



An extended spatiotemporal exposure index for urban racial segregation

Qingsong Liu ^a, Mengmeng Liu ^b and Xinyue Ye ^c

^aDepartment of Geography, Kent State University, OH, USA; ^bGeorgia Tech Tianjin University Shenzhen Institute, Shenzhen, Guangdong, China; ^cDepartment of Landscape Architecture and Urban Planning, Texas A&M University, College Station, TX, USA

ABSTRACT

The Segregation Index quantifies the degree of segregation of social groups or classes. Because of the increasing use of fine-grained spatiotemporal activity and flow data, the conventional segregation measurements' inclusiveness is challenged. We add population flow to the conventional place-based spatial exposure index to identify spatiotemporal segregation changes. Specifically, we considered the population-flow network, hierarchical structure, and time. In Chicago's demonstration case study, we first used the time-dependent Twitter Origin-Destination flow matrices and their hierarchical structure information to estimate interactions between areal units at the neighborhood level. Then we computed the new population composition of units based on their interactions with other units and estimated the proposed spatiotemporal exposure index for different times. Finally, we systematically compared their differences with the conventional indices at global and local scales to see how population-flow patterns affect the exposure index. The results show that the population-flow patterns reflect valuable information in neighborhood interactions in temporal and spatial dimensions, but it is missing information in the conventional segregation computations. Furthermore, we emphasize that the hierarchical structures of flow patterns and the choice of appropriate parameters are also important factors for a rational segregation evaluation.

ARTICLE HISTORY

Received 28 December 2020
Accepted 5 August 2021

KEYWORDS

Spatiotemporal exposure index; flow network; hierarchical structure; spatiotemporal activity and trajectory data; residential segregation

1. Introduction

Spatiotemporal activity and trajectory big data play an increasingly important role in measuring the degree of segregation of social groups or classes (Farber et al., 2015; Wong & Shaw, 2011; Yip et al., 2016). Social segregation implies a lack of communication between groups and indicates an uneven distribution of population or resources and varying interactions. Due to data limitations, conventional segregation studies rely heavily on census data and focus on static residential spaces (Kwan, 2009). Over the past decade, spatiotemporal activity and trajectory data helped researchers to characterize changes in individual or group activity spaces (D. Wang et al., 2012) and flow patterns (Gao et al., 2013; Guo et al., 2012) at different time scales (Silm & Ahas, 2014). The observed people's movement connects the city's various spaces and creates an entire network of interactions. In turn, the population-flow network and its interaction information are instrumental to our understanding of cities (Batty, 2013).

However, flow patterns, which contain interaction information, have not been fully investigated in segregation studies. Our first question focuses on methodology: how can we incorporate flow patterns into conventional segregation calculations with minimal

modifications of conventional formulas? Our second question is about observing indices from a comparative perspective: when comparing the new index to the conventional one, where and when do the two exhibit significant differences? What factors influence the extent of their differences? Answering the above questions will allow us not only to build a comparable system to link traditional research and current research with new data, to form a succession and continuation of research, but also to provide a baseline for using the richer spatiotemporal activity and flow data.

We try to incorporate time-dependent population-flow patterns into conventional segregation indices computation. Specifically, we use the information on the flow network, time, and hierarchy structure to estimate interactions between areal units. By utilizing more flow information, we expect to comprehensively evaluate the interactions and then compute the segregation more accurately. Also, we also systematically compare the new flow-based segregation index with the conventional index. Based on the comparison results, we want to understand better the impact of including population-flow patterns into the segregation indices.

We organize this paper as follows:

- [Section 2](#) briefly reviews the development of the segregation index and the progress of methods using new data.
- [Section 3](#) covers the study area and demonstration data.
- [Section 4](#) illustrates the methodology used in this paper, including how to construct flow patterns, compute the segregation with flow data, and the comparison method.
- [Section 5](#) shows the results. We first conducted a descriptive analysis of flow patterns of different neighborhood types in the Chicago area. We then compared the segregation index changes after incorporating flow patterns at both the global and local scales.
- [Sections 6](#) shows the implications of adding flow data to segregated studies and insights from the comparative results. We also provided some future research plans.
- [Section 7](#) gives a brief conclusion.

2. Literature review

The temporal and spatial perspective provided by geographers is far-reaching for quantifying segregation. Since the dissimilarity index (Duncan & Duncan, 1955) was first designed, many segregation indices have been proposed. Much literature has continually criticized, improved, and generalized these segregation indices by considering areal units' spatial arrangement/structure in the study area when evaluating segregations (Yao et al., 2019). Spatial interactions are the most discussed spatial structure in the literature. Without spatial interactions, the corresponding segregation degree depends only on the areal unit's demographic composition. The segregation index will not change, even if the areal units are shuffled into different spatial arrangements, because the shuffle will not change areal units' own demographic composition (White, 1983). Researchers refer to the problem of disengaging from spatial arrangements as the checkerboard problem (Morril, 1991; Wong, 1993). How to accurately describe these spatial interactions between units becomes crucial to solving the checkerboard problem.

There are two main classes of methods to describe these interactions. One is based on spatial proximity functions, and the other is based on topology or geometry information. The spatial proximity functions follow the principle of distance decay and assume that the farther the distance between two areal units, the weaker their interactions. White (1983) suggested four proximity

functions, including negative exponential functions and inverse distance functions with different parameter settings. With an empirical comparison between these proximity functions, White concluded that the issue of choosing appropriate proximity function forms and parameters is still not yet resolved. Because there is no better way to select a spatial proximity function and its related parameters, Reardon and O'Sullivan (2004) left the choice to users. While generalizing his predecessor's work, they used a notation to represent this spatial proximity function in terms of decreasing distance. They also emphasized that any desired proximity function could be used in their proposed generalized framework. This framework is promising but left subsequent researchers with the difficult task of quantifying the spatial proximity function.

The second class of methods uses topology, or geometry information, to describe spatial interactions between areal units. A simple topology-based approach is to use the binary form of contiguity. If areal unit i and j are neighbors, their corresponding weight, $\omega(i, j)$, is set to 1, and 0 otherwise (Morril, 1991). Subsequently, Wong (1993) incorporates geometry properties, such as shape and area of units, to adjust the interactions. Wong assumes that longer shared boundaries would lead to a stronger interaction between two units. Wong (1998) also proposed a notion of *composite population count (CPC)*, which mixing its neighbors' populations into the target unit population. After the mixing procedure, the new population composition is used in the subsequent calculations. CPC is another way to describe the probability of the target unit population interacts with its neighbors' population. In short, both above classes of methods attempt to provide descriptions of neighbor interactions and eliminate the checkerboard problem.

Although these two estimation methods can solve the checkerboard problem, their estimation of the interactions is subjective. No matter how complex the mathematical forms, there is no guarantee that their interaction estimation reflects the actual social interaction across units. With the development of information technology, the accumulation of massive amounts of spatiotemporal activity and trajectory data provides new insights into modeling and estimating social interactions. First, researchers emphasize expanding static residential space into other socio-geographical spaces (Kwan, 2013; Wong & Shaw, 2011; Yip et al., 2016). Many studies have demonstrated, from different perspectives, that spatiotemporal big data can provide information in addition to traditional survey data. For example, Shelton et al. (2015) examined residents' tweeting activity on both sides of the imaginary "9th

Street Divide” from the perspective of a collective activity distribution. Q. Wang et al. (2018) analyze georeferenced tweets to compare the travel patterns for 50 U.S. cities by racial/ethnic groups. Park and Kwan (2018) proposed the individual segregation index to depict when and how much segregation people experience dynamically throughout a day. Second, researchers emphasize the importance of the interaction between physical and social space in the study of segregation. Farber et al. (2015) resorted to spatiotemporal paths to describe individual trajectories and further assessed interaction-based segregation. Xu et al. (2019) used phone call records to portray interpersonal communication intensity and then estimated social segregation based on friendship networks. These examples show the promise of using new data to describe the distribution of activities or interactions between regions within cities.

However, current conventional segregation studies do not fully explore the dynamic interaction information from population-flow patterns. First, without changing the conventional segregation index formula, population-flow networks have not been fully discussed. We argue that a population-flow network can be easily integrated into conventional segregation indices with only a few uncomplicated transformations. The linkage exists in the notions of the CPC (Wong, 1998) or *population density of the local environment* (Reardon & O’Sullivan, 2004) or *local population intensity* (Feitosa et al., 2007). All notions emphasize the exchanges/interactions between residents and their neighbors. Thus, measuring the intensity of such interaction becomes crucial for segregation computation. In the CPC notion, the interactions between two areal units are estimated based on their contiguity. In comparison, the other two notions underscore the distance decay functions in the estimation. Although researchers have made many modifications to the contiguity or set up many distance decay functions, the assessments are still artificial.

Second, collective-level interactions change with time, and the temporal dynamics need more attention. The traditional definition of interaction is fixed and artificial in not only the spatial dimension but also in the temporal dimension. For example, if using contiguity to construct spatial interactions, one’s neighbors do not change from ten years ago to ten years later. The same problem exists for the method of distance decay functions. As data becomes more abundant, researchers have realized the necessity of focusing on temporal dynamics in segregation studies (Kwan, 2013; Lee & Li, 2017; Silm & Ahas, 2014); however, the temporal variations in interactions between areal units have received little attention. Equally as important, the

temporal interaction variations tell us when interracial contact is lower or higher. This temporal pattern can help us better understand segregations in today’s fast-paced and mobile world and can provide a basis for reviewing integration and planning policies (Silm & Ahas, 2014). If we add population-flow information to estimate interactions, one’s neighbors can change with commuters traveling to work in the short term and change with the city’s development in the long run. Therefore, using population flows to determine spatial interactions can avoid the static problem of segregation index.

Finally, the hierarchical structure of population flow has received little attention in segregation studies. Due to information technology and transportation accessibility, residential segregation becomes less and less important. The contacts with distant neighborhoods or central cities shape the external connections of the residents. Residential segregation cannot represent the segregation status in other activity spaces. We should pay more attention to the mobility of residents and their position in social and physical networks (Browning & Soller, 2014). Some researchers have begun to analyze the relationship between realistic characteristics (e.g. socioeconomic and demographic characteristics) and the structural features of population-flow networks (Dorman et al., 2020; Q. Wang et al., 2018), and then used them to understand actual segregation and isolation (Prestby et al., 2020). The hierarchical structure is one way to evaluate the importance of different neighborhoods because it consists of mobility networks between these neighborhoods in the city. Bassolas et al. (2019) used flow data to assess the hierarchical structures of 301 cities worldwide and identified different hotspot levels for areas in each city. The study found that cities with a more robust flow hierarchy have a higher degree of population-mixing, extensive public transportation, and a higher walkability level. Liu et al. (2018) found that the core-peripheral flow structure is in accordance with the visitors’ distribution in one area and affects the residents’ and visitors’ meeting probability. The hierarchical information facilitates a comprehensive picture of segregation within the urban’s current spatial layout. Thus, the segregation indices with hierarchical structure information can be a meaningful reference for urban planners.

3. Study area and data

The 77 Chicago neighborhoods (Chicago Data Portal, n.d.) have three main racial/ethnic groups, including approximately 29% non-Hispanic Black, 32.5% non-Hispanic White, and 29% Hispanic. Based on the

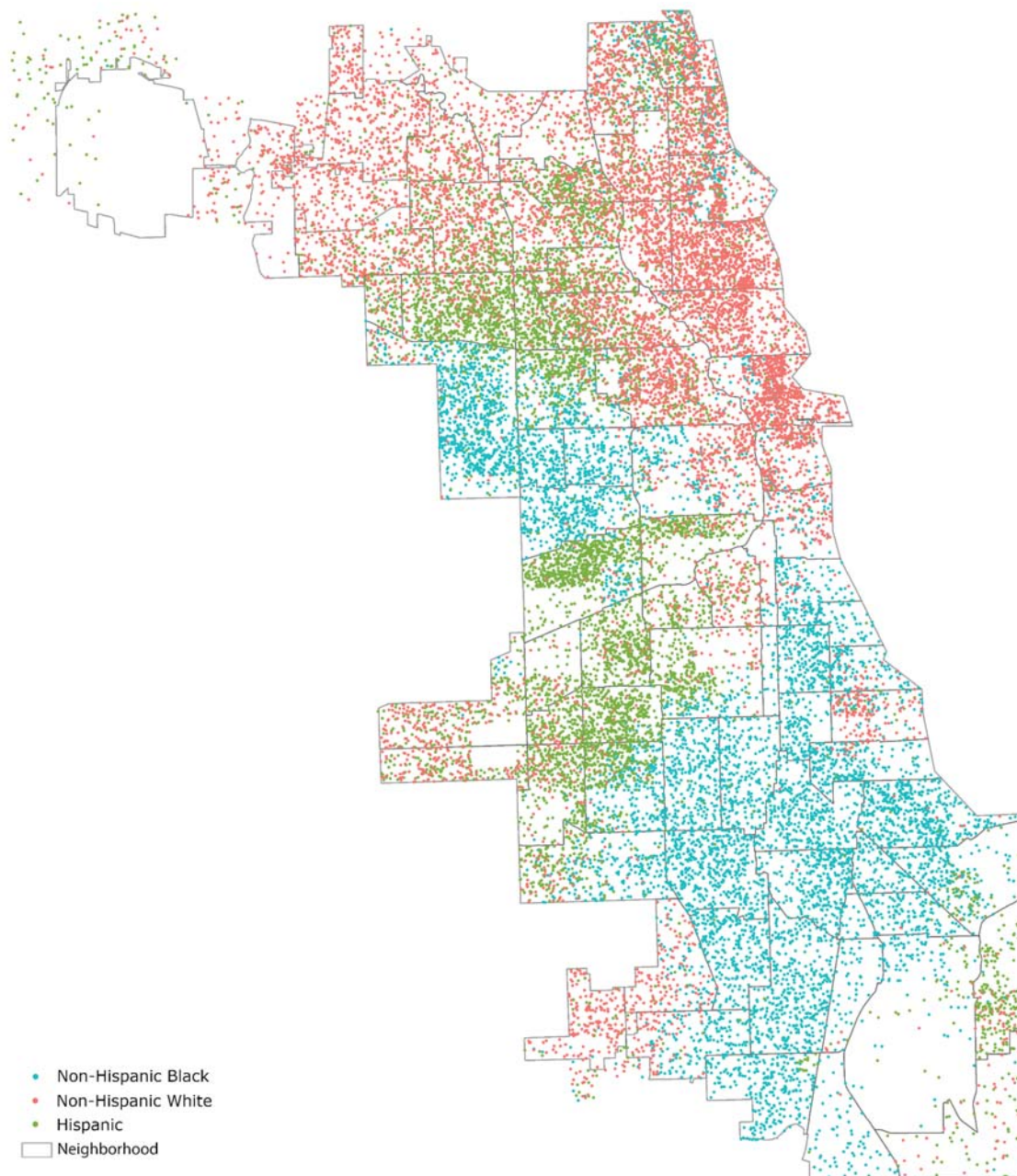


Figure 1. The Population Distribution of Three Groups within Chicago 77 Neighborhoods (100 persons/dot, population data source: 2012–2016 American Community Survey 5-year Estimates).

population information of 2012–2016 *American Community Survey 5-year Estimates*, [Figure 1](#) shows the three groups of the population reside in Chicago. On the map, the colors of aqua, red, and green dots represent non-Hispanic Black, non-Hispanic White,

and Hispanic groups, respectively (100 persons per dot). Like other metropolitan areas in the United States, Chicago's racial diversity has been generally increasing at an overall level. But at the neighborhood level, it still exhibits significant residential segregation

(Walker, 2018). The non-Hispanic Black population group is predominantly located in the South and West. In contrast, the non-Hispanic White population group is primarily located in the North, and the Hispanic population group is interspersed between the other two groups.

Multiple spatiotemporal big data can be used to model the population-flow pattern, such as taxi, phone call records, and social media data (Hawelka et al., 2014; Tang et al., 2015). However, each of these data has its own limitations. For example, Twitter and taxi data are subject to representativeness bias, and phone call records are not easily available. Because of our limited availability of data and funding, we ended up using Twitter data to extract population movements. We used Twitter Streaming Application Programming Interface (API). We collected georeferenced tweets within Chicago city from Oct.09 to Nov.16, 2016, using the Twitter Streaming API. The Twitter users of these tweets include tourists, residents, and commuters near the study area. To compare data with the traditional residential segregation, we must extract the residents' users. We adopt the algorithm in Luo et al. (2016) by checking whether the user's historical activity was clustered within the study area during the evening (8:00 pm-8:00 am). Therefore, we also collect each user's historic tweets by the Twitter Timeline API. We got 3 million tweets for the demonstration study. While some studies have supported the applicability of using Twitter to extract collective population movements (Lenormand et al., 2014), we must always be cautious of the limitations of the Twitter data and the possible impact on the results.

4. Methodology

4.1. Construct the time-dependent flow network based on Twitter data

The time-dependent flow network is represented by an origin-destination (OD) flow matrix with a time restriction. Each element of the matrix represents the number of flows between two areal units at time t . In theory, the time variable t can be continuous, meaning that the OD flow matrix has its specific values for any infinitesimal short time period. In practice, due to the limited amount of flow data, we can use only longer periods to obtain valid estimates of the OD flow matrix, such as 4 hours. For example, for a given time t , we extend 2 hours forward and backward, respectively, and use the flows that happened in this extended period to construct the OD flow matrix for time t .

We used the following steps to extract the OD flow matrix at a given time t based on geotagged tweets. According to Gao's procedure (Gao et al., 2014), we

first sequenced each Twitter user's tweets by time, and then take each of the two temporal consecutive tweets as one valid move record, if the record satisfies two criteria: 1) the time interval between two tweets is less than 4 hours and 2) the straight-line distance between two tweets is longer than 100 meters (Liu et al., 2018). Criterion 1 ensures each move record's integrity as much as possible based on Gao et al. (2014) inter-tweeting time analysis and criterion 2 help eliminate the uncertainty of the GPS signal. After extracting all qualified moves, we used the "points in polygon" operation to determine each move's origin and destination neighborhood and kept only those moves which were across neighborhood boundaries. Finally, we created a time-dependent OD flow matrix by aggregating moves according to their origin, destination, and specified time at the neighborhood level. In many studies, the OD flow matrix is a measure of the interaction between spatial units (Guo et al., 2012). Besides, the adoption of the OD flow matrix as spatial interaction gives us a more general framework that makes it possible to integrate other movement data. We must note that the above-mentioned time sequence construction method is only one possible way to extract population movements, and we do not exclude other flow extraction approaches (Palchykov et al., 2015).

To provide a global overview of the interactions among the three racial/ethnic groups in the case study, we classified the 77 neighborhoods into four types according to their demographic composition: non-Hispanic Black-majority neighborhood, non-Hispanic White-majority neighborhood, Hispanic-majority neighborhood, and Mixed-neighborhood. A neighborhood is classified as racially majority if one racial/ethnic group is larger than the other two and accounts for more than 50% of its population; if there are no racially majority groups, it is classified as Mixed-neighborhood. For simplicity of description, only the initials are used below to represent the majority type: B-neighborhood, W-neighborhood, H-neighborhood, and M-neighborhood.

4.2. Segregation indices with flow patterns

Measures of spatial segregation require defining the spatial interactions between all pairs of units in region R at time t . The difference between our proposed extended spatiotemporal exposure index and spatial indices lies in calculating spatial interactions. To demonstrate our proposed extension method, we applied it to the exposure index, which is an important dimension of segregation assessment. Massey & Denton defined exposure index to "measure the extent to which minority and majority members physically confront one

another by virtue of sharing a common residential area.” For consistency, we used the notation from Reardon and O’Sullivan (2004), but we use the traditional polygon as the basic units (i.e. neighborhoods). Let R denote the entire study region. Assume that R is split into n non-overlapping areal units, which are indexed by i or j . Suppose there are M racial/ethnic groups in R ($M = 3$ in the case study), and we index them by x or y , e.g. Black or White group. Let τ denote population. Use a tilde ($\sim$) to indicate the mixing procedure between the target areal unit and other interacted units. Also, use t to express the specified time. We have:

- $\phi_t(i, j)$: flow size from unit i to unit j at time t .
- $\omega_t(i, j)$: the standardized interaction weight from unit i to unit j at time t .
- $\tau_{i,x,t}$: population size of group x in unit i at time t .
- $\tau_{i,t}$: population size in unit i at time t (note that $\tau_{i,t} = \sum_{x=1}^M \tau_{i,x,t}$).
- $\tilde{\tau}_{i,x,t}$: population size of group x in unit i at time t after a mixing procedure.
- $\tilde{\tau}_{i,t}$: population size in unit i at time t after a mixing procedure.
- $T_{x,t}$: population size of group x in R at time t (note that $T_{x,t} = \sum_{i=1}^n \tau_{i,x,t}$).

We assume that the strength of the spatial interactions at time t is proportional to $\phi_t(i, j)$. If we set the value of $\phi_t(i, i)$ to 0, then the composite population of unit i at time t can be expressed as $\tilde{\tau}_{i,t}$.

$$\tilde{\tau}_{i,t} = \frac{1}{\Phi_t(i, \cdot)} \sum_{j=1}^n \tau_{i,t} \phi_t(i, j), \quad (1)$$

where $\Phi_t(i, \cdot) = \sum_{j=1}^n \phi_t(i, j)$ and $\phi_t(i, i) = 0$. By moving $\Phi_t(i, \cdot)$ into the summation notation in Equation (1), we get:

$$\tilde{\tau}_{i,t} = \sum_{j=1}^n \omega_t(i, j) \tau_{i,t}, \quad (2)$$

where $\omega_t(i, j) = \frac{\phi_t(i, j)}{\Phi_t(i, \cdot)}$, and $\sum_j \omega_t(i, j) = 1$ and $\omega_t(i, i) = 0$. Higher values of $\omega_t(i, j)$ mean stronger interactions from unit i to unit j at time t . Equation (2) is analogous to the spatial lag of population of unit i at specified time t (i.e. the weighted average of neighbors’ population of unit i). Here, $\omega_t(i, j)$ is equivalent to the elements in the row-standardized spatial weight matrix at time t . However, it is worth noting that in the flow-based spatial weight matrix, the number of

elements with values greater than zero (i.e. $\omega_t(i, j) > 0$) is much larger than those in the matrix constructed by the contiguity-based method.

Although we set $\omega_t(i, i)$ to zero, the flow-based matrices we have constructed so far are sufficient to compute the spatial dissimilarity index. The computation process does not involve $\omega_t(i, i)$, and we can see the description of the spatial dissimilarity equation in (Cortes et al., 2020) and (Wong, 1993). However, the determination of $\omega_t(i, i)$ becomes crucial for Exposure and Isolation index. In the literature, the equal weights strategy (Wong, 1998) or the distance decay function strategy (White, 1983) are common ways to determine the $\omega(i, i)$. These two approaches are, to some extent, a compromise in the absence of real interaction data. The resulting $\omega(i, i)$ does not reflect the true weights and does not describe the temporal variation of weights. We propose a hierarchy-based strategy to estimate $\omega_t(i, i)$ to measure the weights more accurately. Our assumption is that the higher the rank of hierarchy the higher the degree of mixing in the area (Bassolas et al., 2019), and the lower the $\omega_t(i, i)$. This strategy has two steps:

Step 1) determines hierarchy-level (hotspot-level) $h_t(i)$, for each unit at time t by the Lorenz curve method (Bassolas et al., 2019). The Lorentz curve method arranges the outflow size of all areal units of all times in ascending order and plots them by normalizing the cumulative number of units (x-axis) vs. the fraction of total flow (y-axis). Then it takes the derivative of the Lorenz curve at (1, 1) and extrapolates it to the point at which it intersects the x-axis to get the threshold. We assign $h_t(i)$ to k -level for the areal units on the x-axis greater than this threshold (k starts from 1, which is the highest hierarchy-level). Then we remove the units of k -level and recalculate the threshold to obtain $(k + 1)$ -level. Repeat this process to assign $h_t(i)$ for all areal units in each time until the threshold is close to zero or the hierarchy-level reaches an upper limit m , and the rest of unassigned units will go to the last level. In the case study, we set the upper limit, m , to 6.

Step 2) derives $\omega_t(i, i)$ for unit i at time t by its hierarchy-level $h_t(i)$. $\omega_t(i, i)$ is obtained by a piecewise function in which each hierarchy-level $h_t(i)$ is mapped to an interval of $[\alpha, \beta]$ ($0 \leq \alpha \leq \beta \leq 1$). In practice, α is the self-mixing rate for the most active unit (at the top hierarchy-level) in the study area among all times. Similarly, β is the self-mixing rate which is assigned to the least active unit (at the bottom hierarchy-level) among all times. According to Equation (3), the $\omega_t(i, i)$ is a value between α and β which represents the self-mixing rate in the population mixing procedure. When target unit i interacts more actively with other units at a time t , $\omega_t(i, i)$ will shift toward α , indicating the

decreasing influence of its own population size in the mixing procedure. Conversely, when the target unit i becomes less active, the self-mixing ratio moves toward β . To get $\omega_t(i, i)$ according to hierarchy-level, we are using a piecewise function as Equation (3):

$$\omega_t(i, i) = \alpha + (h_{t,i} - 1) \frac{\beta - \alpha}{m - 1} \quad (3)$$

where m is the total hierarchy level. Note that, after getting the new $\omega_t(i, i)$, we need to adjust (shrink) other elements in the same row with $\omega_t(i, i)$ to ensure that $\sum_j \omega_t(i, j) = 1$. After obtaining the flow-based spatial weight matrix at time t , we can then compute the

flow-based segregation via Equation (4), referred to as the Flow-based Spatial Exposure Index (FSxPy). For comparison, we also compute 1) the Boundary-based Spatial Exposure Index (BSxPy) (Cortes et al., 2020; Wong, 1993) and 2) the Distance Decay-based Spatial Exposure Index (DDSxPy) (Morgan, 1983). BSxPy and DDSxPy are calculated by Equation (5), and they construct the interaction weights by using the length of shared boundaries and the distance decay function, respectively. In addition to the different methods of constructing spatial weight matrices, they do not have temporal subscripts, meaning that these spatial weight matrices do not change with time. To demonstrate the basic mechanism of the algorithm, we use a hypothetical example to show how flow impacts the BSxPy and FSxPy in a typical weekday. See Table A1 in Appendix.

$${}_xP_y(t) = \sum_{i=1}^n \frac{\tau_{i,x,t}}{T_{x,t}} \cdot \frac{\tilde{\tau}_{i,y,t}}{\tilde{\tau}_{i,t}} \quad (4)$$

$${}_xP_y = \sum_{i=1}^n \frac{\tau_{i,x}}{T_x} \cdot \frac{\tilde{\tau}_{i,y}}{\tilde{\tau}_i} \quad (5)$$

Equations (4) and (5) are global indices, which provide one summarized index for the entire study area. To investigate the local variation of differences, we choose the local variant of spatial exposure index, as shown in Equation (6) (Feitosa et al., 2007; Wong & Shaw, 2011):

$${}_xP_y(i, t) = \frac{\tau_{i,x,t}}{T_{x,t}} \cdot \frac{\tilde{\tau}_{i,y,t}}{\tilde{\tau}_{i,t}} \quad (6)$$

4.3. Comparative inference

We compare the segregation indices mentioned in the previous section. For any set of comparisons, we use the simulation approach to test whether there is a statistical difference. This simulation approach is an extension of the “systematic” inference approach described in (Cortes

et al., 2020). The “systematic” approach adds randomness to the current population distribution and calculates segregation for each simulation. In the step of adding randomness, it draws samples from a multinomial distribution, with the success probabilities being the proportions of the population of our interested racial/ethnic group in each areal unit, and with the number of independent trials being the total population of our interested racial group in the entire study area. For any of the indices in the comparison, we conduct 9999 “systematic” simulations, and eventually, we generate two simulating distributions (each contains 9999 simulated results) for one comparison process. For each simulating distribution, we use the mean value as its point estimate and the Highest Density Interval (HDI) at a certain significance level as its most credible interval (Kruschke, 2014, p. 87). In practice, given a significance level s_α (e.g. 5%), the HDI is defined by the $s_\alpha/2$ and $(1 - s_\alpha/2)$ percentiles of the simulating distribution. During a comparison, if two HDIs intersect, we infer that there is no difference between two calculated indices. Otherwise, the difference of the two indices (referred as *indices distance*) is calculated as the difference of the two point estimates. However, the variance of the difference between the two point estimates is not calculated in this paper because it is not the main concern in the context. In this article, we choose 5% as the significance level and we will use $d_{0.95}$ to denote the indices difference in the following context as in Figure 2.

5. Results

5.1. Descriptive analysis of Twitter flow data

Based on the neighborhood classification method in Section 4.1, we classified the 77 neighborhoods into four types of neighborhoods (i.e. B-, W-, H- and M-neighborhood). Table 1 shows the number, percentage of the total population, and outflow for each type of neighborhoods. We see that a substantially larger share of total Twitter outflows originated in the W- than in the B- and H-neighborhoods, with 28% of the population contributing 63% of Chicago’s total Twitter outflows. It could be a concrete manifestation of the digital divide in different types of neighborhoods or reflect that the White population tends to enable their geocoding in tweets more than other racial/ethnic populations do.

Table 2 shows the row-standardized flow percentage between the four neighborhoods’ types, the row representing the origin and the column representing the destination. We find that outflows most happen between the same type of neighborhoods (diagonal elements are all more than 50%) except for the H-neighborhood. The outflow from H-neighborhood is mainly drawn to the W- and

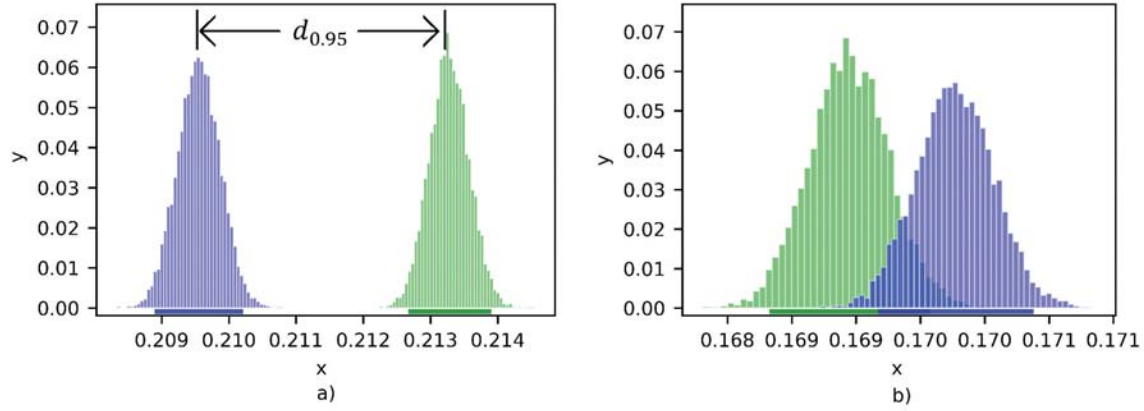


Figure 2. Two scenarios of simulating distribution comparison. The HDI of each distribution is the horizontal line drawn at the bottom: a) Since the two HDIs do not intersect, we say that the point estimate of blue distribution is statistically significantly less than its comparing index (green distribution). Their indices distance is calculated as the difference of the two point estimates, as marked by $d_{0.95}$; b) Since the two HDIs intersect, we say that the point estimate of blue distribution is not statistically significantly different from its comparing index (green distribution).

Table 1. Basic statistics of four types of neighborhoods regarding count, percentage of the population, and percentage of twitter outflows.

	B-	W-	H-	M-	Total
Neighborhood Counts	28	18	16	15	77
Population Size (%)	694,770 (26%)	770,809 (28%)	557,613 (21%)	691,420 (25%)	2,714,612 (100%)
Twitter Outflow Size (%)	17,262 (7%)	156,085 (63%)	20,069 (8%)	54,499 (22%)	247,915 (100%)

Table 2. The row standardized percentage of flow size between four types of neighborhoods.

To From	B-	W-	H-	M-	Total outflow size
B-	59%	14%	13%	14%	17,262 (100%)
W-	2%	83%	4%	11%	156,085 (100%)
H-	10%	36%	29%	25%	20,069 (100%)
M-	4%	33%	9%	54%	54,499 (100%)

M-neighborhoods. Second, the percentage of outflow is asymmetric across different types of neighborhoods. For example, 14% of total outflow from B-neighborhood goes to W-neighborhoods, but conversely, only 2% of total outflow from W-neighborhood is destined for B-neighborhoods. Similarly, there is a significant asymmetry in the proportion of outflows between H- and W-neighborhoods (36% vs. 4%). It is clear that W-neighborhoods are the most popular destinations in Chicago city. The asymmetry pattern is consistent with Q. Wang et al.'s finding (2018).

Figure 3 shows the average outflow size of four neighborhood types on a typical weekday and weekend. Figure 4 shows their average travel distance. We find that more flows came from the W-neighborhoods than from the other three types of neighborhoods at any time

during a typical weekday or weekend (Figure 3(a,b)). In contrast, the distance traveled from the W- neighborhoods are significantly shorter than that of the other three types of neighborhoods (Figure 4(a,b)). This observation is similar to Huang & Wong's finding (Huang & Wong, 2016), which claimed Twitter users of poor communities in D.C. had larger activity spaces than the other wealthier groups. The longer travel distances of the disadvantaged groups are mainly caused by the spatial mismatch of their work location and residence location (Easley, 2018). Since the general activity space represented by Twitter data includes work locations, we observed the effects of long work commutes for residents in the B-neighborhoods.

Although the average outflow size and travel distance on weekdays have similar trends with that on weekends, we also see some differences. For example, weekday outflow curves show two flow peaks (at 14:00, 19:00 in Figure 3(a)) while weekends show only one (at 14:00 in Figure 3(b)). Besides, the weekend maximum peak flow is 44% (i.e. (2.08k-1.16k)/2.08k) less than the weekday peak flow. This discrepancy of weekday and weekend travel patterns is also reflected in taxi and phone call

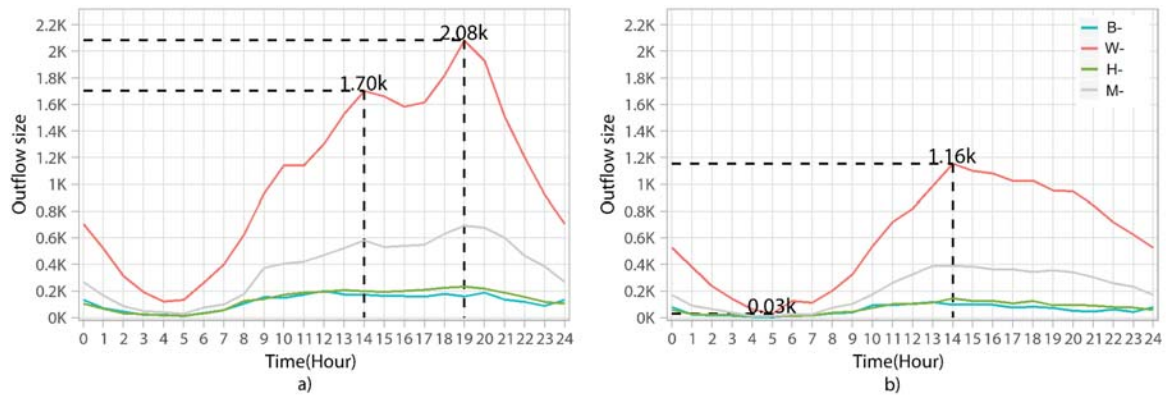


Figure 3. Average Outflow Size on Weekday and Weekend by Origin Neighborhood's Type: a) average outflow size on a typical weekday; b) average outflow size on a typical weekend.

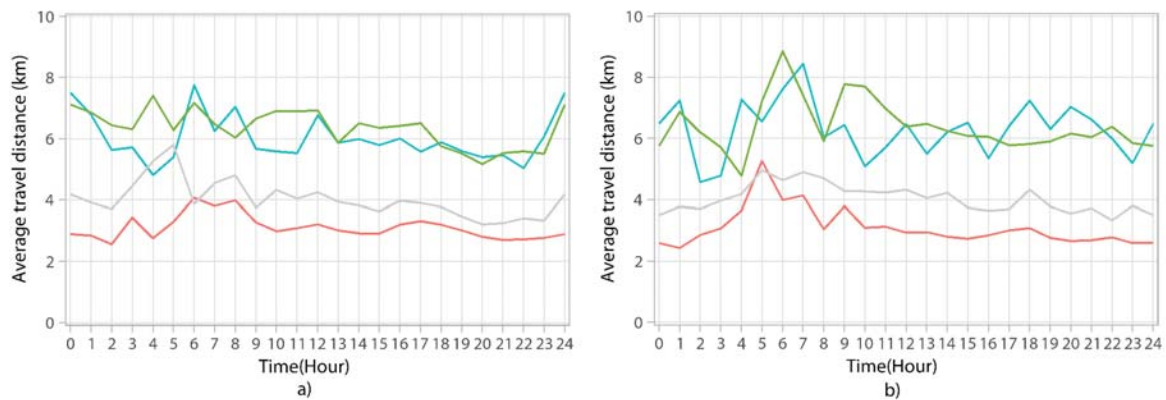


Figure 4. Average Travel Distance in Weekday and Weekend by Origin Neighborhood's Type: a) average travel distance on a typical weekday; b) average travel distance on a typical weekend.

records datasets (Calabrese et al., 2011; Zhu et al., 2017). For demonstration purposes in the Chicago case study, we only use the weekday flow data to estimate the interactions between neighborhoods when calculating the FSxPy.

5.2. Comparison result of global exposure index

According to Section 4.2, we know that the major difference between the global exposure indices of FSxPy, BSxPy, DDSxPy is their ways of constructing the spatial weight matrix. Therefore, we can focus on the parameters of construction methods when comparing FSxPy to BSxPy or FSxPy to DDSxPy. As an example, we calculated only the exposure index of black to white racial/ethnic groups. We first compare black-white FSxPy to BSxPy in Figure 5. From Figure 5(a-f), the x-axis and y-axis represent the

time parameter t of FSxPy and the parameter of $\omega(i, i)$ of BSxPy, respectively. The z-axis represents their indices distance ($d_{0.95}$) with FSxPy as the reference index. We also added the 2-D heat map of the difference surface in Appendix Figure A1. For illustration purposes, we assume that if one unit has the bottom level in the hierarchical flow structure, it has no interactions with other units. So, we set β equal to 1.0 and only change α in FSxPy. Increasing α from 0.0 to 1.0 is the process of weakening the influence of flow data in the weights. Figure 5(a-f) show how the gradually increasing α affects the indices distance between black-white FSxPy and BSxPy. We use gradient colors from red to blue to represent the high (positive) to low (negative) values of indices distance. We also use the same color to plot contours of indices distance on top of each 3-D figure. But we

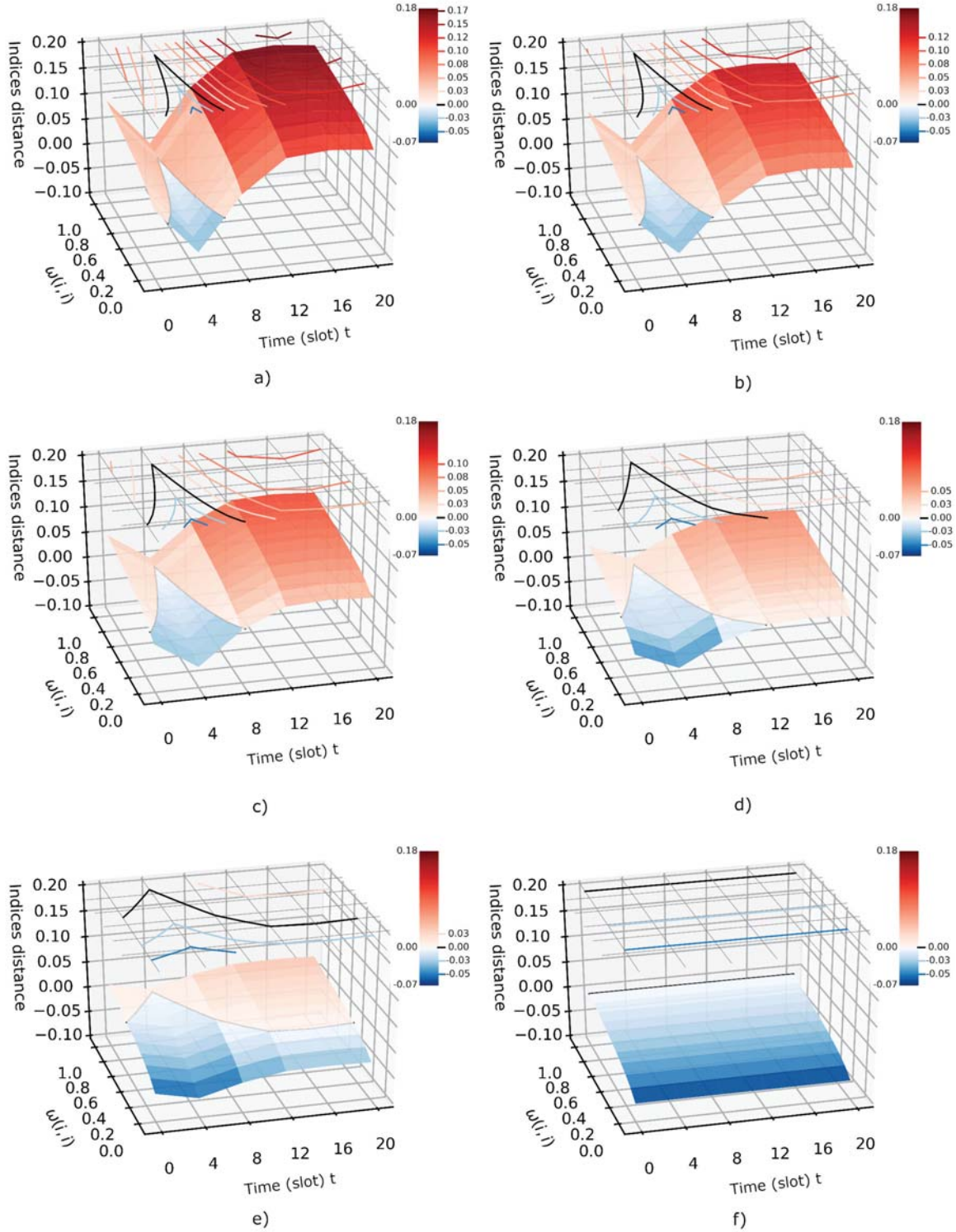


Figure 5. Statistical difference between black-white FSxPy (interested parameters are $\alpha, \beta, \text{Time (slot) } t$) and BSxPy (the interested parameter is $\omega(i, i)$) (FSxPy – BSxPy): a) $\alpha=0.0$ and $\beta=1.0$; b) $\alpha=0.2$ and $\beta=1.0$; c) $\alpha=0.4$ and $\beta=1.0$; d) $\alpha=0.6$ and $\beta=1.0$; e) $\alpha=0.8$ and $\beta=1.0$; f) $\alpha=1.0$ and $\beta=1.0$.

use a black line to indicate $d_{0.95}$ equal to 0.0, and it means the FSxPy index has the same segregation results as BSxPy.

From Figure 5(a–e), we see a clear temporal pattern along the x-axis (time t). For any given $\omega(i, i)$ of BSxPy (x-axis), the *indices distance*, $d_{0.95}$, reaches the lowest value at 4:00 and maximum value at some point between 12:00, 20:00. This finding echoes with the outflow pattern in Figure 3(a). Besides, for any given time t (y-axis), a greater *indices distance* appears when $\omega(i, i)$ increases of BSxPy. From Figure 5(a–f) contour plots, we find that the area of red contour lines (i.e. $d_{0.95} > 0$) is shrinking until it disappears when $\alpha = \beta = 1.0$ in Figure 5(f).

Similarly, we compared black-white FSxPy and DDSxPy (Figure 6(a–f) and Appendix Figure A2 for 2-D heat map). In DDSxPy, we use the Gaussian kernel with the bandwidth parameter to model the distance decay interactions. We let the y-axis represents the bandwidth in Figure 6(a–f); the x- and z-axis have the same meanings as in Figure 5. When bandwidth increases, the Gaussian kernel will become flattened, which corresponds to a smaller $\omega(i, i)$ in BSxPy. Thus, the bandwidth of DDSxPy and $\omega(i, i)$ of BSxPy have an inverse effect on $d_{0.95}$. For example, two surfaces in Figures 5(a and 6(a)) are oriented in different directions. Except for the orientation, the $d_{0.95}$ 3-D surface in Figure 6 exhibits the same V-shaped pattern in the time dimension as in Figure 5. Besides, we could identify the points where three indices have no differences by following the black contours lines.

To better examine the effect of the variation of parameter α , we draw the contours of $d_{0.95} = 0$ for each of the six 3-D plots in Figures 5 and 6 in one plot (see Figure 7(a, b)). We can see that a singularity is formed at 4:00 when the results of FSxPy are always less than or equal to BSxPy and DDSxPy, regardless of what α to be chosen. It corresponds to the flow data of Figure 3(a), in which the flow reaches its minimum at 4:00. Population returns to its place of residence at that time, the exchange between areal units reaches its minimum. Accordingly, residential segregation will dominate when the interaction weights of FSxPy reach a minimum value. However, the weight construction method of using boundary length (BSxPy) or the distance decay function (DDSxPy) ignores these decreasing interactions at 4:00, resulting in a biased (higher positive bias) exposure result.

5.3. Comparison result of local exposure index

We employed Equation (6) to compare the local version of black-white FSxPy and BSxPy. Two maps were created to show the *indices distance* heterogeneity at the local scale at the time 4:00 (2:00–6:00) and 12:00 (10:00–

2:00) (Figure 8). The colors of green to red represents the difference between FSxPy and BSxPy from low to high, and different infill styles describe the neighborhood types.

In Figure 8, we find that the spatial clustering of the negative *indices distance* regions (green clusters in Figure 8(a)) or positive *indices distance* regions (red clusters in Figure 8(b)) are almost identical to the distribution of B-neighborhoods. It means that black-white FSxPy results of B-neighborhoods are significantly less than the corresponding BSxPy results at 4:00 (Figure 8(a)) and significantly larger than BSxPy at 12:00 (Figure 8(b)). In contrast, the *indices distance* between black-white FSxPy and BSxPy was not statistically significant for almost all W-neighborhoods, meaning W-neighborhoods are barely affected by the flow patterns. This seems to contradict the fact that the outflow from W-neighborhoods is significantly larger than that of B-neighborhoods (Figure 3(a)) if we assume that more flow leads to more impact on segregation computation. However, considering that users from B-neighborhoods are more likely to travel to W-neighborhoods and the reverse does not stand (Table 2), it is not surprising that adding flow patterns have more effect on B-neighborhoods than W-neighborhoods.

6. Discussion

It is promising to introduce the population mobility network, hierarchical structure, and temporal information into the conventional segregation indices. This paper provides a general framework to support a timely and dynamic response to segregation changes. The proposed population-flow-based method has unique advantages over interaction estimation methods based on topology, geometry properties, or distance decay functions. It can better reflect the impacts of population movements on the degree of segregation in both the temporal and spatial dimensions. First, the V-shaped curve along the time dimension demonstrates the dynamic nature of the temporal dimension in the case study. It is consistent with our expectation of population movement over time. Second, the spatial clustering patterns in Figure 8 clearly show the heterogeneity of impacts of population flow in the spatial dimension, and this local variation is closely related to the type of the neighborhood and their flow patterns. Besides, the choice of parameters significantly influenced segregation results. For example, with a fixed β , in the process of increasing α from 0.0 to 1.0, we observe a progressively shallower V-shaped curve along the time dimension. And by controlling the bandwidth in the

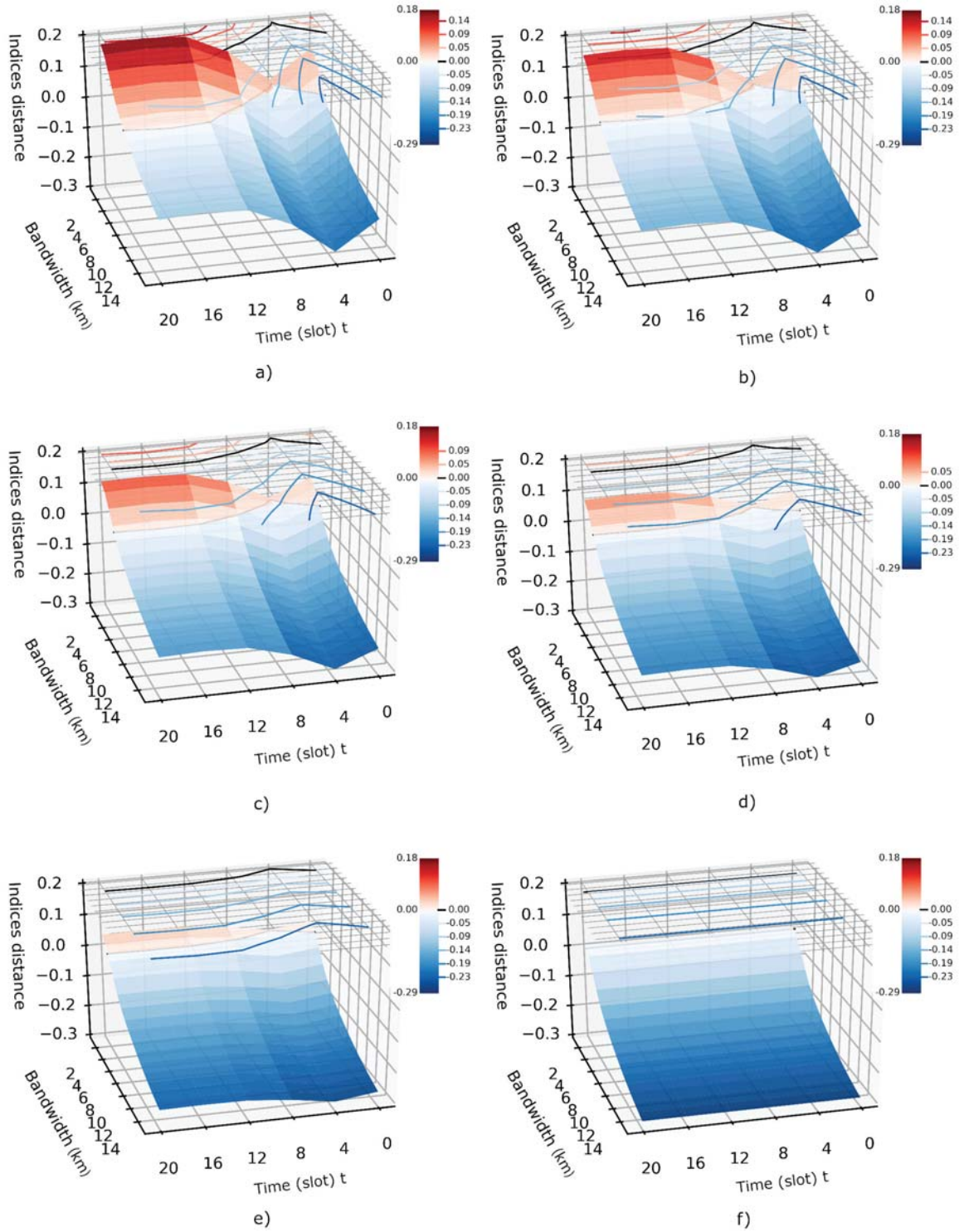


Figure 6. Statistical difference between black-white FSxPy (interested parameters are $\alpha, \beta, \text{Time (slot) } t$) and DDSxPy (the interested parameter is bandwidth) (FSxPy – DDSxPy): a) $\alpha=0.0$ and $\beta=1.0$; b) $\alpha=0.2$ and $\beta=1.0$; c) $\alpha=0.4$ and $\beta=1.0$; d) $\alpha=0.6$ and $\beta=1.0$; e) $\alpha=0.8$ and $\beta=1.0$; f) $\alpha=1.0$ and $\beta=1.0$.

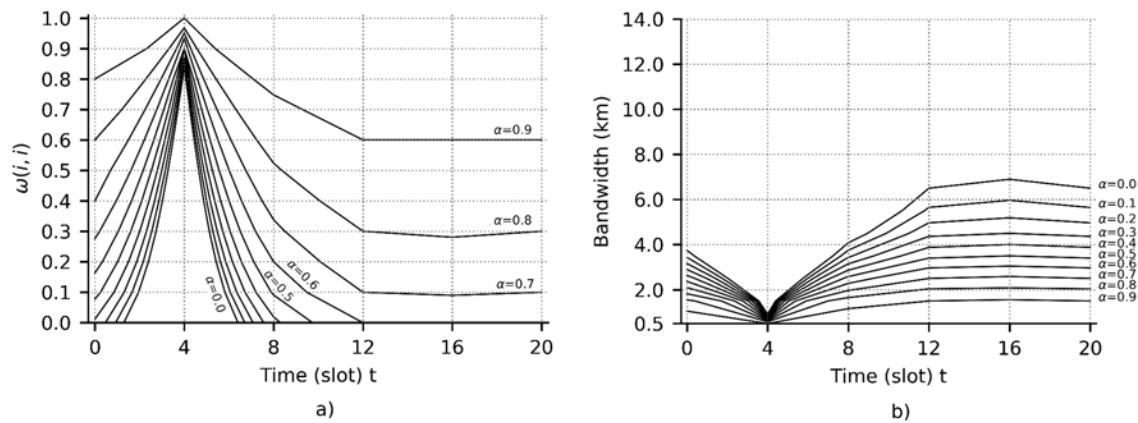


Figure 7. The Equal Lines of the new and old indices: a) FSxPy vs BSxPy; b) FSxPy vs. DDSxPy.

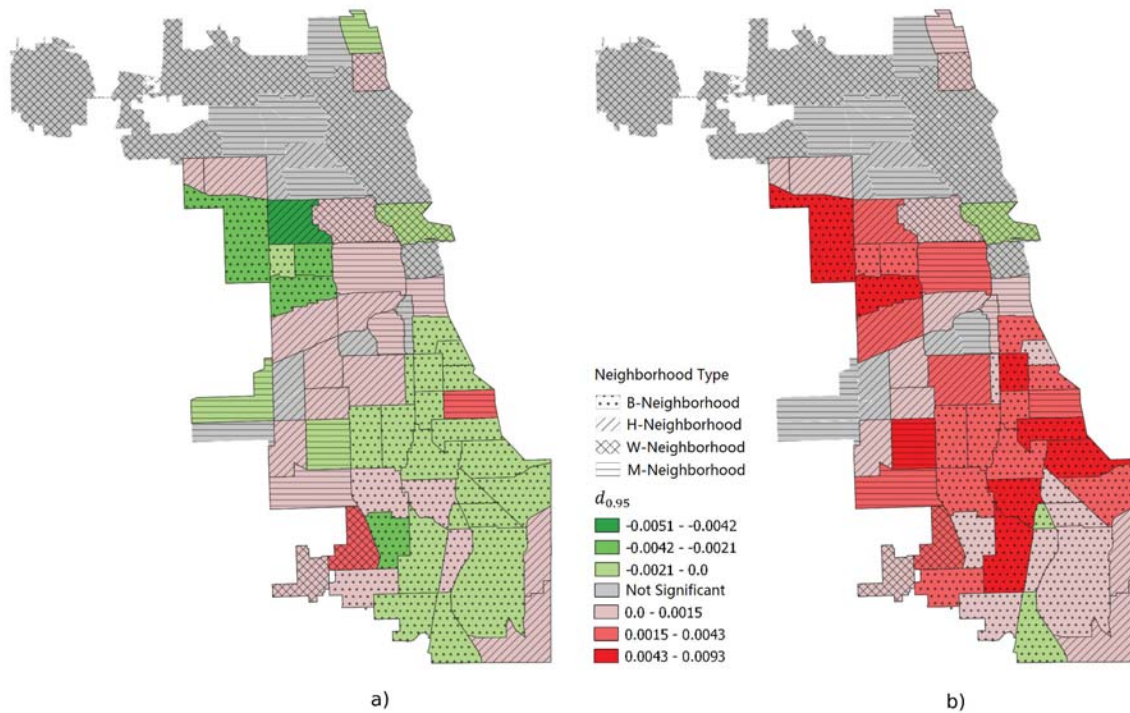


Figure 8. Map of Differences between two black-white exposure index (FSxPy – BSxPy) at the time a) 4:00 and b) 12:00.

distance decay method or the $\omega(i, i)$ in the topology-based method would boost or weaken the difference between FSxPy and the other two conventional segregation indices.

In the case study, we also found that the black-white FSxPy results of B-neighborhoods change significantly over time. But we need to interpret this change behavior with caution. We note that at 12:00, the FSxPy of

B-neighborhoods is significantly elevated on average, which clearly implies that the black-white exposure is improved. If we conclude that the B-neighborhoods are becoming less segregated at 12:00, we may overlook the other side of the inequality. By combining the flow patterns shown in Section 5.1, we can see that reduction in segregation comes at the cost of distance and time for B-neighborhoods users, who must travel a longer distance

to get to other neighborhoods (as shown in Figure 4). Therefore, we cannot rely on one dimension of the index alone but need to integrate other dimensions to study the underlying segregation or inequality mechanism.

We demonstrated how to include the population mobility patterns into segregation indices and how the flow-based segregation indices differ from conventional indices. We also conducted a systematic comparison study and provided a baseline for the future segregation study with flow data. But there are still some limitations, and future work needs to be done. First, we did not determine the α and β for the Chicago area because of the absence of proper data. As mentioned in section 4.2, α and β represent the proportion of the population of their own unit in the population mixing process for the most and least active neighborhoods at all observed times, respectively. In the future, we plan to use the residence and workplace surveys from the U.S. census bureau to estimate α and β . But when determining the value of α and β for one study, we also need to consider the spatial scale (level) of the study area. For example, if our study units change from a neighborhood level to a county level, the impact of population flow on the demographic composition of each unit tends to decrease. Therefore, a reasonable α at the county level should be larger than α at the neighborhood level.

Second, we use only Twitter data to demonstrate the proposed method in Chicago city. Twitter data has limitations. Biases in the data (e.g. the underrepresentation of older populations) and the failure to associate users with socioeconomic attributes (e.g. we do not know the Twitter user's racial/ethnic information) significantly limit its effectiveness and applications. One of the pressing issues is how to mitigate the representative bias of Twitter data. In the future, we want to conduct more studies on the representativeness of Twitter data and verify our method's extendibility using other flow datasets, such as taxi data and phone call records. On the other hand, to get a complete picture of the improvements in integration in Chicago with the integration of mobility data, we need to calculate the exposure index proposed in this paper in more cities to establish a comparative benchmark.

Finally, while we use the mobility data and its structure to enrich the exposure index, we do not extend its inherent concept. According to Massey and Denton (1988)'s definition, the exposure index only depicts the potential contact probability, not the actual contact and engagement of two groups. With the introduction of flow data, we still cannot distinguish whether there is a real interaction between the users. Clearly, there is room for future research in this direction. When measuring segregation, we need to look beyond

single segregation indices to describe the two groups' segregation in a broader context. In the future, we want to find the association between the population flow and the actual engagement of groups.

7. Conclusion

In summary, we proposed a method to add population mobility patterns to conventional segregation indices to portray the dynamics of segregation over time and space. From the systematic comparison of indices, we demonstrated that neighborhood flow networks, hierarchical structure information, time information, and the choice of parameters are all crucial factors to the segregation estimation. The segregation index calculation is a valuable reference to identify segregated populations and areas, but understanding its mechanism needs a multi-dimensional approach. This study facilitates the broader use of flow data in segregation research. It provides a powerful tool for urban planners to gain a more comprehensive understanding of the dimension of segregation dynamics.

Acknowledgments

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of these funding agencies. The authors would also like to thank the four anonymous reviewers and editor for their valuable suggestions and contributions. We also are indebted to Winston Yang for proof-reading the revised paper.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Data Availability Statement

The data that support the findings of this study are openly available in <https://doi.org/10.5281/zenodo.4397525>.

Funding

This work was supported by the National Science Foundation [1739491, 1937908]; Texas A&M University start-up funding [241117].

ORCID

Qingsong Liu  <http://orcid.org/0000-0001-5858-9960>
Mengmeng Liu  <http://orcid.org/0000-0001-7797-0193>
Xinyue Ye  <http://orcid.org/0000-0001-8838-9476>

References

- Bassolas, A., Barbosa-Filho, H., Dickinson, B., Dotiwalla, X., Eastham, P., Gallotti, R., Ghoshal, G., Gipson, B., Hazarie, S. A., Kautz, H., Kucuktunc, O., Lieber, A., Sadilek, A., & Ramasco, J. J. (2019). Hierarchical organization of urban mobility and its connection with city livability. *Nature Communications*, 10(1), 4817. <https://doi.org/10.1038/s41467-019-12809-y>
- Batty, M. (2013). *The new science of cities*. MIT press.
- Browning, C. R., & Soller, B. (2014). Moving beyond neighborhood: Activity spaces and ecological networks as contexts for youth development. *Cityscape (Washington, D.C.)*, 16(1), 165–196. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4121985/>
- Calabrese, F., Di Lorenzo, G., Liu, L., & Ratti, C. (2011). Estimating origin-destination flows using mobile phone location data. *IEEE Pervasive Computing*, 10(4), 36–44. <https://doi.org/10.1109/MPRV.2011.41>
- Chicago Data Portal. (n.d.). *Boundaries-neighborhoods*. Retrieved December 4, 2020, from <https://data.cityofchicago.org/Facilities-Geographic-Boundaries/Boundaries-Neighborhoods/bbvz-uum9>
- Cortes, R. X., Rey, S., Knaap, E., & Wolf, L. J. (2020). An open-source framework for non-spatial and spatial segregation measures: The PySAL segregation module. *Journal of Computational Social Science*, 3(1), 135–166. <https://doi.org/10.1007/s42001-019-00059-3>
- Dorman, M., Svoray, T., & Kloog, I. (2020). How does socio-economic and demographic dissimilarity determine physical and virtual segregation? *Journal of Spatial Information Science*, 21(21), 177–202. <https://doi.org/10.5311/JOSIS.2020.21.587>
- Duncan, O. D., & Duncan, B. (1955). A methodological analysis of segregation indexes. *American Sociological Review*, 20(2), 210. <https://doi.org/10.2307/2088328>
- Easley, J. (2018). Spatial mismatch beyond black and white: Levels and determinants of job access among Asian and Hispanic subpopulations. *Urban Studies*, 55(8), 1800–1820. <https://doi.org/10.1177/0042098017696254>
- 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. <https://doi.org/10.1016/j.jtrangeo.2015.10.009>
- Feitosa, F. F., Câmara, G., Monteiro, A. M. V., Koschitzki, T., & Silva, M. P. S. (2007). Global and local spatial indices of urban segregation. *International Journal of Geographical Information Science*, 21(3), 299–323. <https://doi.org/10.1080/13658810600911903>
- Gao, S., Wang, Y., Gao, Y., & Liu, Y. (2013). Understanding urban traffic-flow characteristics: A rethinking of betweenness centrality. *Environment and Planning B: Planning and Design*, 40(1), 135–153. <https://doi.org/10.1068/b38141>
- Gao, S., Yang, J., Yan, B., Hu, Y., Janowicz, K., & McKenzie, G. (2014). Detecting origin-destination mobility flows from geo-tagged Tweets in greater Los Angeles Area. https://grants.mckenzie.com/academics/McKenzie_DODMF.pdf
- Guo, D., Zhu, X., Jin, H., Gao, P., & Andris, C. (2012). Discovering spatial patterns in origin-destination mobility data. *Transactions in GIS*, 16(3), 411–429. <https://doi.org/10.1111/j.1467-9671.2012.01344.x>
- Hawelka, B., Sitko, I., Beinat, E., Sobolevsky, S., Kazakopoulos, P., & Ratti, C. (2014). Geo-located Twitter as proxy for global mobility patterns. *Cartography and Geographic Information Science*, 41(3), 260–271. <https://doi.org/10.1080/15230406.2014.890072>
- Huang, Q., & Wong, D. W. S. (2016). Activity patterns, socio-economic status and urban spatial structure: What can social media data tell us? *International Journal of Geographical Information Science*, 30(9), 1873–1898. <https://doi.org/10.1080/13658816.2016.1145225>
- Kruschke, J. (2014). *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan*. Academic Press.
- Kwan, M.-P. (2009). From place-based to people-based exposure measures. *Social Science & Medicine*, 69(9), 1311–1313. <https://doi.org/10.1016/j.socscimed.2009.07.013>
- 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>
- Lee, J., & Li, S. (2017). Extending Moran's index for measuring spatiotemporal clustering of geographic Events. *Geographical Analysis*, 49(1), 36–57. <https://doi.org/10.1111/gean.12106>
- Lenormand, M., Picornell, M., Cantú-Ros, O. G., Tugores, A., Louail, T., Herranz, R., Barthelemy, M., Frías-Martínez, E., & Ramasco, J. J. (2014). Cross-checking different sources of mobility information. *PLoS ONE*, 9(8), e105184. <https://doi.org/10.1371/journal.pone.0105184>
- Liu, Q., Wang, Z., & Ye, X. (2018). Comparing mobility patterns between residents and visitors using geo-tagged social media data. *Transactions in GIS*, 22(6), 1372–1389. <https://doi.org/10.1111/tgis.12478>
- Luo, F., Cao, G., Mulligan, K., & Li, X. (2016). Explore spatiotemporal and demographic characteristics of human mobility via Twitter: A case study of Chicago. *Applied Geography*, 70, 11–25. <https://doi.org/10.1016/j.apgeog.2016.03.001>
- Massey, D. S., & Denton, N. A. (1988). The dimensions of residential segregation. *Social Forces*, 67(2), 281. <https://doi.org/10.2307/2579183>
- Morgan, B. S. (1983). A distance-decay based interaction index to measure residential segregation. *Area*, 15(3), 211–217. <http://www.jstor.org/stable/20001935>
- Morril, R. L. (1991). On the measure of geographic segregation. *Geography Research Forum*, 11(1), 25–36. <https://grf.bgu.ac.il/index.php/GRF/article/view/91/87>
- Palchikov, V., Mitrović, M., Jo, -H.-H., Saramäki, J., & Pan, R. K. (2015). Inferring human mobility using communication patterns. *Scientific Reports*, 4(1), 6174. <https://doi.org/10.1038/srep06174>
- Park, Y. M., & Kwan, M. (2018). Beyond residential segregation: A spatiotemporal approach to examining multi-contextual segregation. *Computers, Environment and Urban Systems*, 71(May), 98–108. <https://doi.org/10.1016/j.compenvurbsys.2018.05.001>
- Prestby, T., App, J., Kang, Y., & Gao, S. (2020). Understanding neighborhood isolation through spatial interaction network analysis using location big data. *Environment and Planning A: Economy and Space*, 52(6), 1027–1031. <https://doi.org/10.1177/0308518X19891911>

- Reardon, S. F., & O'Sullivan, D. (2004). Measures of spatial segregation. *Sociological Methodology*, 34(1), 121–162. <https://doi.org/10.1111/j.0081-1750.2004.00150.x>
- Shelton, T., Poorthuis, A., & Zook, M. (2015). Social media and the city: Rethinking urban socio-spatial inequality using user-generated geographic information. *Landscape and Urban Planning*, 142, 198211. <https://doi.org/10.1016/j.landurbplan.2015.02.020>
- Silm, S., & 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. <https://doi.org/10.1016/j.ssresearch.2014.03.011>
- 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(December2017), 140–153. <https://doi.org/10.1016/j.physa.2015.06.032>
- Walker, K. (2018). Locating neighbourhood diversity in the American metropolis. *Urban Studies*, 55(1), 116–132. <https://doi.org/10.1177/0042098016643481>
- Wang, D., Li, F., & Chai, Y. (2012). Activity spaces and socio-spatial segregation in Beijing. *Urban Geography*, 33(2), 256–277. <https://doi.org/10.2747/0272-3638.33.2.256>
- Wang, Q., Phillips, N. E., Small, M. L., & Sampson, R. J. (2018). Urban mobility and neighborhood isolation in America's 50 largest cities. *Proceedings of the National Academy of Sciences*, 115(30), 7735–7740. <https://doi.org/10.1073/pnas.1802537115>
- White, M. J. (1983). The Measurement of Spatial Segregation. *American Journal of Sociology*, 88(5), 1008–1018. <https://www.jstor.org/stable/2779449>
- Wong, D. W. S. (1993). Spatial Indices of Segregation. *Urban Studies*, 30(3), 559–572. <https://doi.org/10.1080/00420989320080551>
- Wong, D. W. S. (1998). Measuring Multiethnic Spatial Segregation. *Urban Geography*, 19(1), 77–87. <https://doi.org/10.2747/0272-3638.19.1.77>
- 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>
- Xu, Y., Belyi, A., Santi, P., & Ratti, C. (2019). Quantifying segregation in an integrated urban physical-social space. *Journal of the Royal Society Interface*, 16(160), 20190536. <https://doi.org/10.1098/rsif.2019.0536>
- Yao, J., Wong, D. W. S., Bailey, N., & Minton, J. (2019). Spatial segregation measures: A methodological review. *Tijdschrift Voor Economische En Sociale Geografie*, 110(3), 235250. <https://doi.org/10.1111/tesg.12305>
- Yip, N. M., Forrest, R., & Xian, S. (2016). Exploring segregation and mobilities: Application of an activity tracking app on mobile phone. *Cities*, 59, 156–163. <https://doi.org/10.1016/j.cities.2016.02.003>
- Zhu, D., Wang, N., Wu, L., & Liu, Y. (2017). Street as a big geo-data assembly and analysis unit in urban studies: A case study using Beijing taxi data. *Applied Geography*, 86, 152–164. <https://doi.org/10.1016/j.apgeog.2017.07.001>