Manuscript

Urban Analytics and City Science

EPB: Urban Analytics and City Science
2022, Vol. 0(0) 1–16

DOI: 10.1177/23998083221140415
journals.sagepub.com/home/epb
SAGE

How do urban services facilities affect social segregation among people of different economic levels? A case study of Shenzhen city

Yuyang Wu
School of Geography and Information Engineering, China University of Geoscience, Wuhan, China

Yao Yao
School of Geography and Information Engineering, China University of Geoscience, Wuhan, China

Center for Spatial Information Science, The University of Tokyo, Kashiwa-shi, Japan

Shuliang Ren and Shiyi Zhang
School of Geography and Information Engineering, China University of Geoscience, Wuhan, China

Qingfeng Guan
School of Geography and Information Engineering, China University of Geoscience, Wuhan, China

National Engineering Research Center of GIS, China University of Geoscience, Wuhan, China

Abstract

Social segregation hinders the development of cities and has become a hot topic in urban research. Existing studies have focused on the difference in the distribution of crowd activities to measure segregation but have ignored the impact of the urban environment on crowd gathering and segregation. To study the impact and understand social segregation more comprehensively, we coupled mobile phone datasets and housing price data to divide city dwellers into three socio-economic levels. Considering that spatial colocation is a necessary condition for interaction among various social groups, spatial colocation probability was proposed to quantitatively describe the degree of social segregation at the community scale. Point-of-interest (POI) data were introduced to represent the urban service facilities. The effect of urban service facilities on the segregation of different groups was analyzed by using geographically weighted regression (GWR). The results indicate three points, as follows. (1) Significant social segregation in Shenzhen mostly occurs in suburban and downtown areas, and the interaction segregation of people mainly occurs between people with high and low socioeconomic levels. (2) More economically inclusive and necessary

Corresponding author:
Yao Yao, School of Geography and Information Engineering, China University of Geosciences, Wuhan 430078, China; Center for Spatial Information Science, The University of Tokyo, Kashiwa-shi, Chiba 277-8568, Japan.
Emails: yaoy@cug.edu.cn; yaoy@csis.u-tokyo.ac.jp

service facilities (e.g., medical and insurance companies) can promote crowd interaction and ease the segregation of social activities. (3) The impact of service facilities on the interaction of various social groups is related to the development of the area where the activities occur, and the most significant impact is in high-tech industrial zones. This study quantitatively calculated the impacts of different service facilities on different groups of people in different communities and times. From the results, detailed and reasonable suggestions were made for city planners.

Keywords
Social segregation, urban environment, mobile phone data, urban planning, geographically weighted regression

Introduction

Social segregation refers to the phenomenon in which urban residents cluster according to certain standards (e.g., age, education level, and economic income) due to economic (Dorman et al., 2020; Iceland and Wilkes, 2006), cultural (Robinson, 2020), and social network differences (Carrasco and Miller, 2006; Xu et al., 2021). These differences ultimately lead to the separation of activity spaces and the unequal use of urban resources (J A Rv et al., 2021; Schnell and Yoav, 2001). Social segregation results from the comprehensive influence of human activities and urban characteristics and is not conducive to the development of cities. A disadvantaged group may live in the cities' edges, with the lowest-quality housing and income (Downs, 2010). In European cities, segregation has led to increasingly severe problems of poverty and crime (Tammaru et al., 2015). In addition, urban decay and deprived neighborhoods may appear in certain parts of cities due to segregation (Andersen, 2019). Therefore, measuring segregation and promoting communication among different groups of people have become important topics of urban research.

The economy is an important indicator to measure the level of regional development within a city and an important factor that affects the pattern of people's activity (Hanson and Hanson, 1981; Pappalardo et al., 2015). Researching social segregation from an economic perspective could help researchers understand the reasons for the emergence of segregation and restrictions on crowd interaction. In early studies, travel surveys were commonly used to study the activities of different socioeconomic groups (Krivo et al., 2013; Le Roux et al., 2017). These studies indicate the existence of social segregation among people of different socioeconomic statuses (SESs). However, limited by the time-consuming nature and information inaccuracy of surveys, these studies cannot analyze people's activities in a timely and comprehensive manner, and their results are likely biased and limited (J A Rv et al., 2021; Xu et al., 2018).

In the era of big data, massive human-tracking data have promoted research on human activity patterns (Wang et al., 2011; Yang et al., 2021). Silm and Ahas (2014) analyzed the travel patterns of different language groups using the trajectory data provided by an Estonian mobile phone operator. They found that language significantly influences individual travel patterns and that there is strong segregation between groups speaking different languages. Xu et al. (2017) divided the relationships among Singaporeans into friends and strangers based on mobile phone call data and explored the different ways they share urban land. Human-tracking data make it possible to study social segregation on a large scale. However, because of privacy issues, most data, such as mobile phone data, lack users' background information (e.g., personal economic level) (Huang and Wong, 2016; Xu et al., 2018). This shortcoming limits the further exploration of human activity patterns and the social segregation of people at different economic levels.

Many scholars have pointed out that the nature of the home neighborhood has profound implications for the well-being of individuals (Brasington et al., 2015; Mouratidis, 2020). As a result, a commonly used method to compensate for the lack of economic information is using housing prices to reflect the SES of residents. For example, Huang and Wong (2016) used the median house value (MHV) of the community from US survey data (ACS) to infer the SES of community residents. Xu et al. (2018) also defined individuals' SES through housing prices provided by private websites in Singapore. Based on the data with additional economic information, an increasing number of scholars have joined in studying social segregation among people of different economic levels. For example, Xu et al. (2019) proposed a framework that couples human mobility and social networks to quantitatively measure the segregation of people of different economic levels. Using social media data with economic attributes, Huang and Wong (2016) analyzed the differences in occupational and residential locations of different SESs. However, although we can observe the segregation of human beings in social and physical space by utilizing these methods, it is still difficult to analyze the reasons behind the segregation. Without the in-depth exploration of the relationship between segregation and the environment where segregation takes place, it is impossible to have a deeper understanding of crowd activities in the city and make scientific suggestions for alleviating social segregation.

The movement of people is the basis for cities; it promotes social, economic and cultural exchanges and shapes cities (Schl A Pfer et al., 2021). With the deepening of research on human behavior patterns, an increasing number of scholars have discovered a correlation between crowd activities and urban functions. The temporal–spatial behavior of the crowd affects the internal planning and development of the city. On the one hand, the functions of urban settings change with variations in human activities (Tu et al., 2017). On the other hand, the service facilities in the city drive the crowd to travel. The complexity of regional functional attributes affects the concentration of people (Yue et al., 2017). These results provide a reference for understanding the relationship between urban functions and social segregation.

Human activities and cities interact and are inseparable (Frank and Engelke, 2001). Segregation is not a simple consequence of social inequality but a product of both social and spatial differentiation (Andersen, 2019). To better understand and find ways to alleviate the social segregation of people of different SESs, we must consider their activity environment: cities. Therefore, exploring the drivers of aggregation or the mixing of people with different SESs from the perspective of the urban environment is a critical issue in facilitating crowd interaction and alleviating social isolation. As the carrying places for crowd activities and the concrete manifestations of urban functions (Lee et al., 2013), service facilities (e.g., hospitals and schools) are the bridges connecting people to the city. They are also the most common and flexible facilities in urban planning. Therefore, to alleviate social segregation it is necessary to quantitatively analyze the impact of urban service facilities on the activities of people of different SESs. However, to the best of our knowledge, there is a lack of relevant studies.

To fill this gap, we propose a new perspective to analyze segregation in cities. Focusing on urban land, we measured the community's ability to gather people of different SESs in one time period. Then, we defined this ability as "gathering capacity". In this study, we quantitatively delineated the social segregation of the study area by analyzing the differences in "gathering capacity" between communities. Finally, point-of-interest (POI) data were introduced to indicate the distribution of service facilities. Using the geographically weighted regression (GWR) analysis method, the correlation between this ability and service facilities was obtained to determine the effect of urban functions on social segregation. Then, based on the research results, some opinions for government urban planning were provided.

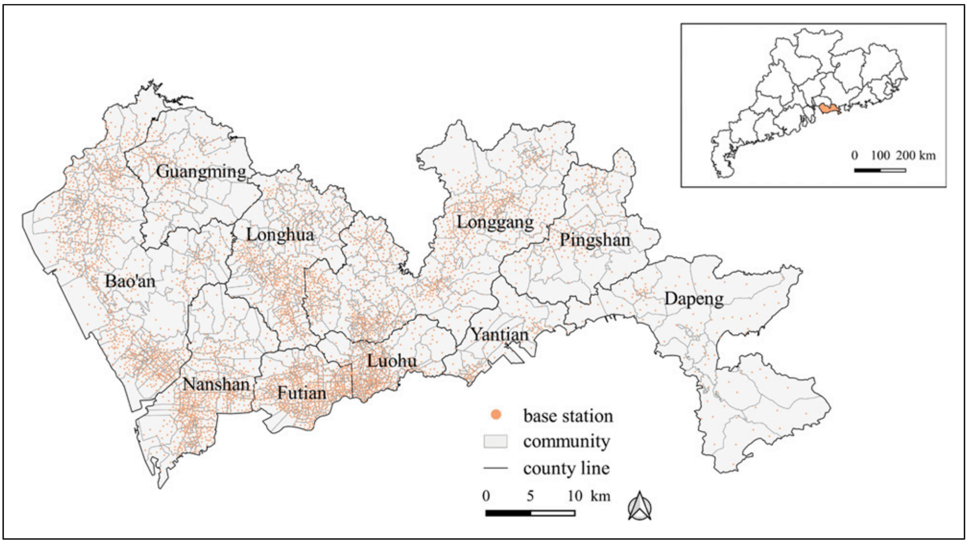

Figure 1. The geographical position of Shenzhen city in Guangdong and the division of mobile phone base stations and community units in the Shenzhen research area.

Study area and dataset

Study area

The study area is Shenzhen, Guangdong Province (Figure 1). Its open policy and inclusive culture have attracted a large influx of people (Yoon and Jongseok, 2016). The complexity of economic development and population composition make it a typical area for research on human activities and social segregation (Zhang et al., 2021). Shenzhen has 10 administrative districts. The 2010–2020 urban master plan of Shenzhen (Bureau, 2018) can be divided into three types of regions: downtown areas (Luohu District, Futian District, Yantian District, and Nanshan District), which have a high level of economic development and are dominated by residential and commercial land; high-tech industrial zones (Longgang District and Bao'an District), which have manufacturing and high-tech industries as the main functions and are mostly industrial land and residential land; and other districts belonging to suburban areas (Longhua District, Pingshan District, Guangming, and Dapeng New District) (Wu and Bing, 2010). Shenzhen can be divided into 781 neighborhood committees (NCs), which are the smallest geographic units in this study. NCs are the official, local community organizations in urban China, and we generally refer to the NC as the community. Information on the size and other aspects of communities can be found in Supp 1.

Dataset

Mobile phone data are used in this study to provide the activity information of residents. These data were collected from 5943 mobile phone base stations in Shenzhen in an active manner (Liu et al., 2011). Regardless of whether the user made or received a call, the latitude and longitude of the user's base station were collected every hour. The data format consists of four parts, namely, a user ID (i), records number (n), time stamp (t_i), and location information (x_i, y_i), where $i \in (0, n]$ and i is an integer. The total time of data collection in this study was one working day (24 hours) in

Table 1. Example of Shenzhen mobile phone data.

User ID	Records	Time	Location	Time	…
S8ff0*******h9r2	18	20120323 00:01:33	114.18** 22.63**	20120323 01:24:36	…
0bhc1*******892k	23	20120322 23:30:12	114.21** 22.61**	20120323 00:30:12	…
8dk32*******pdk6	24	20120322 23:24:36	114.21** 22.61**	20120323 00:12:24	…
472js*******29sj	16	20120322 23:24:28	114.23** 22.60**	20120323 00:27:02	…
…	…	…	…	…	…
dae2a*******w349	18	20120323 10:51:34	114.34** 22.68**	20120323 11:21:39	…

2012, and the number of users was 10.463 million. Table 1 shows the detailed structure of the mobile phone data.

Previous studies have verified that human activity presents a strong, stable pattern, and the probability of people repeating their activities in the same places every day is very high (Song et al., 2010). In addition, our data cover more than 80% of Shenzhen users in total. Therefore, a representative and generalized picture of the travel patterns of the whole Shenzhen population could be extracted. In fact, this sort of approach has been used to study urban vitality, work–housing balance, and the spatiotemporal activities of residents (Wang et al., 2021; Xu et al., 2015; Yue et al., 2017). To better understand the activities of the crowd from the perspective of the urban area, we resampled the mobile phone data at the community scale.

The SES index is sensitive and difficult to obtain directly. Previous studies have confirmed that residential prices can be used to measure people's socioeconomic level (Xu et al., 2018; Zhang et al., 2021). Therefore, this study obtained Shenzhen's fine-scale housing prices based on the method proposed by Yao et al., (2018) and used it as the economic information of users (Figure S1). In particular, since our smallest geographic study unit is the community, the average house price within the community is used as the house price data in our study. In Supp 1.1, we justified the use of average house prices to measure the SESs of the population at the community scale.

POI data can reflect the specific distribution of various service facilities in the city (Jiang et al., 2021), and they form a "thematic map" in each urban region (McKenzie et al., 2015). The powerful semantic information representation capability of POIs has helped many scholars analyze the relationship between urban functional structure and population activities (Jiang et al., 2021; Tu et al., 2017; Wang et al., 2018). In this study, POI data were used to further study the impact of service facilities on segregation. According to the *Classification of Land Use Status* standard and previous studies (Xu et al., 2017), 9 types of POIs (food and beverage, business, shopping, traffic facilities, finance and insurance, science and education, sports and leisure, health care, and hotels and resorts) were selected in the study. The source of POI data is Baidu's data collected in 2012, and in Table S2, more information about each category can be found.

Methodology

As shown in Figure S2, our study can be divided into 3 parts: (1) users' SES class definitions; (2) quantification of the "gathering" capabilities of places; and (3) the relationship between service facilities and "gathering". The next three sections of this chapter will elaborate on each step of the methodology.

Classifying people at different economic levels

Users' activity locations in the residence time window (2:00–5:00, refer to (Gong et al., 2016)) were counted. The community with the highest number of occurrences was selected as the user's residential community. Then, the average house price of the user's residential community was used as the user's SES index. Finally, according to the social stratum model algorithm (Leo et al., 2016) and the principle of equal economic aggregates of different populations, the overall population was divided into 3 different socioeconomic level groups (high, medium and low).

Measuring the "gathering capacity" of the community for two different groups

Spatiotemporal encounters are a precondition for interpersonal meetings, interactions, and shared activities (Crandall et al., 2010; Wang et al., 2011; Xu et al., 2017). This study proposes a method to measure the ability of communities to drive spatiotemporal encounters among people of different economic levels, that is, the "gathering capacity". Referencing the calculation method of colocation probability, our formula is expressed as follows.

Let x be a user within a group A of a certain economic level; then, the probability of group A appearing at community L in time period T is:

$$p_A(L, T) = \frac{1}{N} \sum_{x \in A} m_x(L, T) / n_x(T)$$

where $m_x(L, T)$ denotes the number of times user x appears at community L during T, $n_x(T)$ denotes the total number of locations where user x has activity records during T, and N is the total number of users of economic group A who have activity records during T.

If A and B are two groups with different economic levels, and assuming that the activities between different users do not affect each other, the probability of their simultaneous presence at community L in time period T is:

$$p_{A,B}(L, T) = p_A(L, T) \times p_B(L, T)$$

which is the "gathering capacity" of community L for people of different economic levels A and B in time period T. This reflects the possibility of users sharing urban land and interacting with each other and reflects the magnitude of the "gathering capacity" effect of the community on them.

Analysis of the relationship between service facilities and communities' "gathering capacity"

To further analyze the reasons for differences among communities' "gathering capacity," this study introduced POI data to represent the distribution of service facilities and used GWR to mine the drivers of the "gathering capacity" of the communities. In addition, before the formal construction of the GWR model, collinearity testing of the independent variable factors was carried out.

The *geographically weighted regression* (GWR) model (Brunsdon et al., 1998) combines spatial correlation and linear regression to reflect the spatial heterogeneity among variables. It is widely used in the field of GIS to focus on research objects with unstable spatial and temporal distributions. In this study, the number of each type of POI within communities is used as the influencing factor (independent variable), and the "gathering capacity" is used as the dependent variable. The method can be presented as follows:

$$\gamma_i = \beta_0(u_i, v_i) + \sum_{k=1}^{p} \beta_k(u_i, v_i)x_{ik} + \varepsilon_i$$

The detailed meaning of each parameter in the formula was well explained in Supp 5.

Results

Our study classified the population into three categories (high, medium, and low) according to their SESs (Table S1). The classified populations will be represented by H, M, and L in the rest of the article. In addition, there were 9 different "gathering capacities" of the communities in our research, and their values and distribution can be found in the Section 4.1. For the final sample of the GWR model, Table S1 provides the descriptive statistics; In addition, Table S2 confirms the non-collinearity of the independent variable factors (see more information in Supp 1.1), which enhances the reliability of the interpretation of the regression results in Section 4.2.

The "gathering capacity" of the communities in the study area

Figure 2 shows the degree to which the community gathers people of different economic levels. In general, the capacity of high-tech industrial zones (Bao'an District and Longgang District) is relatively strong, approximately 4 times that of other regions. The communities in the downtown area (Nanshan and Futian) have a strong ability to gather H and M people (denoted by H-M in the

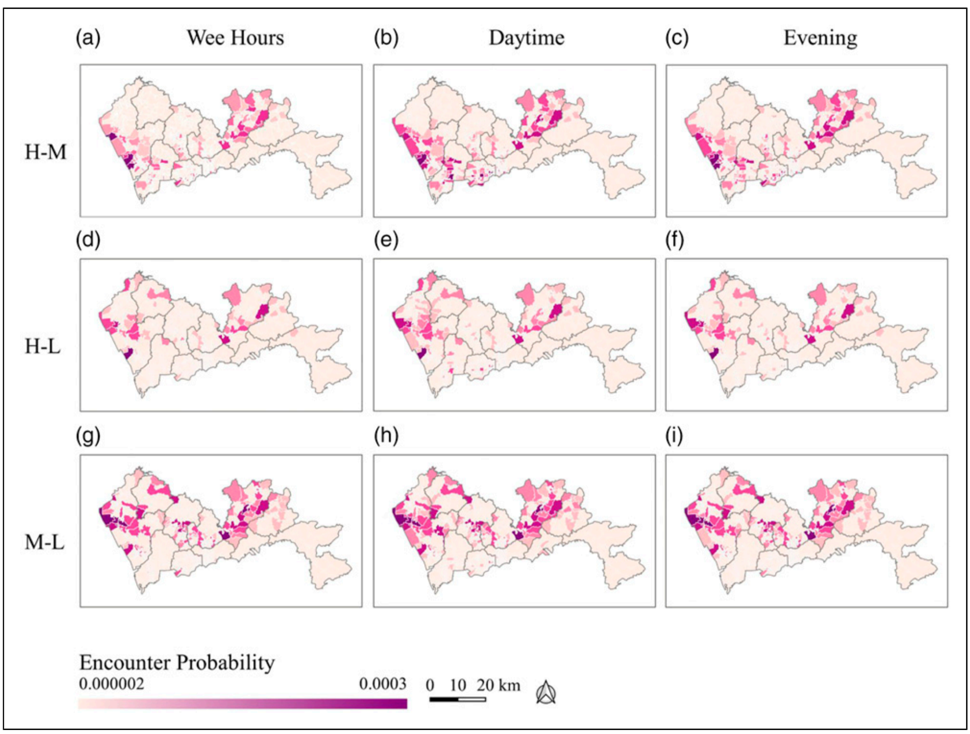

Figure 2. Distribution map of "gathering capacity" of communities, from top to bottom, they are H-M (A–C), H-L (D–F), M-L (G–I), from left to right, they are nighttime (1:00–7:00), daytime (7:00–18:00), evening (18:00–1:00).

Table 2. Results of GWR models.

Groups and time	Bandwidth	R^2	Adjusted R^2	Correlation coefficient	Food and Beverage	Business	Shopping	Traffic Facilities	Finance and Insurance	Science and Education	Sports and Leisure	Health Care	Hotels and Resorts
H-M (nighttime)	97.000	0.722	0.646	Coef	−0.061	0.087*	−0.127	0.089	−0.053	−0.042	0.07	0.302	0.015
H-M (daytime)	97.000	0.707	0.626	Coef	0.014	0.18**	−0.077	0.103	−0.023	−0.045	0.104	0.185	−0.052
H-M (evening)	96.000	0.734	0.660	Coef	−0.037	0.102*	−0.104	0.099	−0.041	−0.049	0.087	0.277	−0.009
H-L (nighttime)	64.000	0.818	0.737	Coef	−0.006***	0.183**	0.03*	0.024	−0.056	−0.099	−0.065	0.282	−0.068
H-L (daytime)	79.000	0.772	0.693	Coef	0.034**	0.271**	−0.01	0.088	0.013	−0.114	−0.013	0.245	−0.078
H-L (evening)	75.000	0.794	0.718	Coef	−0.005**	0.192**	0.009	0.057	−0.03	−0.115	−0.031	0.298	−0.065
M-L (nighttime)	62.000	0.868	0.807	Coef	0.179***	0.182***	0.012*	0.228***	0.028	−0.181	−0.045	0.137	−0.035*
M-L (daytime)	62.000	0.874	0.815	Coef	0.155***	0.303***	−0.032	0.254***	0.061	−0.192	−0.048	0.173	−0.061*
M-L (evening)	62.000	0.874	0.816	Coef	0.166***	0.218***	0.007*	0.233***	0.032	−0.186	−0.04	0.156	−0.039

Coef. = coefficient; Monte Carlo test results: *** < 0.01; ** < 0.05; * < 0.1.

following context, and other groups of people are represented in a similar way). Suburban areas, such as the southern Longhua District and the northern Guangming District, can gather M-L people well. Compared to H-M and M-L, the communities generally show a weak ability to gather H-L. In addition, the "gathering capacity" of most communities is stronger in the daytime than in the nighttime, and the average value during the day is approximately 1.4–1.7 times that of the nighttime.

The correlation between "gathering capacity" and service facilities

Table 2 shows the goodness of fit of the GWR model and the Monte Carlo test results of the impact factor. The adjusted R^2 values were all greater than 0.6. This shows that the service facilities in the community and the community's ability to gather people have a strong correlation, and the former can be used to explain the difference of the latter to a certain extent.

The average coefficients of different types of service facilities are shown in Table 2. The values of the coefficients shown in yellow cells are positive, while those in blue are negative. Additionally, the darker the table cell is, the larger the absolute value of the coefficients. In particular, the values for business, traffic facilities, and health care are all positive, which are approximately 0.19, 0.13, and 0.22, respectively. Both business and traffic facilities have a strong effect on the ability of the community to gather M-L (the average coefficient is 0.23), which is approximately twice that of the other two groups. However, the opposite is true for health care, where the coefficients of H-M and H-L (approximately 0.26) are larger than those of M-L (approximately 0.15). The coefficients of science and education, sports and leisure, and hotels and resorts are basically negative. The first type of facility has a more remarkable impact on crowd interaction, with an average coefficient of −0.11.

Food and beverage as well as shopping show a difference over time. The average coefficient value of shopping during the daytime period is negative, while that of food and beverage is positive. In addition, the coefficients of finance and insurance services and shopping services vary among different groups of people. When considering gatherings of H-L or M-L, the coefficient of finance and insurance is basically positive. In contrast, the positive coefficient of sports and leisure is shown in the case of H-M. Overall, most of the service facilities show a relatively large absolute value of the average coefficient (0.13) when the target is M-L, while that for H-M is relatively small (0.09)

To better understand how the spatial heterogeneity of service facilities influences the interaction of people, we selected 5 types of facilities with relatively high spatial variability. Referring to existing research (Wang and Lu 2021), communities with $p < .05$ were visualized, and local coefficient distribution maps were displayed (Figure 3). Specifically, the coefficients of the companies' business and traffic facilities were both positively significant in the Bao'an and Longgang Districts, but the coefficient value of the former (1.54) was larger than that of the latter (1.33). In addition, business is the only facility that has a strong effect in Dapeng District. Hotels and resorts with a negative average coefficient have a locally significant impact in Bao'an, Longgang, and Guangming, especially for the M-L group. Its local negative coefficient is mainly distributed in the communities close to the downtown area, but for the eastern part that is far away from the downtown area, the coefficient is positive.

Shopping facilities significantly affect the communities in Bao'an, Longgang, and Longhua Districts, but their effects are different among different groups. They promote the interaction of H-L in Bao'an and M-L in Longhua but inhibit that of H-M in Bao'an District. Compared with the other four categories, the local coefficients of food and beverage facilities are all larger (maximum 5.35), which means that such facilities have a stronger effect on crowd activities.

In general, the significant communities are mostly in high-tech industrial zones, followed by suburban areas, and the least significant are in downtown areas. The maximum coefficient of most service facilities is reflected in M-L, while the minimum is mostly in H-M.

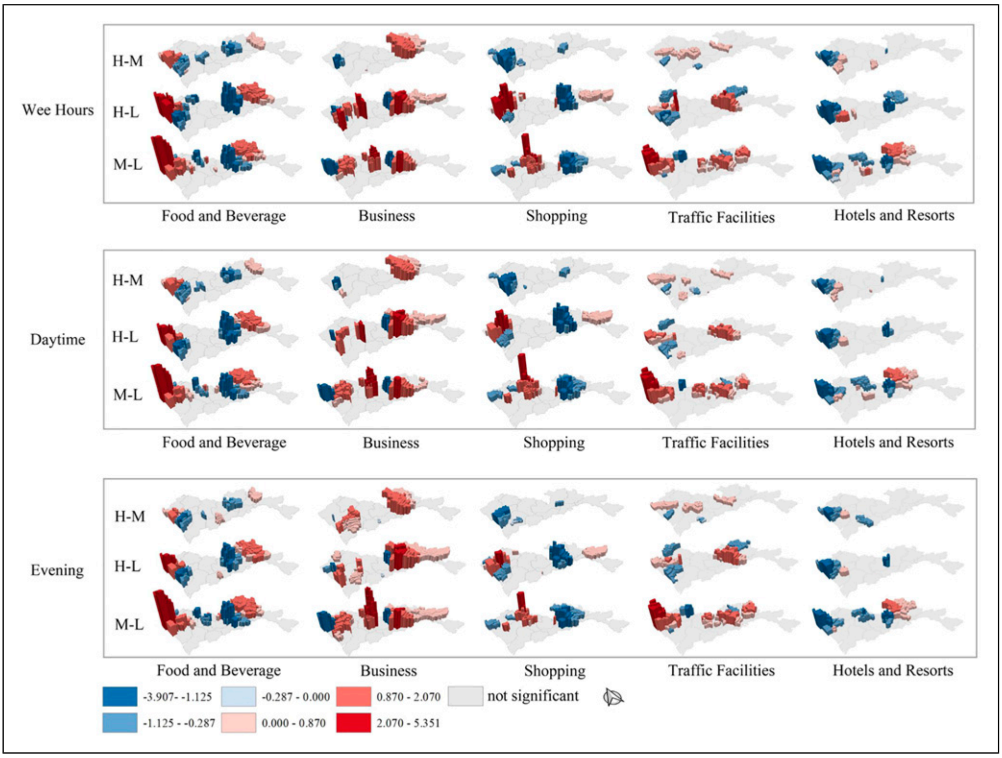

Figure 3. The local distribution map of coefficients of different types of service facilities (the p value of the community shown is less than 0.05).

Discussion and conclusion

The "gathering capacity" of communities and social segregation

The social segregation in the city can be observed from the distribution of "gathering capacity" (Figure 2). Our result indicates that communities in high-tech industrial zones show the most advantaged ability in gathering people, while the "gathering capacity" of communities in suburban areas is very weak. This may be related to the context of development in different areas of the city. In contrast to the downtown areas, where the economic cost of living is high, and suburban areas, where modernization is lagging, high-tech industrial zones supported by the government show a more advantageous character in attracting people of all economic levels. In addition, the number of communities with a strong "gathering capacity" of the H-L group is the lowest, and the distribution range is also the smallest. Furthermore, the "gathering capacity" of communities is generally strongest during the daytime. These results suggest that the most severe segregation existed in the suburban areas of Shenzhen, and segregation between people with high and low economic levels occurs more frequently and seriously than with other groups. Additionally, communities during the daytime are more conducive to interaction between people of different economic levels than in the evening and nighttime.

In addition, due to space and content constraints, this study focused on the impact of service facilities on the mitigation of social segregation in the city. The additional analysis on the presence of crowd segregation in the study area can be found in Supp 6.

The impact of service facilities on the "gathering capacity" of urban communities

The results of the global coefficient (Table 2) show that only service facilities with a strong inclusion of consumers of different economic levels or necessities for people's daily lives have a high probability of enabling people of different groups to work together. Such facilities support crowd interaction and can alleviate social segregation. For example, employees in corporate factories usually have different SESs, transportation facilities have relatively low activity costs, and the population has relatively wide thresholds for medical costs (Zhou et al., 2021). These facilities are all more tolerant of crowd SESs, which are beneficial for crowd communication (Dong and Hong, 2010). In contrast, buying or renting a house is an activity accompanied by high economic expenditures, and there is a large gap in the affordability of people with different economic levels. The housing prices of the downtown area are much higher than those of the high-tech industrial zone and the suburban zone, which intensifies the segregation of their residential behaviors and activities (Blumenberg and King, 2019; Tammaru et al., 2020). In addition, shopping and food and beverage services are varied and complex. Not only are the consumption levels presented quite different across places and occasions (Chua et al., 2020), but individuals' consumption habits and levels also vary significantly with gender and age (Boo, 2017). Therefore, these service facilities have a heterogenous impact on the "gathering capacity" of communities.

The local distribution map (Figure 3) shows that the impact mentioned above is related to the development stage and maturity of urban communities. Most of the service facilities have a significant effect on the interaction of groups in the high-tech industrial zone, but the impact is less significant in the downtown area and suburban area. For example, Dapeng District is basically affected by only business. The shopping service in Guangming District has a significant effect on only the M-L groups. This is associated with the different development stages and plans of various regions in Shenzhen (Ng and Tang, 2004). The downtown area, which developed earlier, belongs to the special economic zone. The urban functional structure system is relatively stable and mature (Deng et al., 2018). Therefore, a single type of service facility has a weaker impact on the "gathering capacity." The high-tech industrial zone, which is in the golden period of concentrated development, is the key area planned by the government (Han-xin, 2002). Service facilities in the high-tech industrial zone are less complete than those downtown, and their distribution structure is more adjustable, so they show a significant positive effect on the "gathering capacity." Suburban areas with relatively late development have a small population density (Yue et al., 2017) and a large area of ecological control land, so the effect of service facilities is insignificant.

The distribution of local coefficients (Figure 3) further reveals that the spatial heterogeneity of the impact of the service facilities is affected by the location of the community and the characteristics of different groups. For example, food and beverage services and traffic facilities are directly related to the daily flow of people (Zhou et al., 2021), and they present a strong effect on promoting crowd interaction in communities with convenient transportation and a complex flow of people. However, suburban areas with lower population mobility are less likely to encourage communication among people. Basic service facilities, such as businesses, are more effective than recreational facilities, such as shopping (Hao et al., 2012; Zhou et al., 2021), in communities that have low land costs and are distant from downtown areas. In our results, the local average coefficients of business and shopping are 1.2 and −0.8, respectively, in Pingshan District. However, when communities are closer to the city center, modern shopping facilities and department stores can attract more people of different economic levels for interaction (Wang et al., 2009). Therefore, there is a case where the coefficient of shopping is greater than 2, in Longhua District.

Implications for urban design and planning

Many urban scores and social problems (e.g., social class contradictions and urban decay and deprived neighborhoods) may emerge due to segregation (Andersen, 2019), and our findings provide some suggestions and evidence for the government when it tries to deal with these problems. By focusing on the services that have a positive impact on the community and the two groups that are most affected, and without destroying the cost of living of the dominant group, we can effectively improve the economic inclusion of the area and ultimately reduce social segregation and other urban problems by adjusting the structure of services within the community.

More specifically, from our results, governments are suggested to consider increasing the construction of companies and enterprises, especially high-tech companies, when planning the land use of regions far away from the city center. With the expansion of work activities, people with higher economic levels gather less in the city center, which helps increase the interaction between them and people with lower economic levels. However, governments are also suggested to pay attention to avoid the rapid growth of regional housing prices, in case the original residents cannot afford the gentrification and ultimately are priced out. In addition, improving transportation service facilities in remote places would greatly facilitate the inflow and outflow of people and enhance the overall vitality and communication of the city. Therefore, transportation construction in suburban areas should be oriented to facilitate commuting and travel to the city center.

For downtown areas and places nearby, city planners should consider optimizing and increasing service industries, such as catering and shopping. Moreover, the specific groups attracted to different service facilities should be considered. For example, shopping service facilities in Longhua have the most remarkable influence on the gathering of the M and L groups. According to the characteristics of these two groups of people, relevant departments can build new facilities matching the economic level of the group if possible, such as houses with appropriate housing prices and restaurants and shopping centers with reasonable consumption levels.

Limitations and future study

There are some shortcomings in this study. There are certain errors between the location of the base station recorded in the signaling data of cell phones and the actual location of user activity. Although the CV distribution of house prices proves that the average house price is statistically representative (Supp 1.1), the error between the coupled housing prices and the actual socioeconomic level of the population cannot be denied. In addition, although the community scale can be very helpful to study the realistic significance of crowd activity, it may have MAUP problems that affect the results of statistical tests (Openshaw, 1981). Simulations with different scales of units can be tried in future studies, and the results can be compared and analyzed to provide support on research scales for similar studies. This study draws on the proximity or spatiotemporal colocation approach to analyze the interaction among people, which is a reasonable but imperfect method. In the next stage, we will try to obtain updated mobile phone data and determine the positioning of users' activity trajectories more precisely and try to collect information about users' SESs in a more accurate way. Meanwhile, we will try to find a more careful way to model the communication between a pair of people on the basis of the spatiotemporal encounter model. In addition, by introducing more comprehensive influencing factors, we plan to obtain more valuable information about social segregation between people of different economic levels in the future.

Declaration of conflicting interests

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

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the National Key Research and Development Program of China (Grant No. 2019YFB2102903), the National Natural Science Foundation of China (Grant No. 41,801,306) and the "CUG Scholar" Scientific Research Funds at China University of Geosciences (Wuhan) (Project No. 2,022,034) and a grant from State Key Laboratory of Resources and Environmental Information System.

ORCID iD

Yao Yao https://orcid.org/0000-0002-2830-0377

Supplemental Material

Supplemental material for this article is available online.

References

Andersen HS, 2019. *Urban Sores: On the Interaction Between Segregation, Urban Decay and Deprived Neighbourhoods*. Routledge.

Blumenberg E and King H, 2019. Low-income workers, residential location, and the changing commute in the United States. *Built Environment*, 45(4), 563–581.

Boo HV, 2017. Service environment of restaurants: findings from the youth customers. *Journal of Asian Behavioural Studies*, 2(2), 67–77.

Brasington DM, Hite D and Jauregui A, 2015. House price impacts of racial, income, education, and age neighborhood segregation. *Journal of Reginal Science*, 55(3), 442–467.

Brunsdon C, Fotheringham S and Charlton M, 1998. Geographically weighted regression. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 47(3), 431–443.

Bureau SS. 2018. *Shenzhen Statistical Yearbook*. China Statistics Press: Beijing, China.

Carrasco JA and Miller EJ, 2006. Exploring the propensity to perform social activities: a social network approach. *Transportation*, 33(5), 463–480.

Chua B, Karim S, Lee S, et al. 2020. Customer restaurant choice: an empirical analysis of restaurant types and eating-out occasions. *International Journal of Environmental Research and Public Health*, 17(17), 6276.

Crandall DJ, Backstrom L, Cosly D, et al. 2010. Inferring social ties from geographic coincidences. *Proceedings of the National Academy of Sciences*, 107(52), 22436–22441.

Deng Y, Fu B and Sun C, 2018. Effects of urban planning in guiding urban growth: evidence from Shenzhen, China. *Cities*, 83, 118–128.

Dong M and Hong G, 2010. *The Empirial Analysis of the Salary and Income Gap Between Executive Office and Ordinary Staff in Listed Company*. East China Economic Management, 9.

Dorman M, Svoray T and Kloog I, 2020. How does socio-economic and demographic dissimilarity determine physical and virtual segregation? *Journal of Spatial Information Science* 2020(21), 177–202.

Downs A, 2010. *Neighborhoods and Urban Development*. Brookings Institution Press.

Frank LD and Engelke PO, 2001. The built environment and human activity patterns: exploring the impacts of urban form on public health. *Journal of Planning Literature*, 16(2), 202–218.

Gong L, Liu X, Wu L, et al. 2016. Inferring trip purposes and uncovering travel patterns from taxi trajectory data. *Cartography and Geographic Information Science*, 43(2), 103–114.

Hanson S and Hanson P, 1981. The travel-activity patterns of urban residents: dimensions and relationships to sociodemographic characteristics. *Economic Geography*, 57(4), 332–347.

Han-xin C, 2002. *Development and Distribution of Shenzhen High-Tech Industry*. Economic Geography, 3.

Hao P, Geertman S, Sliuzes R, et al. 2012. The land-use diversity in urban villages in Shenzhen. *Environment and Planning A*, 44(11), 2742–2764.

Huang Q and Wong DWS, 2016. Activity patterns, socioeconomic status and urban spatial structure: what can social media data tell us? *International Journal of Geographical Information Science: IJGIS*, 30(9), 1873–1898.

Iceland J and Wilkes R, 2006. Does socioeconomic status matter? Race, class, and residential segregation. *Social Problems*, 53(2), 248–273.

Jarv O, Masso A, Silm S, et al. 2021. The link between ethnic segregation and socio-economic status: an activity space approach. *Tijdschrift voor Economische en Sociale Geografie*, 112(3), 319–335.

Jiang R, Song X, Fan Z, et al. 2021. Transfer urban human mobility via poi embedding over multiple cities. *ACM Transactions on Data Science*, 2(1), 1–26.

Krivo LJ, Washington HM, Peterson RD, et al. 2013. Social isolation of disadvantage and advantage: the reproduction of inequality in urban space. *Social Forces* 92(1), 141–164.

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

Lee R, Wakamiya S and Sumiya K, 2013. Urban area characterization based on crowd behavioral lifelogs over Twitter. *Personal and Ubiquitous Computing*, 17(4), 605–620.

Leo Y, Fleury E, Hamelin A, et al. 2016. Socioeconomic correlations and stratification in social-communication networks. *Journal of the Royal Society Interface*, 13(125), 20160598.

Liu Y, et al. 2011. A review of human mobility research based on location aware devices. *Geography and Geo-Information Science*, 27(4), 8–13.

McKenzie G, Janowicz K, Gao S, et al. 2015. POI pulse: a multi-granular, semantic signature–based information observatory for the interactive visualization of big geosocial data. *Cartographica: The International Journal for Geographic Information and Geovisualization*, 50(2), 71–85.

Mouratidis K, 2020. Neighborhood characteristics, neighborhood satisfaction, and well-being: the links with neighborhood deprivation. *Land Use Policy*, 99, 104886.

Ng MK and Tang W, 2004. The role of planning in the development of Shenzhen, China: rhetoric and realities. *Eurasian Geography and Economics*, 45(3), 190–211.

Openshaw S, 1981. The modifiable areal unit problem. *Quantitative Geography: A British View* 23, 60–69.

Pappalardo L, Pedreschi D, Smoreda Z, et al. 2015. Using big data to study the link between human mobility and socio-economic development. *IEEE* 25, 871–878.

Robinson AL, 2020. Ethnic diversity, segregation and ethnocentric trust in Africa. *British Journal of Political Science*, 50(1), 217–239.

Schlapfer M, Dong L, Keeffe KO, et al. 2021. The universal visitation law of human mobility. *Nature*, 593(7860), 522–527.

Schnell I and Yoav B, 2001. The sociospatial isolation of agents in everyday life spaces as an aspect of segregation. *Annals of The Association of American GeographerS*, 91(4), 622–636.

Silm S and Ahas R, 2014. The temporal variation of ethnic segregation in a city: evidence from a mobile phone use dataset. *Social Science Research*, 47, 30–43.

Song C, Qu Z, Blumm N, et al. 2010. Limits of predictability in human mobility. *Science*, 327(5968), 1018–1021.

Tammaru T, et al. 2015. *Socio-Economic Segregation in European Capital Cities: East Meets West*. Routledge.

Tammaru T, Marcin S, Aunap R, et al. 2020. Relationship between income inequality and residential segregation of socioeconomic groups. *Regional Studies*, 54(4), 450–461.

Tu W, Cao J, Yue Y, et al. 2017. Coupling mobile phone and social media data: a new approach to understanding urban functions and diurnal patterns. *International Journal of Geographical Information Science*, 31(12), 2331–2358.

Wang D, Pedreschi D, Song C, et al. 2011. Human mobility, social ties, and link prediction. *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1100–1108.

Wang J and Lu F, 2021. *Modeling the Electricity Consumption by Combining Land Use Types and Landscape Patterns with Nighttime Light Imagery.* Energy, 121305.

Wang Y, Gu Y, Dou M, et al. 2018. Using spatial semantics and interactions to identify urban functional regions. *ISPRS International Journal of Geo-Information*, 7(4), 130.

Wang YP, Wang Y and Wu J, 2009. Urbanization and informal development in China: urban villages in Shenzhen. *International Journal of Urban And Regional Research*, 33(4), 957–973.

Wang Z, Yue Y, He B, et al. 2021. A Bayesian spatio-temporal model to analyzing the stability of patterns of population distribution in an urban space using mobile phone data. *International Journal of Geographical Information Science*, 35(1), 116–134.

Wu H and Bing Z, 2010. *The Strategy of Urban Master Planning in the Urban Transformation Era: A Case Study of the Master Plan of Shenzhen (2009-2020)*. Planners.

Xu Y, Shaw SL, Yin L, et al. 2015. Understanding individual daily activity space based on large scale mobile phone locationdata. *GeoComputation 2015 Conference*.

Xu Y, Belyi A, Bojic I, et al. 2017. How friends share urban space: an exploratory spatiotemporal analysis using mobile phone data. *Transactions in GIS*, 21(3), 468–487.

Xu Y, Belyi A, Bojic I, et al. 2018. Human mobility and socioeconomic status: Analysis of singapore and boston. *Computers, Environment and Urban Systems* 72, 51–67.

Xu Y, Belyi A, Santi P, et al. 2019. Quantifying segregation in an integrated urban physical-social space. *Journal of the Royal Society Interface*, 16(160), 20190536.

Xu Y, Santi P and Ratti C, 2021. Beyond distance decay: discover homophily in spatially embedded social networks. *Annals of the American Association of Geographers* 7, 1935208.

Yang X, Fang Z, Xu Y, et al. 2021. Revealing temporal stay patterns in human mobility using large-scale mobile phone location data. *Transactions in GIS* 25(4), 1927–1948.

Yao Y, Zhang J, Hong Y, et al. 2018. Mapping fine-scale urban housing prices by fusing remotely sensed imagery and social media data. *Transactions in GIS*, 22(2), 561–581.

Yoon and Jongseok. 2016. Rapid urbanization in China : Urban development as dual city and migration city in Shenzhen. *Critical Review of History*, 45–76.

Yue Y, Zhuang Y, Yeh AGO, et al. 2017. Measurements of POI-based mixed use and their relationships with neighbourhood vibrancy. *International Journal of Geographical Information Systems*, 31(4), 658–675.

Zhang T, Duan X, Wong DWS, et al. 2021. Discovering income-economic segregation patterns: a residential-mobility embedding approach. *Computers, Environment and Urban Systems*, 90, 101709.

Zhou X, Chen Z, Yeh AGO, et al. 2021. Workplace segregation of rural migrants in urban China: a case study of Shenzhen using cellphone big data. *Environment and Planning B: Urban Analytics and City Science*, 48(1), 25–42.

Yuyang Wu obtained her undergraduate degree in Geographic Information Science from China University of Geosciences (Wuhan). Now she is pursuing a master's degree at The Bartlett Centre for Advanced Spatial Analysis (CASA) at the University of College London.

Yao Yao (Corresponding author) is a Professor at China University of Geosciences (Wuhan), a researcher from the Center for Spatial Information Center at the University of Tokyo, and a senior algorithm engineer at Alibaba Group. His research interests are geospatial big data mining, analysis, and computational urban science.

Shiyi Zhang obtained her undergraduate degree in Geographic Information Science from China University of Geosciences (Wuhan). Her main research interests are geospatial data mining and computational urban science.

Shuliang Ren received his master's degree from China University of Geosciences (Wuhan). His research interests are geospatial big data mining and geographic information system.

Qingfeng Guan is a Professor at China University of Geosciences (Wuhan). His research interests are high-performance spatial intelligence computation and urban computing.