



Socioeconomic biases in urban mixing patterns of US metropolitan areas

Rafiazka Millanida Hilman¹, Gerardo Iñiguez^{1,2,3} and Márton Karsai^{1,4*} 

*Correspondence: karsaim@ceu.edu

¹Department of Network and Data Science, Central European University, 1100 Vienna, Austria

⁴Alfréd Rényi Institute of Mathematics, 1053 Budapest, Hungary

Full list of author information is available at the end of the article

Abstract

Urban areas serve as melting pots of people with diverse socioeconomic backgrounds, who may not only be segregated but have characteristic mobility patterns in the city. While mobility is driven by individual needs and preferences, the specific choice of venues to visit is usually constrained by the socioeconomic status of people. The complex interplay between people and places they visit, given their personal attributes and homophily leaning, is a key mechanism behind the emergence of socioeconomic stratification patterns ultimately leading to urban segregation at large. Here we investigate mixing patterns of mobility in the twenty largest cities of the United States by coupling individual check-in data from the social location platform Foursquare with census information from the American Community Survey. We find strong signs of stratification indicating that people mostly visit places in their own socioeconomic class, occasionally visiting locations from higher classes. The intensity of this ‘upwards bias’ increases with socioeconomic status and correlates with standard measures of racial residential segregation. Our results suggest an even stronger socioeconomic segregation in individual mobility than one would expect from system-level distributions, shedding further light on uneven mobility mixing patterns in cities.

Keywords: Segregation in mobility; Urban mixing; Socioeconomic inequalities

1 Introduction

Patterns of socioeconomic inequality can be found everywhere in a modern city. Large variations in earned income leading to uneven access to services, healthcare and education [1, 2], as well as spatial and housing segregation [3, 4], are just two of the most drastic examples of socioeconomic disparity. Less studied is the segregation related to mobility mixing, where people from different socioeconomic classes encounter each other less often than what is potentially allowed by the city fabric [5–7].

Big data presents a unique opportunity to analyse the role of human mobility in segregation, from the level of individuals to the scale of societies. Digital data tracing human movements in cities ranges from mobile call detail records (CDRs) [8, 9] and GPS trajectories [10–12], to location-sharing services (LSS) and check-in sequences on social media platforms [13–15]. The analysis of these data sources, providing anonymised individual trajectories with unprecedented spatiotemporal resolution, has proven essential for our

© The Author(s) 2022. This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

growing understanding of the underlying mechanisms of human mobility [16–19], and the associated ability to predict future trajectories [20, 21]. It also offers the possibility to engage in a more comprehensive and nuanced exploration of urban socioeconomic segregation, by combining high-dimensional mobility data with information on the socioeconomic traits of individuals [5, 22, 23].

Earlier studies on human mobility present evidence of characteristic spatial scales [8, 9, 16, 19], as well as a correlation between human spatial behaviour and socioeconomic dynamics [24–26]. Rather than being homogeneously mixed, human mobility (as represented by daily individual trajectories throughout urban spaces) is strongly influenced by socioeconomic preferences. People sharing socioeconomic backgrounds are more likely to visit similar places within their class range and interact amongst themselves [27–30], thus generating stratified mobility and social network patterns. In the presence of homophily mixing [31], spatial exploration is dictated by one's socioeconomic class, reducing the number of visits to locations with different economic status, and thus inducing highly predictable trajectories. However, when people aspire to diversify their experiences by, e.g., visiting lavish areas of the city, where they have never been able to go before, the potential for an upwards bias in visiting patterns appears. Meanwhile, other has studied the effects of segregation of mixing in urban places using location data to detect exploration/exploitation behavioural patterns and their correlations with socioeconomic status [32].

Homophily mixing is not the only mechanism influencing mobility patterns. The variability of socioeconomic traits such as ethnic group, education level, occupation sector, etc. also constrains the possibility of movement in urban spaces via residential segregation [3, 4, 33, 34], where people with similar backgrounds live next to each other and form fragmented areas in the city. Given the potentially complex interplay between human mobility and socioeconomic stratification, it is worth asking whether the presence of biased mobility across tracts of some socioeconomic trait is associated with lower residential segregation. This is particularly relevant given the number of studies reporting mobility as a key pillar in diminishing segregated spaces among people from diverse groups in society [35–38]. People show heterogeneity in many aspects, including their mobility characteristics and socioeconomic capacities, which shape their patterns of movement across urban space.

In related studies, Dong and others [5] investigate segregation in economic and social interactions by using credit card transactions and Twitter data. They find that segregation increases with difference in socio-economic status but is asymmetric for purchase activity. Meanwhile, neighbourhood isolation has been used [29] to observe travel patterns of individuals extracted from Twitter data. These findings show racial differences in the composition of the neighbourhoods visited. Black and Hispanic neighbourhoods, regardless of their socioeconomic status, are less exposed than white neighbourhoods. Moreover, white poor neighbourhoods are substantially isolated from non-poor white neighbourhoods. Morales et al. [30] aim to investigate polarization in shopping, communication, and mobility reflected by online interaction in Twitter. It confirms the theoretical underpinning in which within-group homogenization and between-group differentiation promote social fragmentation. They provide in-depth assessment on polarization of conversations between neighbourhoods and show that the differentiation of online conversations reflects the distribution of wealth.

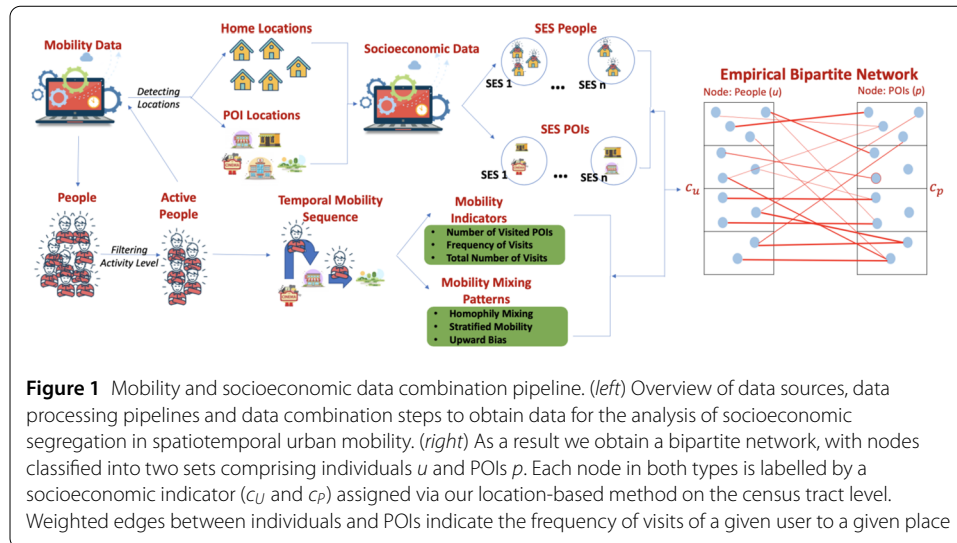
Build on these results, our study is dedicated to reveal the role of visit preference in mobility bias across socioeconomic status at the individual and class levels. We specifically analyse the extent to which mobility may contribute to the emergence of socioeconomic stratification, ultimately leading to urban segregation at large. The scope of our discussion is concentrating on visit preference to detect generic patterns where ‘upwards bias’ increases with socioeconomic status. Additional investigation on ethnic isolation aims to instantiate the entanglement between characteristics and mobility patterns with other socioeconomic features, such as ethnic residential distribution.

In this study, we emphasise the need to query the extent to which behavioural segregation (bias in mobility) is related to residential segregation. We take a step forward in the current analysis of segregation in mobility by asking the following question: How do socioeconomic attributes and geographic constraints affect the spatiotemporal process of individuals moving in urban spaces? To answer this question, we analyse individual check-in trajectories in the twenty largest cities of the United States, coupled with detailed socioeconomic maps indicating the economic status of people and places they visit. After a short data description, in the following we will introduce stratification matrix measures and individual- and class-level mobility bias scores to quantify patterns of mobility segregation, visiting biases, and their variation across cities with wide-ranging socioeconomic and ethnics segregation profiles. We base our analysis on observational behavioural data, which may not be fully representative for the observed populations. To address this shortcoming we carry out a careful analysis about the biases and confounding effects characterising the analysed data set. While the results of this analysis are reported in the discussion and the Supplementary Material of this paper, they confirm the robustness of upward biased visiting patterns of people to places in cities with various socioeconomic stratification profiles.

2 Data description

In order to simultaneously capture the mobility patterns and socioeconomic status of people, we concentrate on two independent sources (mobility and socioeconomic data, described below) and combine them using spatial information.

Mobility data: To construct individual mobility trajectories, we analyse a large, open Foursquare dataset [39], which records how people move from one place to another. Data comes as a sequence of user check-ins to places, or points of interest (POIs), thus providing information on mobility trajectories of individuals and visiting frequencies of places. This dataset is not collected directly through the Foursquare open API, but from Foursquare check-ins via Twitter. The crawling method corresponds to 18 months (549 days) of observations between April 2012 and September 2013 for users with Foursquare-tagged tweets. Using this mobility data, constituted by roughly 26,502 people with nearly 1,830,276 check-ins, we concentrate only on active users (who checked-in from at least two different places during the observation period). Focusing on the 20 largest metropolitan areas in the US, we also infer the home locations of 26,502 users following a conventional pipeline of conditions [40] [for a detailed description of the method and a statistical summary see Additional file 1, Section A]. The Foursquare dataset is not a uniform sample of the population, and as such, it may introduce bias in our analysis of mobility patterns. However, we expect that aggregation and averaging, as well as the length of the observation period (beyond yearly seasonality), decreases this potential for bias. In any case, in



the Supplementary Material we estimate discrepancies between Foursquare data and the real population via a bootstrapping analysis (Additional file 1, Section A), a Kruskal-Wallis H test, and a Dunn's test (Additional file 1, Section H).

Socioeconomic data: To estimate the socioeconomic status of people and places, we rely on the 2012 American Community Survey (ACS) [41] (recorded in the year matching the closest to the Foursquare observation period). After identifying the corresponding ACS census tract where a user's home location lies, we associate the socioeconomic indicators of this location to the individual. In order to estimate the economic status of a place, we follow a similar strategy and assign local socioeconomic status indicators to POIs based on their locations. Although the socioeconomic status of venues could arguably be better estimated from their pricing, this information is at present not available to us. Thus, we assume that the socioeconomic status (SES) of people living at a location is well correlated with the pricing of venues in the same neighbourhood and offered services around (for a summary of our data construction pipeline see Fig. 1).

In order to obtain a proper representation of socioeconomic status in the context of segregation, we consider 78 features from the ACS data. Although such a large number of dimensions in principle provides a rich way of quantifying the socioeconomic status of locations (and people living there), it turns out these variables have high redundancy. We perform a principal component analysis to identify the most relevant ones and find that income features (11 variables) have the largest loading, accounting for most of the socioeconomic variance between places. After implementing three different techniques (mutual information rank [42], decision tree [43], and Gini coefficient [44]), per capita income consistently stands out as the best indicator of individual SES: It accounts for the largest variance and it correlates strongly with other income variables such as earning/wage, wealth, and supplementary source of income (for more details on this analysis see Additional file 1, Section A). By using the average per capita income as the socioeconomic indicator of active users living in a given tract, we sort them in an ascending order. To group them into distinct socioeconomic classes, we then segment this sorted list into 10 equally populated groups with people of the lowest income in class 1 and highest income in class 10.

By means of this procedure we assign a socioeconomic class c_U to each user. In identical fashion, each venue is assigned a value c_P .

3 Results

Our main scientific goal is to study socioeconomic segregation and biases in population mixing in cities by observing correlation patterns between the SES of people and places they visit. Using the collected data, this objective can be addressed by building a network of individuals visiting places. We define a stratified bipartite network $G = (U, P, E)$, where individual u is a node in set U , and place p belongs to set P (with $U \cap P = \emptyset$). People and places are connected by edges $e_{u,p} \in E$ with weights $w_{u,p}$ coding the number of times person u visited place p (see Fig. 1). Further, we stratify U into a set of socioeconomic classes indexed by values from C_U thus assigning a class membership $c_u = i \in C_U$ to each individual.

In the same way we define $c_p = j \in C_P$ classes for places. This network representation captures all information about the socioeconomically stratified visiting patterns of people to venues, coding their possible encounters and giving an aggregated description of the potential mixing patterns of people of different socioeconomic classes.

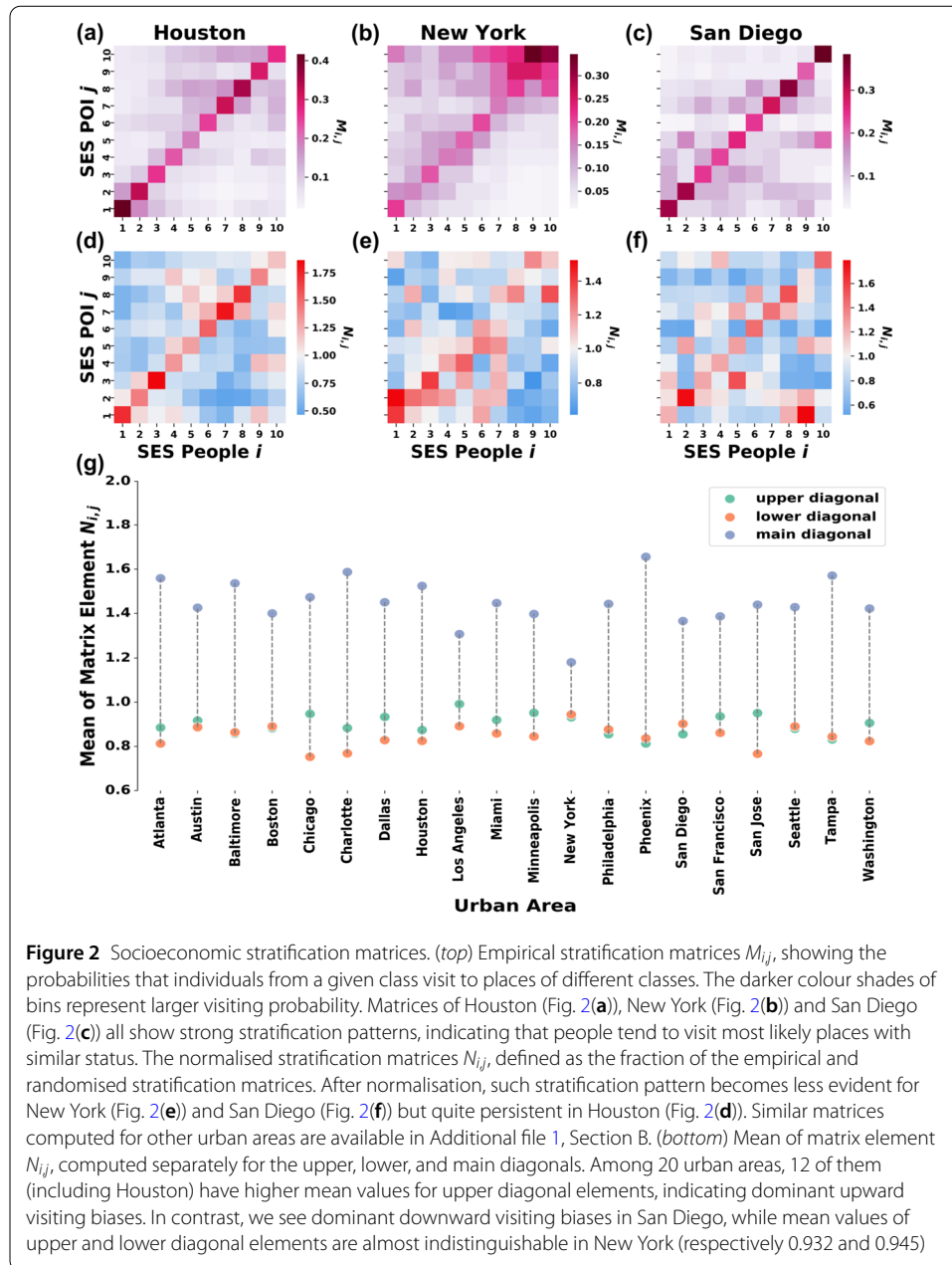
3.1 Matrix measures

Based on the bipartite network representation we can measure the frequency at which people of a given class visit places in different classes. To summarise these visiting patterns we use stratification matrices [23]. An *empirical stratification matrix* gives the probability that a person $u \in U$ from a given socioeconomic class $c_u = i \in C_U$ visits a place $p \in P$ belonging to a class $c_p = j \in C_P$. More formally:

$$M_{i,j} = \frac{\sum_{U, c_u=i} \sum_{P, c_p=j} w_{u,p}}{\sum_{j \in C_P} \sum_{U, c_u=i} \sum_{P, c_p=j} w_{u,p}}, \quad (1)$$

where the numerator counts the number of times people from class i visit places of class j , and the denominator normalises this frequency matrix column-wise to obtain a visiting probability distribution for each individual class $i \in C_U$. Such matrices are shown in Fig. 2 for selected cities (Houston, New York and San Diego). The dominant diagonal elements for Houston and San Diego indicate strongly stratified visiting patterns in these cities. People prefer to visit places of their own or similar socioeconomic class, rather than places from remote classes. Interestingly, for New York this pattern is less evident suggesting weaker socioeconomic preferences in visiting venues.

To decide if these patterns appear as the consequence of population statistics or other confounding effects, we compare the matrix $M_{i,j}$ to a reference matrix, which measures similar stratification patterns in a system where visiting patterns appear uniformly at random with certain constraints. This *randomised stratification matrix* is defined through a random rewiring process of the bipartite network, while constraining the total number and frequency of visits of each individual (i.e. their activity and link weights), the class of individuals and places, but fully randomising links between individuals and visited places otherwise. The randomisation is performed by selecting randomly for each link of an individual u a place to visit from the set of places ever visited by their respected socioeconomic classes c_u , while keeping the link weight intact. This *in-class randomisation* allows us to



compare an individual’s behaviour to similar others, meanwhile distinguishing between socioeconomic classes, which potentially are characterised by very different visiting patterns.

After generating randomised bipartite networks via 100 independent realisations, we compute a similar column-wise normalised stratification matrix R_{ij} , representing the probability of people from class c_i randomly visiting to places of class c_p . To finally obtain whether the empirical mobility patterns appear more frequently than by chance, we compare the empirical and the preference-based null model matrices. We obtain a *normalised stratification matrix* N_{ij} by taking the element-by-element fraction of the empirical and

random contact matrices as:

$$N_{i,j} = \frac{M_{i,j}}{R_{i,j}}. \quad (2)$$

Each element in the matrix $N_{i,j}$, which are $N_{i,j} > 1$ (red bins in Fig. 2(d)–(f)) indicates that the visits made by individuals from class $i \in C_U$ to place of class $j \in C_P$ appeared with higher probability in the empirical observations than it was expected from the random null model.

Otherwise, the blue blocks for $N_{i,j} < 1$ show that the corresponding visits appeared with a smaller probability than expected by chance. In cases red bins dominate the diagonal of the normalised matrix $N_{i,j}$, it indicates patterns of socioeconomic stratification, where people prefer to visit places of similar socioeconomic status as their own, rather than places, which are richer or poorer than them. This is the case of Houston and San Diego (see Fig. 2(d) and (f) respectively) and many other cities listed in Additional file 1, Section B. However, this character is less evident for New York (see Fig. 2(e)), where despite known strong residential segregation, the city fabric mitigates a more homogeneous mixing of people.

These normalised stratification matrices reveal further characters of possible biases of people in choosing places to visit, out of their own class. If in a city people exhibit upward visiting biases, thus they tend to choose more expensive places to visit when they step out of their own class, the upper diagonal matrix elements of $N_{i,j}$ would appear dominantly red. While, if the opposite is true, the lower diagonal elements would reflect similar but downward visiting biases. To simply quantify these patterns, we compute the average values $N_{i,j}$ elements of normalised stratification matrix of cities above, at, and under their diagonals. From Fig. 2(g) it is clear that, in all cities, diagonal elements dominantly concentrate visiting probabilities. However, in terms of off-diagonal averages, in most of the cities (like in Houston in Fig. 2(d)) the upper diagonal average takes a larger value as compared to the lower diagonal average, indicating present upward visiting biases in these metropolitan areas. Meanwhile, in some cases the contrary is true (like in San Diego in Fig. 2(f)) or in some cities these averages are very similar thus indicating no dominant upward or downward visiting biases, as in case of New York (see Fig. 2(e) and Fig. 2(g)).

3.2 Individual bias

The matrix measures presented in Fig. 2 reflect the coexistent socioeconomic configurations derived from visit trajectories. Firstly, the empirical stratification matrices $M_{i,j}$ bring an initial indication of homophily mixing as seen in the dominant frequency visit within own class. Secondly, these results reveal the underlying inclination in visiting places situated in higher SES as depicted by the larger proportion of upper diagonal elements in the normalised stratification matrices $N_{i,j}$ in most of the cities. Taking these two configurations into account, it can be inferred that while individual mobility is dictated by the membership of socioeconomic class most of the time, the embedded motivation to visit upper class places is still present.

We take a technical step ahead in order to adequately quantify this visiting bias that indicates deviations in mixing from the respected c_u socioeconomic class of an individual. We compute a single *empirical individual bias score* for each individual $u \in U$ as

$$B_u = \langle c_p \rangle_u - c_u, \quad (3)$$

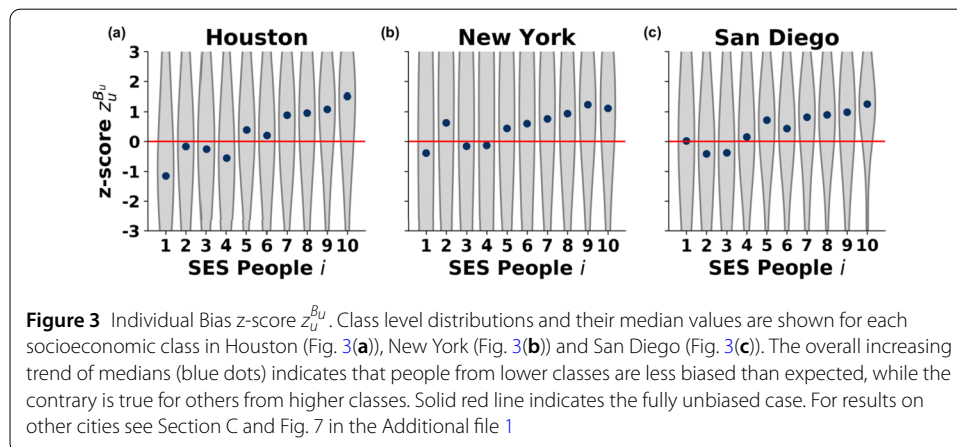
where $\langle c_p \rangle_u = \frac{\sum_{p \in P} w_{u,p} \times c_p}{n_p^u}$ is the average socioeconomic status of places an individual u visited, defined as the fraction of the $\sum_{p \in P} w_{u,p} \times c_p$ sum of socioeconomic status of places in the trajectory of individual u and the $n_p^u = \sum_{p \in P} w_{u,p}$ number of times individual u visited any places. An individual has upward visiting bias if her individual score B_u is positive, meaning that she tends to visit places located in more affluent areas than where she lives. Secondly, an individual with negative score value has downward visiting bias since places she usually visits are situated in lower socioeconomic class than her own. Otherwise, an individual does not have any indication of bias ($B_u = 0$) if she visits places within her own socioeconomic rank. A reference model for this measure can follow a similar logic as the in-class randomisation for the realisations of network reference models explained before. Given the individual trajectory resulted from the random visit generating process, we calculated a *randomised individual bias score* using the same formula as in Eq. (3). Note that in this measure boundary effects may appear, as people from the poorest class cannot exhibit downward bias, and similarly, the highest class cannot be upward biased. Individual bias scores can be fairly compared to null models, which retain these boundary effects. In-class randomisation fulfils this requirement, providing an average randomised bias score $\langle B^{\text{rand}} \rangle_u$ for each individual separately. Note, that the randomised individual bias scores take non-trivial values, different from zero, due to the individual variance of visiting frequencies of individuals to different places. These are represented by the weights $w_{u,p}$ in the bipartite network, which are preserved during the randomisation process.

The comparison of the empirical and in-class normalised individual bias scores can be best quantified by an individual bias z-score as

$$z_{u}^{B_u} = \frac{B_u - \langle B^{\text{rand}} \rangle_u}{\sigma_u^{B^{\text{rand}}}}, \tag{4}$$

where $\langle B^{\text{rand}} \rangle_u$ is the mean and $\sigma_u^{B^{\text{rand}}}$ is the standard deviation of the randomised individual bias scores across 100 independent realisations of the null model. The value of $Z_{u}^{B_u}$ reflects how much the individual bias deviates from the expected bias for an individual who chooses places to visit with the same frequency as before but selects them from a given set of places dictated by others within the same socioeconomic class.

The class distributions of individual z-scores together with their median values are shown in Fig. 3, where the unbiased level is assigned as a flat red line. These distributions



appear broad for each class, indicating that actually people from any class exhibit upward or downward biases in terms of their visiting patterns to other socioeconomic classes. Interestingly, the median z-scores indicate an increasing trend in all the three depicted cities. The people from lower classes appear with slightly negative bias z-score, meaning they have a slightly weaker bias to visit places of different socioeconomic classes than expected from their random visiting patterns. In contrary, the middle and upper classes are evidently biased stronger than expected. This increasing trend of the median of the individual bias z-score with socioeconomic classes surprisingly characterises all the investigated cities as shown in Section C and in Fig. 8 in the Additional file 1.

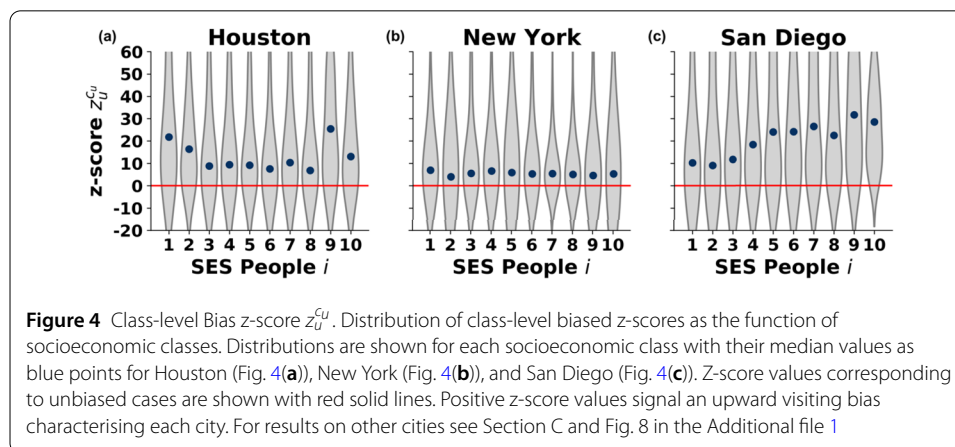
3.3 Class-level bias

The individual bias score B_u compares the average class of visited places of an individual to its own socioeconomic rank inferred from its home location. Meanwhile, its z-score $z_u^{B_u}$ indicates if this individual bias is weaker or stronger than expected from random behaviour. However, this measure is using the class label of the individual as a reference of comparison, and it says less about whether an individual visits higher or lower class places as compared to the random expected behaviour characterising other individuals in its own class. To directly measure this effect we introduce a class level z-score measure

$$z_u^{c_u} = \frac{\langle c_p \rangle_u - \langle c_p \rangle_{c_u}^{rand}}{\sigma_{c_u}^{rand}}, \tag{5}$$

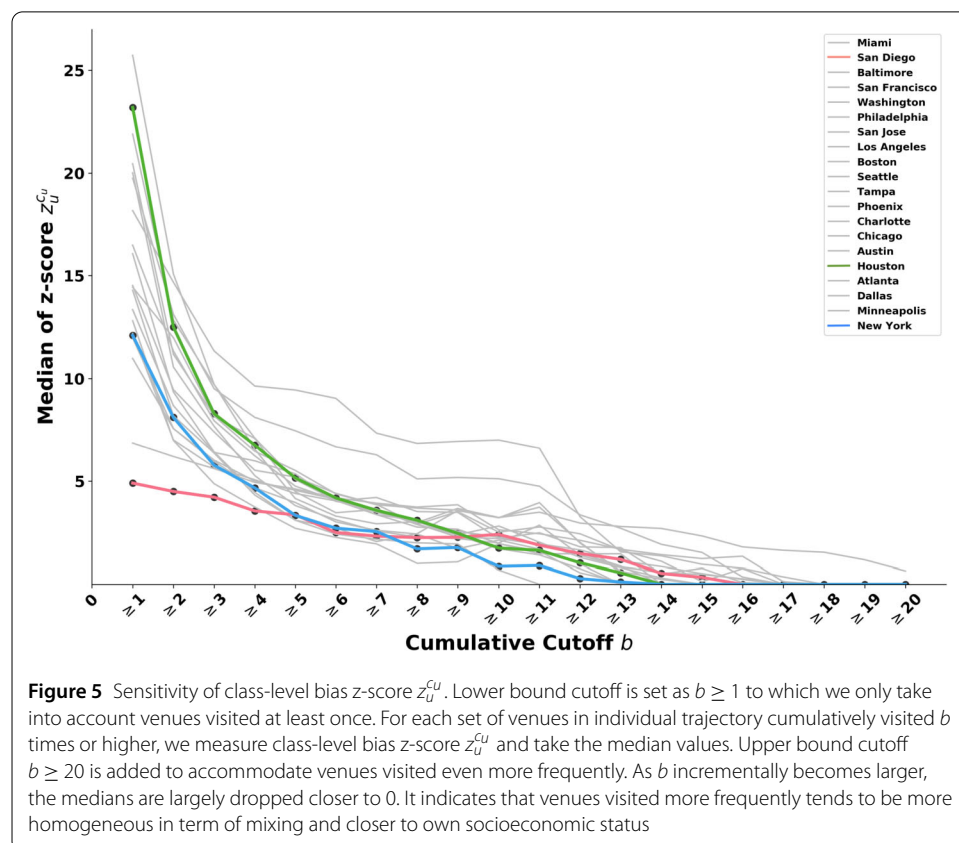
where $\langle c_p \rangle_u$ is the average socioeconomic status of places individual u visited, and $\langle c_p \rangle_{c_u}^{rand}$ and $\sigma_{c_u}^{rand}$ are the average and standard deviation (respectively) of class of places that others from class c_u would visit if behave randomly. This reference measure, just like before, is generated by in-class shuffling to obtain null models over 100 realisations. The value of $z_u^{c_u}$ reflects directly how much the individual behaviour deviates from the expected level, when the individual could choose randomly places to visit from a given set dictated by others from the same socioeconomic class.

Results in Fig. 4 show a different behaviour as compared to the individual bias scores. In case of New York (see Fig. 4(b)), the distributions of the class level bias z-scores indicate that, although the variation is large in each socioeconomic classes, the medians of these distributions are all slightly positive and independent of the socioeconomic class.



This signals a weak upward bias in people’s visiting patterns in New York as compared to the class behaviour that appears for each class. In other cities, we find several other bias patterns during our analysis (see Additional file 1, Section C Fig. 10). In case of San Diego (in Fig. 4(c)) class-level biases are all positive and evidently increasing with the socio-economic classes. This suggests that richer people in San Diego may visit even more affluent places, than one would expect from their random class behaviour. Somewhat the opposite trend can be observed for Houston (Fig. 4(a)), where although the class-level bias z-score is always positive and indicates upward bias for each class, it seems to follow an overall decreasing trend.

Visiting patterns measured by the class-level bias scores suggest that an upward socio-economic bias characterise each cities we study. Although these measures incorporate the visiting frequency distribution of individuals, they do not show evidently that upward biases typically appear due to repeated visits to places with higher class scores, or due to several occasional visits to places out of ones socioeconomic classes. To answer this question, we recompute the median bias scores, excluding places which were visited less number of times by an individual than a given threshold. Results are depicted in Fig. 5 for each city. As expected, the median class-level bias score appears as a decreasing function of the frequency threshold in each city. This suggests that people visit more frequently places, which are closer in terms of socioeconomic status to their own class, while visit more affluent places occasionally only, that in turn causes upward bias patterns characterising their class. Beyond this general decreasing character, this function indicates large variance between different cities. For example, in case of San Diego (red line in Fig. 5), this



curve starts from a high z-score value when all visits are considered but decreases rapidly as repeated visits are taken into account. In case of New York this function starts from a relatively small z-score value and decreases linearly for larger threshold values. This suggests a different visiting behaviour where people typically visit places more than one time, but closer to their own socioeconomic class.

We prefer this particular measure over the one on individual bias, as our objective here is to reveal the source of upward bias, whether it is driven by repeated visits to places with higher class scores, or due to several occasional visits to places out of one's socioeconomic class. The class-level bias z-score serves this purpose as it already incorporates the visiting frequency distribution of individuals compared to their own class and gives positive z-score values.

3.4 Mobility mixing and segregated residences

While there is an expected relation between the mobility mixing patterns and residential segregation in a city, the combined investigation of these phenomena has not received much attention so far. Their relation is important however for several reasons. For example, due to the multitude of correlated socioeconomic factors it is likely that e.g. ethnicity, which strongly correlates with income status in US metropolitan areas, correlates also with residential segregation, as it has been shown in several studies [4, 45]. On the other hand, the daily mobility of people and their visiting patterns to different places are constrained also by these socioeconomic factors, thus they are likely to resemble similar segregation patterns. To investigate these correlations, we focus on different ethnic groups and the likelihood of their mixing in cities, which exhibits different levels of mobility segregation patterns.

To quantify the level of segregation in mobility mixing, we analyse the earlier introduced normalised stratification matrix $N_{i,j}$ for each city. As we have discussed, signatures of segregation can be associated to strong diagonal elements in these matrices, indicating that people of a given SES are the most likely to visit places associated with the same or similar SES, as compared to random visiting patterns. To quantify the strength of diagonal concentration of visiting probabilities, we measure the diagonality index of the normalised stratification matrices [46], which is similar to the assortativity coefficient used by others [5, 47]. It is defined as the Pearson correlation coefficient of matrix entries as

$$r_N = \frac{\sum_{i,j} ijN_{i,j} - \sum_{i,j} iN_{i,j} \sum_{i,j} jN_{i,j}}{\sqrt{\sum_{i,j} i^2N_{i,j} - (\sum_{i,j} iN_{i,j})^2} \sqrt{\sum_{i,j} j^2N_{i,j} - (\sum_{i,j} jN_{i,j})^2}}. \tag{6}$$

Here $i \in c_u$ indicates the socioeconomic class of individuals and $j \in c_p$ is the same for places. The diagonality index takes values between -1 and 1 . In case it is 1 , it indicates perfect assortative mixing corresponding to a fully stratified matrix with non-zero elements in its diagonals and zero anywhere else. Cities with large r_n values are characterised by visiting patterns of people who are strictly bounded to places associated to their own socioeconomic class. On the contrary, if r_n takes smaller than zero values (in extremity $r_n = -1$), it indicates dis-assortative connections between people and places of different socioeconomic status. This corresponds to mobility mixing patterns where people prefer to visit places of different SES rather than places from their own class. In case $r_n = 0$, the

normalised stratification matrix is flat indicating no choice preferences of people to visit places with particular SES.

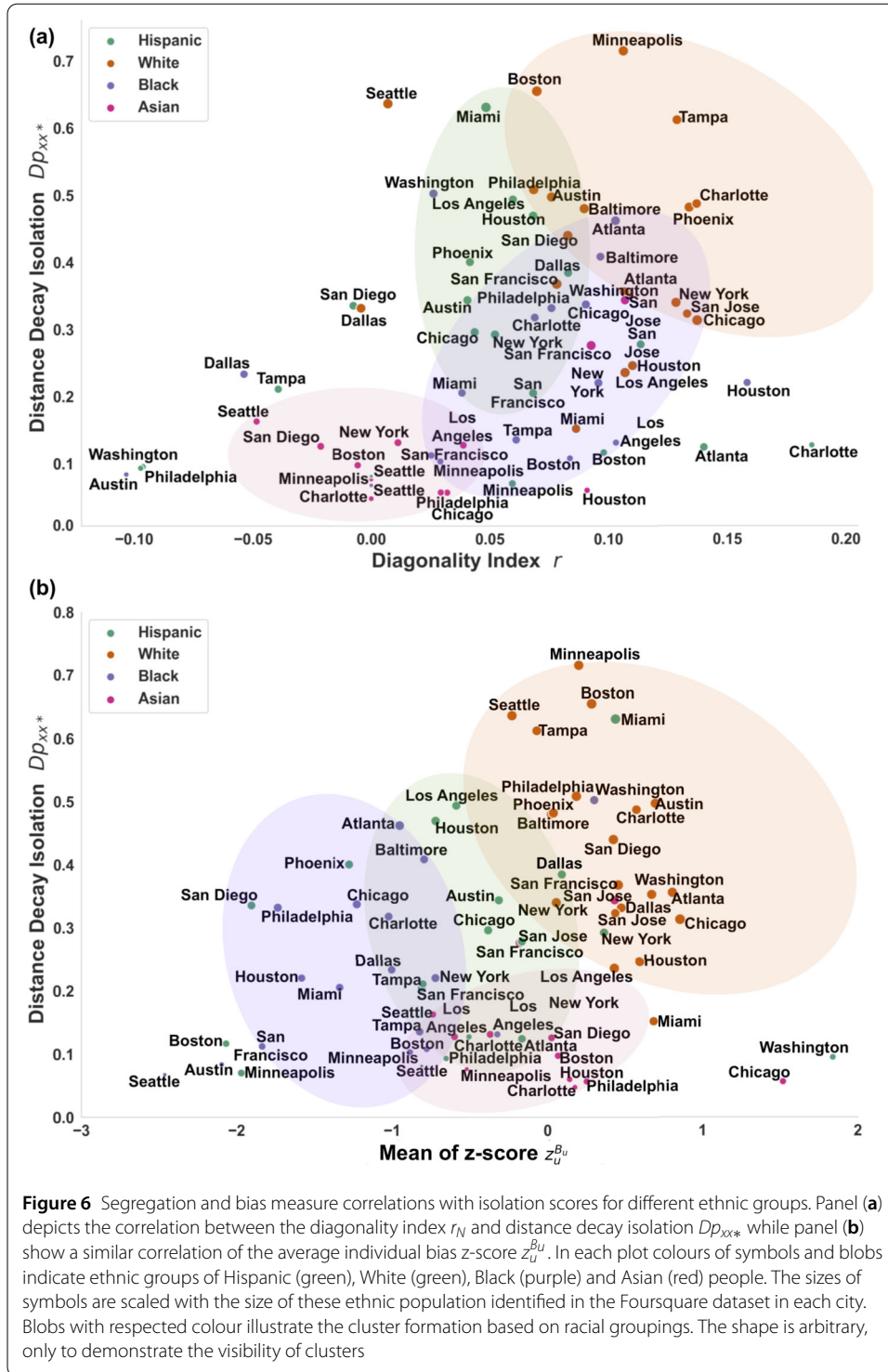
The mixing patterns in a city may not be only determined by the socioeconomic status of people but also by residential segregation. Residential segregation is strongly correlated with the ethnicity of people [45, 48, 49], which in turn, according to Wang and others [29], is an even stronger predictor of mobility mixing than socioeconomic status when it turns to black, Hispanic, and white poor and non-poor populations. This study find that the minority groups – despite their socioeconomic status – have lower exposure to richer or white neighbourhoods, comparing to poor white groups. The fact that they travel across similar distance and frequency to many places, does not change the persistent pattern of their isolation and segregation. Therefore, racial segregation emerges from a higher-order level, not limited to their residential neighbourhood but expanded to their mobility and potential contact.

To address the effects of residential segregation on mobility mixing, we took a similar path than others [29, 50] and considered the ethnic group distribution in a city as a proxy. Residential segregation is indicated by housing clustering tendency of individuals from the same ethnic group. This can be formally quantified by the so-called distance decay isolation [51], which measures the probability that a racial group minority interacts with members of their own group by considering the distance from the racial group minority's housing area. This is measured as:

$$Dp_{xx*} = \sum_{i=1}^n \left(\frac{x_i}{X} \sum_{j=1}^n \frac{k_{ij} x_j}{t_j} \right), \quad (7)$$

where x_i and x_j are the population sizes of a minority group in census tracts i and j (respectively), $X = \sum_i x_i$ is the total population of the minority group, and t_j is the total population of census tract j . The distance decay dimension is reflected by $k_{ij} = \frac{t_j^{-d_{ij}}}{\sum_{j=1}^n t_j^{-d_{ij}}}$, where d_{ij} is the distance between the centroids of census tracts i and j . Hence, higher index suggests higher probability of interaction with people from the same group, inferring isolation from the rest of population. In our case, we use a probabilistic individual profiling to identify the most likely socioeconomic profile of an individual based on the ethnic group with the highest proportion at the respected census tract where one lives. For instance, if an individual u lives at census tract i where the racial composition there is 60% white, 15% Hispanic, 10% black, and 5% Asian, this individual is considered as white. We consider different thresholds at first and we find out that considering a neighbourhood the ethnicity if such people consist of at least the 30% of the given tract is the optimum cut-off because it is the highest threshold with the lowest unidentified census tract ethnicity profiles.

Recalling the above mentioned diagonality index and individual bias, we take the average of z-score of each of these bias measures at the level of ethnic groups in every urban area and correlate them with their distance decay isolation value computed for the same ethnic group in the same city. By considering four ethnic groups (White, Hispanic, Black and Asian) we receive four data points for each cities as shown in Fig. 6. Although the total number of analysed individuals are not proportional to the total population of each city, the in-city fraction of different ethnic groups are similar to the census distributions.



There is a striking correlation emerging between the diagonality index (quantifying assortativity mobility mixing of each ethnic groups) and the distance decay isolation (measuring the isolation of different ethnic groups) with $R = 0.35$ ($p = 0.0$). Notably, almost all diagonality index measures appear with positive values suggesting assortative mixing for most ethnic groups, with a few exceptions. Further, the overall correlation suggests the in-

tuitive picture that the stronger mobility mixing stratification patterns characterises a city (i.e. larger its corresponding diagonality index), the stronger isolation patterns emerge between its ethnic groups. In turn, it indicates that residential segregation (and thus physical proximity) play an important role in determining visiting and mixing patterns of people in a city. More interestingly, the de-coupled ethnic groups for each city show an emerging clustering, which assigns the importance of racial differences in mobility segregation. From Fig. 6(a) it appears that people belonging to the white ethnic group (shown as orange points in Fig. 6(a)) appear to be the most isolated from the rest of the population (with the largest values of Dp_{xx^*}), while they appear with the strongest assortative mobility segregation patterns too (with the highest diagonality indices) consistently in several cities (to lead the eye we coloured this group as an orange blob in Fig. 6(a)). The contrary is true for the members of the Asian ethnic groups (indicated by red points and blob in Fig. 6(a)). In most cities they appear as the least isolated and the most dis-assortative (least stratified) ethnic group, thus mixing well with the rest of the population. In between these two groups, people from the Hispanic ethnic group (green points and blob) seem to be more isolated than people from the black ethnic group (purple points and blob) although they show comparable strength of segregation in mobility mixing, all weaker than white people.

Needless to say that the grouping patterns shown in Fig. 6(a) indicate overall trends only, while several exception exists for each ethnic group. For example, the Hispanic ethnic group of Charlotte appears with the strongest assortative pattern, although this group is not strongly isolated from the rest of the population. Or the black community of Austin appears with the lowest diagonality index, suggesting a strong dis-assortative mixing of these people with the rest of the population, while they also appear as one of the least isolated among any other communities.

A similar positive correlation appears in Fig. 6(b) with $R = 0.25$ ($p = 0.04$) between the average individual bias z-score and distance decay isolation values over all the investigated ethnic groups and cities. Moreover, ethnic groups show certain clustering trends, which suggest ethnic trends in terms of visiting bias patterns. Interestingly, white ethnic groups (shown by orange points and blob), who we have already found the most isolated, show the strongest upward bias to visit more affluent places than their own socioeconomic class. As high SES classes are populated mostly by white people, this pattern derives from our earlier observations in Fig. 3, where we find upward bias to increase with the SES of people. Their high isolation score can be explained by their upward bias towards higher socioeconomic places, which are most likely to be visited by other white people. Meanwhile, it may also indicate that our data have an over-represented white population, as we find upward biases in all of the cities. Strikingly, other racial groups indicate negative visiting biases and lower level of isolation. This effect is the strongest for people from black racial groups all over the country, but also characterises Hispanic and Asian communities although they show more unbiased patterns, with average individual z-score values closer to 0.

Exceptions are again interesting. The Hispanic community of Washington appears as the most upward biased ethnic group, while the black ethnic group of Seattle sits on the other end of the spectrum and being the most downward biased minority among the analysed cities. Both of these communities appear with low level of isolation. Consequently, similar to the conclusion of Wang et al. [29], we observe that beyond socioeconomic sta-

tus, ethnicity (strongly correlated with residential segregation) is another very important factor determining mixing patterns of people.

4 Discussion and conclusions

Mobility patterns are strongly determined not only by the fabric of a city but also by the socioeconomic structure of the population living there. This leads to biased mixing and segregation in mobility, which can be observed as stratification patterns in choices to visit places. We have addressed this complex phenomenon via a mobility analysis of people living in the 20 largest cities in the US, and aimed to quantify segregation patterns in mobility capturing their visit patterns to places of interest. We systematically found upward-biased mobility in all cities, with some variance across metropolitan areas. In one extreme, people living in New York do not exhibit dramatic stratification in their visit patterns but visit places in all kinds of locations, rich or poor, independently of their own socioeconomic status. Meanwhile, in Houston and San Diego people are more stratified and visit places of their own socioeconomic class, and show an upward bias towards richer places to visit. We found that this upward bias, which characterises most cities analysed, is usually induced by single visits of individuals to affluent places, while most visits correspond to their own socioeconomic class. We also revealed distinct patterns of individual mobility in terms of stratified correlations between the bias magnitude and residential segregation based on spatial distribution of racial groups in urban areas. Visual representations of ethnic clusters indicate overall trends of behaviour characterising most studied cities, where segregated mobility is bounded together with residential segregation and broadly contributes to the portrayal of inequality.

It should be taken into account that data for a given socioeconomic class in the population might not be comparable across cities due to sampling in the data collection process. Particularly, Foursquare data over-represents wealthy classes when compared to the underlying population. To understand better the fluctuations of the distribution of SES due to the representativeness of the used dataset, we designed a bootstrapping method (see Additional file 1, Section A). Bootstrapping results suggest that the SES distribution of Foursquare users is sufficiently similar to the SES distribution of the real population.

Multiple sources of data containing digital traces of human movements with higher resolution, such as mobile phone call records and GPS trajectories, may improve the robustness of findings presented in this paper. Methodological improvements to infer individual attributes (like racial group membership) provide a direction of future research. Moreover, algorithms for probabilistic individual profiling could be developed by using machine learning techniques such as Random Forests and Support Vector Machines in the presence of ground truth information from alternative data sources.

One potential confounding factor of the emergent stratification patterns reported here is distance, as people visit places closer to their home more frequently, thus inducing similar correlation motifs. To check the robustness of our methods in investigating segregation in mobility and biased visiting patterns and the magnitude of such distance effect, we re-computed our results on out-of-class data after excluding own census tract visits for each individual's trajectory. Even with this constraint, SES plays a considerable role in shaping mobility (comparative observations for all cities can be found in Additional file 1, Section B and D, along with Additional file 1, Section F Fig. 15 in the case of Houston, New York, and San Diego, in contrast to Fig. 2 above). On the ground of visiting biases, there are

some variations among cities regarding individual bias. The earlier notion of upward visiting bias is also very much present in the case of out-of-class measurements since z-score values are all positive above the red median unbiased line (complete plots are available in Additional file 1, Section C and E, while a deeper exploration for Houston, New York, and San Diego is available in Additional file 1, Section F Fig. 16 and Fig. 17). Therefore, there are no conflicting results from our methodology even after controlling for this confounding factor. Enforcing out-of-class treatment is reasonable in this context because our study aims to analyse and quantify mixing patterns and not yet look for causal links or underlying reasons of their emergence.

Segregation is not an exclusive phenomenon to the quasi-static configuration of housing settlement, but also exists in more dynamic settings such as mobility. Questions about the conceptual relations between segregated mobility and segregated residence stand still in the literature, yet relatively untapped, while scientific investigations should follow this line of inquiry. We take a step forward through empirical data-driven analysis and yield an interaction effect between both types of segregation. Individual attributes (such as racial groups) partly explain the emergence of distinct clusters, beyond income levels. Our findings also highlight the notion that inequality is multidimensional in nature. A comprehensive policy design to address this issue should entail the wider possibility of individual movement across the urban landscape to accommodate larger socioeconomic heterophily and further interaction between socioeconomic classes.

Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1140/epjds/s13688-022-00341-x>.

Additional file 1. Supplementary information (PDF 16.0 MB)

Acknowledgements

GI acknowledges support from AFOSR (Grant No. FA8655-20-1-7020), project EU H2020 Humane AI-net (Grant No. 952026), and CIVICA project 'European Polarisation Observatory' (EPO). MK was supported by the DataRedux ANR project (ANR-19-CE46-0008), the SoBigData++ H2020 project (H2020-871042), and the CIVICA EmoMap project. MK and GI received support from the CHIST-ERA-19-XAI-010 project SAI (grant FWF I 5205-N).

Funding

Open access funding provided by ELKH Alfréd Rényi Institute of Mathematics.

Availability of data and materials

The datasets generated and/or analysed during the current study are available in the Foursquare Dataset repository, <https://sites.google.com/site/yangdingqi/home/foursquare-dataset>.

Declarations

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

RMH, GI and MK designed the research. RMH collected and analysed the data. RMH, GI and MK wrote the manuscript. All authors read and approved the final manuscript.

Author details

¹Department of Network and Data Science, Central European University, 1100 Vienna, Austria. ²Department of Computer Science, Aalto University School of Science, 00076 Aalto, Finland. ³Centro de Ciencias de la Complejidad, Universidad Nacional Autónoma de México, 04510 Ciudad de México, Mexico. ⁴Alfréd Rényi Institute of Mathematics, 1053 Budapest, Hungary.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 14 September 2021 Accepted: 25 April 2022 Published online: 23 May 2022

References

1. Rumberger RW, Palardy GJ (2005) Does segregation still matter? The impact of student composition on academic achievement in high school. *Teach Coll Rec* 107(9):1999–2045
2. Acevedo-Garcia D, Lochner KA (2003) Residential segregation and health. In: *Neighbourhoods and health*. Oxford University Press, London, pp 265–287
3. Taeuber KE, Taeuber AF (2008) Residential segregation and neighborhood change. Transaction Pub., London
4. Iceland J, Weinberg DH, Steinmetz E (2002) Racial and ethnic residential segregation in the United States 1980–2000, vol 8. Bureau of Census, Washington
5. Dong X, Morales AJ, Jahani E, Moro E, Lepri B, Bozkaya B, Sarraute C, Bar-Yam Y, Pentland A (2020) Segregated interactions in urban and online space. *EPJ Data Sci* 9(1):20
6. Moro E, Pentland A, Calacci D, Dong X (2019) Atlas of Inequality. <https://inequality.media.mit.edu>. Accessed 28 Jun 2021
7. Netto VM, Soares MP, Paschoalino R (2015) Segregated networks in the city. *Int J Urban Reg Res* 39(6):1084–1102
8. Gonzalez MC, Hidalgo CA, Barabasi A-L (2008) Understanding individual human mobility patterns. *Nature* 453(7196):779–782
9. Song C, Koren T, Wang P, Barabási A-L (2010) Modelling the scaling properties of human mobility. *Nat Phys* 6(10):818–823
10. Tang J, Liu F, Wang Y, Wang H (2015) Uncovering urban human mobility from large scale taxi gps data. *Phys A, Stat Mech Appl* 438:140–153
11. Gallotti R, Bazzani A, Rambaldi S, Barthelemy M (2016) A stochastic model of randomly accelerated walkers for human mobility. *Nat Commun* 7(1):1–7
12. Alessandretti L, Sapiezynski P, Sekara V, Lehmann S, Baronchelli A (2018) Evidence for a conserved quantity in human mobility. *Nat Hum Behav* 2(7):485–491
13. Hawelka B, Sitko I, Beinart E, Sobolevsky S, Kazakopoulos P, Ratti C (2014) Geo-located Twitter as proxy for global mobility patterns. *Cartogr Geogr Inf Sci* 41(3):260–271
14. Wu L, Zhi Y, Sui Z, Liu Y (2014) Intra-urban human mobility and activity transition: evidence from social media check-in data. *PLoS ONE* 9(5):97010
15. Jurdak R, Zhao K, Liu J, AbouJaoude M, Cameron M, Newth D (2015) Understanding human mobility from Twitter. *PLoS ONE* 10(7):0131469
16. Brockmann D, Hufnagel L, Geisel T (2006) The scaling laws of human travel. *Nature* 439(7075):462–465
17. Baronchelli A, Radicchi F (2013) Levy flights in human behavior and cognition. *Chaos Solitons Fractals* 56:101–105
18. Wang X-W, Han X-P, Wang B-H (2014) Correlations and scaling laws in human mobility. *PLoS ONE* 9(1):84954
19. Alessandretti L, Aslak U, Lehmann S (2020) The scales of human mobility. *Nature* 587(7834):402–407
20. Beiró MG, Panisson A, Tizzoni M, Cattuto C (2016) Predicting human mobility through the assimilation of social media traces into mobility models. *EPJ Data Sci* 5:1
21. Comito C (2018) Human mobility prediction through Twitter. *Proc Comput Sci* 134:129–136
22. Luo F, Cao G, Mulligan K, Li X (2016) Explore spatiotemporal and demographic characteristics of human mobility via Twitter: a case study of Chicago. *Appl Geogr* 70:11–25
23. Leo Y, Fleury E, Alvarez-Hamelin JI, Sarraute C, Karsai M (2016) Socioeconomic correlations and stratification in social-communication networks. *J R Soc Interface* 13(125):20160598
24. Marston SA (2000) The social construction of scale. *Prog Hum Geogr* 24(2):219–242
25. Paasi A (2004) Place and region: looking through the prism of scale. *Prog Hum Geogr* 28(4):536–546
26. Boterman WR, Musterd S (2016) Cocooning urban life: exposure to diversity in neighbourhoods, workplaces and transport. *Cities* 59:139–147
27. Bora N, Chang Y-H, Maheswaran R (2014) Mobility patterns and user dynamics in racially segregated geographies of us cities. In: *International conference on social computing, behavioral-cultural modeling, and prediction*. Springer, Berlin, pp 11–18
28. Yip NM, Forrest R, Xian S (2016) Exploring segregation and mobilities: application of an activity tracking app on mobile phone. *Cities* 59:156–163
29. Wang Q, Phillips NE, Small ML, Sampson RJ (2018) Urban mobility and neighborhood isolation in America's 50 largest cities. *Proc Natl Acad Sci* 115(30):7735–7740
30. Morales AJ, Dong X, Bar-Yam Y, 'Sandy' Pentland A (2019) Segregation and polarization in urban areas. *R Soc Open Sci* 6(10):190573
31. McPherson M, Smith-Lovin L, Cook JM (2001) Birds of a feather: homophily in social networks. *Annu Rev Sociol* 27(1):415–444
32. Moro E, Calacci D, Dong X, Pentland A (2021) Mobility patterns are associated with experienced income segregation in large us cities. *Nat Commun* 12(1):1–10
33. Desu S (2015) Untangling the effects of residential segregation on individual mobility. PhD thesis, Massachusetts Institute of Technology
34. Browning CR, Calder CA, Krivo LJ, Smith AL, Boettner B (2017) Socioeconomic segregation of activity spaces in urban neighborhoods: does shared residence mean shared routines? *Russell Sage Found J Soc Sci* 3(2):210–231
35. Schönfelder S, Axhausen KW (2003) Activity spaces: measures of social exclusion? *Transp Policy* 10(4):273–286
36. Wong DW, Shaw S-L (2011) Measuring segregation: an activity space approach. *J Geogr Syst* 13(2):127–145
37. Farber S, Neutens T, Miller HJ, Li X (2013) The social interaction potential of metropolitan regions: a time-geographic measurement approach using joint accessibility. *Ann Assoc Am Geogr* 103(3):483–504
38. Farber S, O'Kelly M, Miller HJ, Neutens T (2015) Measuring segregation using patterns of daily travel behavior: a social interaction based model of exposure. *J Transp Geogr* 49:26–38

39. Yang D, Zhang D, Qu B (2016) Participatory cultural mapping based on collective behavior data in location-based social networks. *ACM Trans Intell Syst Technol* 7(3):1–23
40. McNeill G, Bright J, Hale SA (2017) Estimating local commuting patterns from geolocated Twitter data. *EPJ Data Sci* 6:1
41. Bureau USC (2012) American Community Survey
42. Kraskov A, Stögbauer H, Grassberger P (2004) Estimating mutual information. *Phys Rev E* 69(6):066138
43. Hastie T, Tibshirani R, Friedman J (2009) *The elements of statistical learning*, vol 2. Springer, New York
44. Raileanu LE, Stoffel K (2004) Theoretical comparison between the Gini index and information gain criteria. *Ann Math Artif Intell* 41(1):77–93
45. Abramovitz M, Smith RJ (2021) The persistence of residential segregation by race, 1940 to 2010: the role of federal housing policy. *Fam Soc* 102(1):5–32
46. Bokányi E, Juhász S, Karsai M, Lengyel B (2021) Universal role of commuting in the reduction of social assortativity in cities. arXiv preprint. [2105.01464](https://arxiv.org/abs/2105.01464)
47. Newman ME (2003) Mixing patterns in networks. *Phys Rev E* 67(2):026126
48. Massey DS, Denton NA (1988) The dimensions of residential segregation. *Soc Forces* 67(2):281–315
49. Logan JR, Stults BJ (2011) The persistence of segregation in the metropolis: new findings from the 2010 census. Census Brief Prepared for Project US2010 24
50. Krivo LJ, Washington HM, Peterson RD, Browning CR, Calder CA, Kwan M-P (2013) Social isolation of disadvantage and advantage: the reproduction of inequality in urban space. *Soc Forces* 92(1):141–164
51. Morgan BS (1983) A distance-decay based interaction index to measure residential segregation. *Area* 15(3):211–217

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Open access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

Submit your next manuscript at ▶ [springeropen.com](https://www.springeropen.com)
