlement areas, commuting rates spread in the shape of concentric rings over the south-central region of Beijing with the core city as the heart; the sub-center commuting rate of the innermost rings exceeds 80%. Unsurprisingly, the sub-center commuting rate of the eco-conserving area is less than 5% due to the limitations imposed by the mountainous geographical environment. There are also some relative isolated pockets separated by rural areas, especially in the surrounding cities beyond the Beijing administrative district, like Baoding city, Zhangjiakou city, and Langfang city (see Fig. 8). For the scattered pockets with higher commuting rates, these commonly aggregate and distribute along the expressways (the red lines); this demonstrates the important role of high grade transportation facilities in the process of urban evolution. For example, Tianjin is a developed city that has strong communication links with Beijing and other cities in BTH-UA. The level of commuting by carpooling between Tianjin and the sub-centers of Beijing, however, is very low, maybe because the Beijing-Tianjin inter-city railway, with its high speeds and high departure frequencies, provides a more attractive option for travelling between these two cities than carpooling.

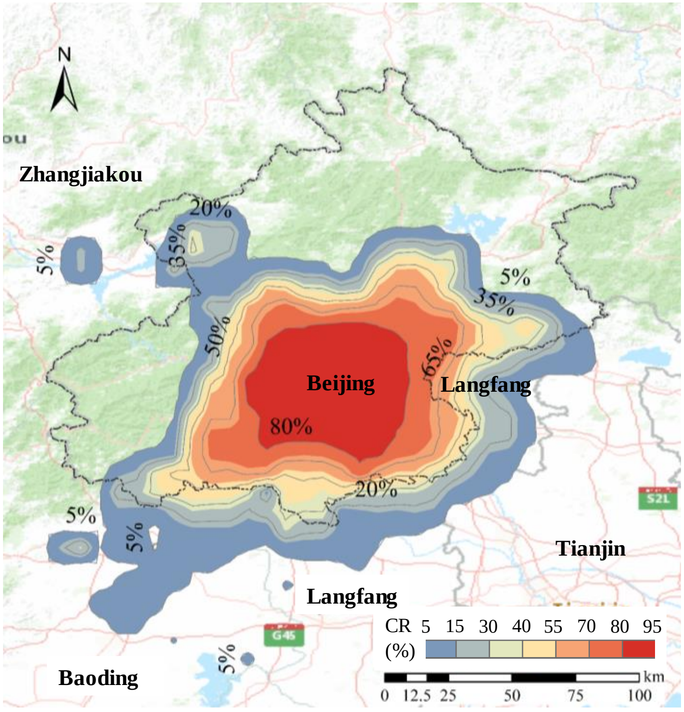



**Fig. 8**. Contour map of commuting rate to the sub-centers of the BMR. The regions with commuting rate beyond 5% are defined as the metropolitan commuting sphere.

For the third judgment criteria of commuting accessibility, the multiple-time-threshold commuting isochrones are calculated and shown in Fig.9. If there are more than half of commuters from a certain grid can reach the sub-centers within 1 hour, we regarded these regions are 1-hour accessible, shown as the dark green grids; similar for the other time thresholds. The travel time thresholds take from 1 hour to 3 hours, step by half hour. It can be seen that the 2.5-hour accessible regions are approximate to the scope of above MCS. Hence we define the overlapping region that are 2.5-hour accessible and with commuting rate beyond 5% as the BMR; it covers about a 100km radius of region around the Beijing core city and can be regarded as the outer commuting circle. BMR excludes the mountainous areas of Beijing and extends beyond the administrative boundary of Beijing and further to the adjacent counties of Baoding and Langfang city, which involves 23 counties in BTH-UA and about 30 million people (in 2016). Furthermore, all of these sub-centers are within the 1.5-hour accessible regions and covering a 50km radius circle and these inner areas can be regarded the core commuting circle of the BMR. Compared with the previous related study with a study duration from 1995 to 2010 (Shi and Cao, 2020), the spatial range of BMR delimited in this paper is broader and radiating more adjacent jurisdictions but not based on the administrative units. This shows that these regional centers have strong attraction and caused the continuous expansion of BMR.

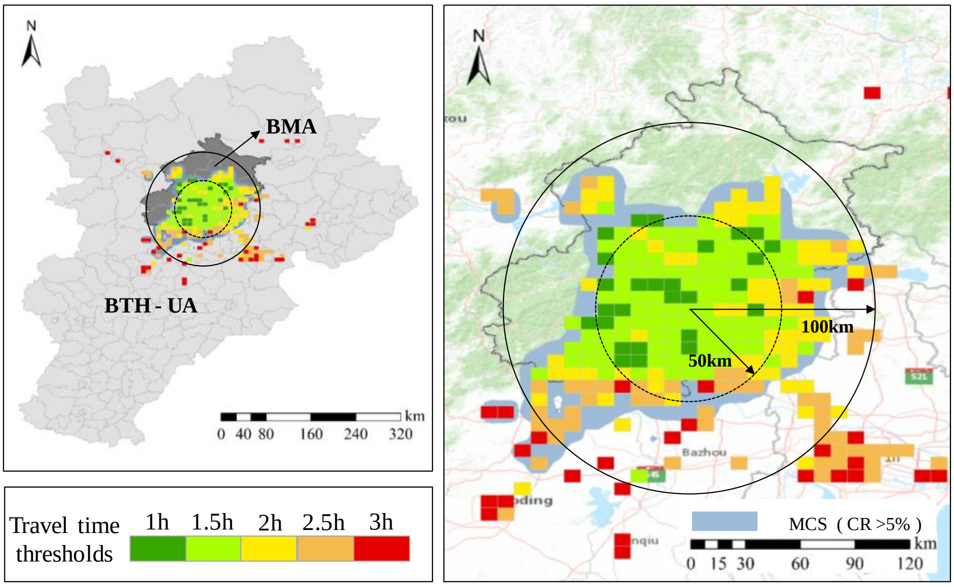



**Fig. 9**. Multiple-time-threshold based commuting isochrones and the influence sphere of BMR. For better presentation, we excluded the grids with commuting trips less than 10.

For further understanding the defined metropolitan influence sphere, we conducted the statistical analysis on the specific commuting accessible trips and regions, shown in Table 2. More than half of commuting trips cannot reach the sub-centers within 1.5 hours in this metropolitan region. When up to 2.5 hours, the majority of commuters (96%) can reach these sub-centers; this also supports our previous decision on selecting the 2.5-hour threshold in defining the BMR. Moreover, from 1-hour to 2-hour, there is a significant gap between the actual accessible trip number and the expected accessible trips number calculated based on the Eq.2; this reflects the strong fluctuation of commuting times within the core circle of BMR because of the serious road congestion. The area of accessible regions is not totally consistent with the area of accessible regions with sufficient commuting intensity and the differences between them become wider along with the ascending time thresholds; this demonstrates that a longer travel time can erode the regional commuting intensity to metropolitan centers, especially for the outer commuting circle area.

**Table. 2**. Multiple-time-threshold commuting accessible trips and regions

| Time thresholds | 1h | 1.5h | 2h | 2.5h | 3h |
|---|---|---|---|---|---|
| Actual accessible trips | 67744 | 538406 | 733581 | 1072096 | 1098538 |
| Expected accessible trips | 128794 | 939406 | 1100748 | 1113519 | 1114804 |
| Actual accessible trip ratio | 6% | 48% | 66% | 96% | 99% |
| Total accessible grid number | 27 | 127 | 192 | 246 | 280 |

| | | | | | |
|---|---|---|---|---|---|
| Accessible grid number within MCS (CA>5%) | 25 | 121 | 176 | 209 | 226 |

### 4.3 Measuring the performance of the metropolitan region

#### 4.3.1 Qualifying the employment aggregation performance

For the morphological patterns of sub-centers, we qualified the employment aggregation performance based on the carpooling big data and three indexes are shown in Table. 3. As expected, the higher-order center, the core city of Beijing is the most important employment agglomeration zone as it has the highest job density and job share (beyond 60%) in the BMR. The core city's JBR of 144.4% shows its serious imbalance between the living and working provision for citizens. In total, 81.9% of commuters take the core city and sub-centers as their destination; this also demonstrates that the hierarchical polycentric structure of BMR with a dominant core center. This total proportion is highly larger than the jobs share of employment centers including the CBD for the 50 largest metropolitan regions in the U.S. (24.6±1.8% in 2000, Angel and Blei, 2016). Compared with the constrained dispersal form of American cities, the BMR still does not have a single, integrated labor market where workers and workplaces are matched at a truly metropolitan scale. Although local government planned Tongzhou to be an administrative sub-center of Beijing, so far it mainly provides housing for people working in the core city, which has the lowest JBR and the second lowest jobs density. Fangshan also performs poorly for local employment attractions with the lowest job density. As the only national Economic-Technological Development Area (ETDA) in Beijing, Yizhuang has these three indexes ranking second only to the core city. The new city built surrounding the Beijing Capital International Airport also attracts plenty of job-seekers from the BMR. Distinctively, Shunyi has a good job-housing balance and a moderate job density.

**Table. 3**. The employment aggregation performance of sub-centers in the BMR

| Centers | Core city | Tongzhou | Daxing | Yizhuang | Shunyi | Airport | Fangshan | Changping |
|---|---|---|---|---|---|---|---|---|
| Job density (per $km^2$) | 918.86 | 327.40 | 332.25 | 731.20 | 455.49 | 524.63 | 231.02 | 383.09 |
| Job share | 60.9% | 3.5% | 2.1% | 7.1% | 1.7% | 4.2% | 1.6% | 0.8% |
| Job-housing ratio | 144.4% | 41.6% | 65.8% | 129.1% | 91.9% | 128.4% | 46.1% | 62.4% |

#### 4.3.2 Discerning the flow interaction performance

For the functional patterns of sub-centers, we discerned the flow interaction between the sub-centers based on the carpooling big data and three indexes including flow-centrality ratio, connection intensity and time-threshold based cumulative trip ratio are calculated; the former two indexes are shown in Table. 4.

Considering the diverse activities and commuting trips, we estimated the multi-flows centrality and commuting-flow centrality, respectively, based on the formula Eq.5.

Considering the regions satisfying FCR>1 as the metropolitan first-order centers, most of sub-centers, even the core city, are of poorly flow-based centrality; these results are highly different with the identification of OD density based centrality. Core city and other two employment centers (Yizhuang and Airport) perform prominent in the commuting-flow centrality, while other centers still cannot reach the threshold value (FCR=1). Especially, affiliating to the core city and without a local employment base, Tongzhou is with the lowest commuting-flow centrality. We conjecture that the forming and growing of BMR's polycentricity can be more of the result a decentralization of employment from a congested core city (or CBD) than the consequence of a coalescence or integration process, like many European metropolitan regions (Veneri, 2013). The decentralization here can be defined as the movement of populations and their activities (residential function, employment, services, administration, etc.) from the core cities to the hinterland. Therefore, density measures based on the former idea can be more appropriate to be used as the centrality indexes. Except for Shunyi, all the outer (lower-order) sub-centers have highly close connections with the core city (CI >70%), which also reflects the dominated function of the core city within in the BMR. As for the connection intensity of core city to outer center secondary centers, beyond one fourth of passenger flows from core city are towards the Tongzhou; this indicates the construction of this administrative sub-center has taken effect and shared the huge population pressure of core city. The Shunyi has barely connection with core city, but has a relative independent status in this metropolitan region.

**Table. 4**. The connections between the core city and lower-order sub-centers within the BMR

| Places | Core city | Tongzhou | Daxing | Yizhuang | Shunyi | Airport | Fangshan | Changping |
|---|---|---|---|---|---|---|---|---|
| Multi-flows centrality | 0.96 | 0.97 | 1.01 | 0.96 | 0.91 | 1.22 | 1.03 | 0.97 |
| Commuting-flow centrality | 2.78 | 0.34 | 0.63 | 1.46 | 0.91 | 1.33 | 0.39 | 0.61 |
| CI (outer centers to core city) | / | 76.7% | 72.8% | 71.9% | 51.2% | 76.4% | 84.1% | 87.0% |
| CI (core city to outer centers) | / | 27.2% | 11.9% | 17.9% | 3.5% | 21.8% | 12.5% | 5.1% |

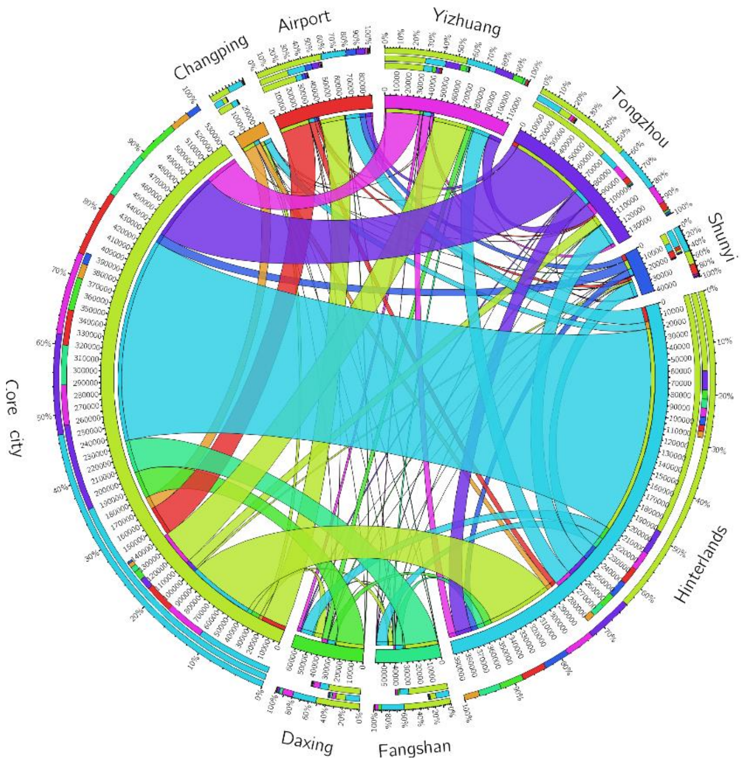



**Fig. 10**. Commuting flow distribution based on commuting carpooling trips within BMR

To further explore the commuting interactions between sub-centers within this metropolitan, we depicted the commuting flows between various orders of sub-centers and the associated hinterlands in a Sankey diagram (Fig.10). 71.4% of the commuting carpooling trips are related to the core city, whose workers mainly come from the broad hinterlands of Beijing, Tongzhou and other cities. Reverse commuting trips from the core city to the secondary centers account for 24.9% of the total commuting trips from the core city. Most of these take new employment sub-centers (Yizhuang and Airport) and the hinterlands as destinations and are the most important part of the local employment sources. This result can be regarded as evidences of metropolitan suburbanization and the polycentric nature, which is accordance with our viewpoint in the forming of BMR. Notably, more than two thirds of the external commuters to Shunyi are from neighboring communities; this embodies Shunyi's function as an employment base for local citizens. Apart from the core city, commuting connections from Shunyi to Airport and from Tongzhou to Yizhuang are also very strong, maybe due to their adjacent geographical locations.

As a measure of flow interaction cost, the time-threshold based cumulative trip ratio (CTR) of carpoolers departing to each center within morning peak hours were computed from 0 to 3 hours by a 15-minute interval. These distribution curves of the eight centers are shown in Fig. 11. At first glance, the cumulative commute time distributions of trips to the eight sub-centers are similar. These each distribution curve is composed of double S-shaped curves of short-distance trips (travel time <1.5h) and long-distance trips (>1.5h) and there is an obvious flat segment neighboring the 1.5-

hour join line. The S-shaped curves of short-distance trips rise sharply at each side of the 30-minute time threshold, while the S-shaped curves of long-distance trips show dramatic changes around the 2-hour time threshold. These characteristics of the curves demonstrate the uneven distribution of carpoolers' commuting times. Most of short-distance commuters need to reach their workplaces within 1 hour, while most of long-distance commuters will finish their trips within 2.5 hours. According to the previous results of commuting isochrones, the short-distance trips to these centers are mainly from the core city and its adjacent centers, while most of trips from the outer suburbs or other cities are the long-distance trips.

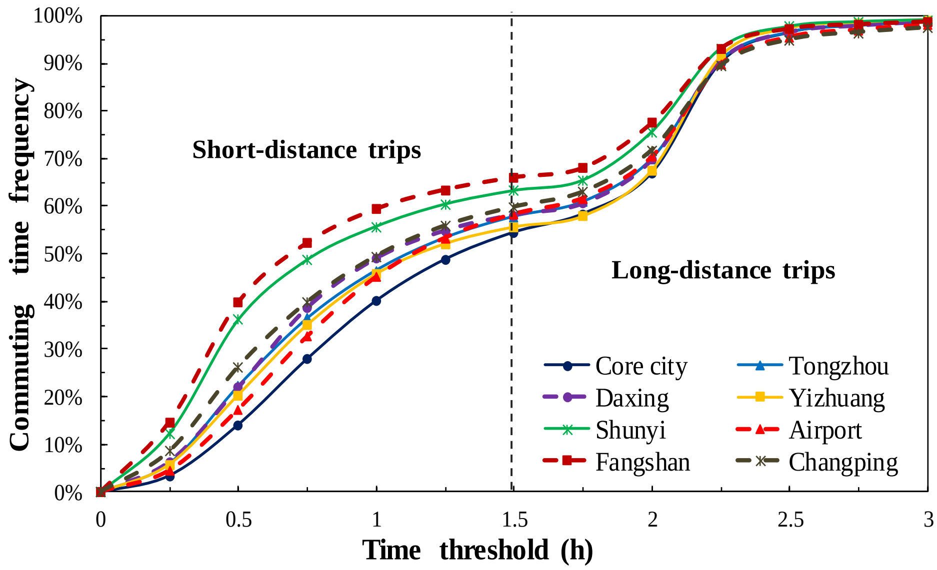



**Fig. 11**. Cumulative commute duration frequency distribution of carpoolers travelling to regional centers with varying time thresholds

Although the cumulative commuting time distributions shown in Fig. 11 are similar, there exist obvious differences among different destinations. Carpoolers working in the core city need the longest commuting time and nearly half of them cannot reach their workplaces within 1.5 hours, while commuters to outer sub-centers spend less time. Carpoolers to Fangshan often need the least time cost; there even is more than a 20% difference between the core city and when the commute time threshold is 30 minutes. Carpoolers travelling to Yizhuang and the Airport settlement, both of which perform well in terms of employment attractions, also spend considerable time commuting. People living and working in the employment centers show a higher tolerance to long-distance commutes. A number of studies have found that a longer commute time is associated with lower levels of both life satisfaction and happiness (Kahneman et al., 2004; Choi et al., 2013). In the developed cities of China like Beijing, these negative correlations are also significant, especially when commute times are more than 1 hour per trip (Nie and Sousa-Poza, 2018; Yin et al., 2019). Beijing government planned to reduce its average commuting time within the Fifth Ring Road (similar with the core city in this paper) from 97 minutes in 2014 to 60 minutes in 2020[3]. However, except for those travelling to the centers of Shunyi and Fangshan, less than 50% of carpoolers can reach a center within one hour during morning peak hours. The average driving speed of carpooling commuters within the BMR is only 22.17 km/h, illustrating the severe traffic congestion problems in this mega metropolitan region. According to previous studies or reports on commute in Beijing, the average commuting time to regional centers are from 30 minutes to 50 minutes (Lin et al., 2015; Hu et al., 2018; BTI, 2018). The obvious differences can be partly due to the longer travel distance of carpooling service and partly due to the broader study area. Overall, the performance of the road network in the BMR seems lower than the expectations of citizens and decision makers.

### 4.3.3 Investigating the role of each sub-center

Using the proposed spatial indexes, including the job-density, job-housing ratio, the workforce source composition, resident employment distribution and connection intensity of each sub-center, the driving force of sub-center forming and primary role of each sub-center in this metropolitan region can be revealed, which are listed in Table 5.

Taking the BMR as an example, the core city has dominant performance in all sorts of indexes due to its strong employment and residence centralization among this metropolitan region. There is no doubt that core city is the primary center of BMR. Reverse commuting trips (trips from core city to secondary centers) account for nearly 50% of total commuting trips to the Yizhuang and Airport (see Fig.10), which means the employment decentralization from the core city is the important cause of forming these two sub-centers. Yizhuang and Airport have the higher employment aggregation performance only inferior to the core city (see Tab.3) and the commuting-flow centrality ratio above the threshold value (see Tab.4), hence they can be regarded two employment sub-centers of the BMR. In contrast to Yizhuang, the Airport (and its associated built-area) has a close connection with the core city, maybe because of its special function as a transportation hub. In contrast, Tongzhou and Fangshan have the lowest local jobs density and JBR (see Tab.3); most of commuters (about 70%, see Fig.10) from these two sub-centers are towards the core city. These places grow and evolve mainly by residence decentralization from core city, maybe because the higher living cost and house price of the latter, which be regarded as commute towns surrounding the core city. There still is a long way for Tongzhou to be the administrative sub-center. As for Daxing and Changping, it is difficult to directly indicate the driving force of regional development and define their functional property considering their mediocre performance in both employment aggregation (see Tab.3) and commuting distribution (see Fig.10). Therefore, we tentatively identify them as mixed-role cities forming by mixed forces, which can evolve by more than one trajectories possible in

[3] http://www.ebeijing.gov.cn/BeijingInformation/BeijingNewsUpdate/t1397427.htm

the future. Considering the longer travel distances to the core city, the residents of Changping need to pay a higher commuting cost for working in the core city, so Changping is more likely to become a satellite city under sustained economic development, while Daxing is more susceptible to becoming another employment sub-centers, if decision-makers adopt powerful measures to improve local employment attraction. Compared with other centers, Shunyi shows its specificity in many quantitative indexes: its job density is not very high, but has relatively balanced Job-housing ratio, commuting-flow centrality close to 1, and less connection with core city (see Tab.3 and Tab.4). As a local employment base, Shunyi with is relatively independent of the core city in terms of both mobility connection and geographic location. 70% of commuters towards Shunyi are from its surrounding hinterlands (see Fig.10). The forming of this center can be a result of spatial coalescence or integration process, by the extension of the metropolitan influence over close systems of small and medium-sized cities. Hence we can consider Shunyi as a satellite city of Beijing downtown.

**Table. 5**. The forming process and main roles of sub-centers in the BMR

| Places | Driving force | Regional role |
|---|---|---|
| Core city | Employment and residence centralization | Primary center |
| Tongzhou | Residence decentralization | Commuter town |
| Daxing | Mixed-forces | Mixed-functions city |
| Yizhuang | Employment decentralization | Employment sub-center |
| Shunyi | Spatial integration | Satellite city |
| Airport | Employment decentralization | Employment sub-center and transportation hub |
| Fangshan | Residence decentralization | Commuter town |
| Changping | Mixed-forces | Mixed-functions city |

# 5. Discussion

5.1 Comparison with studies on metropolitan spatial structure

Comparing with previous works on typical metropolitan regions or urban regions, either in developed country or developing country, the method developed in this study has the advantages of being more reliable, accurate and comprehensive.

First, the advantage of reliability in this research is manifested by the fact that the carpooling big data used in this paper is dynamic, massive and more applicable to metropolitan study. Most of previous studies on metropolitan structure based on survey data or secondary data may be limited by the periodicity and subject of surveys; thus it is difficult to obtain updated and independent conclusions. For example, Angel and Blei (2016) recognized that a number of important recent changes, like revival of city centers and CBDs as centers of employment, have occurred in the intervening 15 years, raising the question as to whether their conclusions still hold. Burger et al. (2011) and Veneri (2013) used commuting flow survey data to uncover the spatial structure of city-regions in British and Italy, respectively. Several authors, however, have pointed out that journey-to-work travel should be used with other indicators to provide realistic insights into the interdependence of places and structure in urban systems (Lambregts et al., 2005; Parr and Hewings, 2007). Studies attempting to reveal the city structure based on other emerging big data, such as taxi trip data (Liu et al., 2015), fail to show the metropolitan characteristics due to lack of information on long-distance commuting within a given metropolitan region. We can extract reliable, up-to-date and consistent information on the urban spatial structure based on carpooling big data, which is vital for numerous applications central to urban planning and land use analysis.

Second, rather than using the administrative divisions of Beijing Municipality, we clustered the polycentric layout under a grid-based, GIS-enabled environment and delimited this metropolitan sphere based on a threefold criterion. The identification methods on study area are more rigorous and the associated results can be more accurate. In fact, there would be a significant difference in the value of some indexes based on different metropolitan spheres, when the inter-city trips were identified incompletely. Within the Beijing Metropolitan Region (BMR), the inter-city trips beyond the municipal boundary of Beijing accounted for 11.2% of total trips. These trips obviously have different spatial-temporal characteristics compared with the trips within Beijing. More specifically, the inter-city trips had much longer travel distances and travel times (31.3km and 92min on average) than the latter (20.3km and 71min on average). If we simply took Beijing Municipality as the case study, we would not only miss the chance of understanding the flow-base patterns of these inter-city trips, but also cause a considerable estimation error of some density-based indexes, especially for the outer sub-centers. For example, the differences of job density and job share would reach 13% and 8% for Tongzhou and be up to 25% and 31% for Fangshan, comparing using the Beijing administrative boundary with using the defined metropolitan sphere. Considering the common existence of inter-city trips in other urban areas, this type of incomplete analysis on metropolitan structure and corresponding estimation errors may exist in other studies. Moreover, this delimiting method can provide effective alternative boundaries for metropolitan planning, especially in highly dynamic cities such as Beijing.

Last, to uncover the metropolitan spatial structure in-depth and comprehensively, we combined the density-based morphological and flow-based functional approach based on a twofold index system. If we only use the employment density-based methods to measure the performance, we may miss the chance to investigate the commuter towns, like Tongzhou in BMR, and the sub-centers without any particularly high employment density, but still as a local center of the metropolitan territory, like Shunyi in BMR. On the other hand, if we measure the polycentric structure only by interaction flows, like some studies (Limtanakool et al., 2009; Veneri 2013), it is difficult to find the metropolitan centers in accordance with the real-world, referring to the flow-centrality indexes in Tab.4. Hence, the combination of a morphological and functional approach can avoid drawing lopsided conclusions to some degree. Compared with the macro-research focusing on the structure of tens of cities, e.g., Burger et al. (2011) in English and Welsh and Angel and Blei (2016) in America, this work first proposed a more delicate method to in-depth investigate each center in a given metropolitan region.

### 5.2 Takeaways for practice

The emerging of on-demand carpooling services generate massive trip data that have commuting function and broader service scope. This provide us a good chance to understand the metropolitan structure better and then support making metropolitan development planning. Based on the results of this paper, some extended suggestions are listed as the takeaways for practice, not only for BMR, but also for other cities.

First, an effective policy change in transportation and land use patterns, including the regulations, taxes and subsidies and public investments, shall focus on helping the great majority of actual travelers, especially the commuters, with the least expense. Hence, we can divide the metropolitan regions with polycentric structure inner and outer two commuting circles to make the associated policies that can facilitate commuting by promoting the transport modes and routes, respectively. For the inner (core) commuting circle covering all centers with higher job density, the authorities should focus on reducing the gaps between expected traveling times and actual ones by relieving the road traffic congestion. For this issue, we can encourage the ridesharing modes, improve the service level and extend the capacity of public transport. For the outer commuting circle covering the broader hinterlands, the authorities should seek to guarantee the mobility demand of longer-range metropolitan travelers to reach their destinations quickly and economically, especially for commuters during the rush hours. For example, we can build the suburban or intercity railways and link them with the inner metro networks to reduce the proportion of long-distance trips by car. Local planners should seek to strike a balance between keeping the attraction of the metropolitan centers and avoiding excessive urban sprawl when developing their polycentric development strategies.

Moreover, when making local policies, we shall consider the specific role of each sub-center within a given metropolitan region. The metropolitan development planning treats all centers without difference can waste the social resources or even hinder the normal development of local city. For the primary center (core city) with the highest job shares and unbalanced job-housing relationship, planner should try to optimize the job-housing distribution (e.g., encouraging local employment in the metropolitan sub-centers and hinterlands) and improve the urban carrying capacity. For the employment sub-centers, to reduce commute travel and to improve quality of life in the long-run, it is important to plan and provide the housing and services suitable for local workers, while for the commuter towns, we shall pay more attention to the construction of local residential infrastructure and the promoting measures on the transport modes and routes from these towns to core city. For the satellite cities with the potential to be a new metropolitan region, policy-makers should focus more on its link with surrounding hinterlands, rather than its connection with core city. For the mix-functions city, the first thing for authorities maybe is to determine a clear regional development orientation before making the associated planning.

Although we take the BMR as a case study, the application of carpooling big data and the proposed method of identifying the polycentric structure would provide a novel perspective for research on other metropolitan regions. Like many emerging metropolitan regions in the developing world, BMR has a polycentric structure, a large but under-developed hinterland, and an ambitious local authority with a strong intention to create a mega-region (Shi and Cao, 2020). For the data availability, carpooling services have been emerging in many large cities and their associated metropolitan regions. Table 6 lists several current online carpooling services provided by the major platforms and their respective development scales. Hundreds of millions of carpooling trips in hundreds of cities generate massive data that can be used in metropolitan studies. More specifically, in the UK, the majority of metropolitan regions is with polycentric forms (Burger et al., 2011), and the local social enterprise Liftshare has more than 500,000 active members, who share more than 1 million journeys each month[4]. In Shanghai, another mega-city with polycentric structure of China, there are about 800 thousand carpooling trips through the Didi Hitch APP during one month (September, 2017). Therefore, the research framework and some conclusions on BMR in this paper may have potentials to be applied to the other metropolitan regions for a similar research purpose, which gives this research a global relevance.

**Table 6.** The characteristics and scales of online carpooling services provided by typical platforms (Data source: the official websites of the respective TNCs)

| Major platforms | Launch time | Trip purpose | Popular regions | Service scale |
|---|---|---|---|---|
| Blablacar | 2006 | Long-distance trip including commuting | 22 countries mainly in Europe and Latin America | 87 million users, 30 billion kilometers shared since 2003 |
| Didi Hitch | 2015 | Diverse, mainly for commuting | 351 major cities in China | 30 million registered drivers; up to 2 million daily orders |
| Waze Carpool | 2016 | Commuting | America, Brazil, Mexico | 60 million users, up to 1 million monthly orders |

## 6. Conclusion

As social, economic and political institutions have changed, contemporary MRs are characterized by more complex spatial structures. Fortunately, the rapid development of big data technology offers us an opportunity to better measure the metropolitan polycentricity and then make targeted metropolitan land use and transport planning. Using carpooling big data, we identified the polycentric layout of Beijing

[4] https://business.liftshare.com/

Metropolitan Region based on a grid-based clustering algorithm. Then we delimited this metropolitan using the overlapping area of higher commuting intensity region with sufficient population and 2.5-hour commuting contour. Lastly, a two-group index system was established to measure the performance of metropolitan polycentricity. This three-step method driven by carpooling big data are more reliable, accurate and comprehensive, based on which we provide some valuable insights to global knowledge.

Regional centers identification and boundary definition shall be the first two necessary steps before conducting in-depth analysis on human activity performances of metropolitan polycentric structure, while the combination of a morphological and functional approach can avoid drawing lopsided conclusion on these performances. The emerging carpooling big data with commuting function on a metropolitan scale can help realize these approaches.

The polycentric metropolitan region represents a hierarchical center system: one primary center interacting with seven surrounding secondary centers. These regional centers have such a strong attraction that results in the continuous spatial expansion beyond the original administrative boundary to radiate more adjacent jurisdictions. The proposed center identification method can help recognize the places where the public resources shall be assigned, while the boundary delimiting method can provide effective alternative boundaries for metropolitan planning. Furthermore, the heterogeneity of human activity performance and role for each regional center is remarkable. The employment sub-centers have higher job density and job-housing ratio, while the commuter towns show reverse trends in employment density indexes, but have closer connections with the core city. An independent satellite city with local employment base perform better in job-housing balance and commuting duration. Travelers working in the core city need the longest commuting time, while commuters to outer sub-centers spend less time. It is necessary to consider the specific role of each sub-center within a given metropolitan area before making more delicate transportation and land use policies.

This study can be regarded as a starting point with respect to researches on metropolitan spatial structure using carpooling data. The limitation needs to be stated. Although we have shown the positive correlation between commuting carpooling trips and employment population, without considering the impact of public transit flows on the structure of the metropolitan region, there will be some differences between the metropolitan spatial structure uncovered using carpooling data and the reality. As mentioned previously, the interaction between the sub-center system and Tianjin is likely to be underestimated due to travel splitting caused by the presence of the inter-city high-speed railways. Therefore, it is necessary to integrate the carpooling data with the data of other transport modes and human activities in metropolitan regions to improve the proposed method and associated results.

The methodological challenge of using unconventional source of data does dominate the paper, hence further work is needed in the development of this research. First, we have indicated that various sub-centers can play different roles in a metropolitan region; then it is interesting to investigate the relationship between different sub-centers by observing the extent to which their functions are complementary or alternative. Second, we illustrated a novel method to explore the metropolitan structure based on the carpooling big data. Due to the limitation in the Beijing case study, it is suggested to apply similar data to the various structural forms of global cities. Considering there are tens of huge cities with millions of carpooling trips annually in China, our further work is to scan the spatial structure of other metropolitan regions and then conduct a comparative analysis to dig the underlying laws and meanwhile demonstrate the wider suitability of the proposed method.

## Acknowledge

This work is financially supported by National Natural Science Foundation (91746201，71621001). Many thanks to the Institute of Policy Studies, DiDi Company for providing the data in this study.

## Conflicts of Interest

The authors declare that there is no conflict of interest in any aspect of the data collection, analysis, or the funding received regarding the publication of this paper.

## Reference

[1]. Allam, Z., & Newman, P. (2018). Redefining the smart city: Culture, metabolism and governance. Smart Cities, 1(1), 4-25.

[2]. Anas, A., Arnott, R., & Small, K. A. (1998). Urban spatial structure. Journal of Economic Literature, 36(3), 1426-1464.

[3]. Angel, S., & Blei, A. M. (2016). The spatial structure of American cities: The great majority of workplaces are no longer in CBDs, employment sub-centers, or live-work communities. Cities, 51, 21-35.

[4]. Beijing Municipal Bureau of Statistics (BMBS). (2018). Beijing statistic yearbook

[5]. Bosker, M., Park, J., & Roberts, M. (2019). Definition Matters: Metropolitan Areas and Agglomeration Economies in a Large Developing Country (No. 142163, pp. 1-54). The World Bank.

[6]. BTI, Beijing transport annual report 2017. (2018). Beijing Transport Institute. 27-28

[7]. Burger, M. J., Goei, B. D., Laan, L. V. D., & Huisman, F. J. M.. (2011). Heterogeneous development of metropolitan spatial structure: evidence from commuting patterns in English and Welsh city-regions, 1981–2001. Cities, 28(2), 160-170.

[8]. Castells, M. (1989). The informational city: information technology, economic restructuring, and the urban-regional process. Cambridge, MA: B. Blackwell, 24.

[9]. Castells,M. (2010). Globalisation, networking, urbanisation: Reflections on the spatial dynamics of the information age. Urban Studies, 47(13), 2737–2745.

[10]. Chen, S., Claramunt, C., & Ray, C. (2014). A spatio-temporal modelling approach for the study of the connectivity and accessibility of the Guangzhou metropolitan network. Journal of Transport Geography, 36, 12-23.

[11]. Choi, J., Coughlin, J. F., & D'Ambrosio, L. (2013). Travel time and subjective well-being. Transportation Research Record, 2357(1), 100-108.

[12]. Clark, W. A. (2000). Monocentric to polycentric: new urban forms and old paradigms. A Companion to the City, 141-154.

[13]. Davoudi, S. (2003). European briefing: polycentricity in European spatial planning: from an analytical tool to a normative agenda. European planning studies, 11(8), 979-999.

[14]. Delclòs-Alió, X., & Miralles-Guasch, C. (2017). Suburban travelers pressed for time: Exploring the temporal implications of metropolitan commuting in Barcelona. Journal of Transport Geography, 65 , 165-174.

[15]. Dong, H. (2013). Concentration or dispersion? Location choice of commercial

developers in the Portland metropolitan area, 2000–2007. Urban geography, 34 (7), 989-1010.

[16]. Dong, Y., Wang, S., Li, L., & Zhang, Z. (2018). An empirical study on travel patterns of internet based ride-sharing. Transportation research part C: emerging technologies, 86, 1-22.

[17]. Elwood, S., Goodchild, M. F., & Sui, D. Z. (2012). Researching volunteered geographic information: Spatial data, geographic research, and new social practice. Annals of the association of American geographers, 102(3), 571-590.

[18]. Ester, M., Kriegel, H., Sander, J., & Xu, X. (1996). A density-based algorithm for sparse representations. In Proc. 2nd Int. Conf. Knowl. Discov. Data Mining (pp. 226-231).

[19]. Fang, C., & Yu, D. (2017). Urban agglomeration: An evolving concept of an emerging phenomenon. Landscape and Urban Planning, 162, 126-136.

[20]. Fernández-Maldonado, A. M., Romein, A., Verkoren, O., & Parente Paula Pessoa, R. (2014). Polycentric structures in Latin American metropolitan areas: Identifying employment sub-centres. Regional Studies, 48(12), 1954-1971.

[21]. Gao, B., Huang, Q., He, C., & Ma, Q. (2015). Dynamics of urbanization levels in China from 1992 to 2012: Perspective from DMSP/OLS nighttime light data. Remote Sensing, 7(2), 1721–1735

[22]. Geurs, K. T., & Van Wee, B. (2004). Accessibility evaluation of land-use and transport strategies: review and research directions. Journal of Transport geography, 12 (2), 127-140.

[23]. Giuliano, G., & Small, K. A. (1991). Subcenters in the Los Angeles region.

[24]. Giuliano, G., Small, K., (1993). Is the journey to work explained by urban structure? Urban Study, 30 (9), 1485–1500.

[25]. Guzman, L. A., Hernandez, D. O. , & Bocarejo, J. P. . (2017). City profile: the bogotá metropolitan area that never was. Cities, 60, 202-2015.

[26]. Hashem, Dadashpoor, Mehdi, & Alidadi. (2017). Towards decentralization: spatial changes of employment and population in tehran metropolitan region, iran. Applied Geography.

[27]. Hu, L., Sun, T., & Wang, L. (2018). Evolving urban spatial structure and commuting patterns: A case study of Beijing, China. Transportation Research Part D: Transport and Environment , 59 , 11-22.

[28]. Huang, D., Liu, Z., Zhao, X., & Zhao, P. (2017). Emerging polycentric megacity in China: An examination of employment subcenters and their influence on population distribution in Beijing. Cities, 69 , 36-45.

[29]. Lee, B., & Gordon, P. (2011). Urban structure: its role in urban growth, net new business formation and industrial churn. Région et Dévelopment , 33 , 137-159.

[30]. Kahneman, D., Krueger, A. B., Schkade, D. A., Schwarz, N., & Stone, A. A. (2004). A survey method for characterizing daily life experience: The day reconstruction method. Science, 306(5702), 1776-1780.

[31]. Lambregts, B., Kloosterman, R. C., & Van der Werff, M. (2005). Polycentricity

and the eye of the beholder: A multi-layered analysis of spatial patterns in the Dutch Randstad. Romanian Economic Journal, 8, 19–34.
[32]. Li, Linna, Goodchild, Michael F., & Xu, Bo (2013). Spatial, temporal, and socioeconomic patterns in the use of Twitter and Flickr. Cartography and Geographic Information Science, 40(2), 61-77.
[33]. Li, H., Wei, Y. D., Wu, Y., & Tian, G. (2019). Analyzing housing prices in Shanghai with open data: Amenity, accessibility and urban structure. Cities, 91, 165-179.
[34]. Limtanakool, N., Dijst, M., & Schwanen, T. (2007). A theoretical framework and methodology for characterizing National urban systems on the basis of flows of people: Empirical evidence for France and Germany. Urban Studies, 44(11), 2123–2145.
[35]. Limtanakool, N., Schwanen, T., Dijst, M., (2009). Developments in the Dutch Urban System on the basis of flows. Reg. Stud. 43 (2), 179–196.
[36]. Lin, D., Allan, A., & Cui, J. (2015). The impact of polycentric urban development on commuting behaviour in urban China: Evidence from four sub-centres of Beijing. Habitat International, 50, 195-205.
[37]. Liu, K., Murayama, Y., & Ichinose, T. (2020). Using A New Approach for Revealing the Spatiotemporal Patterns of Functional Urban Polycentricity: A Case Study in the Tokyo Metropolitan Area. Sustainable Cities and Society, 102176.
[38]. Liu, X., Gong, L., Gong, Y., & Liu, Y. (2015). Revealing travel patterns and city structure with taxi trip data. Journal of Transport Geography, 43, 78-90.
[39]. Liu, X., Yan, X., Liu, F., Wang, R., & Leng, Y. (2019). A trip-specific model for fuel saving estimation and subsidy policy making of carpooling based on empirical data. Applied Energy, 240, 295-311.
[40]. Liu, Y., Yan, X., Wang, Y., Yang, Z., & Wu, J. (2017). Grid mapping for spatial pattern analyses of recurrent urban traffic congestion based on taxi GPS sensing data. Sustainability, 9(4), 533.
[41]. Liu, Z., He, C., Zhang, Q., Huang, Q., & Yang, Y. (2012). Extracting the dynamics of urban expansion in China using DMSP-OLS nighttime light data from 1992 to 2008. Landscape and Urban Planning, 106, 62e72.
[42]. Long, Y., Han, H., Lai, S. K., & Mao, Q. (2013). Urban growth boundaries of the Beijing Metropolitan Area: Comparison of simulation and artwork. Cities, 31, 337-348.
[43]. Longley, Paul A., Adnan, Muhammad, & Lansley, Guy (2015). The geotemporal demographics of Twitter usage. Environment and Planning A, 47(2), 465-484.
[44]. Lv, Y., Zheng, X., Zhou, L., & Zhang, L. (2017). Decentralization and polycentricity: Spatial changes of employment in beijing metropolitan area, china. Sustainability, 9(10), 1880.
[45]. Lynch, C. (2008). Big data: How do your data grow? Nature, 455(7209), 28–29.
[46]. Meijers, E. J., & Burger, M. J. (2010). Spatial structure and productivity in US metropolitan areas. Environment and planning A, 42(6), 1383-1402.

[47]. Moreno-Monroy, A. I., Schiavina, M., & Veneri, P. (2020). Metropolitan areas in the world. Delineation and population trends. Journal of Urban Economics, 103242.

[48]. Morris, E. A., & Guerra, E. (2015). Are we there yet? Trip duration and mood during travel. Transportation research part F: traffic psychology and behavior, 33 , 38-47.

[49]. Najmi, A., Rey, D., & Rashidi, T. H. (2017). Novel dynamic formulations for real-time ride-sharing systems. Transportation research part E: logistics and transportation review, 108, 122-140.

[50]. Nie, P., & Sousa-Poza, A. (2018). Commute time and subjective well-being in urban China. China Economic Review, 48, 188-204.

[51]. Ouředníček, M., Nemeškal, J., Špačková, P., Hampl, M., & Novák, J. (2018). A synthetic approach to the delimitation of the Prague Metropolitan Area. Journal of Maps, 14(1), 26-33.

[52]. Parr, J., & Hewings, G. (2007). Spatial interdependence in a metropolitan setting. Spatial Economic Analysis, 2, 8–22.

[53]. Reggiani, A., Bucci, P., Russo, G., Haas, A., & Nijkamp, P. (2011). Regional labour markets and job accessibility in city network systems in Germany. Journal of Transport Geography, 19 (4), 528-536.

[54]. Riguelle, F., Thomas, I., & Verhetsel, A. (2007). Measuring urban polycentrism: a European case study and its implications. Journal of Economic Geography, 7 (2), 193-215.

[55]. Sánchez-Mateos, H. S. M., Sanz, I. M., Francés, J. M. U., & Trapero, E. S. (2014). Road accessibility and articulation of metropolitan spatial structures: the case of Madrid (Spain). Journal of Transport Geography, 37, 61-73.

[56]. Schleith, D., Widener, M., & Kim, C. (2016). An examination of the jobs-housing balance of different categories of workers across 26 metropolitan regions. Journal of Transport Geography, 57, 145-160.

[57]. Schleith, D., Widener, M. J., Kim, C., & Liu, L. (2018). Assessing the delineated commuter sheds of various clustering methods. Computers, Environment and Urban Systems, 71, 81-87.

[58]. Shi, Q., & Cao, G. (2020). Urban spillover or rural industrialisation: Which drives the growth of Beijing Metropolitan area. Cities, 105, 102354.

[59]. Sorensen, A. (2001). Subcentres and satellite cities: Tokyo's 20th century experience of planned polycentrism. International Planning Studies, 6(1), 9-32.

[60]. Sobolevsky, S., Szell, M., Campari, R., Couronné, T., Smoreda, Z., & Ratti, C. (2013). Delineating geographical regions with networks of human interactions in an extensive set of countries. PloS one, 8(12), e81707.

[61]. Squires, G. D. (Ed.). (2002). Urban sprawl: Causes, consequences, & policy responses. The Urban Insitute.

[62]. Sun, T., & Lv, Y. (2020). Employment centers and polycentric spatial development in Chinese cities: A multi-scale analysis. Cities, 99, 102617.

[63]. Tang, J., Liu, F., Wang, Y., & Wang, H. (2015). Uncovering urban human mobility from large scale taxi GPS data. Physica A: Statistical Mechanics and its Applications, 438, 140-153.
[64]. Sun, T., & Lv, Y. (2020). Employment centers and polycentric spatial development in Chinese cities: A multi-scale analysis. Cities, 99, 102617.
[65]. Tian, G., Wu, J., & Yang, Z. (2010). Spatial pattern of urban functions in the Beijing metropolitan region. Habitat International, 34(2), 249-255.
[66]. UN-Habitat. (2013). State of the World's cities 2012/2013. Prosperity of cities. United Nations Human settlements programme. New York: Routledge.
[67]. US Office of Management and Budget. (2010). 2010 standards for delineating metropolitan and micropolitan statistical areas; Notice. Federal Register, 75(123), 37246-37252.
[68]. Veneri, & Paolo. (2013). The identification of sub-centres in two italian metropolitan areas: a functional approach. Cities, 31, 177-185.
[69]. Wan, L., Gao, S., Wu, C., Jin, Y., Mao, M., & Yang, L. (2018). Big data and urban system model-substitutes or complements? A case study of modelling commuting patterns in beijing. Computers, Environment and Urban Systems, 68, 64-77.
[70]. Wong, D. W., & Huang, Q. (2017). "Voting with Their Feet": Delineating the Sphere of Influence Using Social Media Data. ISPRS International Journal of Geo-Information, 6(11), 325.
[71]. Xing, X., Warden, T., Nicolai, T., & Herzog, O. (2009, September). Smize: a spontaneous ride-sharing system for individual urban transit. In German Conference on Multiagent System Technologies (pp. 165-176). Springer, Berlin, Heidelberg.
[72]. Yin, C., Shao, C., Dong, C., & Wang, X. (2019). Happiness in urbanizing China: The role of commuting and multi-scale built environment across urban regions. Transportation Research Part D: Transport and Environment, 74, 306-317.
[73]. Yue, Y., Wang, H. D., Hu, B., Li, Q. Q., Li, Y. G., & Yeh, A. G. (2012). Exploratory calibration of a spatial interaction model using taxi GPS trajectories. Computers, Environment and Urban Systems, 36(2), 140-153.
[74]. Zhang, P., Zhou, J., & Zhang, T. (2017). Quantifying and visualizing jobs-housing balance with big data: A case study of Shanghai. Cities, 66, 10-22.
[75]. Zhang, Q., & Su, S. (2016). Determinants of urban expansion and their relative importance: A comparative analysis of 30 major metropolitans in China. Habitat international, 58, 89-107.
[76]. Zhang, X., Xu, Y., Tu, W., & Ratti, C. (2018). Do different datasets tell the same story about urban mobility—A comparative study of public transit and taxi usage. Journal of Transport Geography, 70, 78-90.
[77]. Zhao, P., Lü, B., & De Roo, G. (2011). Impact of the jobs-housing balance on urban commuting in Beijing in the transformation era. Journal of transport

geography, 19(1), 59-69.
[78]. Zhen, F., Cao, Y., Qin, X., & Wang, B. (2017). Delineation of an urban agglomeration boundary based on Sina Weibo microblog ‘check-in’data: A case study of the Yangtze River Delta. Cities, 60, 180-191.
[79]. Zhen, F., Wang, B., & Wei, Z. C. (2015). The rise of the internet city in China: Production and consumption of internet information. Urban Studies, 52(13), 2313–2329.
[80]. Zhou, J., Murphy, E., & Long, Y. (2014). Commuting efficiency in the Beijing metropolitan area: An exploration combining smartcard and travel survey data. Journal of Transport Geography, 41, 175-183.

Dear Editor,

Please find the electronic submission of "Characterizing the Polycentric Spatial Structure of Beijing Metropolitan Region Using Carpooling Big Data" by Xiaobing LIU, Xuedong YAN, Helena TITHERIDGE, Wei WANG, Rui WANG, Yang LIU. We would like to have this manuscript reviewed by the ***Cities*** *(Special Issues on Big Data and Urban Planning)*.

For each revision, each of the coauthors has seen and agrees with each of the changes made to this manuscript in the revision and to the way his or her name is listed.

Sincerely,
Xuedong Yan
Beijing Jiaotong University, China