Cities 110 (2021) 103067

Contents lists available at ScienceDirect

Cities

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

# A realistic and multilevel measurement of citywide spatial patterns of economic segregation based on human activities[☆]

Mengling Qiao[a], Yandong Wang[a,b,c,*], Shanmei Wu[a], Xiaokang Fu[a], Yanyan Gu[a], Mingxuan Dou[a]

[a] State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, China
[b] Collaborative Innovation Centre for Geospatial Information Technology, Wuhan University, Wuhan, China
[c] Faculty of Geomatics, East China University of Technology, Nanchang, China

ARTICLE INFO

*Keywords:*
Human activity
Collective activity space
Economic segregation pattern
ICE
TFIDF

ABSTRACT

Research on the realistic and comprehensive identification of citywide spatial patterns of economic segregation is valuable for the sustainable development of cities. The consideration of human activities in segregation research inspires us to develop an alternative method to contribute to this type of research. In our method, we emphasize the combination of collective activity spaces (CASs) and spatial economic data, both of which are obtained from dynamic human activities. We first reveal the realistic use of urban spaces from human mobility patterns to generate multilevel CASs as basic analytical units. Then, we use a type of realistic economic data generated from human activities to measure the segregation level of each CAS. We realize this measurement by tailoring a segregation index, named the Term Frequency-Inverse Document Frequency-Index of Concentration at the Extremes-based (TFIDF-ICE-based) segregation index, for our economic data. Through these methods, we can uncover citywide multilevel spatial patterns of economic segregation realistically and comprehensively. Using Beijing and Wuhan as cases, we demonstrate and discuss the applicability and value of our method.

## 1. Introduction

Segregation refers to differences in the distributions of social groups (James & Taeuber, 1985). Researchers have examined many forms of segregation, such as racial, occupational and economic segregation. Economic segregation, as one type of segregation, can be defined as "the spatial segregation of households by income or social class" (Jargowsky, 1996). Specifically, the rich and poor are concentrated in different neighbourhoods separated by spatial boundaries. In the context of urban sprawl and decentralization, there is a common thought that economic segregation is partly a product of growing social inequality, which can lead to social unrest, an increase in crime, and a decrease in trust between groups in society (Malmberg et al., 2013; Musterd et al., 2017). Therefore, a comprehensive and realistic investigation of citywide spatial patterns of economic segregation is essential to the sustainable development of a city (Ghasemi et al., 2018; Moroke et al., 2019).

Scholars have conducted considerable research on segregation (Miller, 2007; Farber et al., 2015). Part of this work concentrates on people's residential spaces and assumes that people's daily activities occur within these spaces (Ellis et al., 2004). This assumption contributes significantly to understanding the experiences of different social groups in static residential environments, although it seems deficient when considering experiences of segregation outside residential environments (Kwan, 2013; Wong & Shaw, 2011). With the development of human dynamics, there is a new call to extend the research scope to address the possibility of interactions between people. This line of research argues that in addition to residential places, people experience segregation in other places where they undertake daily activities. To address this idea, researchers have constructed individual activity spaces (defined as 'the subset of all locations within which an individual has direct contact as a result of his or her day-to-day activities' by Golledge and Stimson (1997)) to perform related studies and evaluate segregation realistically and comprehensively (Järv et al., 2015; Park & Kwan, 2017; Silm & Ahas, 2014). Research based on individual activity

☆ This work was supported by the National Key Research Program of China under Grant 2016YFB0501403; the National Natural Science Foundation of China under Grant 41271399; and the China Special Fund for Surveying, Mapping and Geoinformation Research in the Public Interest under Grant 201512015.

* Corresponding author at: State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, China.
*E-mail address:* ydwang@whu.edu.cn (Y. Wang).

https://doi.org/10.1016/j.cities.2020.103067
Received 25 December 2019; Received in revised form 20 August 2020; Accepted 27 November 2020
Available online 14 December 2020

spaces illustrates the dynamics of individual trajectories and thus offers a realistic measurement of segregation. However, existing studies have two limitations in the comprehensive and realistic investigation of citywide spatial patterns of economic segregation. First, studies usually identify spatial patterns of segregation by aggregating groups of segregation values in traditional spaces, such as census blocks. This approach falls short in evaluating the reality of the evaluation of segregation and leads to the modifiable areal unit problem (MAUP) (Openshaw & Taylor, 1981). Second, studies that aim to measure citywide economic segregation usually rely on survey or census data, which are not updated quickly enough and are difficult to obtain with regard to a wide range of human trajectories.

This study addresses the two limitations above and aims to design an alternative method for the comprehensive and realistic investigation of citywide spatial patterns of economic segregation. Our method consists of two main parts. In the first part, we take advantage of human mobility in spaces to define multilevel collective activity spaces (CASs) as basic analytical units. In this study, CASs refer to urban spaces that are partitioned based on the mobility and clustering patterns of human activities in space. In the second part, we measure the segregation level of each CAS by designing a segregation index, called the Term Frequency-Inverse Document Frequency-Index of Concentration at the Extremes-based (TFIDF-ICE-based) segregation index, for a type of spatial economic data derived from the agglomerated economic activities of groups. Our method offers three contributions. First, both parts of our method rely on bottom-up data derived from human mobility or human activities, thus supporting realistic investigation. Second, our multilevel results provide the possibility to investigate the segregation of a region at different scales, thus supporting comprehensive investigation. Third, instead of an arbitrary division, our realistic and multilevel CASs can optimize the MAUP to a certain extent. Using Beijing and Wuhan as cases, we demonstrate the applicability and value of our method.

The remainder of this paper is organized as follows. In Section 2, we review related studies concerning the use of traditional spaces and individual activity spaces in segregation research. In Section 3, we describe the study areas and our datasets. In Section 4, we introduce our method and the spatial patterns of urban economic segregation in the case studies. In Section 5, we discuss the role of our methods in achieving realistic and comprehensive results and the describe limitations of our methods. Finally, the conclusions are drawn in Section 6.

## 2. Literature review

Inspired by previous research efforts on segregation, we design an alternative method for the comprehensive and realistic investigation of citywide spatial patterns of economic segregation. In this section, we review research related to two aspects: (1) measuring segregation based on traditional spaces and (2) measuring segregation based on individual activity spaces.

### *2.1. Segregation studies based on traditional spaces*

Most traditional studies on segregation have focused on the combination of residential spaces or administrative units and survey data or census data, producing segregation studies based on traditional spaces. These studies often use mathematical indices to measure certain characteristics of segregation, such as residential, racial, gender, occupation, and income characteristics, in traditional spaces.

We reviewed some related studies, which we describe as follows. Based on census blocks and official census data, Darden and Kamel (2000) revealed that black people were more residentially segregated from white people in the suburbs than in cities. Catney (2018) employed data derived from the 2011 Census of Population of England and Wales to analyse small-area ethnic residential segregation in England and Wales. The findings revealed that the results were highly dependent on the cooperation of the scale and the selected measure of segregation. Leckie and Goldstein (2015) used annual school census data to analyse the changing patterns of ethnic composition and segregation among London secondary schools and found that these patterns remained largely stable during the first decade of the 21st century. Clark et al. (2015) used census data and census block data from 2000 to 2010 for Los Angeles to evaluate the patterns of race and ethnicity at varying scales. They found decreasing segregation between black people and white people and increasing segregation between Asians and Hispanics. Firebaugh and Acciai (2016) utilized census data for all US metropolitan areas in 1980 and 2010 and found that the neighbourhood poverty gap between black people and white people declined substantially while the residential segregation of black people remained very high. Moreover, Chodrow (2017) used data from the Five-Year Estimates of the 2013 American Community Survey to model the structure and dynamics of segregation based on modern information theory and machine learning.

In general, traditional methods evaluate segregation by focusing on residential or official units complemented by survey or census data. This stream of research has two features. First, it contributes greatly to understanding segregation in static residential environments, while it may offer only partial or even biased results when considering human mobility in shaping segregation experiences (Kwan, 2012). Second, it can uncover the spatial patterns of segregation in residential or official units, while it may produce biased results because of the MAUP (Openshaw & Taylor, 1981) in determining the zoning schemes of these units. To address these problems, researchers have proposed using a series of individual activity spaces to obtain more comprehensive and more realistic results.

### *2.2. Segregation studies based on individual activity spaces*

With the increasing attention to the role of human mobility in shaping segregation experiences, a growing body of research argues that segregation studies should be carried out based on individual activity spaces (Kwan, 2013; Östh et al., 2018; Wong & Shaw, 2011). This kind of research utilizes individual activity spaces as analytical units to calculate segregation based on the characteristics of such spaces; thus it emphasizes the comprehensive influence of personal experiences and realistic geographic contexts (Järv et al., 2015; Silm & Ahas, 2014).

We reviewed several related studies. Östh et al. (2018) combined spatial trajectory data with detailed socio-economic residential statistics to study how spatial mobility can shape the segregation experiences of people and change the segregation levels of places. By using patterns of daily travel behaviour and decomposing the social interaction potential (SIP) metric into interactions within and between social groups, Farber et al. (2015) explored spatial variations in segregation and hotspots of segregation. Wang et al. (2012) examined the activity spaces of residents from different types of neighbourhoods and found significant differences in the extensity, intensity, and exclusivity of activity spaces. Tan et al. (2017) examined the effects of ethnicity on people's spatiotemporal behaviours in the Chinese context and concluded that the ethnic characteristics of Hui minorities had a greater independent and significant influence on spatiotemporal behaviour than did the characteristics of Han majorities in Xining. Park and Kwan (2018) proposed a new dynamic notion of segregation that included segregation in various spatiotemporal contexts in people's everyday lives. Considering movements among social groups throughout the day, Le Roux et al. (2017) observed that social segregation within the Paris region decreased during the day and that the upper-class group remained the most segregated at night. Zhang et al. (2019) used a 7-day individual global position system (GPS) tracking dataset, activity diary data and socio-economic attribute data to reveal that the characteristics of segregation varied not only with respect to actual activity spaces and potential activity spaces but also between different days of the week. Based on urban human interaction patterns, Shen (2019) successfully measured the extent to which two trajectories interacted with one another in daily activity spaces and captured the interaction potentials among various social groups. To recognize and assess segregation as a dynamic process, Li and Wang (2017) proposed a regression estimator that measured segregation by assessing the similarity between people and the social environments that they experienced in their daily activity spaces.

In general, this stream of research holds that segregation exists not only in individual residential spaces but also in individual activity spaces. These studies have three features. First, when taking an individual's experience and realistic geographic contexts into account, the evaluation of segregation becomes more comprehensive and more realistic. Second, the results obtained by research using this kind of method are at the individual level, which is valuable for the comprehensive measurement of segregation. In regard to revealing spatial patterns, studies usually aggregate individual-level results into traditional spatial units, which are subject to the MAUP (Kwan, 2012). Third, individual trajectory data used in this type of research are generally from surveys, which are randomly sampled and have lower biases. However, difficulties arise when a large sample and a large spatial range of data are required to reveal citywide spatial patterns of segregation, especially in relation to the economic statuses of individuals.

By considering the advantages and disadvantages of existing methods, this study proposes an alternative method of constructing realistic and multilevel CASs as analytical units and using bottom-up human economic activity data as a realistic indicator of the economic level of a region to comprehensively and realistically reveal citywide spatial patterns of economic segregation. Multilevel CASs can optimize the MAUP to a certain extent. The use of bottom-up human economic activity data can address the problems of generating only partial or even biased results when using static census data. The spatial economic data used in our study come from the agglomerated economic activities of groups. While these data do not reach the individual trajectory level, they are accessible and able to meet our needs in revealing spatial patterns of segregation. We describe the details and applications of our method below.

## 3. Study area and data preparation

This study uses two metropolises—Beijing and Wuhan, which have different geographic characteristics and urban morphology patterns, as cases to verify the applicability and robustness of this study. Beijing, the capital of China, covers an area of 16,410.54 km$^2$ and has a population of 21.542 million people. Developed ring roads are a typical characteristic of this city and play a major role in shaping the urban formation of Beijing. The urban planning of Beijing is mainly based on these ring roads. People living in Beijing have a consensus that in locations closer to the central ring road, transportation is more convenient and the economy is more developed. Therefore, we believe that Beijing is a typical city whose development is greatly promoted by transportation. Wuhan is an essential central city in central China, with an area of 8494.41 km$^2$ and a population of 11.081 million people. Wuhan has a unique landform in that the water area accounts for a quarter of the city's total area. The Yangtze River and its largest tributary, the Han River, intersect in the city, forming a pattern of three towns in Wuhan (Wuchang, Hankou, Hangyang) separated by rivers. There are also numerous lakes embedded in this city. Based on this unique landscape, the urban planning of Wuhan depends greatly on the rivers and lakes. Thus, people living in Wuhan consider that areas around these rivers and lakes are more likely to be developed effectively. Therefore, we believe that Wuhan is a typical city whose development is greatly promoted by landforms. Both of these metropolises have attracted people of various social strata, such as business owners, white-collar workers, institutional workers, low-skilled workers, and residents, to work and live in these cities. These people carry out various activities with characteristics of different social strata, thus providing a large amount of valuable data as the basis for investigating the spatial patterns of urban economic segregation.

In our method, the implementation of the division of analytical units and the measurement of segregation emphasize the use of data relating to human activities to enhance the reality of the results. The development of information and communication technology provides unprecedented possibilities to extract human activity data. In our study, we use Sina Weibo, a twitter-like microblogging site widely used in China, to gather human mobility data for the division of analytical units. Dianping, a popular website on which users comment on and grade commercial facilities, provides spatial economic data to measure the economic segregation level of a city. Dianping data, including geolocations, per capita consumption (PCC) information and other detailed descriptions of shops, can provide clues to the economic level of a region according to PCC information derived from averaging the amount spent during consumers' activities in a shop. In this paper, shops include all the commercial facilities uploaded onto the Dianping website within a city.

By building a data crawler frame based on the application programming interfaces (APIs) provided by Sina Weibo, we collected abundant social media data with embedded geolocation information. These data came from the real-time voluntary release of users during the period of data collection. To ensure the validity of the data, we removed the data collected from advertising accounts, marketing accounts, virtual person accounts, and users with fewer than three messages posted during our collection period. The items used in our study that are embedded in Sina Weibo data include the ID of the data, geographic coordinates, date, time, and ID of the user posting the data. Table 1 shows detailed descriptions of the valid datasets.

We crawled Dianping data via webpage parsing because of the absence of Dianping APIs. To ensure the validity of the data, we removed data without PCC information. The ID, geographic coordinates, grade, and PCC information were attached to each shop. Table 1 provides detailed descriptions of the valid datasets. Dianping data are POI-like data with economic characteristics. The generation of shops on Dianping is not in real time, which is in line with the reality that the existence of a shop will not change in a short time. Thus, once we completed data collection, we could obtain the information of all the uploaded shops of a city on Dianping as of our collection period.

To understand the spatial distribution of the data, we estimated the kernel density value of the Weibo data and Dianping data on fixed cells because of data anonymization. We averaged the hourly amount of Weibo data on weekdays and weekends to analyse the temporal patterns of these Weibo data. Because the generation of shops on Dianping is not in real time, the temporal analysis of Dianping is missing in our study. To better understand the Dianping data, we counted the frequency distribution of the PCC of the Dianping data. Considering that Dianping data mainly apply inside the Sixth Ring Roads of Beijing and the Belt Expressway of Wuhan, the study areas of these two cases focus on these areas.

Fig. 1 shows the spatial distribution of the Weibo data and Dianping data in Beijing. We used the default classification method, the Natural Breaks (Jenks), provided by ArcGIS to classify the value of the kernel density estimation into five classes in order to enhance the differences between different regions. As one might expect, the downtown area contains more Weibo and Dianping data, while areas within the Fifth Ring Road and Sixth Ring Road have fewer data. Furthermore, the northern and eastern parts have more Weibo data and Dianping data than do the southern and western parts.

Fig. 2 shows the spatial distributions of the Weibo data and Dianping data in Wuhan. It is evident that areas along the Yangtze River are significant data gathering places, and the data spread outward from both sides of the Yangtze River. Interestingly, the northwest side of the Yangtze River has fewer Weibo data but more Dianping data compared to the southeast side of the Yangtze River. These characteristics of the two cases are the basis for our investigation of the spatial patterns of urban economic segregation.

The temporal variation of the Weibo data on weekdays and weekends is shown in Fig. 3. The two temporal patterns of Weibo data in

**Table 1**
Description of datasets.

| Data | City | | | |
|---|---|---|---|---|
| | Beijing | | Wuhan | |
| Weibo | Date | 2014.02.01–2014.09.30 | Date | 2017.07.01–2019.05.31 |
| | Amount | 9,015,051 | Amount | 3,455,557 |
| Dianping | Date | 2018.06 | Date | 2019.06 |
| | Amount | 75,776 | Amount | 43,391 |
| | Category | 111 | Category | 94 |

Beijing and Wuhan are highly consistent, which indicates that the temporal patterns of group activities are stable. Moreover, we summarized two characteristics by comparing the temporal patterns between weekdays and weekends. First, most of the time, except between 6 am and 9 am, people have more Weibo activities on weekends than on workdays. Second, the morning peak of people's activities on weekends is approximately 1–2 h behind that on workdays. These two characteristics are consistent with the actual patterns of human activities, whereby people have more time to carry out activities and are more likely to wake up later on weekends than on weekdays. Therefore, our Weibo data in both Beijing and Wuhan include reasonable human activities, which is the basis for our study.

The frequency distribution of the PCC of Dianping data is shown in Fig. 4. We carried out statistical analysis on both all Dianping data and several types of Dianping data that were randomly selected. There is a long tail in these frequency distributions and we can fit them well by power law functions. This indicates the existence of the gap between rich and poor, which encourages us to investigate the patterns of economic segregation.

## 4. Comprehensive and realistic investigation of citywide spatial patterns of economic segregation

In this study, we suggest that the comprehensive and realistic investigation of citywide spatial patterns of economic segregation can be achieved based on the combination of multilevel CASs and spatial economic data obtained from human activities, which aligns the results with the realistic development of a city. To implement this approach, we propose an alternative method that includes two parts. The first part focuses on dividing multilevel CASs in a bottom-up and realistic way by applying community detection algorithms to improved human activity networks obtained from Weibo data with embedded geographic coordinates. The second part designs a segregation index, named the TFIDF-ICE-based segregation index, for Dianping data embedded PCC information that represents the average amount spent during consumers' activities in a shop to measure economic segregation. Finally, by using Beijing and Wuhan as cases, we reveal the realistic and comprehensive spatial patterns of urban economic segregation in these two cities and demonstrate the applicability of our method.

### *4.1. Dividing multilevel CASs based on human mobility*

The first part of our method focuses on generating multilevel CASs as the spatial analytical units in a bottom-up and realistic way. For this purpose, we regard people's mobility flows as proxies to connect discrete places into cohesive units, which are called CASs in our paper (Gao et al., 2013; Guo et al., 2018; Kallus et al., 2015; Liu et al., 2015; Zhong et al., 2014). The principle behind making the CASs multilevel is the hierarchy of human mobility patterns, which further reinforces the role of realistic human mobility in dividing analytical units. Both Tang et al. (2015) and Qiao et al. (2019) proved that there are two levels of human mobility patterns: long-distance mobility and short-distance mobility. Therefore, on the premise that our CASs follow the reality and reasonability of human mobility patterns, we divide people's mobility flows into long-distance flows and short-distance flows to finally connect discrete places into two-level CASs.

Technically, the combination of network analysis and spatial analysis (Gao et al., 2013; Guo et al., 2018; Kallus et al., 2015; Liu et al., 2015; Zhong et al., 2014) makes it possible to convert human mobility flows into two-level CASs. The principle of this combination is that places near each other in geographic space have more connections in network space and are therefore more likely aggregated into a CAS within geographic space. Based on this principle, using human mobility flows obtained from Weibo data to construct networks, we employ the method of partitioning urban spaces proposed by Qiao et al. (2019). There are two steps to realize this partitioning. The first aims to find the hierarchy of human mobility patterns to determine the basis for two-level partitioning, and the second focuses on improving the construction of networks and detecting communities to realize the partitioning of two-level CASs. The diagram in Fig. 5 shows the operational details.

The results of two-level CASs reveal the bottom-up division and use of urban spaces, which are realistic spatial analytical units and contribute significantly to addressing the MAUP, i.e., the non-arbitrary division can mitigate the zoning effect of MAUP, and the multi-scale units with realistic attributes can mitigate the scale effect of MAUP. The hierarchical characteristics of CASs provide the ability to measure segregation at different levels, which contributes to a comprehensive evaluation of segregation. To evaluate the economic statuses of these areas in a realistic way, we continue to consider the role of human activities in the measurement of economic segregation levels.

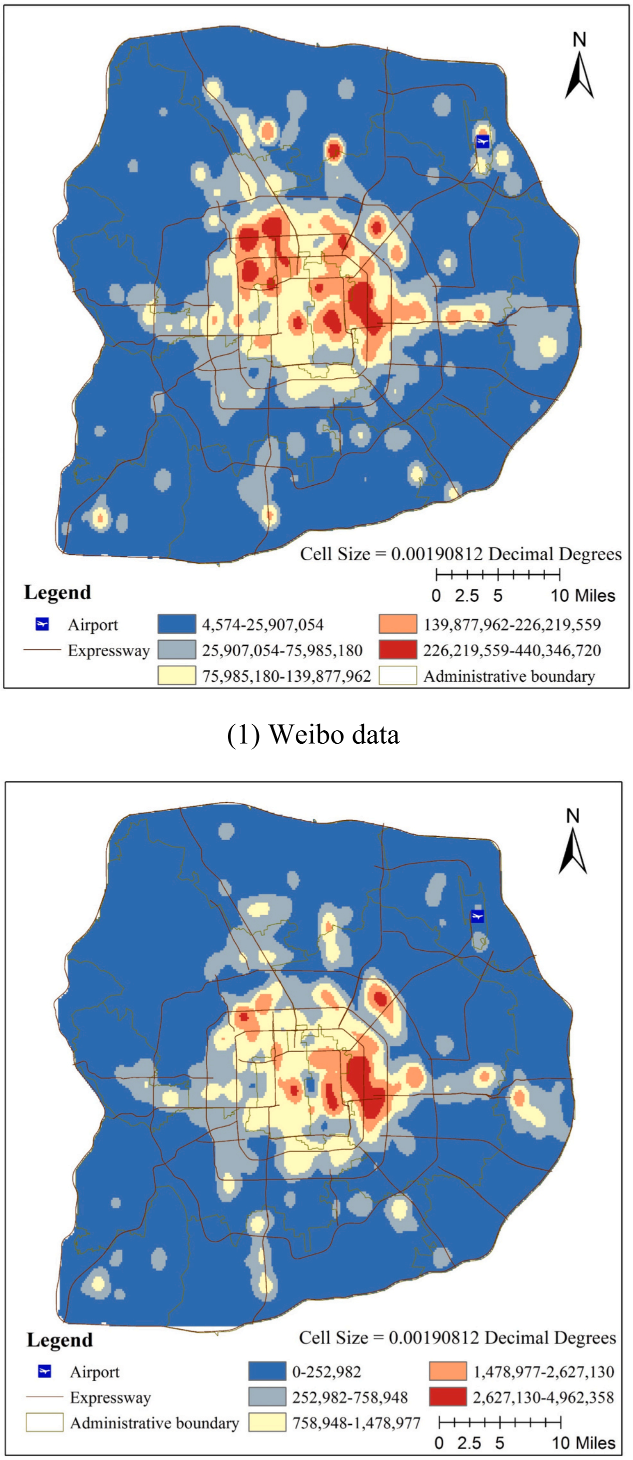
### *4.2. Tailoring a segregation index for spatial economic activity data*

The partitioning of the two-level CASs emphasizes the realistic ways in which people use spaces. By using these CASs as basic analytical units, we further emphasize the reality of the measurement of economic segregation based on human economic activities. Dianping, a popular website on which users comment on and grade shops, provides sufficient PCC information derived from averaging the amount spent during consumers' activities in each shop to support this approach. Despite to the validity and abundance of the data, there are some challenges in measuring segregation using traditional segregation indices for Dianping data, especially the challenge posed by its multi-type nature (Reardon & Firebaugh, 2002). Therefore, tailoring a segregation index for data such as those from Dianping data is needed. In our study, we design a segregation index, named the TFIDF-ICE-based segregation index, by using the term frequency-inverse document frequency (TFIDF) (Aizawa, 2003; Robertson & Jones, 1976; Salton & Buckley, 1988) to weight the Index of Concentration at the Extremes (ICE) (Scally et al., 2018).

ICE is a new type of index that reflects the extent of segregation as the level to which a CAS's economic level is concentrated into extremes of deprivation or privilege. Eq. (1) shows the conventional calculation of the ICE in the context of this paper.

$$ICE = \frac{(x\ shops\ with\ PCC \geq P) - (y\ shops\ with\ PCC \leq D)}{Total\ T\ shops\ with\ PCC\ data} \tag{1}$$

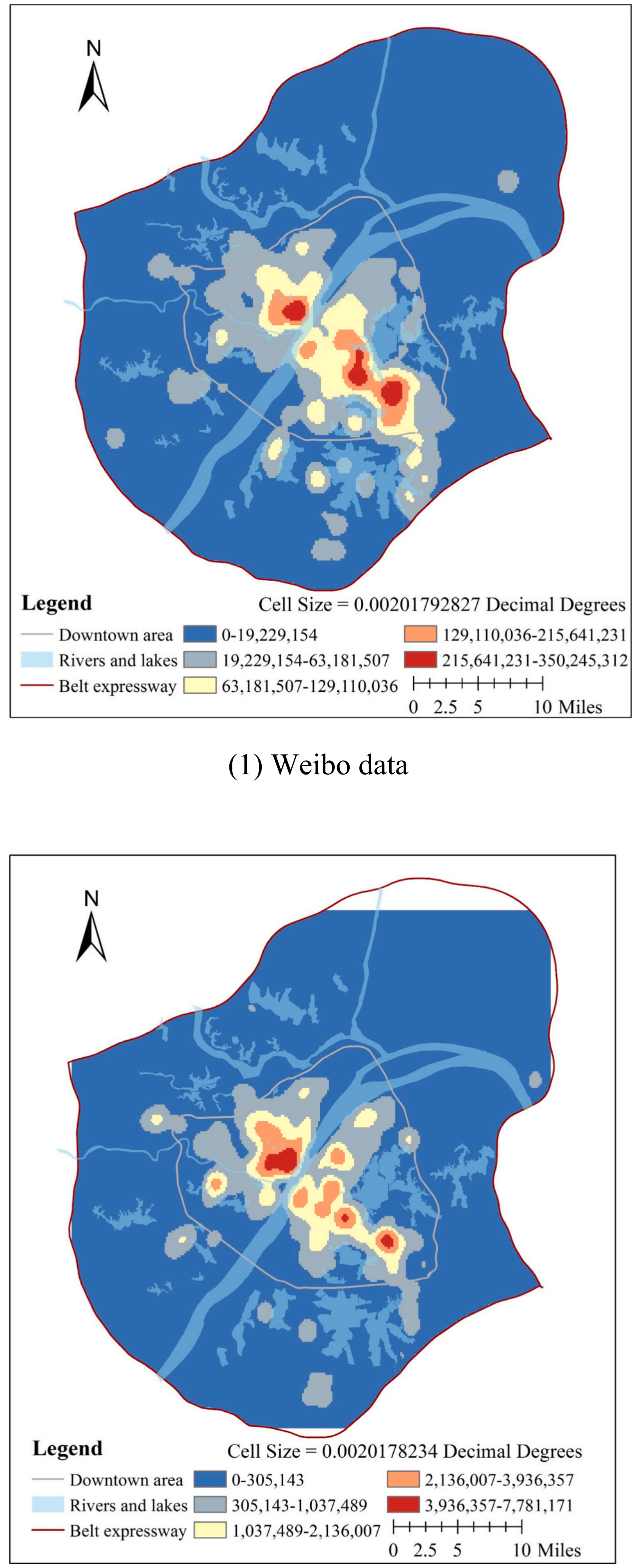
Here, $P$ and $D$ indicate the thresholds of the PCC values used to classify shops with high and low consumption levels, respectively; $x$ and $y$ represent the number of shops with economic privilege and economic




(1) Weibo data

(2) Dianping data

**Fig. 1.** Kernel density distributions of the experimental data from Beijing.




(1) Weibo data

(2) Dianping data

**Fig. 2.** Kernel density distributions of the experimental data from Wuhan.

deprivation, respectively; and $T$ is the total number of shops. Based on the conventional definition, there is a symmetry in the value of $ICE$, and it ranges from $-1$ to 1. When $ICE = 0$, there is an absolute balance between the rich and the poor; when $ICE$ approaches 1, it indicates extreme privilege; and when $ICE$ approaches $-1$, it indicates extreme deprivation.

Regarding the conventional calculation of the ICE in Eq. (1), challenges occur because the calculation is not entirely applicable to this study when we consider the multi-type nature of the Dianping shop categories. To manage these challenges, we combine TFIDF with ICE to generate the TFIDF-ICE-based segregation index. Since the mathematical foundations for the choice of index are valuable (Olteanu et al.,

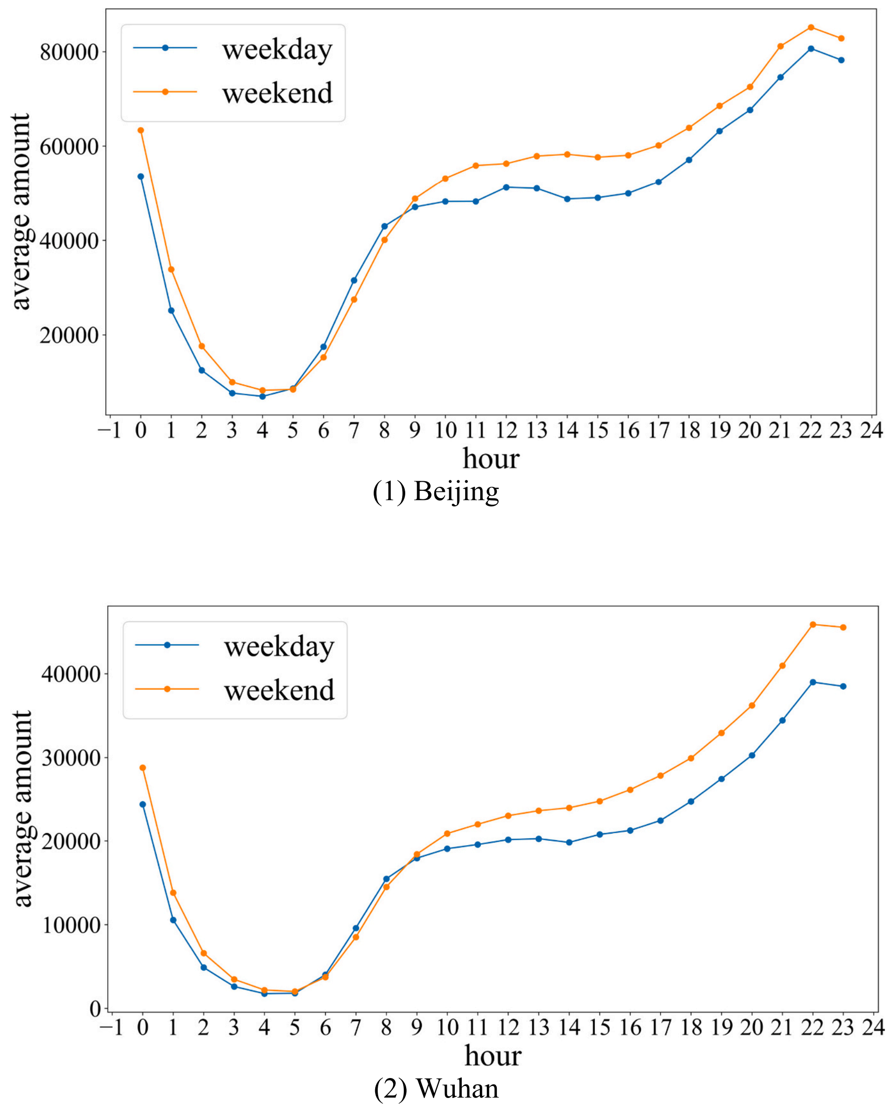



**Fig. 3.** Temporal variation of the Weibo data on weekdays and weekends.

2019), we add a mathematical and practical description of the significance of the TFIDF in our index in SI Appendix. Our main text focuses on the description of its calculation as follows.

(1) The challenge of unequal contribution.

There are two problems arising from unequal contribution. First, data that are within the same category but located in different CASs have unequal contributions to the assessment of the segregation of each CAS. Second, data located in the same CAS but belonging to different categories have unequal contributions to the assessment of the segregation of this CAS. To resolve the challenges associated with the unequal contributions described above, our study introduces the TFIDF (Salton & Yu, 1973). The TFIDF is a weighting factor that is often used in information retrieval and text mining to evaluate the importance of a word to a document in a corpus. The essence of the TFIDF is that the importance of a word correlates positively with the frequency with which the word appears in a document but correlates negatively with the number of documents that contain the word. This essence is consistent with the principle of segregation evaluation, indicating that the rarity of a category corresponds to a more significant impact of the category on segregation. Therefore, by employing the TFIDF, the word, document, and corpus in the field of information retrieval are replaced with the category, CAS, and collection for all CASs, respectively, in our study. Eq. (2) shows the TFIDF calculation in our study.

$$\begin{cases} TF_{ij} = \dfrac{n_{ij}}{n_i} \\ IDF_j = lg\dfrac{|S|}{|S_j|} \\ TFIDF_{ij} = TF_{ij} \times IDF_j \end{cases} \quad (2)$$

Here, $n_i$ represents the total number of Dianping shops in CAS $i$; $n_{ij}$ represents the number of Dianping shops belonging to category $j$ in CAS $i$; $|S|$ refers to the total number of all CASs; $|S_j|$ refers to the total number of CASs that include Dianping shops belonging to $j$; $TF_{ij}$ indicates the frequency of Dianping shops belonging to category $j$ in CAS $i$; and $IDF_j$ indicates the weight of CASs with Dianping shops belonging to category $j$, which is inversely correlated with the frequency of CASs with shops belonging to category $j$ relative to all CASs.

(2) The challenge of significant deviations in the number of shops.

Significant deviations exist among the number of shops in different CASs and among the number of shops belonging to different categories, which will cause a significant deviation of the denominator in Eq. (1)

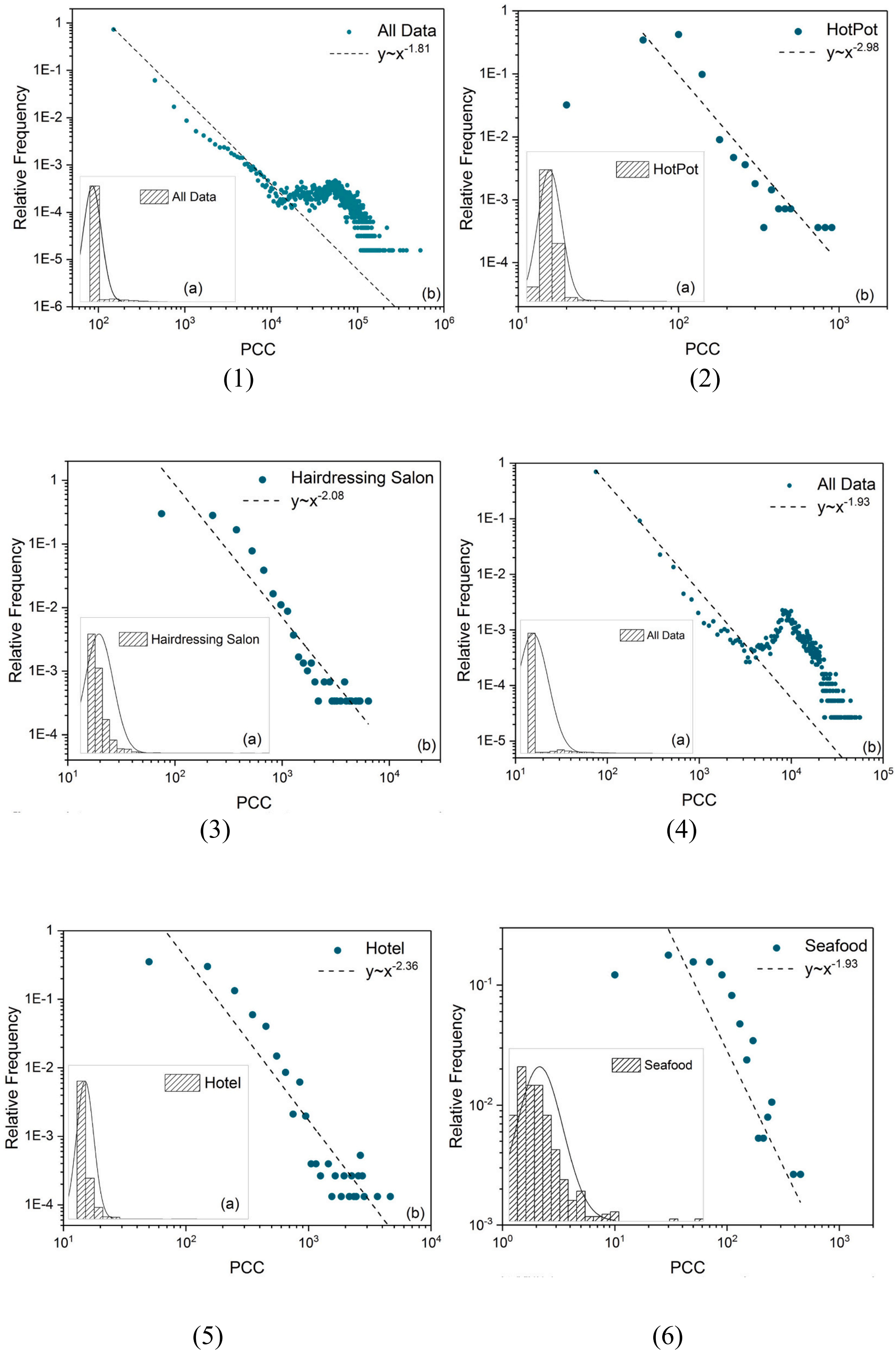



**Fig. 4.** Frequency distribution of PCC of Dianping data. (1)–(3) show the distribution in Beijing and (4)–(6) show the distribution in Wuhan.

and further affect the ICE values of different categories. To mitigate the influences of the significant deviation among the number of shops in different CASs, we take the total number of shops belonging to a category as the denominator in the calculation of segregation. Furthermore, we take the logarithm of the denominator in Eq. (3) to mitigate the significant deviation among the number of shops belonging to different categories.

(3) The challenge of deviation in PCCs.

The PCCs among different categories vary greatly; thus, each category should have an exclusive $P$ and $D$ instead of an arbitrary value (Louf

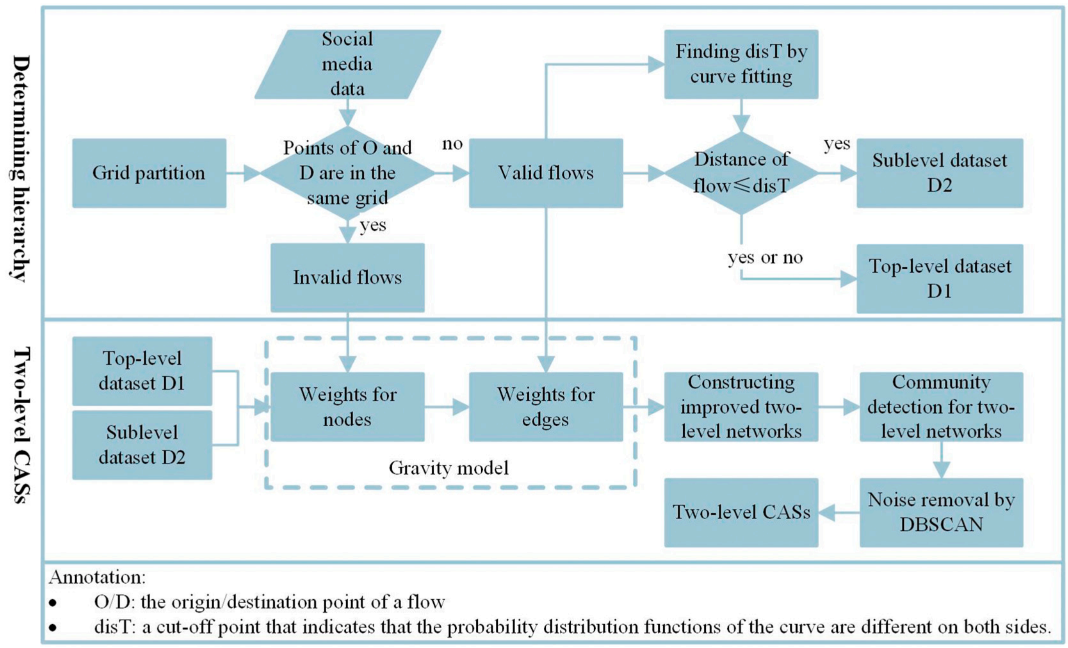



**Fig. 5.** Diagram of the method of two-level CAS partitioning.

& Barthelemy, 2016). We address this challenge by calculating quintiles to determine the exclusive $P$ and $D$ for each category, which is inspired by the 80/20 principle (i.e., the Pareto principle) (Pareto, 1896–97). The 80/20 principle states that approximately 80% of wealth is concentrated in approximately 20% of a population. Based on this principle, the frequency distribution of the PCC of Dianping data is a power law distribution, which indicates that we can regard the first quantile of the quintiles as the threshold of $P$ to classify shops with high consumption levels. To follow the symmetry in the conventional calculation of ICE in Eq. (1), we define the last quantile of the quintiles as the threshold of $D$ to classify shops with low consumption levels. The calculation of the segregation value derived from the shops belonging to category $j$ in CAS $i$ is as shown in Eq. (3).

$$ICE_{ij} = \frac{(x_{ij}\ shops\ with\ PCC \geq P_j) - (y_{ij}\ shops\ with\ PCC \leq D_j)}{lgn_j} \times TFIDF_{ij} \tag{3}$$

Here, $x_{ij}$ and $y_{ij}$ represent the number of shops belonging to category $j$ in CAS $i$ with economic privilege and with economic deprivation, respectively; $n_j$ denotes the total number of shops belonging to category $j$; and the values of $P_j$ and $D_j$ correspond to the highest and lowest quintiles, respectively, with the quintiles obtained by dividing the number of shops belonging to category $j$ into five groups of equal size.

(4) The challenge of integrating a set of segregations.

The integration of a set of segregation values derived from data belonging to different categories is needed when evaluating the overall segregation level of a CAS. After weighting the segregation derived from category $j$ in CAS $i$ by the TFIDF value described above, we consider the mean operation by which to integrate the segregation values belonging to the same CAS. In addition, there is a situation in which both positive values indicating economic advantages and negative values indicating economic disadvantages coexist in the set of segregation values in each CAS. The proportions of positive and negative values should also be taken into account; otherwise, the sizes of the values will be the only determinant of the degree of segregation. Finally, we complete the construction of the TFIDF-ICE-based segregation index, as shown in Eq. (4).

$$ICE_i = \frac{p_i \sum_{j,ICE_{ij}>0} ICE_{ij} + q_i \sum_{j,ICE_{ij}<0} ICE_{ij}}{m_i} \tag{4}$$

Here, $p_i$ and $q_i$ denote the numbers of positive and negative values of $ICE_{ij}$, respectively, and $m_i$ represents the number of categories contained in CAS $i$.

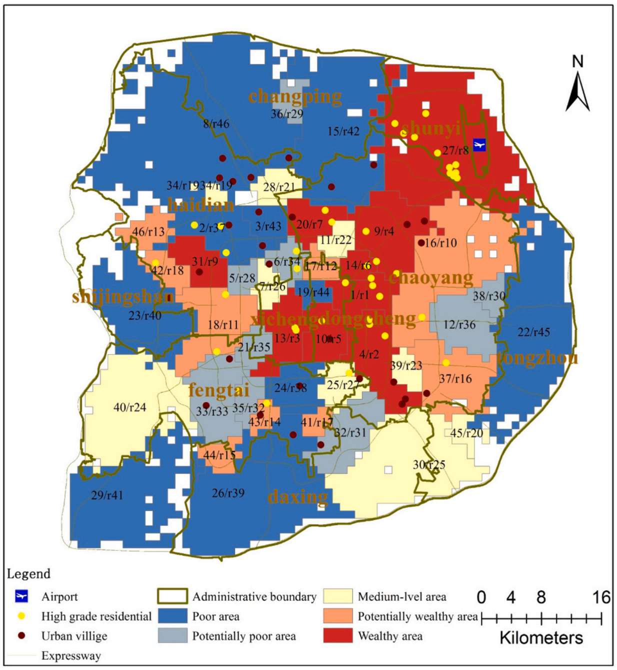
Thus far, we have completed the construction of our TFIDF-ICE-based segregation index. This index is tailored for Dianping data, which involve complex and diverse characteristics that pose several challenges for calculating segregation levels. In addition, this index makes it possible to calculate segregation based on multi-type data as well. By applying the TFIDF-ICE-based segregation index to the combination of Dianping data and two-level CASs, we can perform a comprehensive and realistic investigation of citywide spatial patterns of economic segregation.

### 4.3. *Spatial patterns of urban economic segregation*

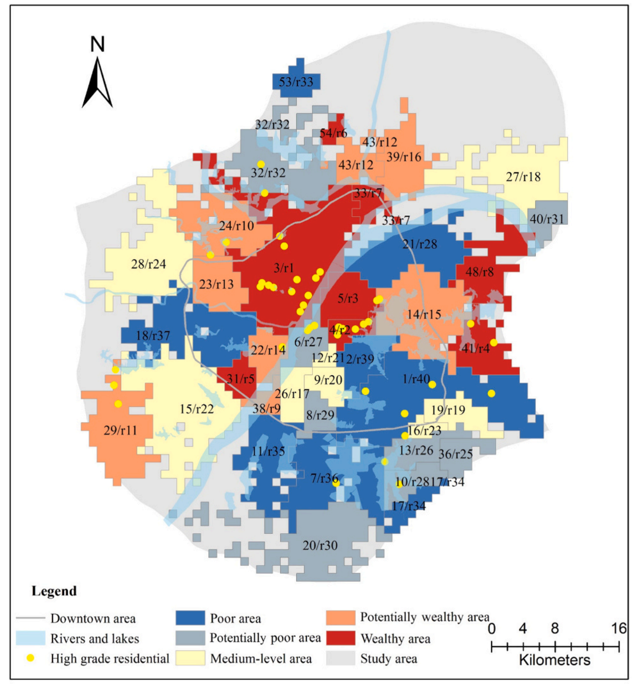
By calculating the TFIDF-ICE-based segregation index using the PCCs of Dianping shops, we obtain the segregation values of the two-level CASs. To reveal the spatial patterns of urban economic segregation, we stratify the values of segregation by quintiles and then map them to the geographic space. The CASs with the highest quintile represent areas with concentrated economic privilege, while CASs with the lowest quintile represent areas with concentrated economic deprivation. To evaluate the results, we collected and marked some high-grade residential and urban villages from the survey. We used the ‘ID/rank’ format to label each CAS, where ‘ID’ represents the index of a CAS and ‘rank’ represents the rank of the ICE value of a CAS. Figs. 6–7 and Tables S1–S2 in the SI Appendix show the details of the results in the two cases.

#### 4.3.1. *Spatial patterns of urban economic segregation in Beijing*

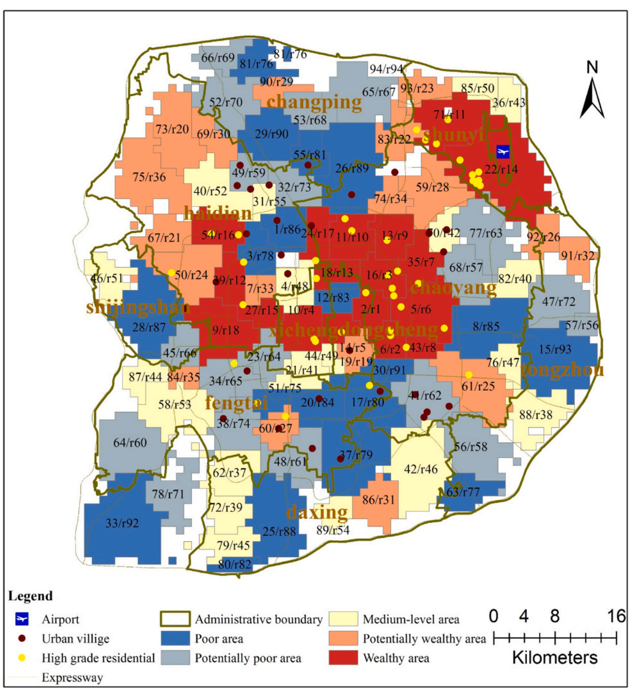
In general, Beijing shows an asymmetric spatial distribution pattern of economic segregation, as revealed in Fig. 6. Although the layout of Beijing is symmetrical around the city centre, the development of the regional economy is not symmetrical around the city centre. Taking the Forbidden City as the centre of symmetry, the eastern part of Beijing is more luxurious than the western part, and the northern part is more luxurious than the southern part. Since antiquity, an adage in Beijing has



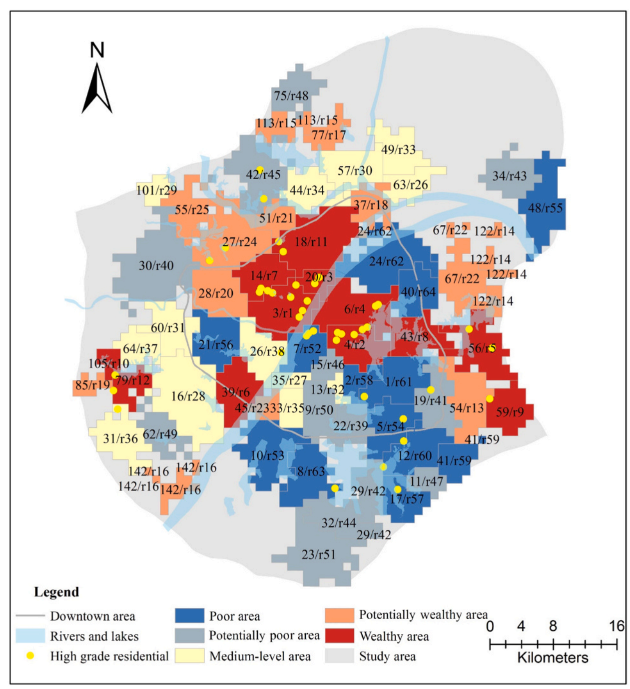

(1) Top-level spatial patterns of Beijing economic segregation




(1) Top-level spatial patterns of Wuhan economic segregation.




(2) Sublevel spatial patterns of Beijing economic segregation.




(2) Sublevel spatial patterns of Wuhan economic segregation.

**Fig. 6.** Two-level spatial patterns of Beijing economic segregation. The locations of famous high-grade residential and urban villages are labelled.

**Fig. 7.** Two-level spatial patterns of Wuhan economic segregation. The locations of famous high-grade residential areas are labelled.

stated that the east is luxurious, the west is noble, the south is impoverished, and the north is humble; the real development pattern is generally but not exactly consistent with this adage. The details of the investigation are described below.

- The east is luxurious. In antiquity, the east was a gathering place for many supplies and granaries and was where wealthy businesspersons settled. Because of the developed transportation infrastructure since then, this area evolved and today includes various office buildings, entertainment venues, and shopping centres, such as the famous trading area of Sanlitun in t-CAS (top-level CAS) 1 and the Beijing Workers Stadium in s-CAS (sublevel CAS) 2, making it the wealthiest area.
- The west is noble. In antiquity, the west was a gathering place for princes and prime ministers, including sites such as Prince Gong's Mansion and Prince Chun's Mansion. Currently, it is a place that

gathers state-owned enterprises and dignitaries and officials, who live a modest and comfortable life.

- The south is impoverished. In antiquity, the south was the place of residence of working people and folk artists, and the south exhibited poor development. There are fewer traffic facilities in this region than in the north; therefore, the flows of material and the development of areas have been less promoted.
- The north is humble. In antiquity, the north was the place of residence for people who served the aristocracy. Historically, traffic conditions were poor and impeded the development of this region. However, changes recently occurred due to preparation for the Beijing 2008 Olympic Games. The Beijing Olympic Park (contained in t-CAS 20) and the Wangjing plate (contained in s-CAS 13) areas along the north Fourth Ring Road were recently developed. In addition, s-CAS 54, which is dominated by both the Summer Palace and Yuquan Mountains, is surrounded by many villas along the north Fifth Ring Road and represents a habitable area for the rich. Since the Beijing 2008 Olympic Games, the development of the north has dramatically improved. This region has attracted the settlement of many wealthy people; therefore, the north is no longer humble but is well developed and prosperous.

In short, our method can reveal the realistic spatial patterns of economic segregation in Beijing. In addition to the details of the investigation described above, we can recognize the role of transportation in promoting urban development in Beijing. Specifically, traffic in the south is not as well managed as that in the north, making its economic development slower than the north's; the closer to the central ring road, the more convenient the transportation and the more developed the economy. In addition, the airport greatly promotes the economy of areas along the route to Beijing Capital International Airport. These findings can provide implications for cities in which transportation has a great effect on promoting economic development.

#### 4.3.2. Spatial patterns of urban economic segregation in Wuhan

The spatial patterns of Wuhan's economic segregation show a strong dependence on the layout of rivers and lakes. Taking the Yangtze River as the main axis and taking the Han River as the secondary axis, the spatial patterns of Wuhan's economic segregation show a 'cross' structure, which is consistent with the Urban Master Plan of Wuhan 2017-2035 (UMPW). The details of the investigation are as follows.

- The spatial patterns of Wuhan's economic segregation along the Yangtze River. In general, areas along the Yangtze River have high concentrations of economic privilege. In particular, the Old Hankou district (contained in t-CAS 3), areas around Dong Lake (composed of s-CAS 4, 6, and 43) and the new ecological town Sixin (contained in t-CAS 31) are areas with concentrated economic privilege. There is a consensus in Wuhan that having a river-view house or a lake-view house is a symbol of wealth. For example, s-CAS 20 consists of the most famous houses with views of the river, including Wuhan-Tiandiyunting and Wuhan-Tiandiyujiangyuan. The s-CAS 3 has many expensive lake-view houses around the Northwest Lake. Similarly, many famous lake-view houses and private villas, such as Shijijiangshang and Donghutianyue, are around Dong Lake. These luxurious houses attract many high-grade facilities that lead to concentrated economic privilege. In addition, according to UMPW, Sixin is planned to become a nationally important exhibition centre. It has successfully attracted many advanced facilities over the past decade, which has led to the concentration of economic privilege.
- The spatial patterns of Wuhan's economic segregation of areas along the Han River to the Guanggu sub-district. In addition to the Old Hankou districts and areas around Dong Lake, the Guanggu sub-district (contained in s-CAS 43) and areas around Yanxi Lake (contained in s-CAS 56) are areas of concentrated economic privilege. As a high-tech industrial zone and with the inclusion of Ma'anshan Forest Park, Guanggu is a vibrant area for business executives and high-level managers. In addition, with the pleasant living and ecological environmental advantages of Yanxi Lake, s-CAS 56 gradually attracts the rich.
- The spatial patterns of Wuhan's economic segregation of areas of concentrated economic deprivation. Urban villages in Wuhan have been rebuilt or are in the process of being rebuilt. In this case, suburbs and universities have become areas with concentrated economic deprivation. Universities with extensive coverage areas mainly attract students without jobs or income, and as a result, these areas contain shops with low PCCs and are areas of economic deprivation. For example, due to the cluster of universities, our results show that the planned Gedihu business residential area (contained in s-CAS 7) and planned Qingshan business area (contained in s-CAS 24) manifest as areas of concentrated economic deprivation. In addition, areas around Tangxun Lake (composed of s-CAS 5, s-CAS 12, and s-CAS 17) are gathering areas for the rich, but clustered university towns in these areas have weakened their economic advantages. Although the development in these areas is up-and-coming, many poor facilities are located around the university town in these areas, which places them at an economic disadvantage.

In short, our method is applicable to evaluating the realistic spatial patterns of economic segregation in Wuhan. In addition to the details of the investigation described above, the role of landforms in promoting urban development in Wuhan is obvious, such as the wealthy areas along the Yangtze River and Han River and near Dong Lake and Northwest Lake. These findings can provide implications for cities in which landforms greatly promote economic development.

In summary, we can identify some characteristics in these areas of concentrated economic status.

(1) The spatial patterns of areas of concentrated economic privilege.

Areas with better humanistic environments, such as areas near downtown areas, important traffic stations, and commercial centres, and areas with better natural environments, such as areas near lakes, are more likely to have surrounding high-grade residences and show concentrated economic privilege. For example, the downtown areas including the Dongcheng and Xicheng districts in Beijing and the Old Hankou district in Wuhan show concentrated economic privilege. Beijing Capital International Airport and the special economic zone of the Beijing Capital International Airport contained in areas corresponding to t-CAS 27 and s-CAS 22 show concentrated economic privilege, as shown in Fig. 6. The location of the Hankou railway station in s-CAS 14 in Fig. 7 shows concentrated economic privilege. Most of the areas around rivers and lakes in Fig. 7 show concentrated economic privilege.

(2) Spatial distribution of areas of economic deprivation.

Areas of economic deprivation are mainly located in suburbs, university areas, and urban villages. For example, t-CAS 22, 26, and 29 in Fig. 6 and s-CAS 23, 41, and 48 in Fig. 7 belong to suburban districts; t-CAS 2 and 3 in Fig. 6 are the major locations of Peking University and Tsinghua University; s-CAS 1 in Fig. 7 is the major location of Central China Normal University and Huazhong University of Science and Technology; and t-CAS 24 in Fig. 6 is the area of Yangqiao Village and Guoyuan Village (urban villages in Beijing).

In conclusion, we have demonstrated the applicability of our method and verified the reasonability of the case study results in this section. Below, we discuss some details of the findings.

## 5. Discussion

The spatial patterns of urban economic segregation obtained by our method can reveal the economic statuses of areas in a bottom-up and realistic way. To further explore the effectiveness and value of our results, we discuss the role of our methods in achieving realistic and comprehensive results. Our study has several limitations, and we will discuss them in this section as well.

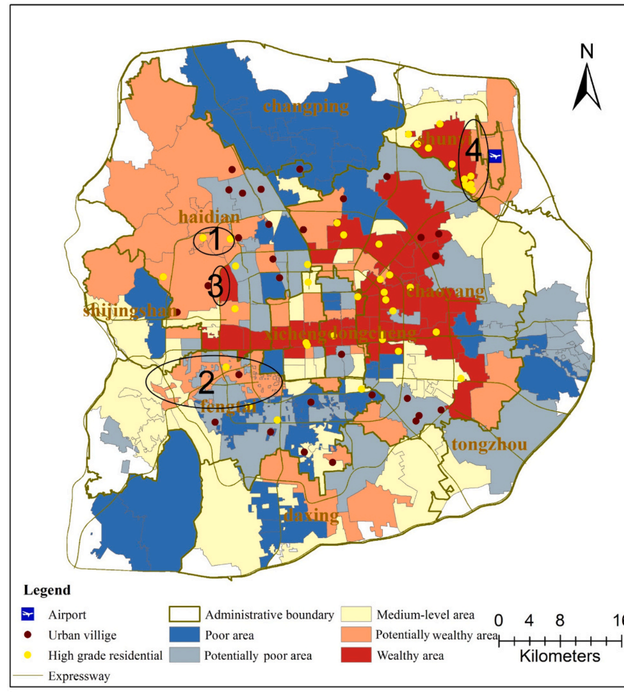
### *5.1. Discussion of the role of our methods in achieving realistic results*

This paper proposes an alternative method by which to uncover realistic spatial patterns of urban economic segregation based on a combination of realistic CASs and spatial economic data obtained from human activities. To prove the effectiveness of this combination in providing realistic results, we use Beijing as a case to compare our results with two conventional spatial patterns of urban economic segregation. In detail, to demonstrate the effectiveness of our CASs in achieving realistic results, we compare our results to results obtained by the combination of census blocks and Dianping data, as shown in Fig. 8 (1). To prove the effectiveness of Dianping data in achieving realistic results, we compare our results to results obtained by the combination of sublevel CASs and house price data, as shown in Fig. 8(2).

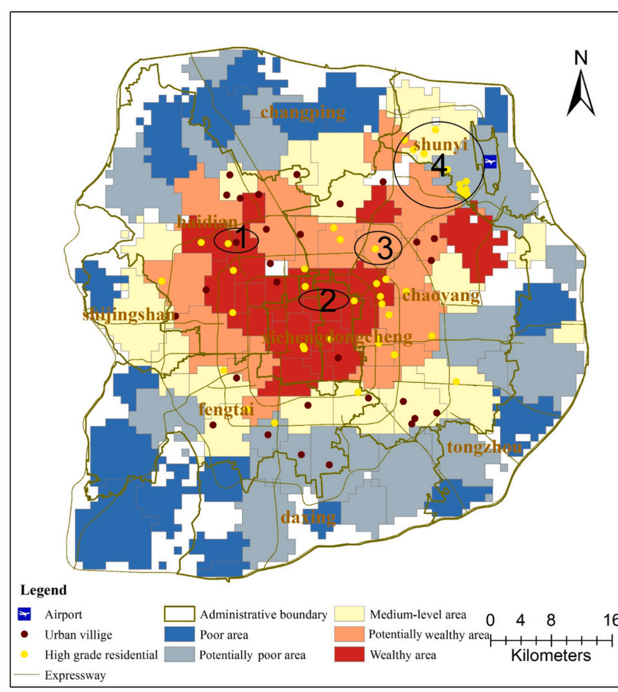
Overall, the spatial patterns of urban economic segregation shown in Fig. 8(1) are consistent with those in Fig. 6(2), which means that the TFIDF-ICE-based index is practical even over conventional spaces; nevertheless, the details reveal several inaccuracies. By analysing Figs. 6 (2) and 8(1), we disclose three disadvantages of using census blocks as basic spatial analytical units compared to using CASs. First, some areas with concentrated economic statuses remain undiscovered. For example, the area labelled by ellipse 1 in Fig. 8(1) corresponds to an area dominated by the Summer Palace and Yuquan Mountains, which we have identified as an area of concentrated economic privilege. Second, inaccurate results inevitably occur regarding enclaves, such as the areas labelled by ellipse 2 in Fig. 8(1). Third, biases may occur when a bustling street is the shared boundary of two census blocks. For example, ellipse 3 and ellipse 4 designate the area of the Zhongguancun military-civilian integration industrial park and the area of Beijing Capital International Airport and the special economic zone of Beijing Capital International Airport, respectively. Businesses on both sides of the shared boundary of the census blocks are flourishing, which means that biases may emerge when evaluating the economic statuses of areas based on census blocks.

Compared to the dynamic agglomerated economic data obtained from human activities, we regard house price data as a type of static indicator of economic status. In the current study, we collected 392,054 house price records from FANG (fang.com), a popular real estate trading platform. Fig. 8(2) shows the results obtained based on the combination of CASs and house price data. Compared to Fig. 6(2), we can reveal two disadvantages of using static indicators of economic status. First, this method cannot uncover fine-grained segregation patterns. The overall level of housing prices in a region depends mainly on location. Downtown areas and historical spots, such as the area of the Summer Palace labelled by ellipse 1, are more likely to have concentrated economic privilege. However, the classification is not without errors. For example, the area labelled by ellipse 2 is a gathering place of quadrangle dwellings that have high house prices and poor facilities. Wangjing and the Beijing Olympic Park are labelled by ellipse 3, which have lower house prices than do the downtown areas but also have excellent facilities for the rich. Second, the estimation of villas is inaccurate. The unit price of a villa is lower than that of a residence in the downtown area, which does not mean that the economic statuses of the people living in villas are not as high as those of the people living downtown.

In short, the combination of CASs and spatial economic data obtained from human activities contributes to a more realistic evaluation of spatial patterns of urban economic segregation.




(1) Census blocks and Dianping data




(2) CASs and house price data

**Fig. 8.** The advantage of the CASs and actual economic data obtained from human activities. (1) shows the spatial patterns of urban economic segregation measured by the combination of census blocks and Dianping data. (2) shows the spatial patterns of urban economic segregation measured by the combination of sub-level CASs and house price data. The black ellipses with labels are examples of several inaccuracies.

### 5.2. Discussion of the role of our methods in achieving comprehensive results

The combined examination of our multilevel results provides a comprehensive evaluation of the patterns of economic segregation. As determined by the reality and hierarchy of human mobility patterns, we have revealed two-level results in our cases, which contribute to a comprehensive evaluation of the spatial patterns of urban economic segregation. The top-level results reveal the patterns from a macroscopic perspective, while the sublevel results reveal the patterns from a microscopic perspective. For example, from the macroscopic perspective, areas along the route to Beijing Capital International Airport show concentrated economic privilege. From the microscopic perspective, s-CAS 59, 74, and 70 in Fig. 6(2) are removed from wealthy areas and have become slightly wealthy or medium-level economic areas. In fact, several villages mainly dominate these areas. The top-level results are the combination of t-CAS 9 and t-CAS 27 in Fig. 6(1), which are dominated by the Wangjing business circle and the special economic zone of the Beijing Capital International Airport, respectively. Influenced by the development of the business circle and the airport's special economic zone, the economic statuses of these areas are not poor but are weaker than those of the business circle and the airport's special economic zone. Therefore, by using two-level CASs, the economic segregation level of an area can be evaluated comprehensively by considering economic positions in both the top-level CAS and sublevel CAS.

### 5.3. Limitations of our methods in uncovering spatial patterns of economic segregation

We demonstrate the applicability of our method in the cases of Beijing and Wuhan; nevertheless, we should discuss some limitations associated with our method. The PCCs of Dianping shops that mostly represent the levels of commercial facilities are one of the major contributors to our method, and they can reflect the development level of the region to a certain extent but not completely. Thus, our results display some deviation. We have summarized five limitations of our results below.

First, inaccurate results may exist in areas with a low diversity of land-use types. For example, t-CAS 46 in Fig. 6 is a tourist area that is the dominant location of the Fragrant Hills and Beijing Botanical Garden. Because this area is slightly distant from downtown, major commercial facilities are gathered within the tourist area and have slightly higher prices than normal, which causes the economic level in this area to be overestimated.

Second, most, but not all, commercial facilities have uploaded their information to the Dianping platform. Shops that are more popular and located in urban areas are more likely to join the Dianping platform, while information for shops located in rural areas, such as villages situated far from cities, may be lost. Therefore, we take Dianping data as a sample dataset to investigate the economic segregation patterns of urban areas based on realistic human consumption information; however, these data may not be applicable to rural areas for which Dianping data are lacked.

Third, the realistic economic statuses of some areas, such as the Haidian and Chaoyang districts in our study, may be inconsistent with our results because of differences in economic structures. According to the Beijing Bureau of Statistics, the annual per capita gross domestic production (PGDP) of the Chaoyang district was higher than that of the Haidian district, but it has become lower by comparison in several recent years. Our two-level results visually show that economic privilege in the Chaoyang district is more concentrated than that in the Haidian district. By investigating the economic structures of these two districts, we note that the Chaoyang district has a commercial and trade economic structure, while education and high-tech industries dominate the economic structure of the Haidian district. The economic structure of Chaoyang district is closely related to people's daily lives, while Haidian district is more suitable for people to work. Therefore, based on human activity data, the results obtained by our method emphasize the investigation of the comprehensive quality of commercial facilities and human life, which is useful and valuable for reducing urban segregation and assisting in sustainable city development.

Fourth, our study has a limitation in revealing refined temporal patterns of segregation because the generation and updating of shops on Dianping is not in real time. Specifically, although different customers of a shop spend different amounts, the PCC of the shop generally does not change because a shop targets a consumer group at the same consumption level. Therefore, our current research can focus only on the spatial patterns of segregation.

Fifth, social media data are not a random sample, which means some kinds of biases such as demographic biases exist in social media data and these biases are difficult to mitigate (Liu et al., 2020; Yuan et al., 2020). For example, Yuan et al. (2020) found that the senior population (age sixty-five and older) in China is systematically underrepresented on Weibo. Nevertheless, these data still provide valuable insights in urban studies (Salganik, 2018).

In general, our method shows advantages in the comprehensive and realistic investigation of citywide spatial patterns of economic segregation. In addition, there are some limitations to be addressed in the future.

## 6. Conclusions

A comprehensive and realistic investigation of citywide spatial patterns of economic segregation has important implications for the configuration of cities and urban planning. Our study contributes to this emerging body of research by proposing an alternative method based on the reality of human mobility and human activities. We first define multilevel CASs as basic spatial analytical units by analysing the hierarchical patterns of human activity mobility. This approach provides analytical units that follow the reality and comprehensiveness of human mobility and mitigate the MAUP to a certain extent. Based on these CASs, we tailor the TFIDF-ICE-based segregation index for human economic activity data, which are Dianping data, to measure the degree of economic segregation. The measurement relies on data generated from human economic activities, which further embodies the consideration of the reality of human activity. By using this method, we can ultimately perform a realistic and comprehensive evaluation of citywide economic segregation.

In the cases of Beijing and Wuhan, by analysing the reliability of our experimental results, we proved the applicability and value of our method in the comprehensive and realistic investigation of citywide spatial patterns of economic segregation. In the case of Beijing, taking the Forbidden City as the centre of symmetry, we have revealed that the eastern part of Beijing is richer than the western part, while the northern part is richer than the southern part. This case demonstrated the applicability of our research in cities with a development mode promoted by transportation. In the case of Wuhan, taking the Yangtze River as the central axis and the Han River as the secondary axis, the spatial patterns of economic segregation show a 'cross' structure. This case demonstrated the applicability of our research in cities with a development mode affected by landforms. Moreover, the common patterns of these two cases show that areas near downtown, important traffic stations, commercial centres, and areas with extensive natural environments show concentrated economic privilege and are usually surrounded by high-grade residential areas. Areas of economic deprivation are mainly located in suburbs, university areas, and urban villages. These common patterns further proved the robustness of our method.

By comparison with census blocks and house price data, we discussed the advantages of using CASs and spatial economic data obtained from human activities in making the evaluation more realistic. We also discussed the function of multiple levels in making the evaluation of urban economic segregation more comprehensive. Moreover, we discussed the limitations of our methods. Biases exist in the evaluation of some areas, such as rural areas where Dianping data are lacking, areas with an economic structure mainly consisting of non-commercial activities, and areas that typically have a low diversity of land-use types. An additional limitation of our study is related to revealing refined temporal patterns of segregation because the generation and updating of shops on Dianping is not in real time.

The main finding in this article reminds us that examining the reality of human activity can help to identify the realistic spatial patterns of economic segregation. The facilities are the link that connects human activity and the spatial patterns of economic segregation. If a place is equipped with enough high-end facilities, the rich will come to settle down to enjoy the amenities, such as in areas with business districts and high-grade residences. In contrast, people with low budgets will gather in areas with poor facilities, such as in areas with university towns. Furthermore, areas with the same economic level tend to congregate into larger areas, as shown in Figs. 6–7 and Table S1–S2, which in turn aggravates the economic segregation of the city. To embrace the findings of this study, we suggest reducing agglomeration of areas at the same economic level and considering the combination of areas at low and high economic levels to reduce economic segregation. Specifically, planning a business district or a scenic area in a university town will promote economic development in the university town, which is valuable to reducing economic segregation.

In conclusion, the multilevel results obtained by our method present a comprehensive and realistic investigation of citywide spatial patterns of economic segregation. Despite its drawbacks, the method we propose is effective, and the replicability and open-source features of our study demonstrate its great value and implications for the official management and sustainable development of cities. In future research, we will consider more factors, such as the population, differences in the economic structures of different areas, mixtures of land use types, and vitality of commercial facilities, in this method.

## CRediT authorship contribution statement

**Mengling Qiao**: Conceptualization, Methodology, Writing - Original Draft, Data Curation, Validation, Investigation. **Yandong Wang**: Conceptualization, Writing - Review & Editing, Funding acquisition. **Shanmei Wu**: Software, Data Curation, Investigation. **Xiaokang Fu**: Conceptualization. **Yanyan Gu**: Data Curation. **Mingxuan Dou**: Conceptualization.

## Declaration of competing interest

We have no conflicts of interest to disclose.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cities.2020.103067.

## References

Aizawa, A. (2003). An information-theoretic perspective of TF–IDF measures. *Information Processing & Management, 39*(1), 45–65.

Catney, G. (2018). The complex geographies of ethnic residential segregation: Using spatial and local measures to explore scale-dependency and spatial relationships. *Transactions of the Institute of British Geographers, 43*(1), 137–152. https://doi.org/10.1111/tran.12209.

Chodrow, P. S. (2017). Structure and information in spatial segregation. *Proceedings of the National Academy of Sciences, 114*(44), 11591. https://doi.org/10.1073/pnas.1708201114.

Clark, W. A. V., Anderson, E., Östh, J., & Malmberg, B. (2015). A multiscalar analysis of neighborhood composition in Los Angeles, 2000–2010: A location-based approach to segregation and diversity. *Annals of the Association of American Geographers, 105*(6), 1260–1284. https://doi.org/10.1080/00045608.2015.1072790.

Darden, J. T., & Kamel, S. M. (2000). Black residential segregation in the city and suburbs of Detroit: Does socioeconomic status matter? *Journal of Urban Affairs, 22*(1), 1–13. https://doi.org/10.1111/0735-2166.00036.

Ellis, M., Wright, R., & Parks, V. (2004). Work together, live apart? Geographies of racial and ethnic segregation at home and at work. *Annals of the Association of American Geographers, 94*(3), 620–637. https://doi.org/10.1111/j.1467-8306.2004.00417.x.

Farber, S., O'Kelly, M., Miller, H. J., & Neutens, T. (2015). Measuring segregation using patterns of daily travel behavior: A social interaction based model of exposure. *Journal of Transport Geography, 49*, 26–38.

Firebaugh, G., & Acciai, F. (2016). For blacks in America, the gap in neighborhood poverty has declined faster than segregation. *Proceedings of the National Academy of Sciences, 113*(47), 13372. https://doi.org/10.1073/pnas.1607220113.

Gao, S., Liu, Y., Wang, Y., & Ma, X. (2013). Discovering spatial interaction communities from mobile phone data. *Transactions in GIS, 17*(3), 463–481. https://doi.org/10.1111/tgis.12042.

Ghasemi, K., Hamzenejad, M., & Meshkini, A. (2018). The spatial analysis of the livability of 22 districts of Tehran Metropolis using multi-criteria decision making approaches. *Sustainable Cities and Society, 38*, 382–404. https://doi.org/10.1016/j.scs.2018.01.018.

Golledge, R. G., & Stimson, R. J. (1997). *Spatial behavior: A geographic perspective.* New York: The Guilford Press.

Guo, D., Jin, H., Gao, P., & Zhu, X. (2018). Detecting spatial community structure in movements. *International Journal of Geographical Information Science, 32*(7), 1326–1347. https://doi.org/10.1080/13658816.2018.1434889.

James, D. R., & Taeuber, K. E. (1985). Measures of segregation. *Sociological Methodology, 15*, 1–32. https://doi.org/10.2307/270845.

Jargowsky, P. A. (1996). Take the money and run: Economic segregation in U.S. metropolitan areas. *American Sociological Review, 61*(6), 984–998. https://doi.org/10.2307/2096304.

Järv, O., Müürisepp, K., Ahas, R., Derudder, B., & Witlox, F. (2015). Ethnic differences in activity spaces as a characteristic of segregation: A study based on mobile phone usage in Tallinn, Estonia. *Urban Studies, 52*(14), 2680–2698. https://doi.org/10.1177/0042098014550459.

Kallus, Z., Barankai, N., Szüle, J., & Vattay, G. (2015). Spatial fingerprints of community structure in human interaction network for an extensive set of large-scale regions. *PLoS One, 10*(5), e0126713. https://doi.org/10.1371/journal.pone.0126713.

Kwan, M.-P. (2012). The uncertain geographic context problem. *Annals of the Association of American Geographers, 102*(5), 958–968. https://doi.org/10.1080/00045608.2012.687349.

Kwan, M.-P. (2013). Beyond Space (As We Knew It): Toward temporally integrated geographies of segregation, health, and accessibility. *Annals of the Association of American Geographers, 103*(5), 1078–1086. https://doi.org/10.1080/00045608.2013.792177.

Le Roux, G., Vallée, J., & Commenges, H. (2017). Social segregation around the clock in the Paris region (France). *Journal of Transport Geography, 59*, 134–145.

Leckie, G., & Goldstein, H. (2015). A multilevel modelling approach to measuring changing patterns of ethnic composition and segregation among London secondary schools, 2001–2010. *Journal of the Royal Statistical Society, Series A, 178*(2), 405–424. https://doi.org/10.1111/rssa.12066.

Li, F., & Wang, D. (2017). Measuring urban segregation based on individuals' daily activity patterns: A multidimensional approach. *Environment and Planning A, 49*(2), 467–486.

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. https://doi.org/10.1016/j.jtrangeo.2015.01.016.

Liu, Y., Yuan, Y., & Zhang, F. (2020). Mining urban perceptions from social media data. *Journal of Spatial Information Science, 20*. https://doi.org/10.5311/josis.2020.20.665.

Louf, R., & Barthelemy, M. (2016). Patterns of residential segregation. *PLoS One, 11*(6), e0157476. https://doi.org/10.1371/journal.pone.0157476.

Malmberg, B., Andersson, E., & Östh, J. (2013). Segregation and urban unrest in Sweden. *Urban Geography, 34*(7), 1031–1046. https://doi.org/10.1080/02723638.2013.799370.

Miller, H. (2007). Place-based versus people-based geographic information science. *Geography Compass, 1*(3), 503–535. https://doi.org/10.1111/j.1749-8198.2007.00025.x.

Moroke, T., Schoeman, C., & Schoeman, I. (2019). Developing a neighbourhood sustainability assessment model: An approach to sustainable urban development. *Sustainable Cities and Society, 48*, 101433. https://doi.org/10.1016/j.scs.2019.101433.

Musterd, S., Marcińczak, S., van Ham, M., & Tammaru, T. (2017). Socioeconomic segregation in European capital cities. Increasing separation between poor and rich. *Urban Geography, 38*(7), 1062–1083. https://doi.org/10.1080/02723638.2016.1228371.

Olteanu, M., Randon-Furling, J., & Clark, W. A. V. (2019). Segregation through the multiscalar lens. *Proceedings of the National Academy of Sciences, 116*(25), 12250–12254.

Openshaw, S., & Taylor, P. J. (1981). The modifiable areal unit problem. In N. Wrigley, & R. J. Bennett (Eds.), *Quantitative geography: A British view* (pp. 60–70). London: Routledge & Kegan Paul.

Östh, J., Shuttleworth, I., & Niedomysl, T. (2018). Spatial and temporal patterns of economic segregation in Sweden's metropolitan areas: A mobility approach. *Environment and Planning A: Economy and Space, 50*(4), 809–825. https://doi.org/10.1177/0308518X18763167.

Pareto, V. (1896–97). *Cours d'économie politique.* Lausanne: Rouge.

Park, Y. M., & Kwan, M.-P. (2017). Individual exposure estimates may be erroneous when spatiotemporal variability of air pollution and human mobility are ignored. *Health & Place, 43*, 85–94.
Park, Y. M., & Kwan, M.-P. (2018). Beyond residential segregation: A spatiotemporal approach to examining multi-contextual segregation. *Computers, Environment and Urban Systems, 71*, 98–108.
Qiao, M., Wang, Y., Wu, S., Luo, A., Ruan, S., & Gu, Y. (2019). Fine-grained subjective partitioning of urban space using human interactions from social media data. *IEEE Access, 7*, 52085–52094. https://doi.org/10.1109/ACCESS.2019.2911664.
Reardon, S. F., & Firebaugh, G. (2002). Measures of multigroup segregation. *Sociological Methodology, 32*, 33–67.
Robertson, S. E., & Jones, K. S. (1976). Relevance weighting of search terms. *Journal of the American Society for Information Science, 27*(3), 129–146.
Salganik, M. J. (2018). *Bit by bit: Social research in the digital age*. Princeton, NJ: Princeton University Press.
Salton, G., & Buckley, C. (1988). Term-weighting approaches in automatic text retrieval. *Information Processing & Management, 24*(5), 513–523.
Salton, G., & Yu, C. T. (1973). On the construction of effective vocabularies for information retrieval. *SIGIR Forum, 9*(3), 48–60. https://doi.org/10.1145/951761.951766.
Scally, B. J., Krieger, N., & Chen, J. T. (2018). Racialized economic segregation and stage at diagnosis of colorectal cancer in the United States. *Cancer Causes & Control, 29*(6), 527–537. https://doi.org/10.1007/s10552-018-1027-y.
Shen, Y. (2019). Segregation through space: A scope of the flow-based spatial interaction model. *Journal of Transport Geography, 76*, 10–23.
Silm, S., & Ahas, R. (2014). Ethnic differences in activity spaces: A study of out-of-home nonemployment activities with mobile phone data. *Annals of the Association of American Geographers, 104*(3), 542–559. https://doi.org/10.1080/00045608.2014.892362.
Tan, Y., Kwan, M.-P., & Chai, Y. (2017). Examining the impacts of ethnicity on space-time behavior: Evidence from the City of Xining, China. *Cities, 64*, 26–36.
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. https://doi.org/10.1016/j.physa.2015.06.032.
Wang, D., Li, F., & Chai, Y. (2012). Activity spaces and sociospatial segregation in Beijing. *Urban Geography, 33*(2), 256–277. https://doi.org/10.2747/0272-3638.33.2.256.
Wong, D. W. S., & Shaw, S.-L. (2011). Measuring segregation: An activity space approach. *Journal of Geographical Systems, 13*(2), 127–145. https://doi.org/10.1007/s10109-010-0112-x.
Yuan, Y., Lu, Y., Chow, T. E., Ye, C., Alyaqout, A., & Liu, Y. (2020). The missing parts from social media–enabled smart cities: Who, where, when, and what? *Annals of the American Association of Geographers, 110*(2), 462–475.
Zhang, X., Wang, J., Kwan, M.-P., & Chai, Y. (2019). Reside nearby, behave apart? Activity-space-based segregation among residents of various types of housing in Beijing, China. *Cities, 88*, 166–180.
Zhong, C., Arisona, S. M., Huang, X., Batty, M., & Schmitt, G. (2014). Detecting the dynamics of urban structure through spatial network analysis. *International Journal of Geographical Information Science, 28*(11), 2178–2199. https://doi.org/10.1080/13658816.2014.914521.

**Mengling Qiao** is currently working toward the PH.D. degree in the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing. Her research interests include spatiotemporal data mining, and urban analysis.

**Yandong Wang** is currently a Professor is in the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University. His research interest mainly concentrates in space information intelligent service, GIS theory and engineering application, and spatiotemporal information knowledge mining.

**Shanmei Wu** is studying for a M.S. degree at Wuhan university of Cartography and Geographical Information System. Her research interests lie in the fields of Geographical Information Science, data mining.

**Xiaokang Fu** is currently working toward the PH.D. degree in the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing. His research interests include text mining, deep learning and disaster analysis.

**Yanyan Gu** is currently working toward the PH.D. degree in the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing. His research interests include text mining, deep learning and transportation planning.

**Mingxuan Dou** is currently working toward the PH.D. degree in the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing. Her research interests include spatiotemporal data mining, and disaster analysis.