Interaction networks reveal that large, dense cities facilitate segregation

Hamed Nilforoshan[*,1], Wenli Looi[*,1], Emma Pierson[*,1], Blanca Villanueva[4], Nic Fishman[1], Yiling Chen[1], John Sholar[1], Beth Redbird[2,3], David Grusky[2,3,4,5,6], Jure Leskovec[†,1,7]

[1] Department of Computer Science, Stanford University, Stanford, CA 94305, USA

[2] Department of Sociology, Northwestern University, Evanston, IL 60208, USA

[3] Institute for Policy Research, Northwestern University, Evanston, IL 60208, USA

[4] Department of Biomedical Data Science, Stanford University, Stanford, CA 94305, USA

[*] These authors contributed equally to this work

[†] Corresponding author. Email: jure@cs.stanford.edu

A long-standing hope and expectation is that large, dense, cosmopolitan areas support diverse interactions and socioeconomic mixing [1,2]. However, it has been difficult to assess this hypothesis because standard approaches to measuring segregation rely on static data about where people live [3,4], not whom they actually meet at work, in places of leisure, and in home neighborhoods. Here we develop a new measure of *Interaction Segregation*, which captures the diversity of the set of people that a given person actually comes into contact with in their day to day lives. Leveraging cell-phone mobility data, Interaction Segregation captures 1.6 billion interactions between 9.6 million people in the United States, measuring segregation across 382 Metropolitan Statistical Areas (MSAs) and 2829 counties. The results reveal that, contrary to the cosmopolitan hypothesis, large, dense, cities are associated with high segregation: Interaction Segregation in the 10 largest Metropolitan Statistical Areas (MSAs) is 67% higher than in small MSAs with fewer than 100,000 residents. We explain our discovery by the observation that large cities offer a greater choice of highly socioeconomically differentiated spaces, at multiple levels of scale (e.g. from leisure sites to neighborhoods), which facilitate homophilous behavior, producing higher levels of Interaction Segregation. Additionally, we show that aspects of urban design, such as commercial centers (e.g. shopping malls) which are accessible to diverse individuals, may mitigate Interaction Segregation in large cities. Our findings have implications for human geography and urban design, highlighting the role of built environment in mediating diverse human interactions.

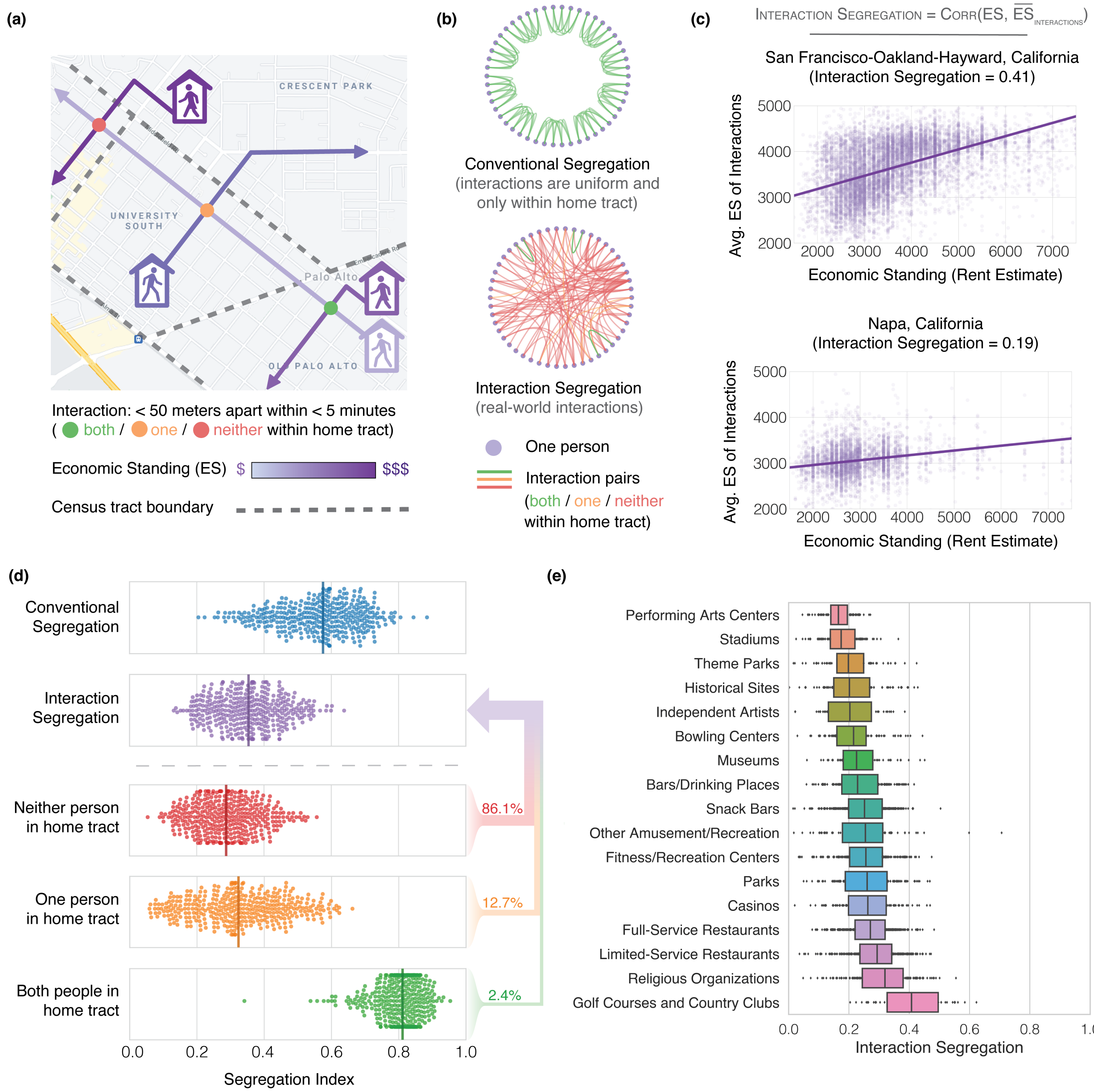


Figure 1: A more realistic measure of economic segregation. (a-c) Illustration of our analytical pipeline **(a)** For all 9.6 million individuals, we infer economic standing (estimated rent) from home address. We then leverage anonymized cell phone mobility data to identify *interactions* between pairs of individuals. **(b)** We construct a nation-wide interaction network of 1.6 billion interactions across 2829 counties and 382 Metropolitan Statistical Areas (MSAs). Our interaction network contrasts the conventional measure of economic segregation (Neighborhood Sorting Index) which assumes that people interact uniformly and only with other residents of their home Census tract. We show a real example of a community of 50 individuals from San Francisco, CA residing in 10 different census tracts, whose cross-tract interactions are undetected by conventional segregation measures. Nodes are individuals; edges are interactions. **(c)** For each geographic region (e.g. MSA, county) we apply a statistical model to the interaction network to estimate *Interaction Segregation*: the correlation between an individual's economic standing (ES) and the mean ES of those they interact with. 1 signifies perfect segregation; 0 signifies no segregation. This definition is equivalent to Neighborhood Sorting Index (NSI), with the key difference that it leverages real-life interactions from mobility data, instead of synthetic interactions from individuals grouped by Census tracts. For two MSAs, we show the inputs and fit of our model; each point represents one individual; San Francisco-Oakland-Hayward, CA is 2.2× more segregated than Napa, CA. **(d)** Top: Interaction Segregation is 38% lower than the conventional segregation measure NSI. Each point represents the Interaction Segregation estimate in one MSA; vertical colored lines represent median across MSAs. Bottom: breaking down Interaction Segregation into component parts: interactions where both people are within their home Census tract (green) are 2.7× more segregated than interactions where at least one person is outside their home Census tract (orange and red) reflecting the *homophily effect*, in which people preferentially interact with those of similar ES. Out-of-tract interactions (orange and red) are less segregated, reflecting the *visitor effect*, in which visiting other tracts exposes individual's to economically diverse individuals. Because the vast majority (97.8%) of interactions pairs are out-of tract, the visitor effect dominates the homophily effect and Interaction Segregation is lower than conventional measures. **(e)** Interaction Segregation varies by leisure site. Each point represents segregation in one MSA using only interactions occurring in a given leisure site; boxes indicate the interquartile range across MSAs. Segregation is highest at golf courses and religious organizations, and lowest at performing arts centers and stadiums.

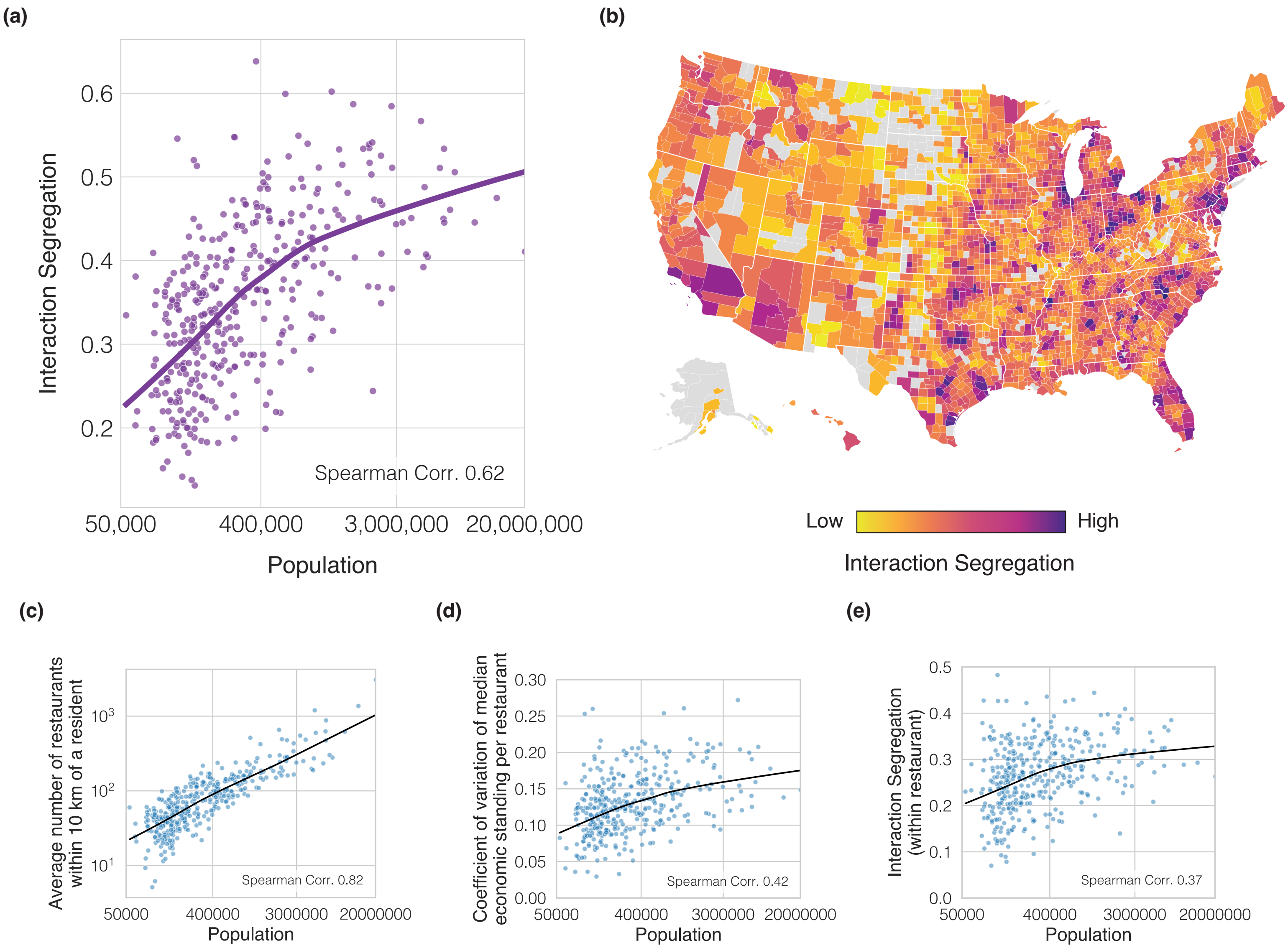


Figure 2: Large, dense, metropolitan areas are more segregated (a-b) Contrary to conventional wisdom that large, dense, metropolitan areas support diverse interactions and socioeconomic mixing, we find that: **(a)** Larger MSAs are more segregated (Spearman Correlation 0.62). The top 10 largest MSAs, by population size, are 67% more segregated than small MSAs with less than 100,000 residents. We show a scatterplot of the relationship between population size and Interaction Segregation across all 382 MSAs in the United States; each point represents one MSA; purple line indicates LOWESS fit; vertical axis plots Interaction Segregation; horizontal axis plots MSA population logarithmically due to the large range of population sizes. Population density is similarly positively associated with Interaction Segregation (Spearman Correlation 0.46). **(b)** To visualize this phenomenon, we show a choropleth of Interaction Segregation across the 2829 USA counties with at least 50 individuals represented within our dataset. We observe that Interaction Segregation varies significantly across counties in the United States. Moreover, similar to at the MSA-level, at the county-level Interaction Segregation is also associated with both large population size (Spearman Correlation 0.45) and high population density (Spearman Correlation 0.45). **(c-e)** We use full-service restaurants as a case study to understand the mechanism for why large, densely populated geographic areas are more segregated. We find that large, densely populated metropolitan areas are associated with increased options and economic differentiation of venues (e.g. in New York City, one can spend $10, $100, or $500 on a sushi meal, depending on the choice of sushi restaurant), which may lead to higher self-segregation. We highlight restaurants because they are the leisure site with the largest number of interactions; results for other leisure sites are similar and are reported in Figure S20. **(c)** Larger MSAs have more restaurants within 10 kilometers of the average resident, giving residents more options to self-segregate. **(d)** Moreover, restaurants in larger MSAs vary more in terms of the median ES of their visitors, offering a greater choice of highly socioeconomically differentiated restaurants. The coefficient of variation across restaurant ES in 10 largest MSAs is 63% more than the coefficient of variation in small MSAs with fewer than 100,000 residents). **(e)** Consequently, overall interaction segregation at restaurants is higher in larger MSAs. Data is shown for restaurants with at least 10 people who interact in the restaurant; black lines indicate the LOWESS fit.

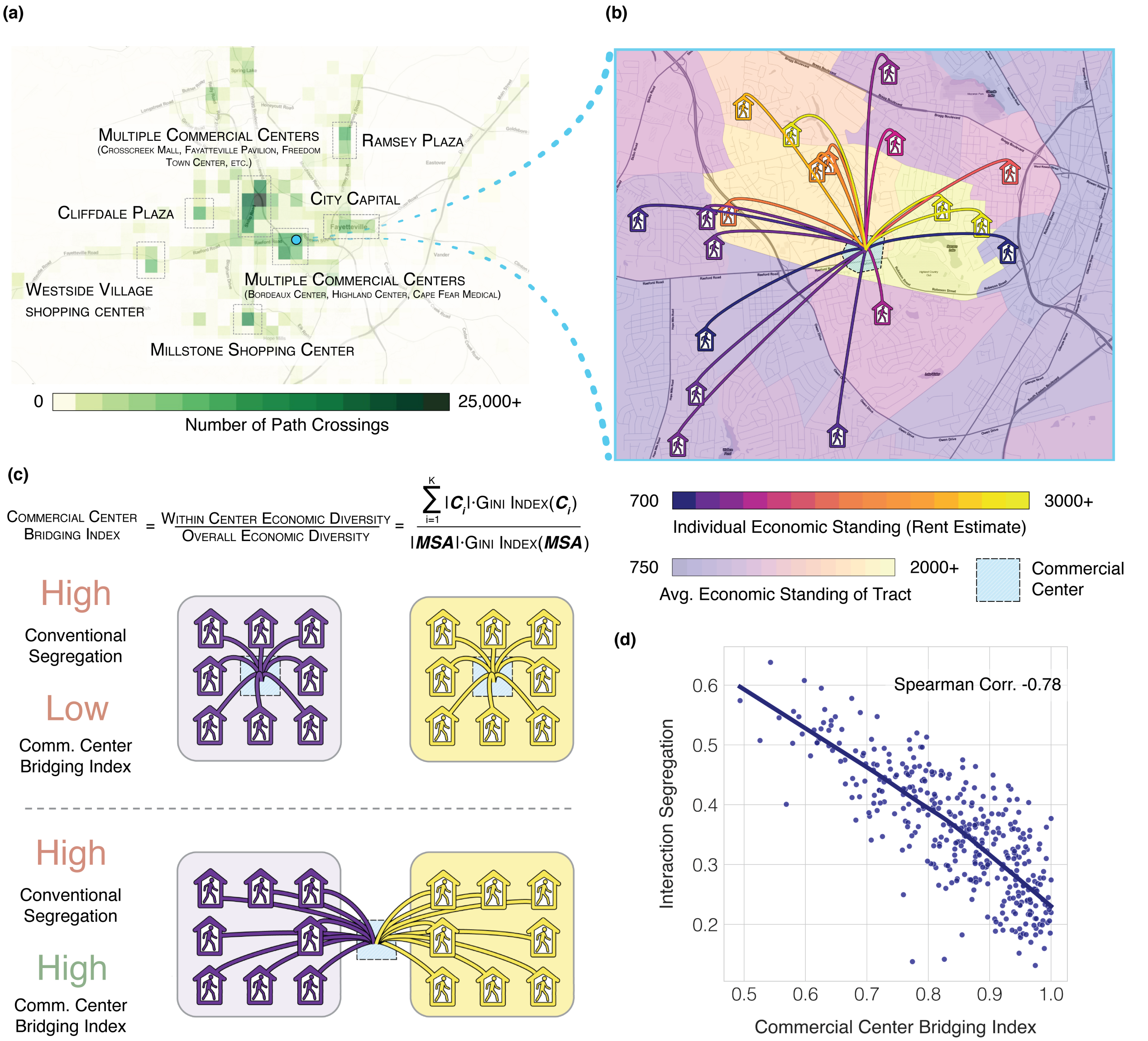


Figure 3: Aspects of urban design, such as accessibility of commercial centers (e.g. shopping malls, plazas, etc.) to individuals of diverse economic standings, may mitigate Interaction Segregation (a-b) A case study of Fayetteville, NC, an MSA with low (21st percentile) Interaction Segregation despite having above-median population size (64th percentile) and income inequality (60th percentile). **(a)** Commercial centers (e.g. shopping malls, plazas, etc.) are associated with high density of interactions. We illustrate this phenomenon in Fayetteville, NC, visualizing a heat map of the number interactions across the MSA. All visually discernible hotspots are associated with one or more commercial centers. More generally, across all 382 MSAs, we find that the majority (56.9%) of interactions occur in close proximity (within 1km) of a commercial center, even though only 2.5% of land area of all MSAs is within 1km of a commercial center. **(b)** In Fayetteville, NC, commercial centers are located in accessible proximity to both high and low ES census tracts, leading to diverse interactions. As an illustrative example, we show a zoomed-in map of one commercial center (Highland Center) in Fayetteville, NC, and display a random sample of 10 interactions occurring inside of it. Home icons demarcate individual home location (up to 100 meters of random noise added to preserve anonymity); home colors denote individual ES; arcs indicate an interaction inside of the commercial center; background colors indicate mean census tract ES. **(c)** We aim to build a generalized measure of the *Commercial Center Bridging* phenomenon observed in Fayetteville, NC: the extent to which commercial center locations encourage interactions between diverse individuals. We propose the *Commercial Center Bridging Index* (CCBI), a simple metric constructed by clustering homes by nearest commercial center, then measuring the within-cluster diversity of ES (Methods M3). We show two examples to illustrate that CCBI is distinct from conventional residential measures of segregation (i.e., Neighborhood Sorting Index). In both figures, residential segregation is constant: high and low-ES individuals are highly segregated by neighborhood (denoted by purple and yellow bounding box). Top: When there is a commercial center at the center of each census tract, and tracts are perfectly residentially segregated, CCBI is 0, denoting no bridging. Bottom: When there is a commercial center (blue box) in between both neighborhoods, then CCBI is 1, denoting perfect bridging. **(d)** Commercial Center Bridging Index strongly predicts Interaction Segregation (Spearman Correlation −0.78). The top 10 MSAs with the highest CCBI are 53.1% less segregated than the 10 MSAs with lowest CCBI. CCBI predicts segregation more accurately than population size, ES inequality, NSI, and racial demographics, and is significantly associated with segregation ($p < 0.01$) after controlling for all aforementioned variables (Extended Data Table 2).

Introduction

Segregation reduces economic mobility[5–7], harms mental and physical health[8,9], and increases political polarization[10]. Consequently, understanding how to mitigate segregation remains an active area of research[11]. One commonly touted mitigator of segregation is large, dense, cosmopolitan cities[1,2] ; a long-standing hope and expectation among urbanists is that cities facilitate diverse social interactions by constraining space and pushing diverse individuals into contact with each other (e.g., the New York City subway is often touted as a mixing bowl of diverse individuals[12]). In a continuously urbanizing world, with 68% of the world's population expected to be living in urban areas by 2050[13], the validity of the "cosmopolitan city" hypothesis becomes increasingly relevant to understanding the future of segregation.

This long-standing hypothesis is challenging to assess empirically because it is difficult to measure real-world interaction between individuals. Thus, standard measures of segregation rely on static data about where people live as a proxy for interaction[3,4,14,15]. However, grouping individuals by residential neighborhood is a weak proxy for estimating diversity of interactions as it ignores the time that individuals spend outside of home, meeting and interacting with others during work and leisure. To fully understand the relationship between urbanization and segregation, it is necessary to develop a more realistic measure of segregation which accurately captures where people go, when they go there, and whom they come into contact with.

To address these needs, we leverage anonymized cell phone geolocation data to construct a fine-grained, dynamic network which captures 1.6 billion interactions between 9.6 million people in the United States. We then propose *Interaction Segregation*, a new measure of economic segregation which extends a traditional measure of segregation[3] to accommodate an interaction network in a given geographic area. We estimate Interaction Segregation in 382 MSAs and 2829 counties to conduct a nationwide study of the relationships between urbanization, built environment, and economic segregation. Our measure of segregation builds on prior research that surveys different sub-populations and compares the areas they visit [16,17]; separately, other studies have analysed mobility data to estimate segregation of different categories of POIs (e.g. restaurants vs. hotels), focusing on a small subset of cities[18], as well to contrast the experiences of coarse groups of people[19] (e.g. all people from majority non-White zip codes).

Counterintuitively, we find that large, dense cities have higher levels of Interaction Segregation. The 10 largest MSAs by population size are 67% more segregated than small MSAs with less than 100,000 residents. We explain our discovery by observing that large cities offer a greater choice of highly socioeconomically differentiated spaces, at multiple levels of scale (e.g. from leisure sites to neighborhoods), which facilitate homophilous behavior, producing higher levels of Interaction Segregation. Additionally, we show that aspects of urban design, such as commercial centers (e.g. shopping malls) which are accessible to diverse individuals, may mitigate Interaction Segregation in large cities. The top 10 MSAs according to our metric of commercial center accessibility are 53.1% less segregated than the 10 MSAs with lowest commercial center accessibility. Our findings have implications for human geography and urban design, highlighting the role of built environment in encouraging diverse human interactions.

Results

Developing a more realistic measure of socioeconomic segregation

We develop a more realistic measure of socioeconomic segregation, which we term *Interaction Segregation*. We leverage data from SafeGraph, a company which provided anonymized data tracking the movements of 9,567,559 people in the United States over the course of three months in 2017 (Methods). Our metric aims to estimate, for each geographic area (i.e. Metropolitan Statistical Area or county), the extent to which residents come into contact with diverse individuals of different economic status. We use location data, consisting of the latitude and longitude of an individual at a specific point in time, to construct a dynamic *interaction network*, with 9,567,559 nodes (representing individuals) and 1,570,782,460 edges (representing interactions), across 382 MSAs and 2829 counties in the United States. Specifically, we connect two users with an interaction edge if the GPS pings from both two individuals indicate the possibility of having interacted because they were within 50 meters of each other within less than 5 minutes. Figure 1 illustrates our analytical pipeline and the resulting interaction network .

To estimate each person's economic standing (ES) , we first infer their home location from their cell phone mobility data, and use the estimated monthly rent value of this home location (Methods). This method provides us with a more fine-grained estimate of individual ES than previous measures which use aggregate Census data [3,18,19]. To measure the economic segregation of each geographic region (e.g. MSA, county), we filter the network for all individuals inferred to live within that area, as well as those they are connected to. We then apply a linear mixed effects model to the interaction network to estimate the correlation between a person's ES and the mean ES of everyone they interact with (Methods). Thus, our measure Interaction Segregation ranges from 0 (perfect integration) to 1 (complete segregation), with higher values indicating higher levels of socioeconomic segregation. Our model can be interpreted as the extent to which an individual's ES predicts the ES of those they interact with. Figure 1c shows the model segregation estimates for two MSAs in California: Napa and San Francisco-Oakland-Hayward.

Interaction Segregation is derived from extending a classic definition residential segregation, the Neighborhood Sorting Index [3] (NSI), which is equivalent to the correlation across people between each person's ES and the mean ES in their Census tract. The NSI is an intuitive definition of segregation and can be calculated from Census data on the ES of people living in each tract. A disadvantage is that it only incorporates information about the Census tract where people *live*[4], an artificial boundary which fails to capture interactions at work and at leisure. We observe that NSI is equivalent to the correlation between each person's ES and the mean ES of everyone they interact with, under the unrealistic assumptions that a) people only interact with those in their home Census tract and b) they do so uniformly at random. These assumptions limit the real-world applicability of NSI, because people may also interact with more heterogeneous populations as they visit other Census tracts for work, leisure, or other activities, a phenomenon we refer to as the *visitor effect*. Furthermore, even within home tract, individuals may interact non-uniformly as they seek out people of similar economic standing; we refer to this as the *homophily effect*. Thus, to design a metric which captures both the *visitor effect* and *homophily effect*, Interaction Segregation leverages cellphone mobility data to capture the extent of contact between diverse individuals throughout the day, both outside and inside of individuals' home tract. Furthermore, Interaction Segregation allows for direct comparability to NSI, because both measures are of the same underlying statistical quantity, but differ in their definition of the interaction network.

Interaction Segregation is lower than conventional segregation. The median interaction segregation across all MSAs is 38% lower than conventional segregation (NSI; Figure 1d top). To explain why, we examine how interaction segregation differs when both people are within their home Census tract, as opposed to when at least one person is outside their home Census tract (Figure 1d bottom). Within-home-tract interactions are 41% more segregated than conventional segregation, illustrating the *homophily effect*: within their own neighborhoods, people tend to interact with those more similar in ES than would be predicted if they crossed paths at random. However, within-home-tract interactions constitute less than 2.5% of interactions (Figure 1d bottom) and interactions where at least one person is outside their home tract are less segregated, reflecting the *visitor effect*. Because these out-of-tract interactions are more frequent, the visitor effect dominates the homophily effect and interaction segregation is lower than conventional segregation.

Interaction Segregation varies significantly across leisure sites We observe significant variability in interaction segregation across different categories of leisure sites such as golf courses, restaurants, stadiums, etc. (NSI; Figure 1e). For instance, Interaction Segregation within golf courses and country clubs is over 2× more than Interaction Segregation within performing arts centers in the median MSA. We compute Interaction Segregation within a specific leisure site by filtering the interaction network to consist of interactions which occurred only within that category of leisure site, and re-applying our statistical model to estimate Interaction Segregation for each MSA (Methods M3). We identify two primary factors, both related to socioeconomic differentiation of POIs, which explain between-activity heterogeneities in segregation: (1) localization and (2) within-activity stratification. (1) We observe that leisure sites which which are high in number and embedded locally within residential communities (e.g. churches) are, on average, more segregated (Spearman Corr. -0.77, Extended Data Figure 2). Conversely, leisure sites which are fewer in number, such as stadiums, are more integrated because they are designed to serve an entire city (e.g. to see the NY Giants play football, individuals of all economic standing will attend MetLife stadium). (2) We find that in rare cases, POIs may be few in number but highly stratified within-activity, resulting in high segregation. For instance, golf courses and country clubs are the most segregated category of venue despite few venues. We attribute this to high stratification of golf courses and country clubs in most cities; in NYC, golf courses range from extremes of free public golf courses to high-end private golf courses with initiation fees of up to $250,000 (Extended Data Figure 2d). These results show that POIs, depending on the extent of their economic differentiation, may positively or negatively contribute to interaction segregation.

Large, dense metropolitan areas facilitate segregation

We find that Interaction Segregation is higher in large, dense MSAs (Figure 2). Specifically, we find that the Spearman correlation between MSA population and MSA segregation is 0.62, and the 10 largest MSAs by population size are 67% more segregated than small MSAs with less than

100,000 residents. Population density is similarly positively associated with Interaction Segregation (Spearman Correlation 0.46). The discovery that larger, denser MSAs and counties are more segregated runs counter to prior literature which identifies dense cities as hubs for diverse interactions[1,2]. Thus, we validate this counter-intuitive finding through extensive robustness checks including controlling for relevant co-variates (e.g. political, racial, and economic demographics, as well as aspects of built environment; Extended Data Table 1 and Supplementary Table S7), varying granularity of analysis (across 2829 USA counties; Figure 2b, Extended Data Figure 3), and a variety of specifications of Interaction Segregation (Supplementary Table S6, Supplementary Figure S2-S8). To better understand why larger, denser MSAs are more segregated, we examine the mechanisms which link population size and segregation.

Mechanisms producing higher interaction segregation in larger metropolitan areas. To understand the mechanisms which contribute to higher segregation in larger metropolitan areas, we use interaction segregation within leisure POIs as a case study, first focusing on full-service restaurants they are the leisure site with the largest number of interactions (Figure 2c,d,e); our results for other leisure sites are analogous (Figure S20). Larger MSAs offer their residents a greater choice of restaurants to visit: the average resident of the 10 largest MSAs will have 149$\times$ as many restaurants within 10 kilometers of their home as a resident of small MSAs (Supplementary Figure 2c). Restaurants in large MSAs are also more socioeconomically differentiated: defining the ES for each restaurant as the median ES of all people who cross paths there, the coefficient of variation of restaurant ES (i.e., the standard deviation in restaurant ES divided by the mean restaurant ES) in the 10 largest MSAs is 63% larger than that in small ones (Figure 2d). Thus, large MSAs offer both a larger choice of restaurants, and restaurants which are more socioeconomically differentiated. As a consequence, Interaction Segregation at restaurants is 29% higher at the 10 largest MSAs than at small MSAs (Figure 2e). Overall, this indicates that by offering their residents greater choice of socioeconomically differentiated leisure sites, large MSAs enable segregation by allowing residents to self-stratify within each activity. This finding is consistent with our earlier observation that types of leisure site which offer more differentiated options also have higher segregation: for example, the typical MSA has relatively few stadiums and performing arts centers, and these sites, in turn, are less segregated because residents have no choice but to mix there (Figure 1e, Extended Data Figure 2). We also search for evidence of alternative mechanisms that might produce higher leisure segregation, *constant homophily*: that homophily remains constant throughout cities, and that the increase in segregation is due to difference in income distribution in larger cities; and *between-activity homophily*: that differences in choices in activity drive the increase in segregation, e.g. that high-ES individuals visit golf courses/country clubs instead of restaurants more frequently in larger cities. We do not find evidence of either effect (Extended Data Figure S9).

Multi-scale economic differentiation of built environment in large metropolitan areas. We conduct an analysis analogous to segregation within fine-grained leisure POIs (e.g. restaurants) at higher levels of scale (i.e. at the granularity of commercial centers and neighborhoods). We find that the pattern of increased differentiation, leading to increase segregation, is consistent across these multiple scales. Commercial centers (e.g. plazas, malls, boardwalks) are higher-level POIs which encapsulate fine-grained POIs such as restaurants, grocery stores, etc. For commercial centers, we conduct an analogous analysis to that for restaurants (Extended Data Figure 4) and find that large, densely populated metropolitan areas are associated with increased options and economic differentiation of commercial centers (Spearman Corr. 0.58), resulting in higher segregation within commercial centers (Spearman Corr. 0.64).

Similarly, at the highest level of scale, neighborhoods, larger MSAs have more Census tracts (Extended Data Figure 5a), affording their residents a greater choices of neighborhoods to live in. Census tracts in large MSAs are also more socioeconomically differentiated (Extended Data Figure 5b), consistent with previous findings [20]: defining the ES for each Census tract as the median ES of its residents, the coefficient of variation of Census tract ES is 2.5$\times$ higher at the 10 largest MSAs than at small MSAs (Spearman correlation between MSA population and coefficient of variation of Census tract ES, 0.58). As a consequence, home tract interaction segregation is higher in larger MSAs (Spearman correlation, 0.35; Extended Data Figure 5c). We also search for evidence of a separate mechanism that might produce higher home tract segregation, *within-tract homophily*:

that is, if residents preferentially associate with people of similar socioeconomic status even *within their home tract*. We do not find evidence of this effect (Extended Data Figure 5d). Overall, these analyses indicate that, at multiple levels of scale (i.e. individual POIs, commercial centers, and neighborhoods), larger MSAs offer their residents a greater choice of more socioeconomically differentiated options, facilitating the homophily effect and producing higher levels of interaction segregation.

Mitigating segregation via urban design

The trend towards increased socioeconomic differentiation, and consequently, increased Interaction Segregation, poses a challenging question: how can Interaction Segregation be reduced in an increasingly urbanizing world? To better understand how policies can mitigate Interaction Segregation, we first conduct a case study of one city which has low segregation despite high population size. Fayetteville, North Carolina is a city with low (21st percentile) Interaction Segregation, despite having above median population size (64th percentile) as well as income inequality (60th percentile). We find that in Fayetteville, Interaction Segregation is mitigated by commercial centers (e.g. plazas, malls, shopping centers, boardwalks, etc.) which are located in between high and low ES neighborhoods, and which facilitate diverse interactions. Our case study leads us to the more general discovery that commercial center locations, in relation to the locations of low and high ES neighborhoods, strongly predicts Interaction Segregation. As commercial center locations are a modifiable extrinistic factor of urban design; these results suggest that policies which encourage the construction of commercial centers bridging diverse residential neighborhoods (e.g. zoning laws or subsidies) may mitigate Interaction Segregation.

Commercial centers are associated with high density of interactions. We observe that in Fayetteville, NC. areas with high-density of interactions are associated with commercial centers such as plazas, malls, shopping centers, boardwalks, etc. (Figure 3a). We discover that, more generally, across all 382 MSAs, the majority (56.9%) of interactions occur in close proximity (within 1km) of a commercial center, even though only 2.5% of the land area of all MSAs is within 1km of a commercial center. Commercial centers are associated with high density of interactions, not only because of the diverse POIs that they encapsulate, but because other POIs will frequently be built in close proximity to commercial centers to attract foot traffic (e.g. all of Walmart Supercenters in Fayetteville, NC are within 1km of a commercial center).

Accessibility of commercial centers may mitigate Interaction Segregation. In Fayetteville, NC, we observe that commercial centers are located in close proximity to both high and low ES census tracts, facilitating interactions between diverse individuals, a phenomenon which we term *commercial center bridging*. Figure 3b illustrates a random sample of 10 interactions between pairs of individuals in the Highland Center, one such commercial center in Fayetteville. To quantify the extent to which commercial centers facilitate diverse interactions across all 382 MSAs, we propose the *Commercial Center Bridging Index* (CCBI). CCBI is computed from home locations and commercial center locations, and does not directly incorporate mobility data (Figure 3c, Methods). CCBI is based on the simple assumption that individuals are most likely to visit the commercial center closest to their homes. A CCBI of 1.0 indicates perfect bridging, meaning the individuals expected to visit each commercial center are as economically diverse as the overall MSA; a CCBI of 0.0 indicates no bridging, meaning the individuals expected to visit each commercial center have no diversity in ES (e.g. uniform income). To measure income diversity, we use Gini Index, a common measure of income dispersion; findings are robust to other measures of income diversity, such as variance (SI).

We find that Commercial Center Bridging Index strongly predicts Interaction Segregation (Spearman Correlation -0.78), and thus may be an modifiable extrinistic factor which could mitigate Interaction Segregation (Figure 3d). The top 10 MSAs with the highest CCBI are 53.1% less segregated than the 10 MSAs with lowest CCBI. CCBI predicts segregation more accurately than population size, ES inequality, NSI, walkability, and racial demographics, the number of commercial centers, and is significantly associated with segregation ($p < 0.01$) after controlling for all aforementioned variables (Extended Data Table 2) . We perform a number of sensitivity analyses to confirm the robustness of this finding (Supplementary Table S6, Supplementary Figure

S2-S8), including varying the definition of ES, distance, time, and tie strength needed to constitute and interaction, up-weighting pairs of people who cross paths more frequently, including only interactions within leisure POIs; excluding people who live in the same home; excluding interactions which occur on roads. Furthermore, we show that CCBI strongly explains Interaction Segregation because it captures the locations of their commercial centers, their quantity, as well residential segregation (Extended Data Figure 6), all of which independently contribute to its predictive performance (Supplementary Figure S11). This suggests that policies which encourage the construction of commercial centers bridging diverse residential neighborhoods (like zoning laws or subsidies) may mitigate Interaction Segregation.

Discussion

Our measure of Interaction Segregation has limitations. The most significant limitation is that Interaction Segregation uses two individuals being in close proximity as a proxy for a possible interaction. We do not know whether pairs of individuals have actually interacted, and furthermore cannot distinguish between the depth of the interaction, e.g. a dinner conversation at a restaurant, and standing next to another individual at a grocery store are both counted equally as a single interaction pair. Nevertheless, we show external validity of our measure in that it is more predictive of external outcomes such as upward economic mobility than conventional measures of segregation (Extended Data Figure 1), and that our results are robust to varying definitions of an interaction, e.g. using stricter time and distance proximity thresholds, and by requiring multiple days of interactions to assume a connection between a pair of individuals (Supplementary Figure S5-S8). SafeGraph mobility data also has limitations: it is a convenience sample, and has been shown to under-represent the elderly and ethnic minorities[21]. However, SafeGraph has been shown to be approximately geographically representative, as well an unbiased sample along the dimensions of income and education, and is widely used in studies of human mobility [19,22–24]; its fine-grained location data represent a significant improvement to static census census used in conventional studies of segregation. Furthermore, SafeGraph data is anonymous and cannot be directly linked to individual economic status directly. Thus, we infer home locations, and consequently estimate rent via Zillow Zestimate, which are imperfect measures of economic standing because they are not individual-level and are subject to differences in willingness to spend on housing. However, our findings are robust to different definitions of economic standing (Supplementary Figure S3). Furthermore, our network-based measure of economic segregation naturally extends to other social network datasets , and these represent interesting directions for future work. Due to the cross-sectional nature of the study, we were not able to make any causal inferences between population size, built environment, and segregation, as unobserved city-level and individual demographic and social characteristics could lead to confounding. However, we controlled for observed confounders such as ES distribution, racial demographics, city walkability, etc. in our analyses.

Our results provide policymakers with a more realistic picture of segregation, allowing them to understand its underlying mechanisms and design cities which mitigate it. Our key insight is that rather than grouping individuals by where they live, efforts to quantify, understand, and mitigate segregation are more realistic when approached from the perspective of where people move throughout their daily lives, e.g. at work, during leisure activities, and within their home neighborhoods. We show that these behaviors are associated with built environment; segregation is higher in larger cities due to increased differentiation of neighborhoods as well as POIs, but can be mitigated by commercial centers which connect high and low income neighborhoods and thereby facilitate diverse interactions. Increasing socioeconomic mixing within commercial centers may complement other approaches which seek to increase socioeconomic mixing within neighborhoods (i.e. building affordable housing within high-income neighborhoods). Moreover, increasing the accessibility of higher-level groups of POIs such as commercial centers may prove more feasible than fine-grained interventions that target individual POIs (e.g. price ceilings in restaurants, or reducing the number of parks in a city) to be less differentiated–as the majority of interactions are associated with commercial centers. Estimating the casual impact of policies which seek to implement such commercial center bridges, either via natural experiments or real-life policy interventions, represents an interesting direction for future work. Our findings thus have implications for human geography and urban design, highlighting the role of built environment in mediating diverse human interactions.

References

1. Jacobs, J. *Death and life of great american cities* (The Bodley Head, 1961).

2. Roche, M. P. Taking innovation to the streets: microgeography, physical structure, and innovation. *Review of Economics and Statistics* **102**, 912–928 (2020).

3. Jargowsky, P. A. Take the money and run: Economic segregation in us metropolitan areas. *American Sociological Review* 984–998 (1996).

4. Jargowsky, P. A. & Kim, J. A measure of spatial segregation: The generalized neighborhood sorting index. *National Poverty Center Working Paper Series* (2005).

5. Massey, D. & Denton, N. A. *American apartheid: Segregation and the making of the underclass* (Harvard university press, 1993).

6. Chetty, R., Hendren, N. & Katz, L. F. The effects of exposure to better neighborhoods on children: New evidence from the moving to opportunity experiment. *American Economic Review* **106**, 855–902 (2016).

7. Chetty, R. & Hendren, N. The impacts of neighborhoods on intergenerational mobility ii: County-level estimates. *The Quarterly Journal of Economics* **133**, 1163–1228 (2018).

8. Bor, J., Cohen, G. H. & Galea, S. Population health in an era of rising income inequality: Usa, 1980–2015. *The Lancet* **389**, 1475–1490 (2017).

9. Do, D. P., Locklar, L. R. & Florsheim, P. Triple jeopardy: The joint impact of racial segregation and neighborhood poverty on the mental health of black americans. *Social psychiatry and psychiatric epidemiology* **54**, 533–541 (2019).

10. Brown, J. R., Enos, R. D., Feigenbaum, J. & Mazumder, S. Childhood cross-ethnic exposure predicts political behavior seven decades later: Evidence from linked administrative data. *Science Advances* **7**, eabe8432 (2021).

11. Shlay, A. B. The dream revisited: Contemporary debates about housing, segregation and opportunity, edited by ingrid gould ellen and justin peter steil: New york, ny, columbia university press, 2019 (2020).

12. Ocejo, R. E. & Tonnelat, S. Subway diaries: How people experience and practice riding the train. *Ethnography* **15**, 493–515 (2014).

13. of Economic, U. N. D. *World urbanization prospects: The 2018 revision*, vol. 216 (United Nations Publications, 2018).

14. Massey, D. S. & Denton, N. A. The dimensions of residential segregation. *Social forces* **67**, 281–315 (1988).

15. Brown, J. R. & Enos, R. D. The measurement of partisan sorting for 180 million voters. *Nature Human Behaviour* 1–11 (2021).

16. Matthews, S. A. & Yang, T.-C. Spatial polygamy and contextual exposures (spaces) promoting activity space approaches in research on place and health. *American behavioral scientist* **57**, 1057–1081 (2013).

17. Zenk, S. N. *et al.* Activity space environment and dietary and physical activity behaviors: a pilot study. *Health & place* **17**, 1150–1161 (2011).

18. Moro, E., Calacci, D., Dong, X. & Pentland, A. Mobility patterns are associated with experienced income segregation in large us cities. *Nature communications* **12**, 1–10 (2021).

19. Athey, S., Ferguson, B. A., Gentzkow, M. & Schmidt, T. Experienced segregation. Tech. Rep., National Bureau of Economic Research (2020).

20. Krupka, D. J. Are big cities more segregated? neighbourhood scale and the measurement of segregation. *Urban Studies* **44**, 187–197 (2007).

21. Coston, A. *et al.* Leveraging administrative data for bias audits: Assessing disparate coverage with mobility data for covid-19 policy. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 173–184 (2021).

22. Squire, R. F. What about bias in the safegraph dataset? *Safegraph* (2019).

23. Chang, S. *et al.* Mobility network models of covid-19 explain inequities and inform reopening. *Nature* **589**, 82–87 (2021).

24. Chen, M. K. & Rohla, R. The effect of partisanship and political advertising on close family ties. *Science* **360**, 1020–1024 (2018).

Methods

In Methods M1, we explain the datasets used in our analysis; in Methods M2, we explain data processing procedures, including inference of economic standing and interactions; and in Methods M3, we explain the analysis underlying our main results.

M1 Datasets

SafeGraph

Our primary mobility and location data comprise GPS locations from a sample of adult smartphone users in the United States. The data are anonymous GPS location pings from smartphone applications which are collected and transmitted to SafeGraph by participating users[25]. While the sample is not random sample, prior work has demonstrated that SafeGraph data is geographically representative (e.g. an approximately unbiased sample of different census tracts within each State), and well-balance along the dimensions of race, income, and education[22,23]. Furthermore, SafeGraph data is a widely used standard in large-scale studies of human mobility across many different areas including COVID-19 modeling[23], political polarization[24], and tracking consumer preferences[29]. All data provided by SafeGraph was anonymized, does not contain any identifying information, and was stored on a secure server behind a firewall. Data handling and analysis was conducted in accordance with SafeGraph policies and with the guidelines of the Stanford University Institutional Review Board.

The raw data consists of 91,755,502 users and 61,730,645,084 pings (one latitude and longitude for one user at one timestamp) from three evenly spaced months in 2017: March, July, and November. The mean number of raw pings associated with a user is 667 and the median number of pings is 12. We apply several filters to improve the reliability of the SafeGraph data. To ensure locations are reliable, we remove pings whose location is estimated with accuracy worse than 100 meters as recommended by SafeGraph[30]. We filter out users with fewer than 500 pings, as these are largely noise. Since we incorporate a user's home value and rent in measuring their economic standing, we filter out users for whom we are unable to infer a home. Finally, to avoid duplicate users, we remove users if more than 80% of their pings have identical latitudes, longitudes, and timestamps to those of another user; this could potentially occur if, for example, a single person in the real world carries multiple mobile devices. After the initial filters on ping counts and reliability, we are able to infer home locations for 12,805,490 users in the United States (50 states and Washington D.C.), leveraging the CoreLogic database. Of users for whom we can infer a home location, we are able to successfully link 9,576,650 to an estimated rent value via the Zillow API. Section Methods M2 provides full details on the use of CoreLogic database to infer home locations and the use of the Zillow API to link these home locations to estimated rent values. Finally, after removing users where $> 80\%$ of their pings are duplicates with another user, we reduce the number of users from 9,576,650 to 9,567,559 (i.e., we remove about 0.1% of users through de-duplication).

CoreLogic

We use the CoreLogic real estate database to link users to home locations[31]. The database provides information covering over 99% of US residential properties (145 million properties), over 99% of commercial real estate properties (26 million properties), and 100% of US county, municipal, and special tax districts (3141 counties). The CoreLogic real estate database includes the latitude and longitude of each home, in addition to its full address: street name, number, county, state, and zip code.

Zillow

We use the Zillow property database to derive rent estimates[32] (our primary measure of economic standing), as well as home values (used in robustness checks of different measures of economic standing). The Zillow database contains home value data (”Zestimate”) for 135 million US residential properties and rent data (”rent Zestimate”) for 119 million US residential properties. We use the first CoreLogic address for which Zillow data are available. We were able to determine a rent Zestimate, the primary Zillow feature used in our analysis, for 9,576,650 out of 12,183,523 inferred SafeGraph user homes (a 79% hit rate).

SafeGraph Places

Our database of US business establishment boundaries and annotations comes from the SafeGraph Places database[25], which indexes the names, addresses, categories, latitudes, longitudes, and geographical boundary polygons of 5.5 million US points of interest (POIs) in the United States. SafeGraph includes the NAICS (North American Industry Classification System) category of each POI, which is standard taxonomy used by the Federal government to classify business establishments[33]. For instance, the NAICS code 722511 indicates full service restaurants. We identify relevant leisure sites using the prefixes 7, which includes arts, entertainment, recreation, accommodation, and food services, and supplement these POIs with the prefix 8131 to include religious organizations such as churches. We restrict our analysis of leisure sites to the top most frequently visited POI categories within these NAICS code prefixes (Figure 1d): full service restaurants, snack bars, limited-service restaurants, stadiums, etc. SafeGraph Places also includes higher-level "parent" POI polygons which encapsulate smaller POIs. Specifically, we identified commercial centers with the NAICS code 531120 (lessors of non-residential real estate) which we find in practice corresponds to commercial centers such as shopping malls, plazas, boardwalks, and other collections of businesses which lease property from a single private or public entity. We provide an illustrative of commercial centers in Supplementary Figures S14-S16.

US Census

We extract demographic and geographic features from the 5-year 2013-2017 American Community Survey (ACS)[34]. This allows us, as described below, to link cell phone locations to geographic areas including census block group, census tract, and Metropolitan Statistical Area (MSA), as well as to infer demographic features corresponding to those demographic areas including median household income.

A census block group (CBG) is a statistical division of a census tract. CBGs are generally defined to contain between 600 and 3,000 people. A CBG can be identified on the national level by the unique combination of state, county, tract, and block group codes.

A census tract is a statistical subdivision of a county containing an average of roughly 4,000 inhabitants. Census tracts range in population from 1,200 to 8,000 inhabitants. Each tract is identified by a unique numeric code within a county. A tract can be identified on the national level by the unique combination of state, county, and tract codes.

Census tracts and block groups typically cover a contiguous geographic area, though this is not a constraint on the shape of the tract or block group. Census tract and block group boundaries generally persist over time so that temporal and geographical analysis is possible across multiple censuses.

Most census tracts and CBGs are delineated by inhabitants who participate in the Census Bureau's Participant Statistical Areas Program. The Census Bureau determines the boundaries of the remaining tracts and block groups when delineation by inhabitants, local governments, or regional organizations is not possible [35].

A Metropolitan Statistical Area (MSA) is a US geographic area defined by the Office of Management and Budget (OMB) and is one of two types of Core Based Statistical Area (CBSA). A CBSA comprises a county or counties associated with a core urbanized area with a population of at least 10,000 inhabitants and adjacent counties with a high degree of social and economic integration with the core area. Social and economic integration is measured through commuting ties between the adjacent counties and the core. A Micropolitan Statistical Area is a CBSA whose core has a population of between 10,000 and 50,000; a Metropolitan Statistical Area is a CBSA whose core has a population of over 50,000. In our primary analysis, we follow Athey et al. [36] and focus on Metropolitan Statistical Areas, excluding Micropolitan Statistical Areas due to data sparsity concerns.

TIGER

Road and transportation feature annotations come from the Census-curated Topologically Integrated Geographic Encoding and Referencing system (TIGER) database[37]. The TIGER databases are an extract of selected geographic and cartographic information from the U.S. Census Bureau's Master Address File / Topologically Integrated Geographic Encoding and Referencing (MAF/TIGER) Database (MTDB). We use the MAF/TIGER Feature Class Code (MTFCC) from the TIGER Roads and TIGER Rails databases to identify road and railways. TIGER data is in the format of Shapefiles, which provide the exact boundaries of roads and railways as latitude/longitude coordinates.

M2 Data processing

For each individual, we first infer their home location, in order to estimate economic standing based on their home rent value (see *Inferring home location* and subsequently *Inferring economic standing*). We then calculate all interactions between individuals (see *Constructing interaction network*), which we then annotate based on the location, i.e. if the interaction occurred in both, one, or neither individual's home tract, and whether it occurred inside of a fine-grained POI such as a specific restaurant or a "parent" POI such as a commercial center (see *Annotating interactions*). Details on all inferences and interaction calculations are provided below.

Inferring home latitude and longitude

We first infer a user's home latitude and longitude using the latitude and longitude coordinates of their pings during local nighttime hours . We first remove users with fewer than 500 pings to ensure that we have enough data to reliably infer home locations. We then interpolate each person's location at each hour (eg, 6 PM, 7 PM, and 8 PM) using linear interpolation of latitudes and longitudes, to ensure we have timeseries at constant time resolution. We filter for hours between between 6 PM and 9 AM where the person moves less than 50 meters until thee next hour; these stationary nighttime observations represent cases when the person is more likely to usually at home. We filter for users who have stationary nighttime observations on at least 3 nights and at least 60% of observations within a 50 meter radius and define the inferred home latitude and longitude as the median latitude and longitude of the nighttime home locations (after removing outliers outside the 50 meter radius). We choose the thresholds above because they yield a good compromise between 1) inferring home latitudes and longitudes for most users and 2) inferring home latitudes and longitudes with high confidence. Overall, we are able to infer home latitudes and longitudes for 70% of users with more than 500 pings, and these locations are inferred with high confidence; 89% of stationary nighttime observations are within 50 meters of the inferred home latitude and longitude .

Inferring economic standing from home latitude and longitude

Having inferred home latitude and longitude from nighttime user locations, we link the latitude and longitude to housing databases to infer the estimated rent of the user's home, using this as a proxy for economic standing. We do this in two steps. First, we link the inferred user's home latitude and longitude to the CoreLogic property database (Methods M1), a comprehensive database of properties in the United States, by taking the closest CoreLogic residential property (single family residence, condominium, duplex, or apartment) to the user's inferred home latitude and longitude. Second, we use the CoreLogic address to query the Zillow database, which provides estimated home rent and price for each user. (We cannot query the Zillow database using the latitude and longitude, which is why we first link to CoreLogic to get an address.) We use Zillow's estimated rent for the user's home as our main measure of economic standing. We apply several quality control filters to ensure that the final set of users we use in our main analyses have reliably inferred home locations and economic standings: 1) we remove a small number of users whose inferred nighttime home latitude and longitude are identical to another user's, since we empirically observe that these people have unusual ping patterns; 2) we remove users for whom we are lacking an Zillow rent estimate, since this constitutes our primary economic standing measure; 3) we winsorize Zillow rent estimates which are greater than $20,000 to avoid spurious results from a small number of outliers; 4) we remove a small number of users who are missing Census demographic information for their inferred home location; 5) we remove users whose Zillow home location is further than 100 meters from their CoreLogic home location, or whose CoreLogic home location is further than 100 meters from their nighttime latitude and longitude; 6) we remove a small number of users in single family residences who are mapped to the exact same single family residence as more than 10 other people, since this may indicate data errors.

The set of users who pass these filters constitute our final analysis set. We confirm that the Census demographic statistics of these users' inferred home locations are similar to those of the US population in terms of income, age, sex, and race, suggesting that our inference procedure yields a demographically representative sample.

Any individual quantitative measure provides only a partial picture of a person's economic standing. Recognizing this, we conduct robustness checks in which rather than using the Zillow estimated rent of the user's home as a proxy for economic standing, we use 1) the median Census Block Group household income in that area; and 2) the percentile-scored rent of the home, to account for long-tailed rent distributions. Our main results are robust to using these alternate measures of economic standing (Supplementary Figure S3).

Constructing interaction network

We construct a fine-grained, dynamic interaction network $\mathcal{G}$ between all 9,567,559 individuals across 382 MSAs and 2829 counties, which is represented as an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with time-varying edges. Each node $v_i \in \mathcal{V}$ in the graph represents one of the $N = 9,567,559$ individuals in our study, such that the set of nodes is $V = \{v_1, v_2, ..., v_N\}$. Each node v_i has a single attribute x_i, representing the inferred economic standing (estimated rent) of the individual.

Individuals v_i and v_j are connected by one edge $e_{i,j,k} \in \mathcal{E}$ per interaction, with k indicating the kth interaction between individuals v_i and v_j. Each edge $e_{i,j,k}$ has three attributes $t_{i,j,k}$, $lat_{i,j,k}$, $lon_{i,j,k}$ indicating the timestamp, latitude, and longitude of the interaction respectively. We now focus our discussion on explaining how each of the interactions edges of the network is calculated.

We define an *interaction* to occur when two users have GPS pings which are close (according to a fixed threshold) in both physical proximity and time. Specifically if user v_i has a GPS ping with t_i, lat_i, lon_i (indicating the timestamp, latitude, and longitude of the ping respectively), and user v_j has a GPS ping with t_j, lat_j, lon_j, then we users are said to have shared an interaction if $|t_i - t_j| < T$ and $distance((lat_i, lat_i), (lat_j, lat_j)) < D$, where T represents the time threshold (i.e. maximum time distance the two pings can be apart to count as an interaction) and D represents the distance threshold (i.e. maximum physical distance the two pings can be apart to count as an interaction). We filter for both distance and time simultaneously to ensure that our interaction network only includes pairs of users who are likely to have come into contact with each other. This contrasts to other methods which consider all individuals to visit the same location, irrespective of time[18], to have an equal likelihood of interaction, an assumption which may prove unrealistic in many cities (e.g. demographics of individuals visiting public parks varies starkly by time of day[39]).

We use a threshold T of 5 minutes, which is a stringent threshold on time as the mean number of pings per person per hour during day time is approximately one ping; we use a distance threshold D of 50 meters, following prior work which shows that even exposure to individuals from afar is linked to long term outcomes[10].

Furthermore, we show through a series of robustness checks that our key results in Figure 1, Figure 2, and Figure 3 are robust to varying thresholds (i.e. 1 minute or 2 minutes time threshold, as well as 10 meters or 25 meters distance threshold), as well as additional criteria to increase tie strength (i.e. requiring multiple consecutive interactions, or multiple interactions on unique days)—and under all circumstances the main findings remain consistent (Supplementary Information *Robustness Checks*).

To efficiently calculate the interactions between all users, we implement our interaction threshold a k-d tree[41], a data structure which allows one to efficiently identify all pairs of points within a given distance of each other. In total, we identify 1,570,782,460 interactions. The timestamp $t_{i,j,k}$ of the interaction is the minimum ping timestamp in the pair of individuals' ping timestamps (t_i,t_j), and the location $lat_{i,j,k}$, $lon_{i,j,k}$ of the interaction is the average latitude and longitude of pair of pings belonging to the two individuals (lat_i,lat_j) and (lon_i,lon_j).

Annotating interactions

Interactions are annotated to indicate whether they occurred at or near features of interest: e.g., near a user's home. Annotations are not mutually exclusive in that an interaction may be simultaneously tagged as having occurred near multiple features from multiple data sources. We describe the specific annotations below.

Annotating User Home and Work We annotate a user's interaction as having occurred in their home if it occurs within 50 meters of the user's home location. We annotate a interaction as having occurred in a user's place of work if it occurs within 200 meters of their work location, to account for the possibility of work meetings with co-workers in nearby POIs (e.g. a lunch meeting in a nearby cafe).

Annotating TIGER Roads/Railways and Safegraph POIs An interaction is annotated with a TIGER road/railway if it occurs within 20 meters from that feature. An interaction is annotated as having occurred within a SafeGraph Places point-of-interest (POI) if the interaction occurs within the polygon defined for the POI. Polygons are provided by the SafeGraph Places database for both fine-grained POIs (e.g. individual restaurants) as well as "parent" POIs (e.g. commercial centers). We focus our analysis of fine-grained POIs (Figure 3e, Extended Data Figure 2) on the most visited fine-grained POIs: full-service restaurants, snack bars, limited-ervice restaurants (e.g. fast food), stadiums, etc (see 3e for full list). These categories roughly align with those used by prior work [36].

M3 Analysis

Interaction Segregation

We define the Interaction Segregation of a specified geographical area (i.e. Metropolitan Statistical Area, County) as the Pearson correlation between the economic standing (ES) of individuals residing in that geographical area, and the mean ES those that they come into contact with.

$$Interaction\ Segregation = Corr(ES, \overline{ES}_{interactions})$$

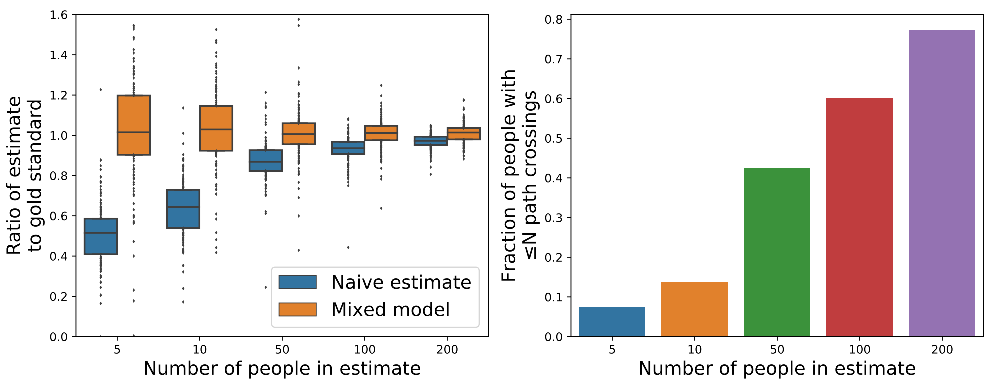
Our metric captures the extent to which an individual's ES explains, or predicts, the ES of their immediate interaction network. Thus, in a perfectly integrated area in which individuals interact randomly with others regardless of ES, Interaction Segregation would equal 0.0. In a perfectly segregated area in which individuals interaction with only those of the exact same ES, Interaction Segregation would equal 1.0.

Interaction Segregation is derived from extending a classic definition residential segregation, the Neighborhood Sorting Index [3,4] (NSI), which is equivalent to the correlation across people between Pearson each person's ES and the mean ES in their Census tract. The NSI is an intuitive definition of segregation and can be calculated from Census data on the ES of people living in each tract. A disadvantage is that it only incorporates information about the Census tract where people *live*[4], an artificial boundary which fails to capture interactions at work and at leisure. We observe that NSI is equivalent to the correlation between each person's ES and the mean ES of everyone they interact with, under the unrealistic assumptions that a) people only interact with those in their home Census tract and b) they do so uniformly at random. These assumptions limit the real-world applicability of NSI, because people may also interact with more heterogeneous populations as they visit other Census tracts for work, leisure, or other activities, a phenomenon we refer to as the *visitor effect*. Furthermore, even within home tract, individuals may interact non-uniformly as they seek out people of similar economic standing; we refer to this as the *homophily effect*. Thus, to design a metric which captures both the *visitor effect* and *homophily effect*, Interaction Segregation leverages cell-phone mobility data to capture the extent of contact between diverse individuals throughout the day, both outside and inside of individuals' home tract. Furthermore, Interaction Segregation allows for direct comparability to NSI, because both measures are of the same underlying statistical quantity, but differ in their definition of the interaction network.

To calculate the Interaction Segregation of a specified geographical area (i.e. Metropolitan Statistical Area, County), we first select the set of all individuals who reside in area: $\mathcal{V}_{MSA} \subset \mathcal{V}$. For instance, to calculate Interaction Segregation for Napa, California (Figure 1c Top), we first filter for the 3707 individuals with home locations inside the geographical boundary of the Napa, CA MSA: $\mathcal{V}_{MSA}$. Subsequently, for each individual resident of the area $v_i \in \mathcal{V}_{MSA}$ we query the population interaction network ($\mathcal{G} = (\mathcal{V}, \mathcal{E})$) for the set of individuals they interact with, $\mathcal{I}_i$: $\{v_j \in \mathcal{I}_i | e_{i,j,k} \in \mathcal{E}\}$. We then aim to estimate the Pearson correlation between the ES of each individual x_i and the mean ES of their network, $y_i = mean(\mathcal{Y}_i)$, where each person's network $\mathcal{Y}_i$ is defined as the economic standing values of their connections in the interaction network: $\{x_j \in \mathcal{Y}_i | v_j \in \mathcal{I}_i\})$.

Mixed model

We use a linear mixed effects model to accurately estimate Interaction Segregation. Our linear mixed effects model is an unbiased estimator of the Pearson correlation. A statistical model is required to estimate Interaction Segregation because, while we define Interaction Segregation as the Pearson correlation coefficient between a person's ES and the mean ES of the people they cross paths with, naively computing the correlation based on limited data (in counties or MSAs with low population sizes) results in estimates that are downward biased [1]. This effect is shown in Methods Figure 1. To illustrate why this effect is observed, imagine that we compute the correlation between a person's ES and the "true" mean ES of the people they interact with. Now, we add noise to the mean ES values, which represents the noisy mean estimates given limited data. As the noise is increased, the correlation is decreased. Thus, because estimates of each person's mean ES will be more noisy in geographical areas with less data, there will be a downward bias to naive estimates of the Pearson correlation in these areas.



Methods Figure 1: Mixed model estimate compared with naive estimates of the Pearson correlation. We took people who crossed paths with at least 500 other people and computed the Pearson correlation coefficient (the "gold standard estimate"). Then, for each person we randomly sampled 5, 10, 50, 100, and 200 people from the 500+ people and computed segregation estimates based on the reduced sets of people. The top plot shows the ratio of the estimates to the gold standard, for each MSA. The bottom plot shows the overall number of people in the dataset with $\leq N$ interactions.

Our mixed model models the distribution of datapoints (x_i, y_{ij}) through the following equation:

[1] By "naive" estimation of the Pearson correlation, we intend to convey calculating the correlation using the sample:

$$\frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2 (y_i - \overline{y})^2}}$$

$$
\begin{aligned}
y_{ij} &= ax_i + b + \epsilon_i^{(1)} + \epsilon_{ij}^{(2)} \\
\text{where } x_i &= \text{ES of person } i \\
y_{ij} &= \text{ES of person } j \text{ who has interacted with person } i \\
a, b &= \text{model parameters} \\
\epsilon_i^{(1)} &= \text{person-specific noise term} \\
\epsilon_{ij}^{(2)} &= \text{noise for each data point}
\end{aligned}
$$

The Pearson correlation coefficient between person i's ES and the mean ES of the people they cross paths with is then computed as follows. We assume that x_i has a variance of 1 through data preprocessing and that x_i is uncorrelated with $\epsilon_i^{(1)}$.

$$
\begin{aligned}
\operatorname{corr}\left(x_i, ax_i + b + \epsilon_i^{(1)}\right) &= \operatorname{corr}\left(x_i, ax_i + \epsilon_i^{(1)}\right) \\
&= \frac{\operatorname{cov}\left(x_i, ax_i + \epsilon_i^{(1)}\right)}{\sqrt{\operatorname{Var}\left(x_i\right) Var\left(ax_i + \epsilon_i^{(1)}\right)}} \\
&= \frac{\operatorname{cov}\left(x_i, ax_i\right)}{\sqrt{\operatorname{Var}\left(ax_i + \epsilon_i^{(1)}\right)}} \\
&= \frac{a}{\sqrt{a^2 + \operatorname{Var}\left(\epsilon_i^{(1)}\right)}}
\end{aligned}
$$

We estimate a and $\operatorname{Var}\left(\epsilon_i^{(1)}\right)$ by fitting the mixed model using the R lme4 package, optimizing the restricted maximum likelihood (REML) objective.

Decomposing segregation by time

Each interaction edge ($e_{i,j,k}$) in our interaction network is timestamped with a time of interaction $t_{i,j,k}$. This allows us to decompose our overall Interaction Segregation into fine-grained estimates of segregation during different hours of the day, by filtering for interactions that occurred within a specific hour. In Supplementary Figures S19, we partition estimates of segregation by 3 hour windows to illustrate how segregation varies throughout the day (see Supplementary Information).

Decomposing segregation by activity

Each interaction edge ($e_{i,j,k}$) in our interaction network occurs at a specific location $lat_{i,j,k}$, $lon_{i,j,k}$. Thus, it is possible to annotate interactions by the fine-grained POI (e.g. specific restaurant) they occurred in, as well as the by the higher-level "parent" POI (e.g. commercial center) in which the POI was located (Methods M2). This allows us to decompose our overall Interaction Segregation into fine-grained estimates of segregation by specific leisure activity. We do so by filtering the network for all interactions that occurred in a specific POI category, and re-calculating Interaction Segregation for the MSA or county, using only those interactions. In Figure 1d, we show the variation in Interaction Segregation by leisure site, and further explain these variations (i.e. why are golf courses and country clubs the most segregated, and performing arts centers the least segregated?) in Extended Data Figure 2.

Commercial Center Bridging Index

We seek to identify a modifiable, extrinsic aspect of a city's built environment which may mitigate Interaction Segregation. One promising candidate is the locations of commercial centers, because commercial centers are strongly associated with high density of interactions (Figure 3a). Specifically, the majority (56.9%) of interactions happen inside of or within 1km of a commercial center (e.g. shopping mall, plaza, etc.) even though only 2.5% of the land area of MSAs is within 1km of a commercial center.

We define a new measure, the *Commercial Center Bridging Index* (CCBI), which measures the extent to which the commercial centers in a geographical area facilitate the integration of individuals of diverse economic standing. Intuitively, CCBI aims to capture, if individuals visit the commercial center closest to them, how economically diverse of a population they will be exposed to. A CCBI of 1.0 indicates that if an individual visits their local commercial center, they will be exposed to a set of people *as economically diverse as the overall city they reside in*. Thus, a CCBI of 1.0 signifies perfect commercial center bridging, i.e. even if individuals live in segregated neighborhoods, the commercial centers of each neighborhood are located such that individuals must leave their neighborhoods and interact with diverse others during leisure activity and shopping. On the other hand, a CCBI of 0.0 signifies the opposite extreme; a city with a CCBI of 0.0 is one in which, if an individual visits their local commercial center, they will be exposed to people of the exact same economic standing.

$$Commercial\ Center\ Bridging\ Index\ (CCBI) = \frac{\sum_{i=1}^{K} |\mathcal{C}_i| \cdot Gini\ Index(\mathcal{C}_i)}{|\mathcal{V}_{MSA}| \cdot Gini\ Index(\mathcal{V}_{MSA})}$$

Formally, we define the Commercial Center Bridging Index by first clustering all individuals who live in a specific MSA ($\mathcal{V}_{MSA}$) into K clusters ($\mathcal{C}_1, \mathcal{C}_1, ..., \mathcal{C}_K$) according to the commercial center closest to their home location; K is the number of commercial centers present in the MSA. Thus, each individual is grouped with all other individuals who reside next to the same local commercial center. We quantify the economic diversity of each commercial center $\mathcal{C}_i$ using the Gini Index: $Gini\ Index(\mathcal{C}_i)$, a well-established measure of economic statistical dispersion[44]. We define the Commercial Center Bridging Index as the weighted average of the income diversity (Gini Index) of each commercial center cluster, relative to the overall income diversity (Gini Index) of the MSA. We divide by the overall overall income diversity (Gini Index) as we aim to measure the extent to which malls facilitate diverse interactions relative to the baseline diversity of interactions within the MSA, i.e. we need to normalize for the baseline economic diversity of the MSA, as that will also influence the economic diversity within each cluster, and there is a positive correlation between Gini Index and Interaction Segregation (Extended Data Table 1). We discover that CCBI strongly predicts Interaction Segregation (Spearman Correlation -0.78, Figure 33d). The top 10 MSAs with the highest CCBI are 53.1% less segregated than the 10 MSAs with lowest CCBI. CCBI predicts segregation more accurately than population size, ES inequality, NSI, and racial demographics, and is significantly associated with segregation ($p < 0.01$) after controlling for all aforementioned variables (Extended Data Table 2). .

Methods References

25. Safegraph. Safegraph dataset (2017).

26. Squire, R. F. What about bias in the safegraph dataset? *Safegraph* (2019).

27. Chang, S. *et al.* Mobility network models of covid-19 explain inequities and inform reopening. *Nature* **589**, 82–87 (2021).

28. Chen, M. K. & Rohla, R. The effect of partisanship and political advertising on close family ties. *Science* **360**, 1020–1024 (2018).

29. Athey, S., Blei, D., Donnelly, R., Ruiz, F. & Schmidt, T. Estimating heterogeneous consumer preferences for restaurants and travel time using mobile location data. In *AEA Papers and Proceedings*, vol. 108, 64–67 (2018).

30. SafeGraph. Determining point-of-interest visits from location data: A technical guide to visit attribution. Tech. Rep. (2020).

31. CoreLogic, I. Corelogic real estate database (2017).

32. Zillow, I. Zillow rent zestimate (2017).

33. Kelton, C. M., Pasquale, M. K. & Rebelein, R. P. Using the north american industry classification system (naics) to identify national industry cluster templates for applied regional analysis. *Regional Studies* **42**, 305–321 (2008).

34. Bureau, U. C. American community survey 5-year estimates (2017). Accessed August 30, 2017.

35. US Census Bureau. Glossary.

36. Athey, S., Ferguson, B., Gentzkow, M. & Schmidt, T. Experienced segregation. *NBER* (2020).

37. Bureau, U. C. Tiger/line shapefiles (2017). Accessed August 30, 2017.

38. Moro, E., Calacci, D., Dong, X. & Pentland, A. Mobility patterns are associated with experienced income segregation in large us cities. *Nature communications* **12**, 1–10 (2021).

39. Madge, C. Public parks and the geography of fear. *Tijdschrift voor economische en sociale geografie* **88**, 237–250 (1997).

40. Brown, J. R., Enos, R. D., Feigenbaum, J. & Mazumder, S. Childhood cross-ethnic exposure predicts political behavior seven decades later: Evidence from linked administrative data. *Science Advances* **7**, eabe8432 (2021).

41. Bentley, J. L. Multidimensional binary search trees used for associative searching. *Communications of the ACM* **18**, 509–517 (1975).

42. Jargowsky, P. A. Take the money and run: Economic segregation in us metropolitan areas. *American Sociological Review* 984–998 (1996).

43. Jargowsky, P. A. & Kim, J. A measure of spatial segregation: The generalized neighborhood sorting index. *National Poverty Center Working Paper Series* (2005).

44. Dorfman, R. A formula for the gini coefficient. *The review of economics and statistics* 146–149 (1979).

45. Bailey, M., Cao, R., Kuchler, T., Stroebel, J. & Wong, A. Social connectedness: Measurement, determinants, and effects. *Journal of Economic Perspectives* **32**, 259–80 (2018).

46. Bailey, M., Farrell, P., Kuchler, T. & Stroebel, J. Social connectedness in urban areas. *Journal of Urban Economics* **118**, 103264 (2020).

47. Bailey, M. *et al.* International trade and social connectedness. *Journal of International Economics* **129**, 103418 (2021).

48. Kuchler, T., Li, Y., Peng, L., Stroebel, J. & Zhou, D. Social proximity to capital: Implications for investors and firms. Tech. Rep., National Bureau of Economic Research (2020).

49. Kuchler, T., Russel, D. & Stroebel, J. Jue insight: The geographic spread of covid-19 correlates with the structure of social networks as measured by facebook. *Journal of Urban Economics* 103314 (2021).

50. Chetty, R., Hendren, N., Kline, P. & Saez, E. Where is the land of opportunity? the geography of intergenerational mobility in the united states. *The Quarterly Journal of Economics* **129**, 1553–1623 (2014).

51. Economist, T. The big mac index (2021). Accessed Sep 29th, 2021.

52. to Dine Out at a Top Michelin-Starred Restaurant, H. M. Tiger/line shapefiles (2021). Accessed Sep 29th, 2021.

53. Walkscore. Walkscore walkability estimates (2020). Accessed August 30, 2020.

54. Merry, K. & Bettinger, P. Smartphone gps accuracy study in an urban environment. *PloS one* **14**, e0219890 (2019).

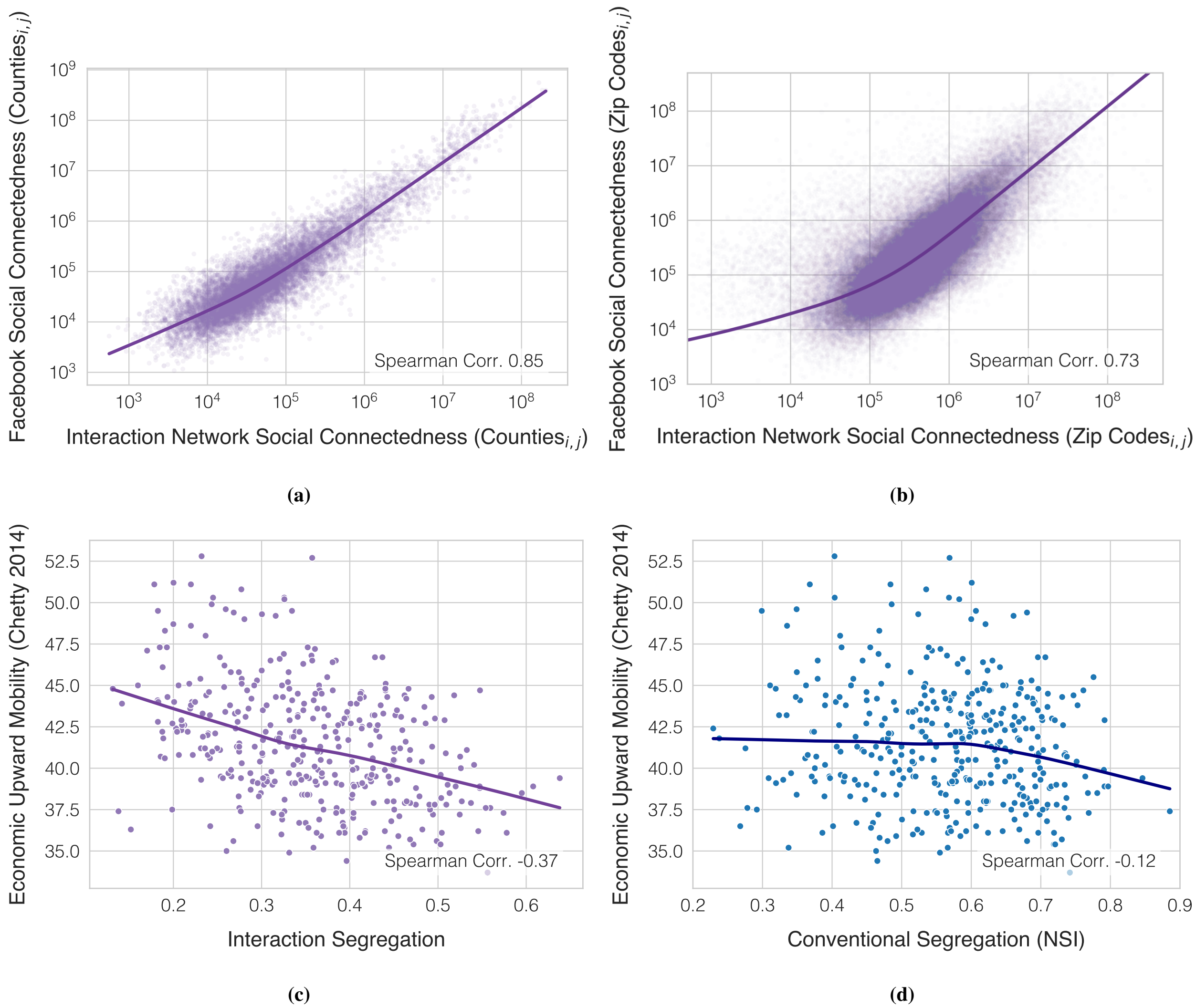
55. Menard, T., Miller, J., Nowak, M. & Norris, D. Comparing the gps capabilities of the samsung galaxy s, motorola droid x, and the apple iphone for vehicle tracking using freesim_mobile. In *2011 14th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, 985–990 (IEEE, 2011).

56. Barabási, A.-L. Network science. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **371**, 20120375 (2013).

57. Maslov, S. & Sneppen, K. Specificity and stability in topology of protein networks. *Science* **296**, 910–913 (2002).

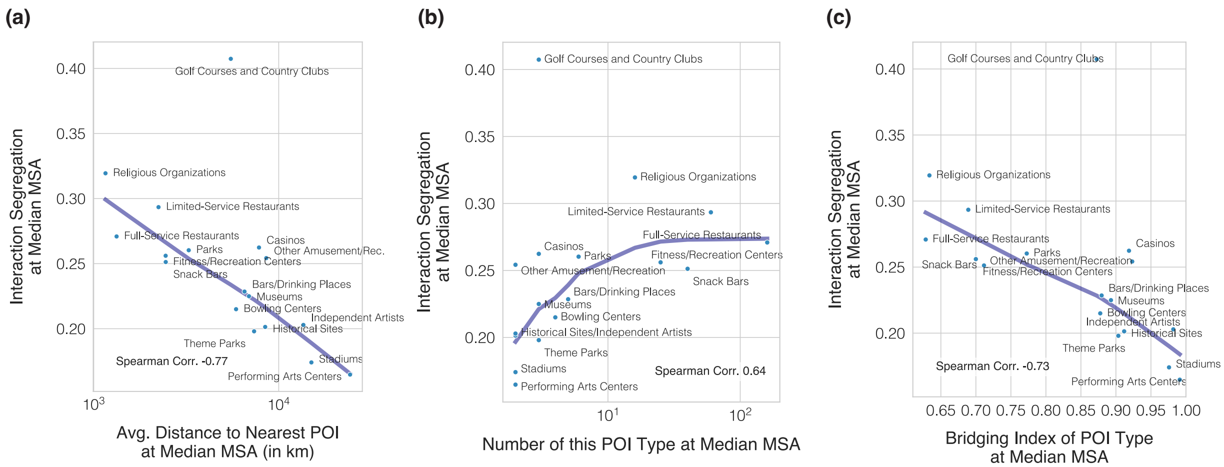
58. Dunbar, R. *How many friends does one person need?: Dunbar's number and other evolutionary quirks* (Faber & Faber, 2010).

Extended Data



Extended Data Figure 1: This studies' interaction network predicts population-scale friendship formation and upward economic mobility outcomes. We measure the external validity of our definition of interaction, by linking our interaction network to outcomes across two gold-standard, large-scale, data-sets. We find at the zip code, county, and MSA-level, our interaction network mirrors population-scale outcomes resulting from dynamic human processes: **(a-b)** the Facebook Social Connectedness Index[45] measures the relative probability of a Facebook friendship link between a given Facebook user in location i and a given user in location j. FB Social Connectedness Index has been used social segregation[46], and has also been linked to economic[47,48] and public health outcomes[49]. We reproduce the Social Connectness Index using our interaction network ($\frac{\#InteractionPairs_{i,j}}{\#Individuals_i \cdot \#Individuals_j}$) at the county **(a)** and zip code **(b)** level, and find strong correlations across county pairs (Spearman $\rho = 0.85$, $N = 121,595$, $p < 10^{-10}$) and zip code pairs (Spearman $\rho = 0.73$, $N = 1,053,539$, $p < 10^{-10}$). **(c-d)** The Chetty et al. Intergenerational Mobility data-set quantifies upward economic mobility from federal income tax records for each MSA as the mean income rank of children with parents in the bottom half of the income distribution [50]. We find that Interaction Segregation at the MSA-level **(c)** correlates to upward economic mobility (Spearman $\rho = -0.37$, $N = 379$, $p < 10^{-10}$), and does so significantly more strongly ($p < 0.0001$, Fisher's z-test) than **(d)** the conventional segregation measure NSI (Spearman $\rho = -0.12$, $N = 379$, $p < 0.05$)

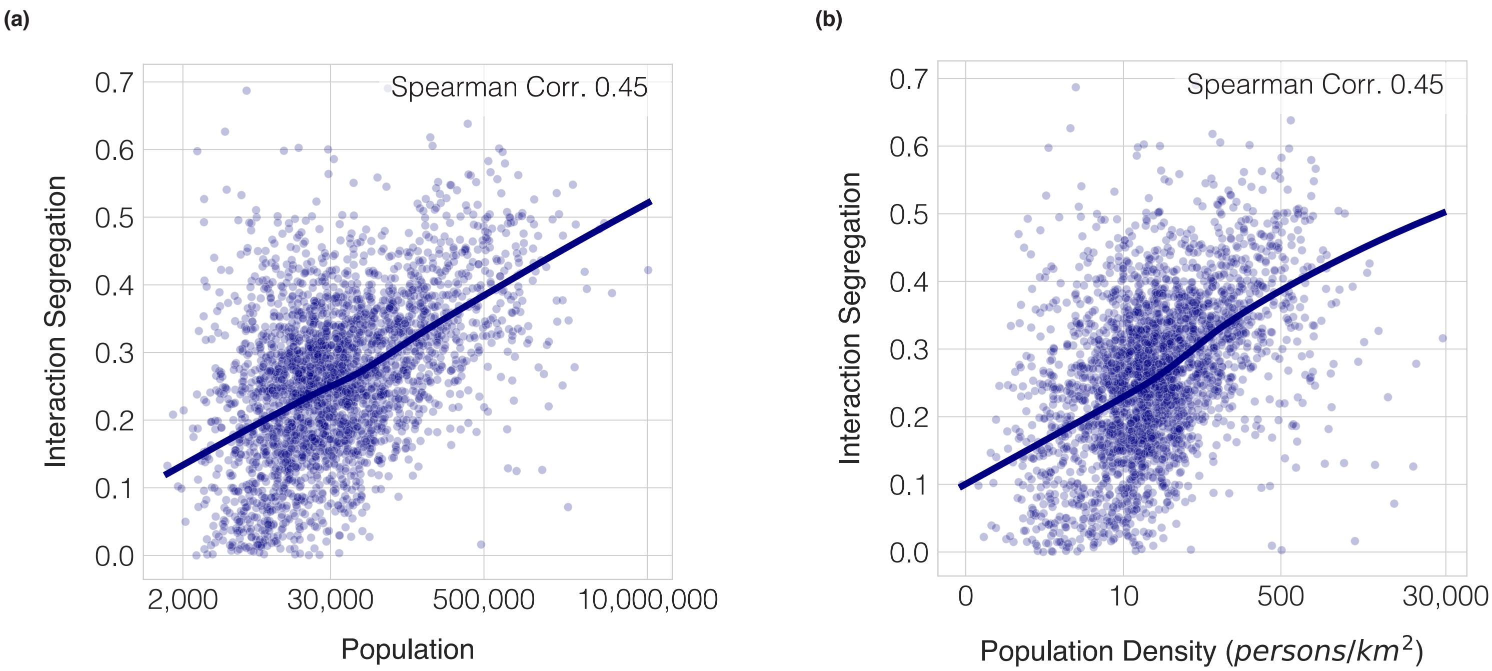
.



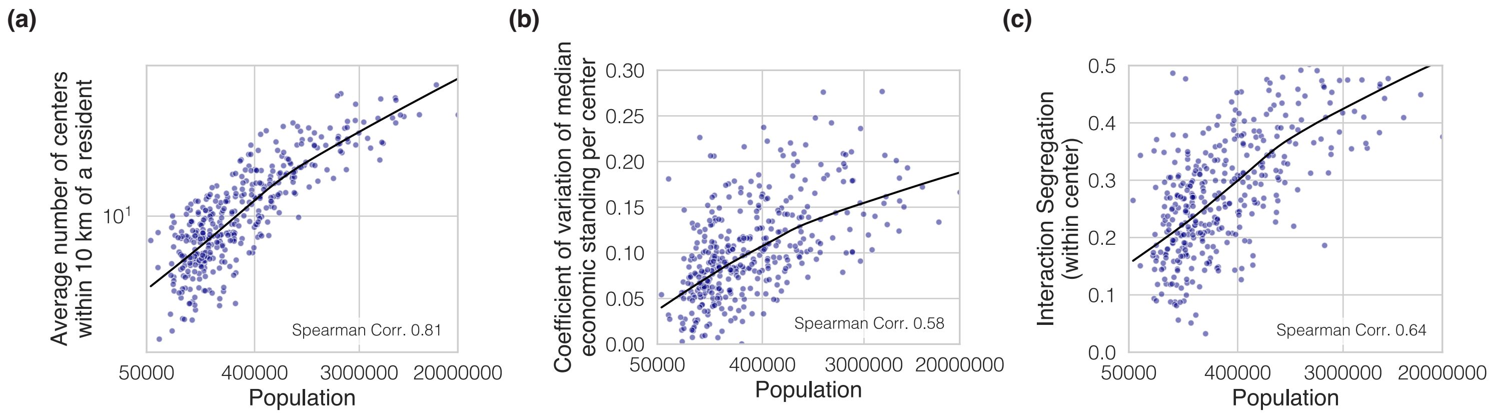
(d)

Metropolitan Statistical Area (MSA)	POI Name	ES	Minimum Cost of Entrance ($)
Chicago-Naperville-Elgin, IL-IN-WI	Wynstone Golf Club	High	40,000
	Flagg Creek Golf Course	Low	19
Los Angeles-Long Beach-Anaheim, CA	Coto de Caza Golf & Racquet Club	High	45,000
	Rancho Vista Golf Club	Low	129
Miami-Fort Lauderdale-West Palm Beach, FL	Old Palm Golf Club	High	199,200
	Eco Golf Club	Low	17
New York-Newark-Jersey City, NY-NJ-PA	Scarsdale Golf Club	High	8,900
	Weequahic Park Golf Course	Low	35
Phoenix-Mesa-Scottsdale, AZ	The Estancia Club	High	150,000
	Peoria Pines	Low	28

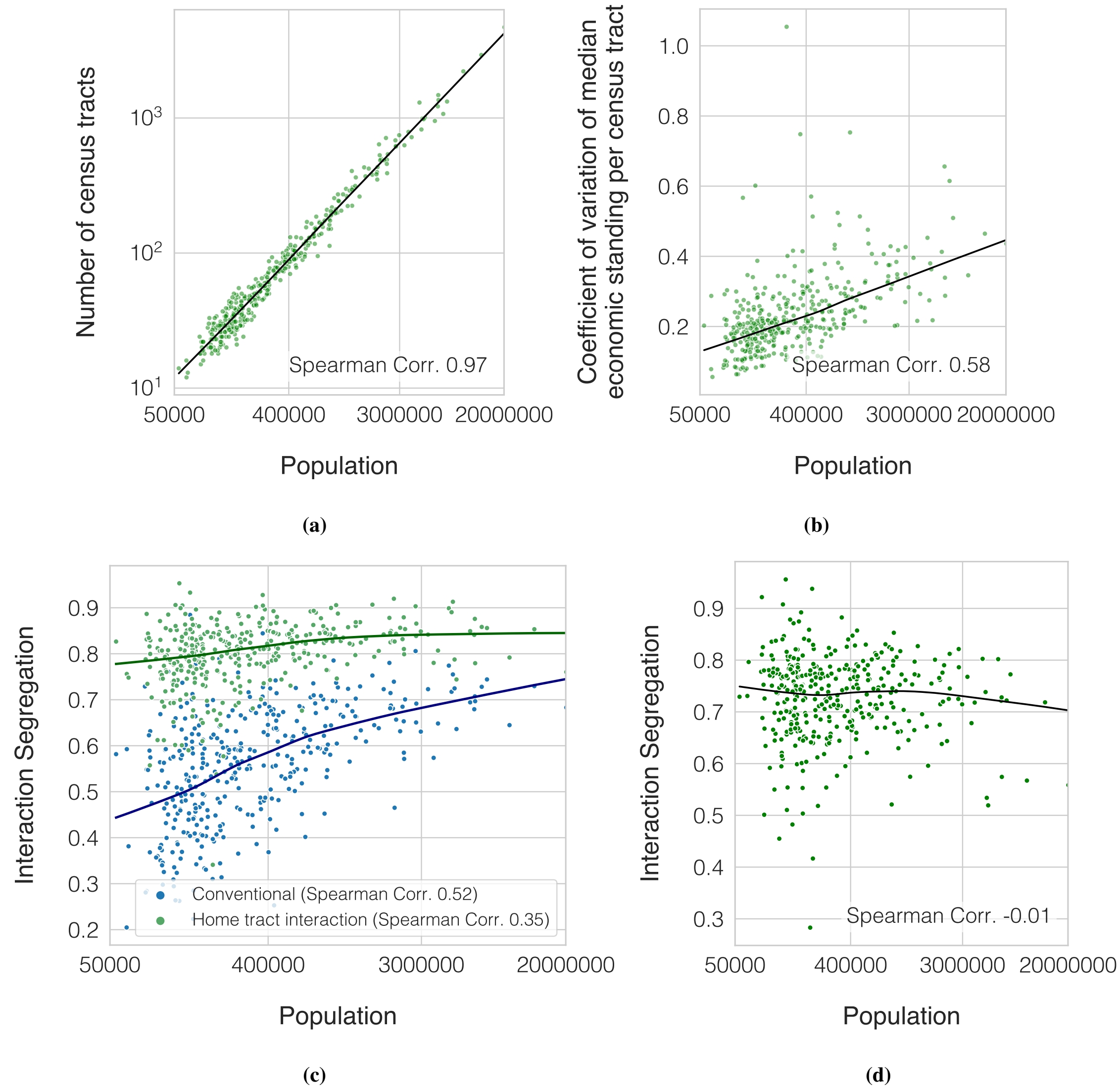
Extended Data Figure 2: Understanding why Interaction Segregation varies significantly across leisure sites. We identify two primary components of POI differentiation which explain the differences between the heterogenous segregation levels of different leisure POIS (Figure 1e): differentiation due to **(a-c)** localization **(d)** within-activity stratification. **(a-c)** We find that activities which are more localized, and in which there are many differentiated POIs which cater to different demographics (e.g. religious organizations) are more segregated than activities in which POIs are small in number and cater to the overall city (e.g. stadiums). **(a)** We operationalize localization as the average distance from each indvidual in the MSA to the nearest POI of a category, show that localization strongly predicts segregation across all POI categories (Spearman ρ=-0.77, N=17, p<0.01). **(b)** The raw number of POIs of one category can also be used to as a proxy to measure localization; leisure activities with more POIs are more localized and consequently more segregated (Spearman ρ=0.64, N=17, p<0.01), **(c)** We validate our measure of Commercial Center Bridging Index, by calculating the Bridging Index for each POI category (e.g. for Restaurant Bridging Index, we perform the same process shown in 7, except we cluster by nearest restaurant instead of by nearest commercial center) and show that, as an aggregate measure of differentiation, it also captures localization (Spearman ρ=-0.73, N=17, p<0.01). However, we find that CCBI is a stronger measure of differentiation as it captures other aspects beyond localization (Supplementary Figure 6), and thus is a stronger predictor of Interaction Segregation than localization alone (Supplementary Figure S9). **(d)** Golf courses and country clubs (golf clubs) are an anomaly in that they have a small number of unlocalized POIs, but are highly segregated. We conduct a case study in which look at top and bottom golf clubs by mean visitor ES in 5 cities. We find that the high segregation of golf clubs is due to extreme stratification between venues; for instance the minimum cost to play at the high-ES golf course in Miami, FL is 11717$\times$ higher than at the lowest-ES golf course. By contrast, the average cost of a MacDonalds Big Mac ($5.65[51]) is only 63$\times$ higher than the average cost of a Michelin 3-star restaurant ($357[52]).



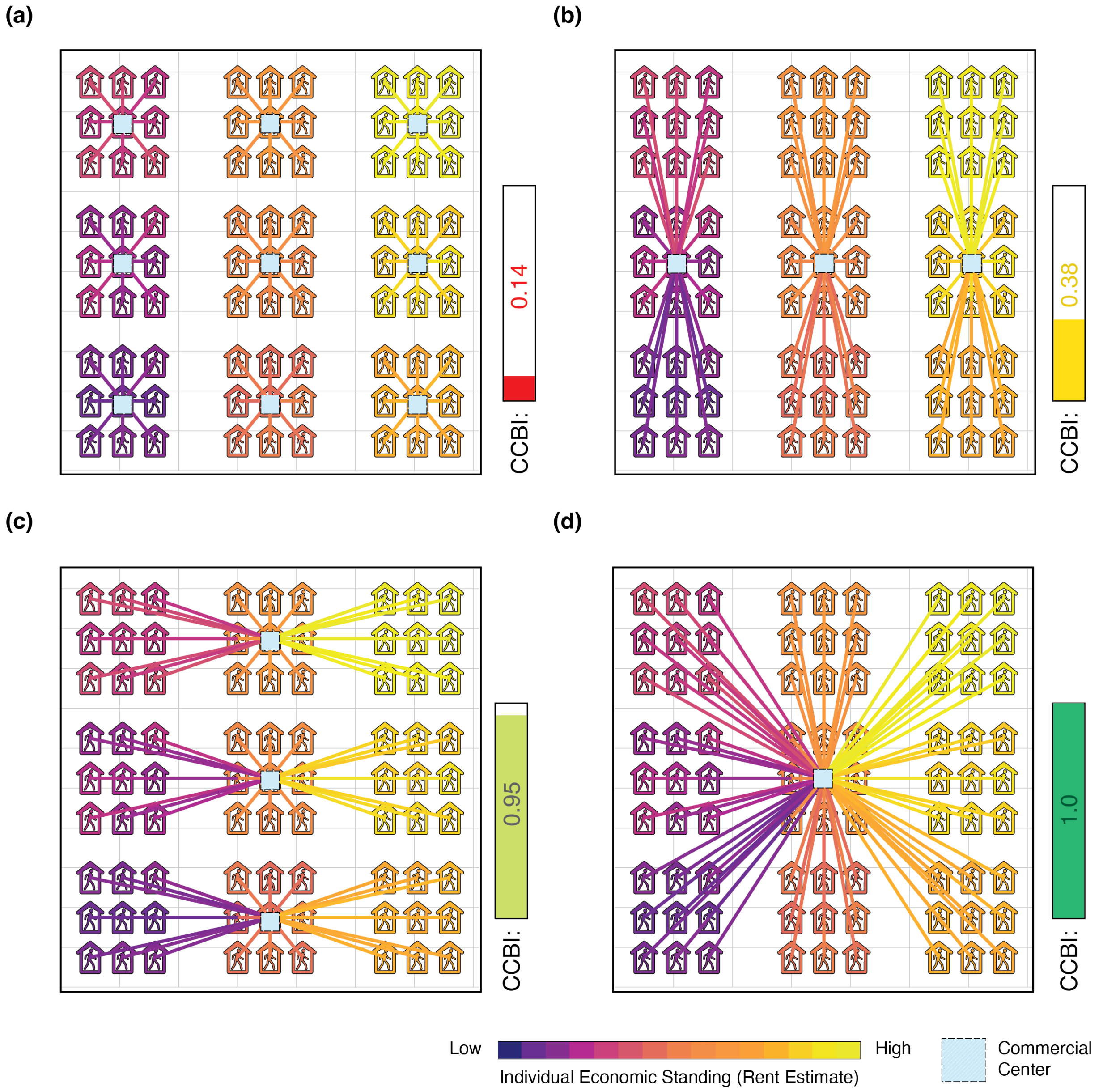
Extended Data Figure 3: Large, dense counties are more segregated. We compute Interaction Segregation across 2829 USA counties (94% of the counties in the USA), excluding counties in which there are less than 50 individuals in our data-set. We find that at the county-level, Interaction Segregation is also positively correlated with population size (Spearman ρ=0.45, N=2829, p$< 10^{-10}$) and population density (Spearman ρ=0.45, N=2829, p$< 10^{-10}$). These correlations reveal that the association between large, dense cities and Interaction Segregation (Figure 2a) is likely an artifact of city boundaries, and may in fact be an emergent property from dynamics of individuals residing highly populated, dense geographic areas, which persists across multiple scales of granularity.



Extended Data Figure 4: At multiple levels of scale, POIs in large cities are more differentiated and consequently segregated: commercial centers. Commercial centers are higher-level "parent" POIs which contain individual POIs such as restaurants, stores, and fitness centers (Supplementary Figures S9-7) **(a-c)** For commercial centers, we conduct an analogous analysis to that for restaurants in Figure 3c-e. We find that large, densely populated metropolitan areas are associated with increased options and economic differentiation of commercial centers, which may lead to higher self-segregation: **(a)** Larger MSAs have more commercial centers within 10 kilometers of the average resident, giving residents more options to self-segregate (Spearman ρ=0.81, N=382, p$<$ 10^{-10}). **(b)** Moreover, commercial centers in larger MSAs vary more in terms of the median ES of their visitors, offering a greater choice of highly socioeconomically differentiated commercial centers (Spearman ρ=0.58, N=382, p$<$ 10^{-10}). **(e)** Consequently, overall interaction segregation within commercial center is higher in larger MSAs (Spearman ρ=0.64, N=382, p$<$ 10^{-10}). Data is shown for commercial centers with at least 10 people who interact in the commercial center; black lines indicate the LOWESS fit.



Extended Data Figure 5: At multiple levels of scale, POIs in large cities are more differentiated and consequently segregated: home neighborhoods. **(a-c)** Conducting an analogous analysis to that for restaurants in Figure 3c-e and commercial centers in Extended Data Figure 4, we find that higher segregation is driven by an increase in highly differentiated choice of neighborhoods in large cities: (**a**) Larger MSAs have more census tracts, giving residents more options to self-segregate (Spearman ρ=0.97, N=382, p$<10^{-10}$). **(b)** Consequently, census tracts in larger MSAs vary more in terms of the mean ES of their residents (Spearman ρ=0.58, N=382, p$<10^{-10}$) and as a result, (**c**) both conventional NSI and interaction segregation are higher (Spearman ρ=0.52 and ρ=0.35, N=382, both p$<10^{-5}$). However, **(c)** also shows that interaction segregation (green series) rises more slowly with population than conventional segregation (blue series), suggesting that within-home-tract homophily, which increases interaction segregation but not conventional segregation, is *not* more pronounced in large MSAs. Substantiating this, **(d)** shows that when home tract interaction segregation is computed using an alternate ES measure so it captures only within-home-tract-homophily, it is no higher in large MSAs (Spearman ρ=-0.01, p>0.1). (The alternative ES measure is computed by subtracting the mean ES in each Census tract; see Methods.) Overall, this analysis suggests that the higher home tract segregation in large MSAs is driven by people's greater choice of neighborhoods of varying ES in which to live, but not by a greater tendency to interact homophilously within their own neighborhood.



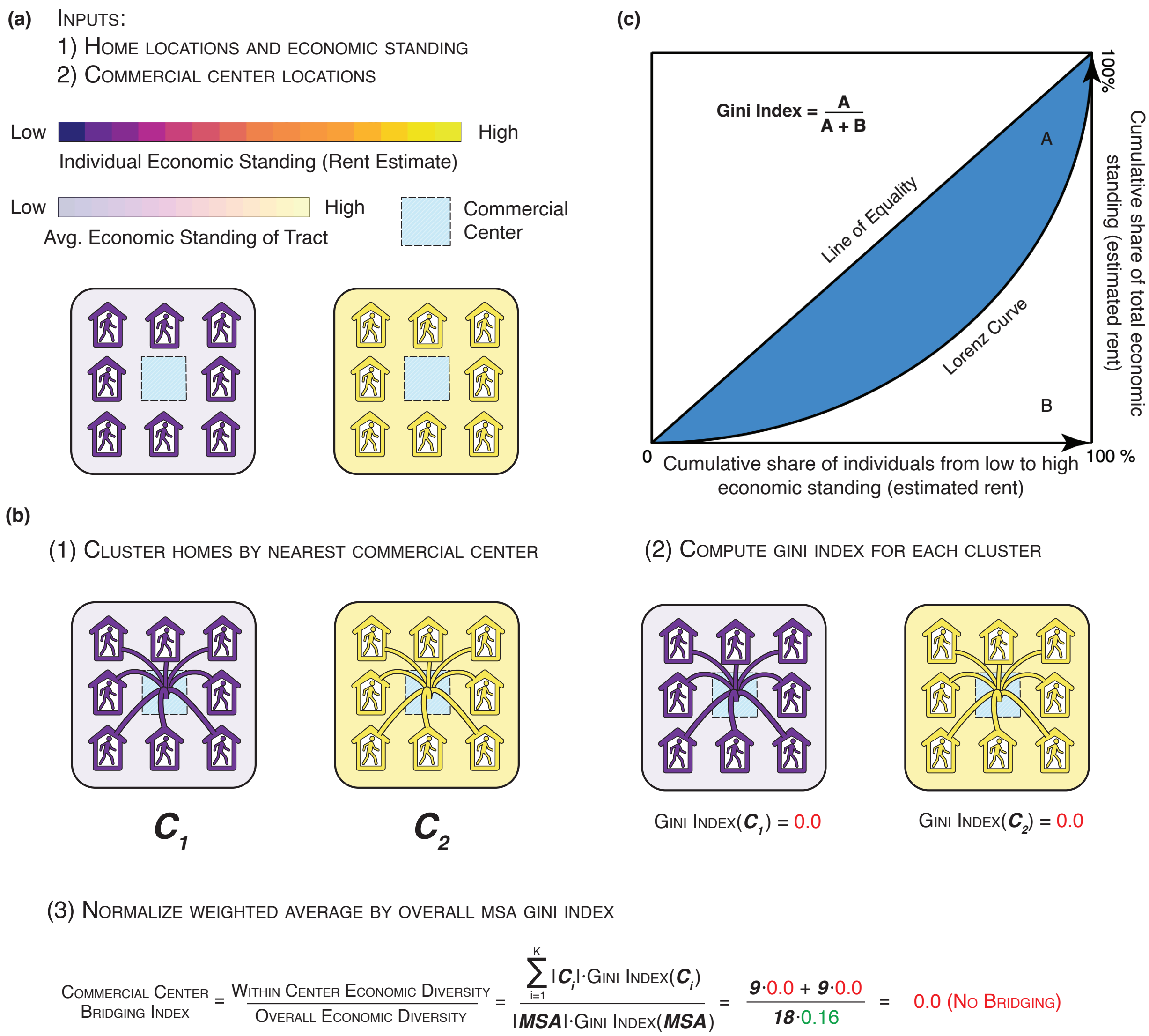
Extended Data Figure 6: Understanding the determinants of CCBI. The Commercial Center Bridging Index (CCBI) is a single metric which captures three important factors of built environment (see Supplementary Figure S11 for contributions of these factors to explaining Interaction Segregation):

(1) The locations of commercial centers — If commercial centers are located in between diverse neighborhoods, CCBI will be high as commercial centers will bridge together diverse individuals.

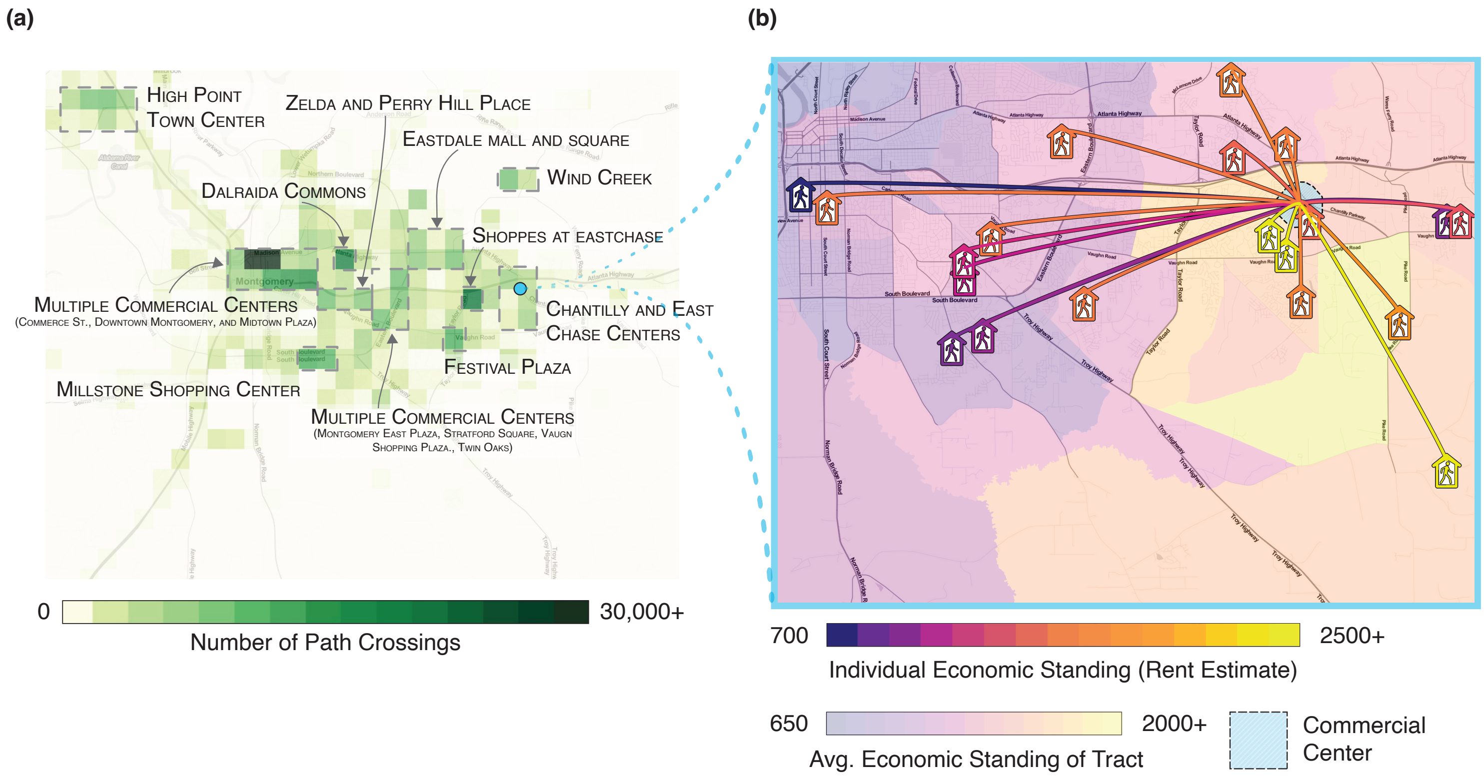
(2) The # of commercial centers — as # of commercial centers decreases, CCBI increases (e.g if there is only 1 commercial center in a city, CCBI will be 1.0 as all individuals are unified by a single commercial)

(3) Residential segregation, i.e. the locations of homes and their associated economic standing — as residential segregation decreases we can expect that individuals residing near each commercial center will be more diverse.

This figure builds intuition for CCBI by showing how CCBI may vary for a single simulated city, consisting of highly segregated neighborhoods. We hold residential segregation (3) constant, and vary the location (1) and number (2) of commercial centers across panels **(a)**, **(b)**, **(c)**, **(d)**, in order of increasing CCBI. Note that CCBI in **(c)** is substantially higher than CCBI in **(b)**, because commercial centers in **(c)** are better positioned to bridge diverse neighborhoods—even though the number of commercial centers remains constant.



Extended Data Figure 7: Computing Commercial Center Bridging Index (CCBI). Illustration of our analytical pipeline for calculating CCBI. **(a)** CCBI is computed from the locations and number of commercial centers in the MSA, as well the locations and economic standing values of all homes within MSA boundaries. We intentionally develop CCBI without using mobility data, with the intention of identifying a modifiable extrinsic aspect of an MSA that can be intervened on to *impact* mobility patterns and decrease Interaction Segregation **(b)** In order, we (1) cluster all homes by nearest commercial center (using straight line distance from home to commercial center), partitioning all homes into K clusters, where K is the number of commercial centers in the MSA (2) compute the Gini Index for each commercial center cluster, which aims to measure the aggregate economic diversity of those residing near the commercial center and (3) take the weighted average (by number of homes) of each cluster's Gini Index, normalized by the overall Gini Index of the MSA, to control for the positive association between Gini Index and segregation (Extended Data Table 1) **(c)** The graphical definition of Gini Index is provided, which is a standard measure of economic dispersion[44]. While we operationalize economic diversity by Gini Index, the strong correlation between CCBI and Interaction Segregation is robust to the definition of economic diversity, and holds true when using variance in ES instead of Gini Index (Supplementary Figure S12).



Extended Data Figure 8: Montgomery, AL. We conduct an analogous analysis to Figure 3a,b but for Montgomery, AL, which has nearly identical population (374K vs 385K residents) and income inequality (55th vs 60th percentile Gini Index) to Fayetteville, NC but is 74% more segregated (88th percentile vs. 21st percentile Interaction Segregation). We find that the difference in segregation is explained by Montgomery, AL having a significantly higher CCBI compared to Fayetteville, NC (65th vs. 13th percentile), i.e. in Montgomery, AL commercial center locations are differentiated by ES which results in high-ES individuals and low-ES individuals visiting separate commercial centers and prevents them from engaging in cross-ES interactions. (a) shows that, as with all MSAs, commercial centers account for the vast majority of interactions Commercial centers (e.g. shopping malls, plazas, etc.) are associated with high density of interactions. We illustrate that in Montgomery, AL all visually discernible hotspots are associated with one or more commercial centers. More generally, across all 382 MSAs, we find that the majority (56.9%) of interactions occur in close proximity (within 1km) of a commercial center, even though only 2.5% of land area of all MSAs is within 1km of a commercial center. (b) In Montgomery, AL, commercial centers are located in different locations which cater separate to high and low ES residents, leading to segregated interactions. As an illustrative example, we show a zoomed-in map of one commercial center (Chantilly Center) in Montgomery, AL, and display a random sample of 10 interactions occurring inside of it. Chantilly Center in Montgomery, AL is located accessibly for high ES individuals but is far apart from low-ES tracts. As a result, the sample shows that the majority of interactions are middle-upper ES, and only a few low-ES individuals visit Chantilly Center and interact with these high-ES individuals. Home icons demarcate individual home location (up to 100 meters of random noise added to preserve anonymity); home colors denote individual ES; arcs indicate an interaction inside of the commercial center; background colors indicate mean census tract ES.

	Dependent variable:				Interaction Segregation	
	(1)	(2)	(3)	(4)	(5)	(6)
Intercept	0.355***	0.355***	0.355***	0.355***	0.355***	0.355***
	(0.004)	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)
Log(Population Size)	0.059***		0.041***	0.044***	0.026***	0.028***
	(0.004)		(0.004)	(0.004)	(0.004)	(0.004)
Gini Index (Estimated Rent)		0.064***	0.050***	0.051***	0.045***	0.047***
		(0.004)	(0.004)	(0.004)	(0.003)	(0.003)
Political Alignment (% Democrat in 2016 Election)				0.004		0.004
				(0.004)		(0.004)
Racial Demographics (% non-Hispanic White)				0.001		0.006*
				(0.004)		(0.003)
Mean ES (Estimated Rent)				-0.012***		-0.005
				(0.004)		(0.004)
Walkability (Walkscore)					0.002	0.001
					(0.003)	(0.003)
Commutability (% Commute to Work)					-0.011***	-0.010***
					(0.003)	(0.004)
Conventional Segregation (NSI)					0.042***	0.041***
					(0.003)	(0.003)
Observations	382	382	382	376	382	376
R^2	0.350	0.419	0.567	0.578	0.704	0.705
Adjusted R^2	0.348	0.417	0.565	0.573	0.701	0.698

*p<0.1; **p<0.05; ***p<0.01

Extended Data Table 1: Population size is significantly associated with Interaction Segregation, after controlling for MSA income inequality (Gini Index), political alignment (% Democrat in 2016 election), racial demographics (% non-Hispanic White), mean ES, walkability (Walkscore[53]), commutability (% of residents commuting to work), and residential segregation (NSI). Here we show the coefficients (after normalizing via z-scoring to have mean 0 and variance 1) from the primary specifications estimating the effect of population size on Interaction Segregation across all MSAs. Columns (1-5) are models specified with different subsets of covariates; Column 6 shows model specification with all covariates. Differences between sample size in models is due to missing data for several covariates in a small number of MSAs (Walkscores were not available for all MSAs). (*p $<$ 0.1; **p $<$ 0.05; *** p $<$ 0.01).

	Dependent variable:			Interaction Segregation	
	(1)	(2)	(3)	(4)	(5)
Intercept	0.355***	0.355***	0.355***	0.355***	0.355***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Commercial Center Bridging Index	-0.078***	-0.059***	-0.058***	-0.035***	-0.036***
	(0.003)	(0.005)	(0.005)	(0.006)	(0.006)
Log(Population Size)		0.003	0.008*	0.010**	0.017***
		(0.004)	(0.005)	(0.004)	(0.006)
Gini Index (Estimated Rent)		0.031***	0.032***	0.035***	0.036***
		(0.003)	(0.003)	(0.003)	(0.003)
Political Alignment (% Democrat in 2016 Election)			0.001		0.002
			(0.004)		(0.004)
Racial Demographics (% non-Hispanic White)			0.003		0.005
			(0.003)		(0.003)
Mean ES (Estimated Rent)			-0.009**		-0.005
			(0.004)		(0.003)
Walkability (Walkscore)				0.002	0.001
				(0.003)	(0.003)
Commutability (% Commute to Work)				-0.011***	-0.009**
				(0.003)	(0.004)
Conventional Segregation (NSI)				0.028***	0.026***
				(0.004)	(0.004)
# of Commercial Centers					-0.006
					(0.005)
Observations	382	382	376	382	376
R^2	0.620	0.686	0.693	0.733	0.736
Adjusted R^2	0.619	0.684	0.688	0.729	0.729

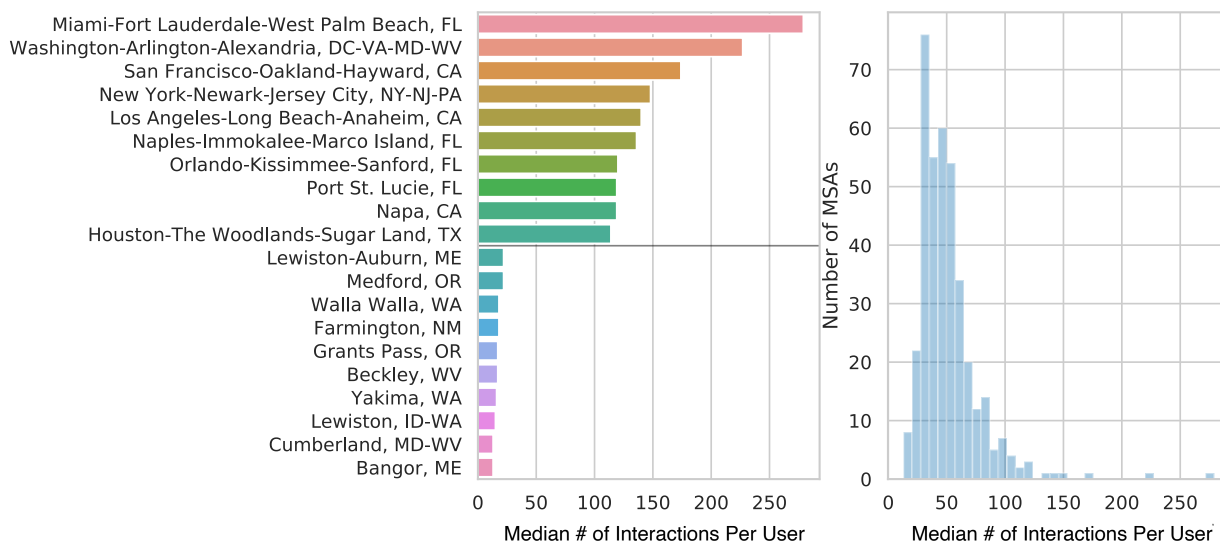
*p<0.1; **p<0.05; ***p<0.01

Extended Data Table 2: Commercial Center Bridging Index (CCBI) is significantly associated with Interaction Segregation, after controlling for population size, # of commercial centers, MSA income inequality (Gini Index), political alignment (% Democrat in 2016 election), racial demographics (% non-Hispanic White), mean ES, walkability (Walkscore[53]), commutability (% of residents commuting to work), and residential segregation (NSI). Here we show the coefficients (after normalizing via z-scoring to have mean 0 and variance 1) from the primary specifications estimating the effect of population size on Interaction Segregation across all MSAs. Columns (1-4) are models specified with different subsets of covariates; Column 5 shows model specification with all covariates. Differences between sample size in models is due to missing data for several covariates in a small number of MSAs (Walkscores were not available for all MSAs). (*$p < 0.1$; **$p < 0.05$; *** $p < 0.01$).

Supplementary Information

	Accurate pings	Unique days	Distinct user pairs interact	Interactions	Accurate pings	Distinct user pairs who interact	Interactions
count	8,609,406				382		
mean	3,273	35	184	363	73,757,695	2,577,322	4,845,144
std	16,507	20	374	1,073	163,848,305	8,872,464	16,838,938
min	11	2	1	1	2,196,084	27,326	53,350
10%	570	13	8	17	8,398,875	140,251	313,803
50%	1,471	30	76	141	22,054,930	504,525	1,031,691
90%	5,857	63	436	785	175,295,175	4,573,152	8,954,800
max	4,755,081	95	42,323	193,193	1,605,070,032	94,140,015	215,183,409

Supplementary Table S1: Combined descriptive statistics for all individuals residing in 382 Metropolitan Statistical Areas (MSAs). 8,609,406 individuals reside in a Metropolitan Statistical Area (90% of the overall 9,567,559 individuals in our study). The remaining 958,153 users live outside of MSAs, influencing the Interaction Segregation of an MSA by coming into contact with MSA residents. Descriptive statistics are grouped by individual (left) and MSA (right). At least one of two users in each interaction pair must live in an MSA to be included in this table.



Supplementary Figure S1: Descriptive statistics of path crossings.
(a) Ten Metropolitan Statistical Areas (MSAs) with the highest and lowest median users crossed per user.
(b) Overall distribution of median users crossed per user over MSAs.

display_name	# POIs (25%)	# POIs (50%)	# POIs (75%)	# POIs (max)	# POIs (mean)	# POIs (min)	# POIs (std)
Full-Service Restaurants	75.5	160.0	424.0	24,689.0	609.8	12.0	1,820.05
Snack Bars	18.0	40.0	110.0	6,266.0	169.76	1.0	511.17
Limited-Service Restaurants	33.0	60.0	145.5	4,847.0	192.14	5.0	434.78
Stadiums	1.0	2.0	4.0	43.0	3.67	1.0	4.32
Performing Arts Centers	1.0	2.0	4.0	28.0	3.25	1.0	3.41
Fitness/Recreation Centers	10.0	25.0	72.0	4,877.0	126.6	1.0	414.26
Historical Sites	1.0	2.0	7.0	206.0	9.16	1.0	21.65
Theme Parks	1.0	3.0	6.0	158.0	7.64	1.0	16.77
Bars/Drinking Places	2.0	5.0	13.0	447.0	19.52	1.0	45.91
Parks	3.0	6.0	17.0	793.0	28.69	1.0	80.44
Religious Organizations	7.0	16.0	41.25	2,644.0	63.28	1.0	196.97
Bowling Centers	2.0	4.0	8.0	204.0	9.77	1.0	20.52
Museums	1.0	3.0	6.0	137.0	6.78	1.0	13.99
Casinos	1.0	3.0	7.0	188.0	8.05	1.0	17.22
Independent Artists	1.0	2.0	5.0	130.0	7.55	1.0	17.42
Other Amusement/Recreation	1.0	2.0	7.0	525.0	10.17	1.0	36.13
Golf Courses and Country Clubs	2.0	3.0	7.0	101.0	8.07	1.0	13.77

Supplementary Table S2: POI descriptive statistics (# of POIs in each MSA) for each of the fine-grained POI categories in Figure 1e.

display_name	POI ES (25%)	POI ES (50%)	POI ES (75%)	POI ES (max)	POI ES (mean)	POI ES (min)	POI ES (std)
Full-Service Restaurants	1,210.96	1,395.0	1,674.27	3,628.06	1,493.04	763.0	430.99
Snack Bars	1,229.73	1,412.35	1,684.69	3,621.34	1,513.61	788.12	433.86
Limited-Service Restaurants	1,174.84	1,351.64	1,587.4	3,501.19	1,440.15	771.34	410.95
Stadiums	1,310.0	1,500.0	1,775.0	3,585.25	1,593.21	795.0	424.77
Performing Arts Centers	1,395.0	1,583.1	1,832.4	3,632.78	1,659.56	875.0	431.06
Fitness/Recreation Centers	1,230.03	1,431.79	1,703.94	3,749.05	1,528.73	700.0	453.4
Historical Sites	1,325.0	1,527.94	1,793.75	3,618.58	1,627.62	757.5	452.96
Theme Parks	1,300.0	1,498.75	1,750.0	3,900.0	1,612.58	700.0	501.79
Bars/Drinking Places	1,220.02	1,420.25	1,676.4	3,656.17	1,505.86	750.0	440.08
Parks	1,279.82	1,470.15	1,748.12	3,748.11	1,562.62	725.0	454.13
Religious Organizations	1,269.27	1,459.86	1,677.08	3,670.38	1,529.02	754.0	428.42
Bowling Centers	1,180.08	1,368.75	1,621.15	3,504.36	1,457.96	725.0	434.56
Museums	1,275.0	1,490.83	1,775.36	3,606.66	1,585.92	800.0	474.37
Casinos	1,200.0	1,400.0	1,655.54	3,606.17	1,503.88	725.0	469.68
Independent Artists	1,374.38	1,611.5	1,904.6	3,691.68	1,725.42	850.0	528.33
Other Amusement/Recreation	1,266.0	1,450.0	1,700.74	4,053.39	1,549.13	758.0	462.03
Golf Courses and Country Clubs	1,399.06	1,648.4	1,964.19	4,248.5	1,765.92	900.0	542.13

Supplementary Table S3: POI descriptive statistics (average POI economic standing in an MSA) for each of the fine-grained POI categories in Figure 1e. POI economic standing is operationalized as the median visitor ES of the POI.

display_name	IS (25%)	IS (50%)	IS (75%)	IS (max)	IS (mean)	IS (min)	IS (std)
Full-Service Restaurants	0.22	0.27	0.32	0.48	0.27	0.08	0.07
Snack Bars	0.2	0.25	0.31	0.5	0.25	0.01	0.08
Limited-Service Restaurants	0.24	0.29	0.34	0.47	0.29	0.04	0.08
Stadiums	0.14	0.17	0.22	0.36	0.18	0.02	0.06
Performing Arts Centers	0.14	0.16	0.19	0.27	0.17	0.05	0.05
Fitness/Recreation Centers	0.2	0.26	0.31	0.47	0.25	0.03	0.08
Historical Sites	0.15	0.2	0.27	0.43	0.21	0.0	0.09
Theme Parks	0.16	0.2	0.25	0.42	0.2	0.02	0.08
Bars/Drinking Places	0.18	0.23	0.3	0.42	0.23	0.06	0.08
Parks	0.19	0.26	0.33	0.47	0.26	0.05	0.09
Religious Organizations	0.24	0.32	0.38	0.55	0.31	0.05	0.1
Bowling Centers	0.16	0.21	0.26	0.44	0.22	0.03	0.08
Museums	0.18	0.22	0.28	0.45	0.24	0.06	0.08
Casinos	0.2	0.26	0.32	0.47	0.26	0.02	0.09
Independent Artists	0.13	0.2	0.27	0.39	0.21	0.02	0.09
Other Amusement/Recreation	0.18	0.25	0.31	0.71	0.25	0.02	0.12
Golf Courses and Country Clubs	0.33	0.41	0.5	0.62	0.4	0.2	0.11

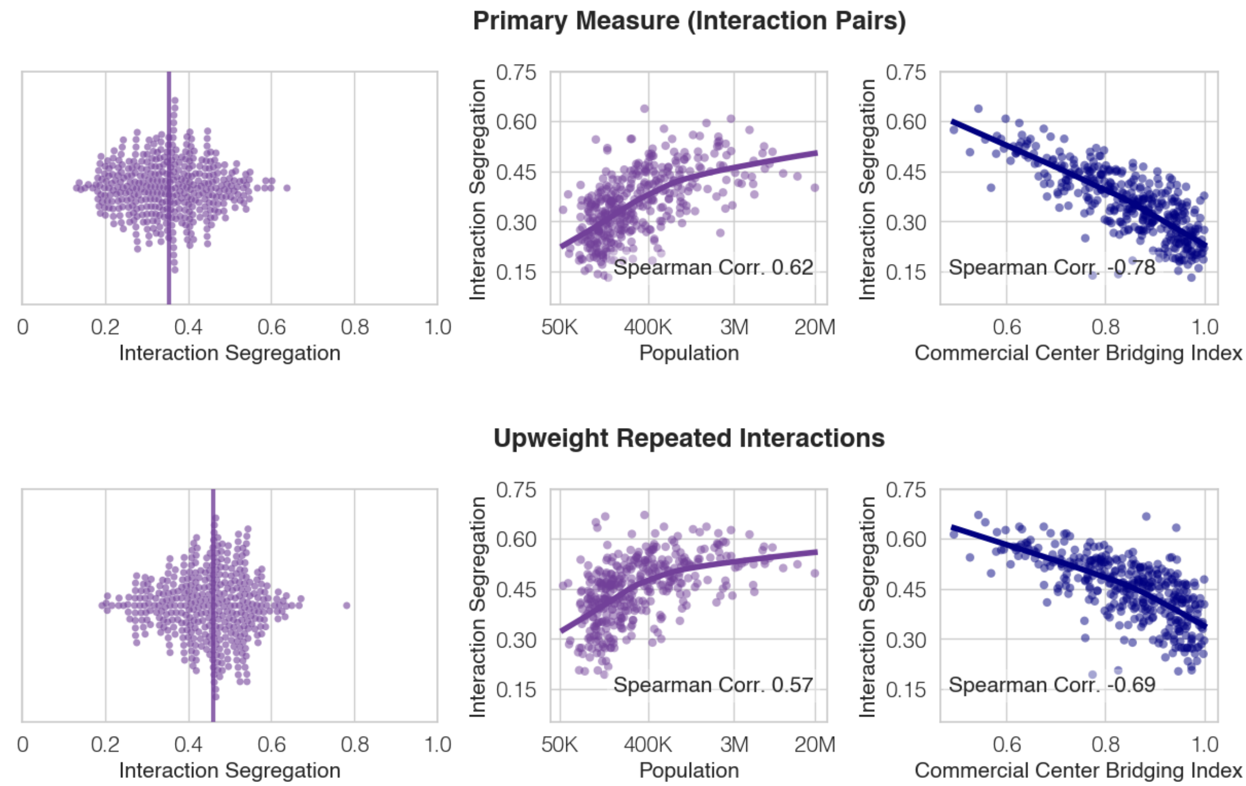
Supplementary Table S4: POI descriptive statistics (Interaction Segregation within-category) for each of the fine-grained POI categories in Figure 1e. Interaction Segregation is calculated for each POI category by filtering for only interactions which occurred inside of the POI category, before estimating Interaction Segregation (Methods).

display_name	# Interactions (25%)	# Interactions (50%)	# Interactions (75%)	# Interactions (max)	# Interactions (mean)	# Interactions (min)	# Interactions (std)
Full-Service Restaurants	23,060.75	54,219.5	156,645.0	19,540,673.0	398,304.04	6,112.0	1,634,147.68
Snack Bars	15,582.0	38,954.0	120,225.0	14,128,466.0	291,523.01	5,233.0	1,205,873.07
Limited-Service Restaurants	16,485.5	38,444.0	106,515.0	10,453,353.0	227,378.5	4,122.0	878,243.0
Stadiums	53,077.5	96,487.5	336,479.75	8,942,618.0	348,920.85	17,024.0	988,719.72
Performing Arts Centers	66,712.0	120,256.0	384,770.0	6,972,326.0	403,378.22	27,589.0	932,589.88
Fitness/Recreation Centers	11,165.25	21,541.5	62,380.5	5,630,299.0	158,025.53	3,740.0	573,788.93
Historical Sites	18,351.5	47,793.0	98,385.5	6,362,665.0	187,978.09	5,147.0	684,470.65
Theme Parks	34,989.0	61,553.0	135,744.0	1,883,136.0	157,622.73	14,460.0	290,329.93
Bars/Drinking Places	11,592.5	21,401.0	63,929.0	1,266,235.0	84,752.14	4,553.0	181,978.58
Parks	10,050.75	22,301.0	61,492.75	1,520,092.0	88,789.84	5,383.0	193,888.94
Religious Organizations	6,014.0	13,002.0	35,157.25	2,206,316.0	60,683.34	2,739.0	212,948.46
Bowling Centers	16,423.5	26,517.0	65,563.5	1,030,970.0	92,515.06	5,874.0	174,876.21
Museums	14,807.5	27,802.0	66,729.75	681,994.0	87,797.02	4,310.0	146,495.57
Casinos	13,844.0	23,109.0	60,222.0	826,676.0	68,012.52	7,474.0	124,063.14
Independent Artists	8,795.0	23,789.5	56,736.75	1,106,402.0	87,662.95	3,951.0	198,394.53
Other Amusement/Recreation	6,923.0	14,349.0	42,793.0	365,436.0	38,640.52	2,929.0	56,964.16
Golf Courses and Country Clubs	4,765.75	7,836.5	15,555.0	58,348.0	13,047.42	2,636.0	13,348.76

Supplementary Table S5: POI descriptive statistics (number of interactions occurring inside POI category) for each of the fine-grained POI categories in Figure 1e.

Interaction Segregation Measure	Pearson Corr. w/ Primary	Spearman Corr. w/ Primary	Median	Mean
Primary Measure	—	—	0.35	0.35
Primary Measure (+ Up-weight Multiple Interactions)	0.89	0.91	0.46	0.45
ES Definition: Rent Zestimate Percentile	0.78	0.80	0.46	0.45
ES Definition: Within-MSA Rent Zestimate Percentile	0.81	0.83	0.54	0.53
ES Definition: Census Median Household Income	0.75	0.77	0.47	0.46
Exclude Pri/Sec Roads	0.99	0.99	0.37	0.37
Exclude Roads	0.98	0.98	0.37	0.37
Exclude Same-home interactions	0.98	0.98	0.34	0.34
Work/Leisure (Neither in Home Tract)	0.93	0.93	0.31	0.31
Leisure (inside POI)	0.85	0.84	0.28	0.29
Minimum Distance Between Pings: $<$ 25 meters	0.98	0.99	0.36	0.36
Minimum Distance Between Pings: $<$ 10 meters	0.95	0.96	0.37	0.36
Minimum Time Between Pings: $<$ 2 minutes	0.99	0.99	0.36	0.36
Minimum Time Between Pings: $<$ 60 seconds	0.99	0.99	0.36	0.36
Minimum Tie Strength: 2 consecutive interactions	0.94	0.95	0.35	0.35
Minimum Tie Strength: 3 consecutive interactions	0.83	0.83	0.37	0.37
Minimum Tie Strength: 2 unique days of interaction	0.88	0.90	0.47	0.46
Minimum Tie Strength: 3 unique days of interaction	0.73	0.76	0.56	0.54
Dist. $<$ 25 meters, Time $<$ 2 min., $>=$ 2 consec. interactions	0.93	0.94	0.35	0.35
Dist. $<$ 25 meters, Time $<$ 2 min., $>=$ 2 unique days	0.88	0.89	0.46	0.45
Dist. $<$ 10 meters, Time $<$ 60 sec., $>=$ 3 consec. interactions	0.80	0.80	0.38	0.37
Dist. $<$ 10 meters, Time $<$ 60 sec., $>=$ 3 unique days	0.73	0.75	0.52	0.50

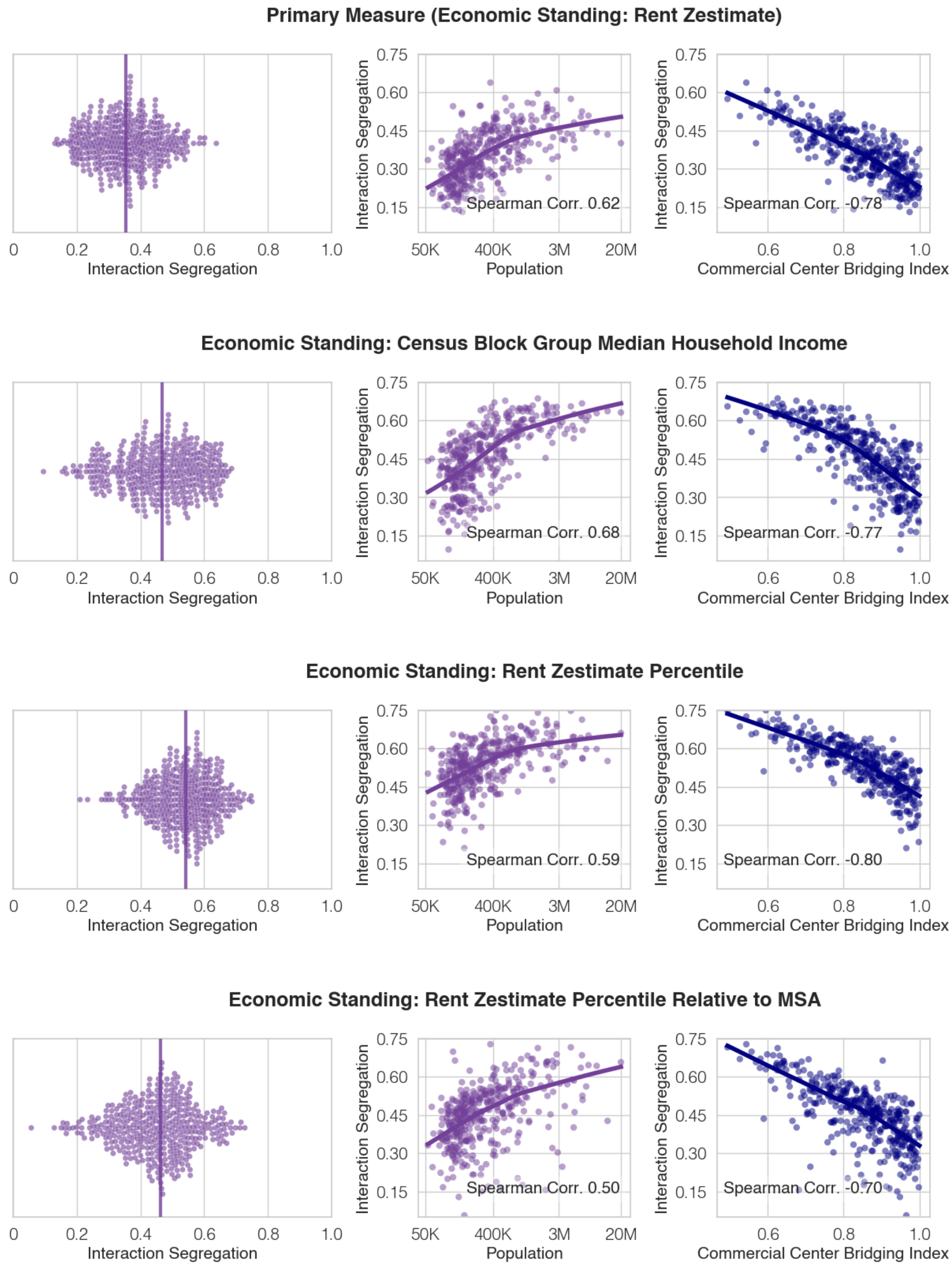
Supplementary Table S6: Robustness checks overview. We find that our definition of Interaction Segregation is robust to varying many parameters: weighting of repeated interactions between the same users, definition of economic standing, inclusion/exclusion of roads and same-home interactions, filtering location of interaction, minimum distance, minimum time, and minimum tie strength (as well as the intersection of distance, time, and tie strength). The above variants all are strongly correlated to our primary measure (all have Spearman Corr. $>=$ 0.75). We also find that our primary findings that (1) large, dense cities facilitate segregation and (2) commercial center locations accessible to diverse individuals may mitigate segregation are robust across all definitions of Interaction Segregation (Supplementary Figures S2-S7). Note that we exclude same-home interactions in robustness checks that vary minimum time, distance, or require repeated interactions, to ensure that results are not influenced by interactions with members of the same household (these interactions ordinarily have minimum influence on Interaction Segregation, as shown by the robustness check which excludes same-home interactions and results in virtually identical metric (Spearman Corr. 0.98); however, the influence of same-home interactions is higher after more conservative filters are applied to the definition of interactions, such as requiring a minimum tie strength of 3 consecutive interaction).



Supplementary Figure S2: Robustness of primary study findings to weighting of repeated interactions. We find that our primary study findings that, (1) large, dense cities facilitate segregation and (2) commercial center locations accessible to diverse individuals may mitigate segregation, are robust to the choice of whether to upweight repeated interactions in our interaction network. We compare the results of:

Primary Measure: Interactions are defined as pairs of users who have ever interacted within the study observation window (three months of 2017). We deduplicate repeated interactions, as frequency of pings varies across smartphone users, to reduce bias from users with a higher frequency of pings. For instance, if an individual A with an individual B (ES $1000) two times and individual C (ES $2000) one, we compute the mean ES of individual A's network as $1500.

Upweight Repeated Interactions: Repeated interactions are unweighted when calculating the mean ES of an individual's interaction network. For instance, if an individual A with an individual B (ES $1000) two times and individual C (ES $2000) once, we compute the mean ES of individual A's network as $1333.



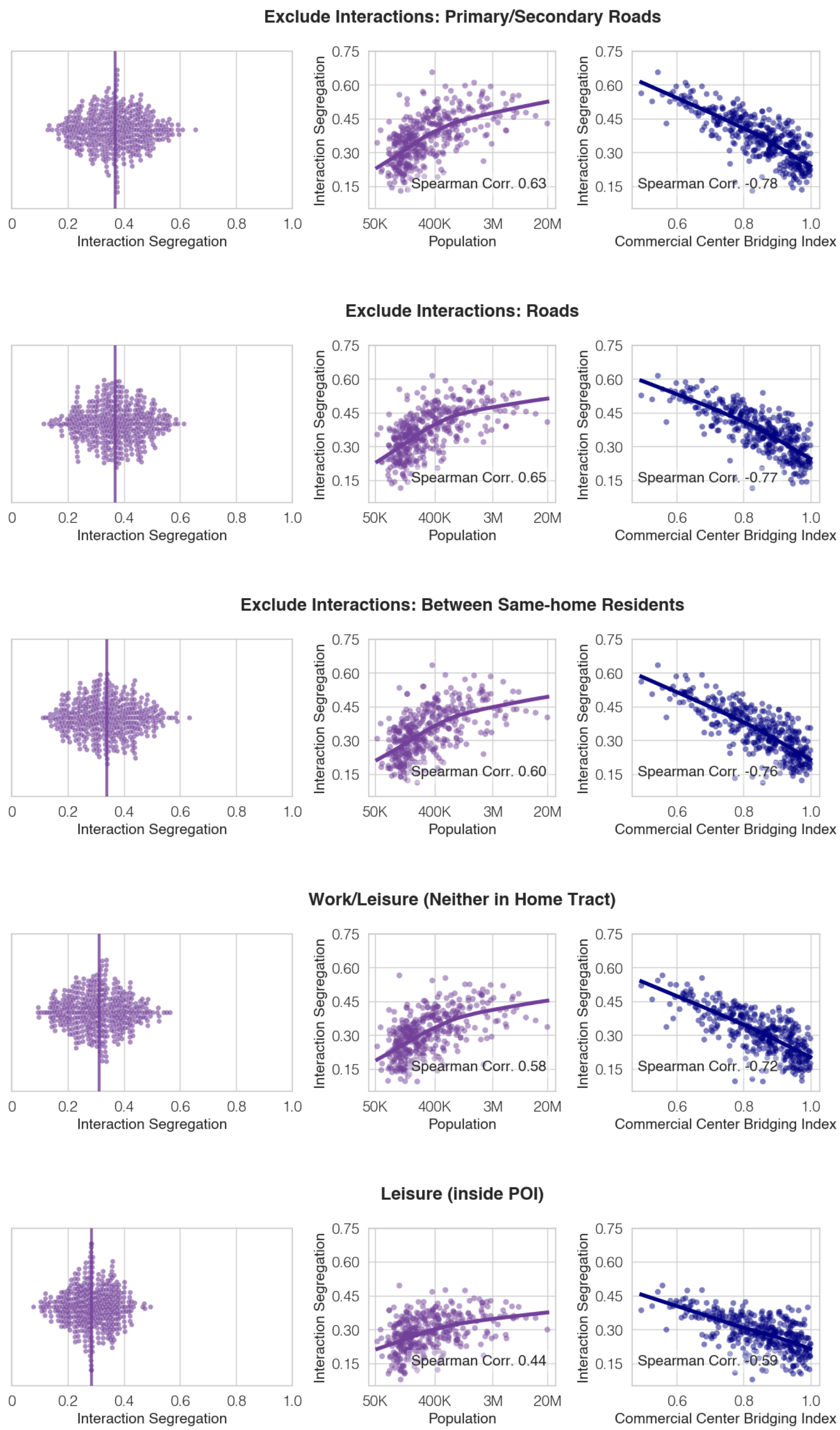
Supplementary Figure S3: Robustness of primary study findings to definition of economic standing. We find that our primary study findings that, (1) large, dense cities facilitate segregation and (2) commercial center locations accessible to diverse individuals may mitigate segregation, are robust to the definition of economic standing. We compare the results of:

Primary Measure: Our primary measure leverages estimated monthly rent value (Zillow Rent Zestimate).

Census Block Group (CBG) Median Household Income: We define the ES of an individual is the median household income in the CBG in which they reside.

Rent Zestimate Percentile: We normalize Rent Zestimate values across all individuals.

Primary measure Relative to MSA: We normalize Rent Zestimate values across all individuals within an MSA, independent of other MSAs, to account for differences in cost of living across cities.



Supplementary Figure S4: Robustness of primary study findings to exclusion of interactions within roads, exclusion interactions with residents of the same home, and exclusion of non-work/leisure interactions. We find that our primary study findings that, (1) large, dense cities facilitate segregation and (2) commercial center locations accessible to diverse individuals may mitigate segregation, are robust to filtering for a subset of interactions. We compare the results of:

Primary Measure: Our primary measure includes all interactions, aiming to give a complete account of an individual's interaction network including path crossings on roads as well as those they share a home with.

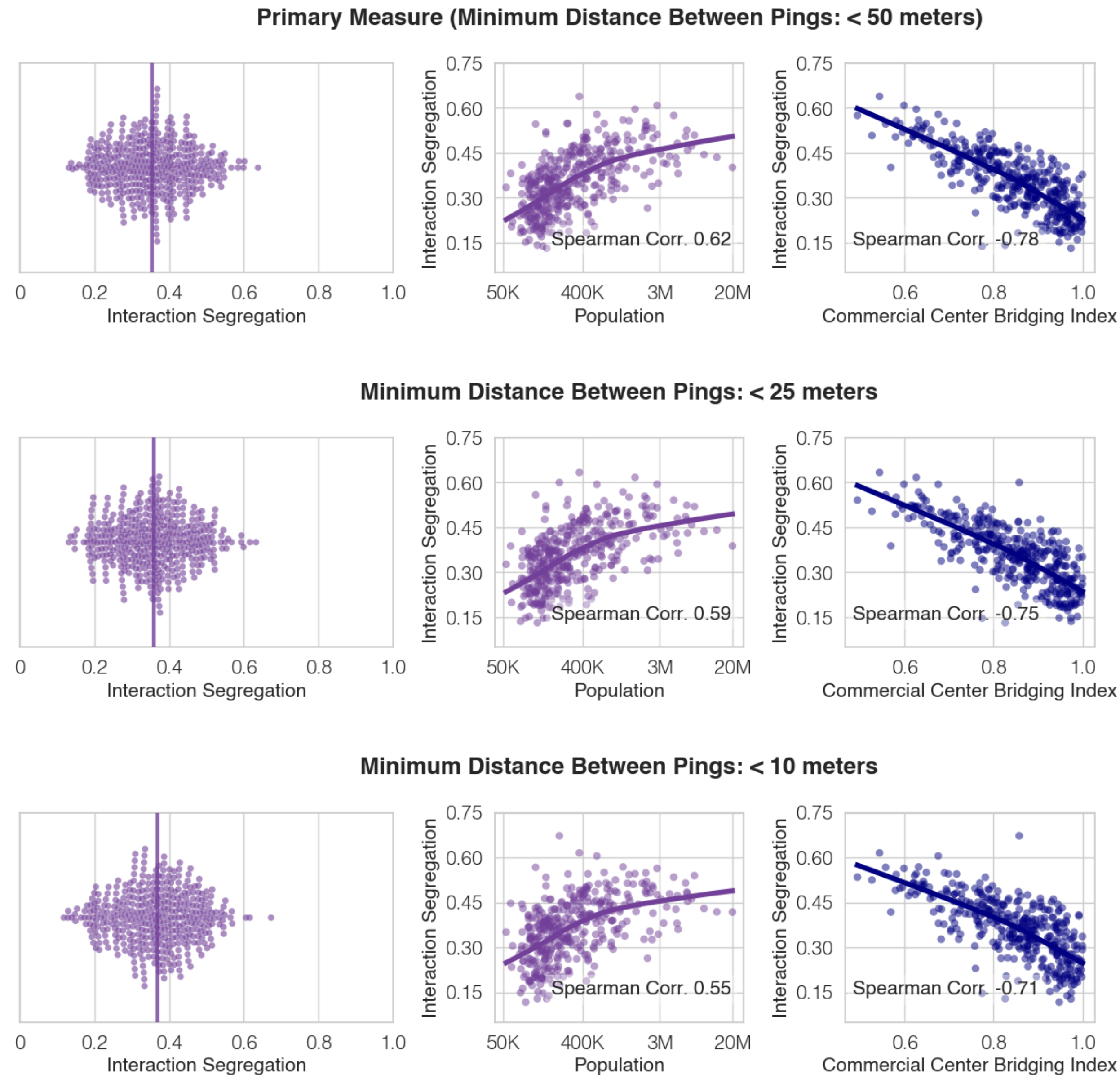
Excluding roads: We define the ES of an individual is the median household income in the CBG in which they reside.

Rent Zestimate Percentile: We normalize Rent Zestimate values across all individuals.

Primary measure Relative to MSA: We normalize Rent Zestimate values across all individuals within an MSA, independent of other MSAs, to account for differences in cost of living across cities.

Work/Leisure: We filter to include only interactions likely to take place in the context of work or leisure, by excluding interactions which occurred when either individuals were located within their home tracts.

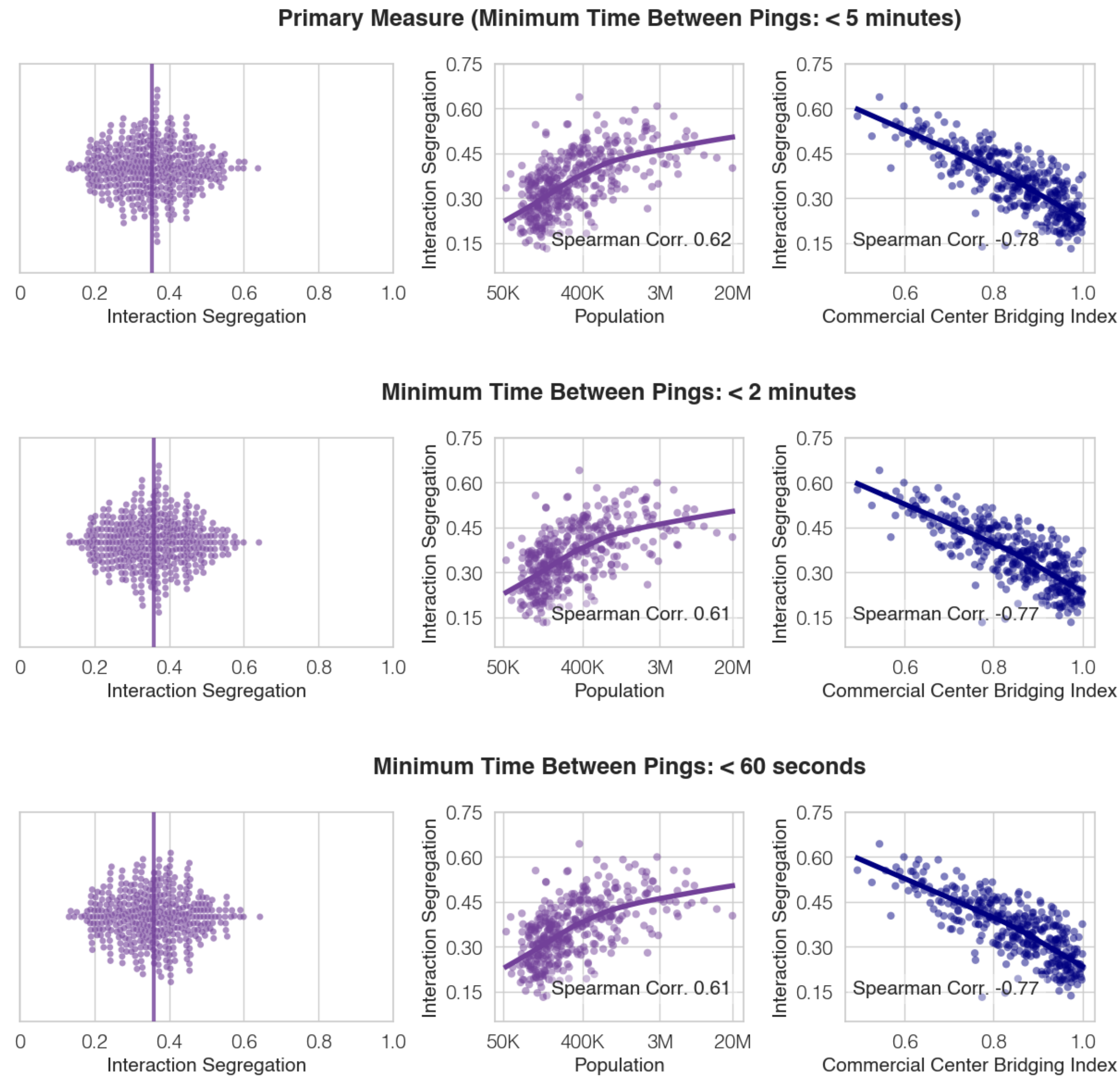
Leisure: We filter for leisure interactions by including only interactions ocurring inside of the POIs categorized as related to leisure (Figure 1e).



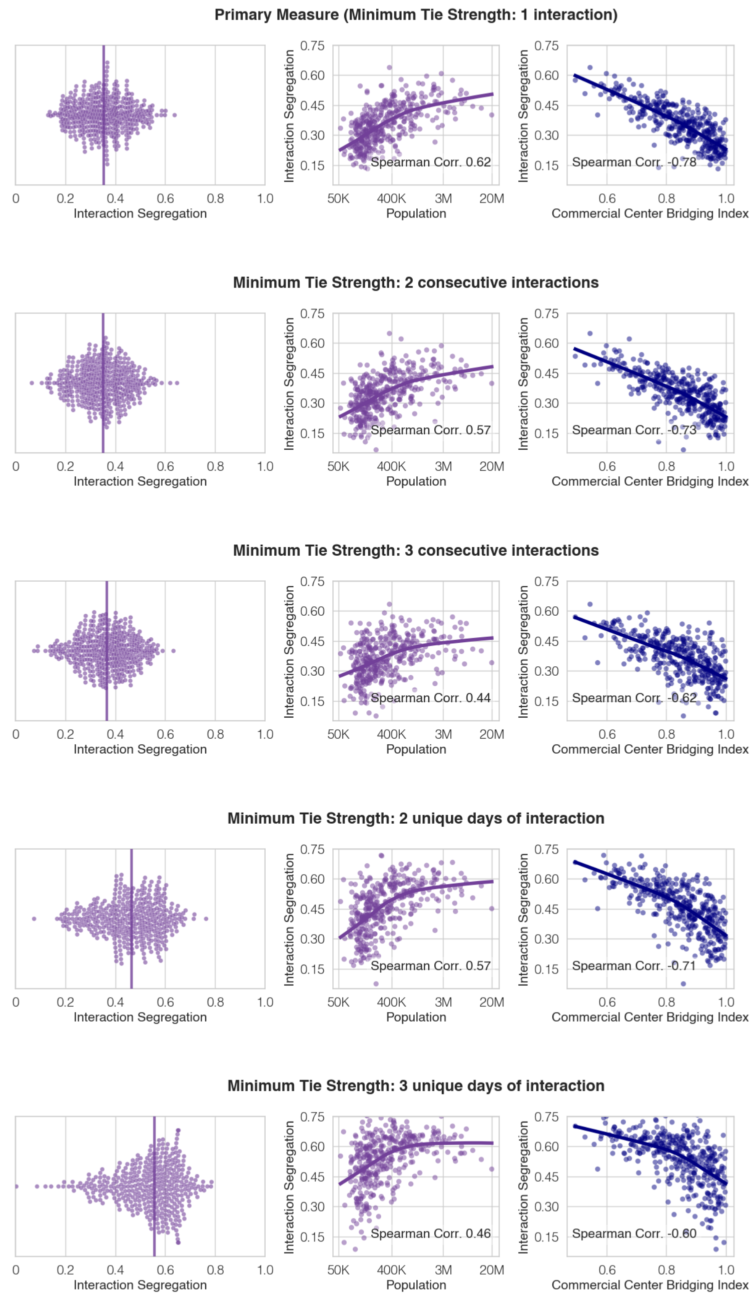
Supplementary Figure S5: Robustness of primary study findings to minimum distance required between two GPS pings for individuals to be considered interacting. We find that our primary study findings that, (1) large, dense cities facilitate segregation and (2) commercial center locations accessible to diverse individuals may mitigate segregation, are robust to the time threshold used in our definition of interaction:

Primary Measure: Our primary measure uses a threshold of 50 meters, based on prior literature which shows that even distant exposure to diverse individuals is predictive of long-term behaviors[10].

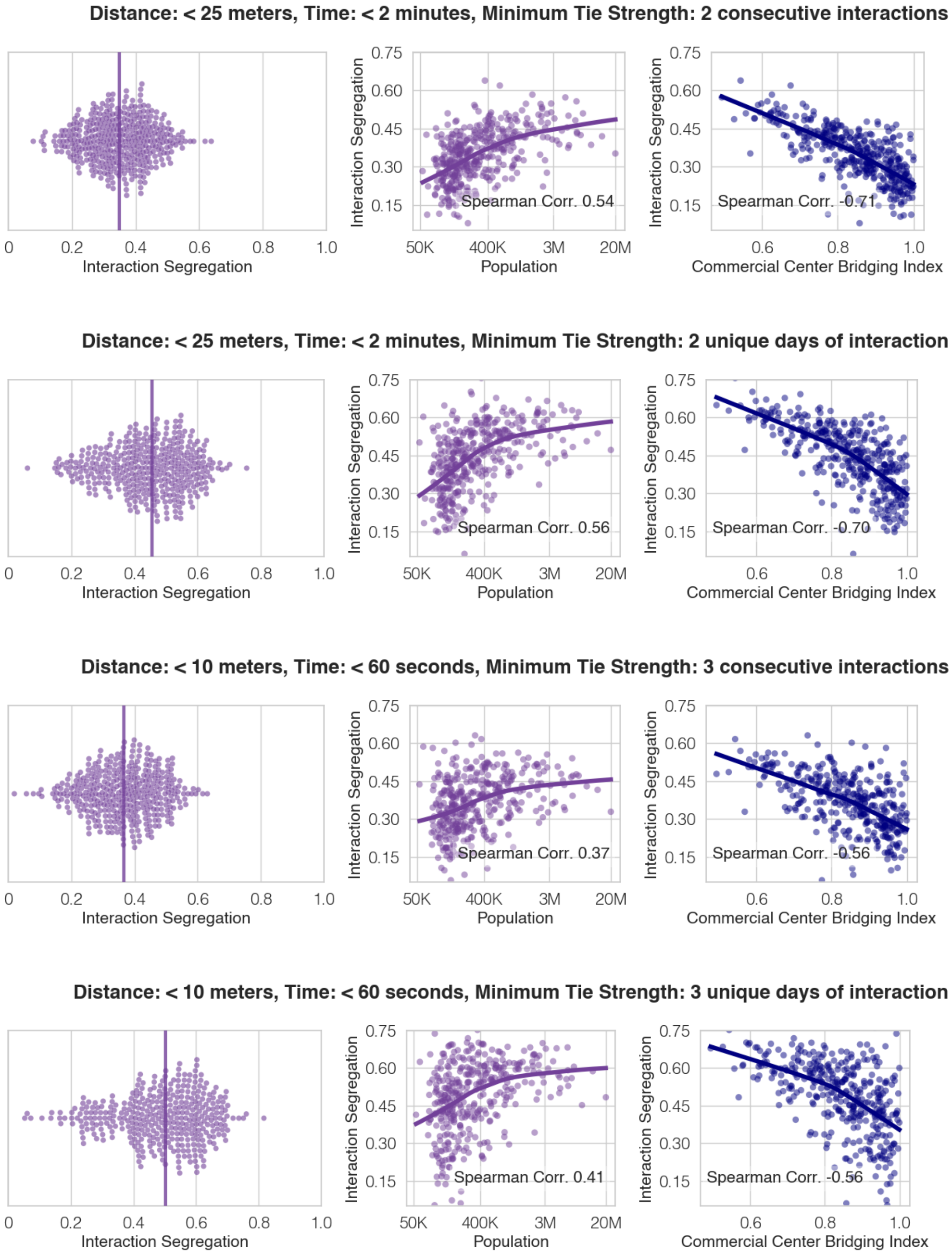
Alternative measures: We alternatively consider more conservative thresholds of 25 meters and 10 meters, with 10 meters being the lowest threshold due to limitations of GPS ping accuracy[54,55].



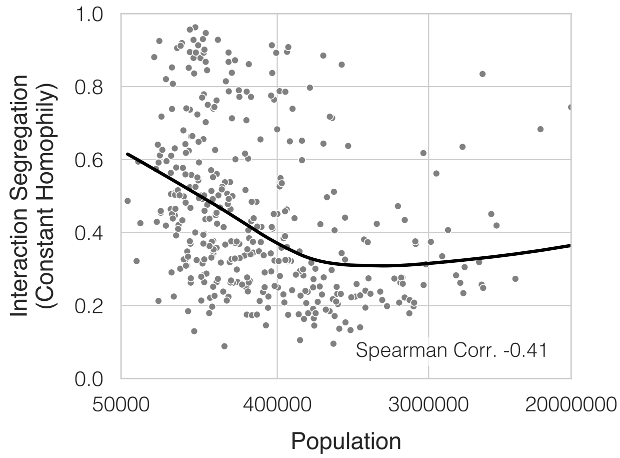
Supplementary Figure S6: Robustness of primary study findings to minimum time elapsed between two pings to constitute an interaction. We find that our primary study findings that, (1) large, dense cities facilitate segregation and (2) commercial center locations accessible to diverse individuals may mitigate segregation, are robust to the time threshold used in our definition of interaction:
Primary Measure: Our primary measure uses a threshold of 5 minutes, to be inclusive of users with sparse pings (e.g., for a subset of users, we only have 1 ping per day, while for others we have 100+ pings per day) while maintaining a reasonable confidence that an interaction may have occurred.
Alternative measures: We alternatively consider more conservative thresholds of 2 minutes and 1 minute.



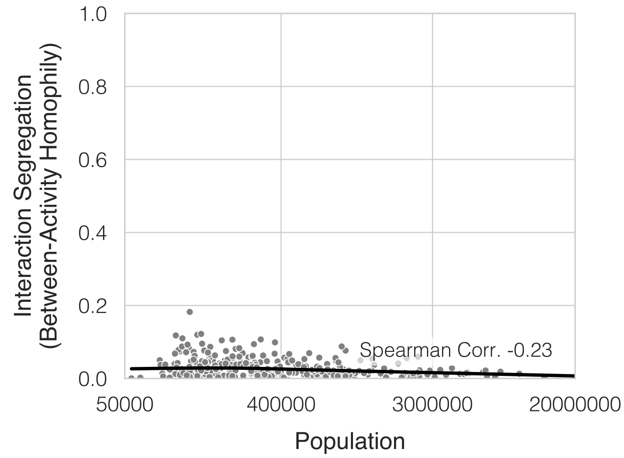
Supplementary Figure S7: Robustness of primary study findings to minimum tie strength required to constitute an interaction. We find that our primary study findings that, (1) large, dense cities facilitate segregation and (2) commercial center locations accessible to diverse individuals may mitigate segregation, are robust regardless of the minimum tie strength threshold between two individuals to be constitute an interaction:
Primary Measure: Our primary measure only requires a single pair of pings between users to constitute an interaction, to be inclusive of users with sparse pings (e.g., for a subset of users, we only have 1 ping per day, while for others we have 100+ pings per day).
Alternative measures: We alternatively consider more conservative thresholds of 2 or 3 consecutive interactions, as well as 2 or 3 interactions across unique days. Requiring consecutive interactions increases the likelihood that individuals actually came into contact together; interactions across unique days increases the likelihood that interactions are not merely path crossings, but social interactions between individuals who are familiar with each other. **(continued on next page)**



Supplementary Figure S8: (continued from previous page). We find that our primary study findings that, (1) large, dense cities facilitate segregation and (2) commercial center locations accessible to diverse individuals may mitigate segregation, are *robust to the combination of the minimum time, minimum distance, and minimum tie strength threshold parameters*. To account for interactions between threshold parameters, we also consider combinations of parameter variants. For instance, the most conservative robustness check defines an interaction as two individuals being < 10 meters apart within a < 60 second window, and for this to have occurred either for either 3 consecutive minutes (second figure from the bottom) or across 3 unique days (bottom figure).



(a)

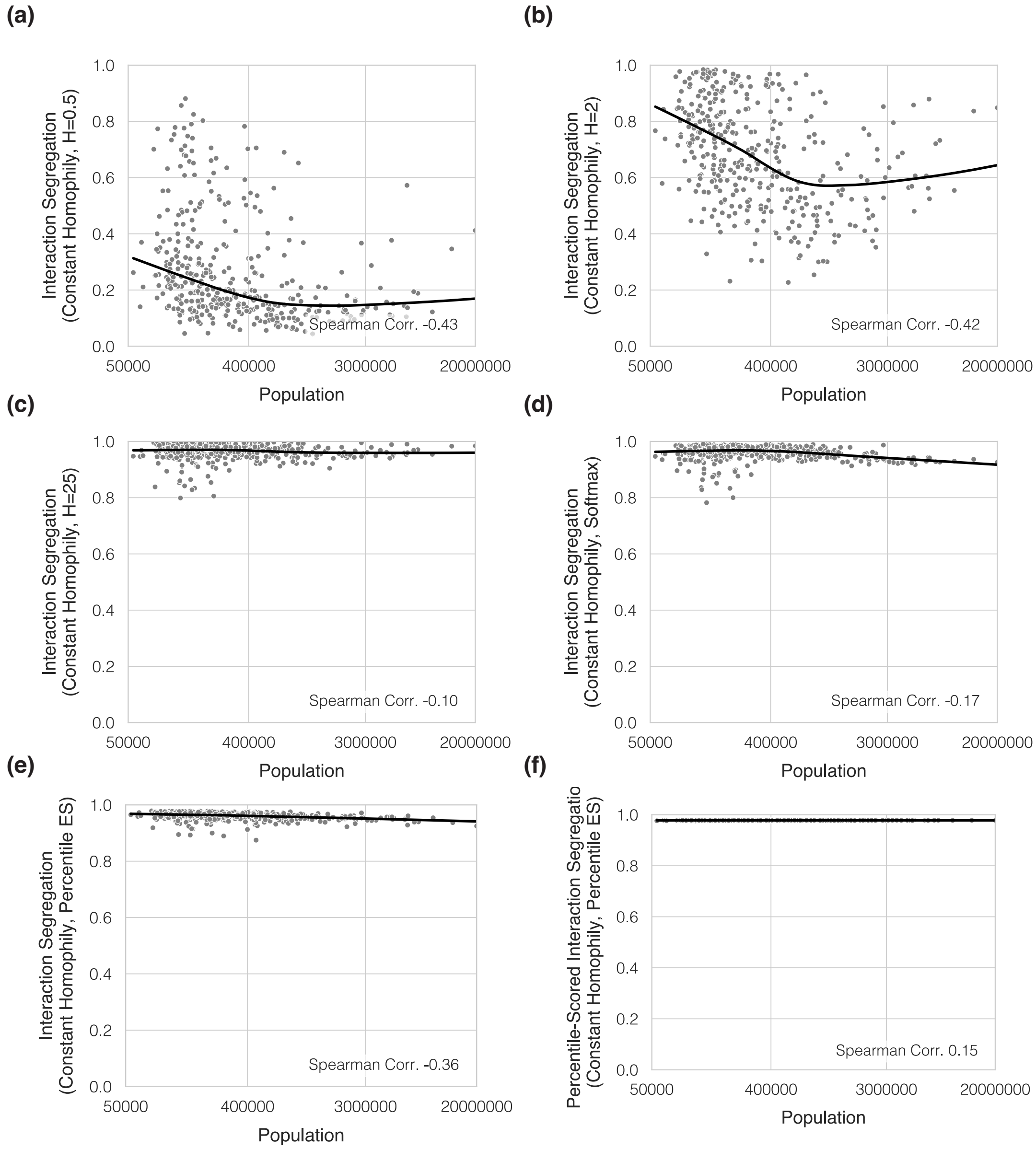


(b)

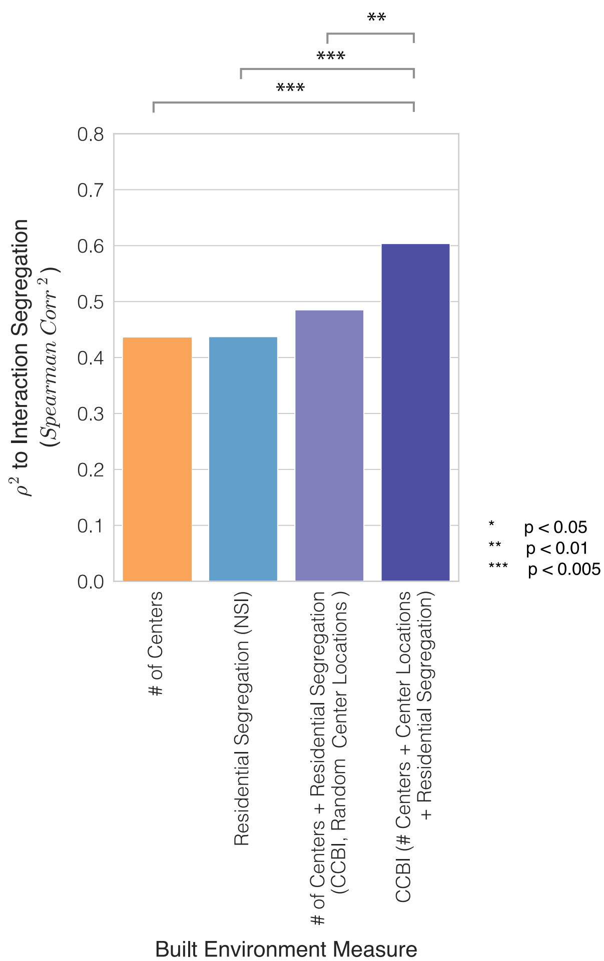
Supplementary Figure S9: Alternative homophily mechanisms do not explain segregation in large cities. We consider the possibility of two alternative hypotheses which may explain the trend towards high segregation in large cities

(a) *Constant Homophily*: i.e. individuals have the same proclivity for interacting with individuals of similar ES regardless of if they live in large or small cities, and it is instead change in distribution of economic standing that drives segregation in large cities (e.g. in large cities there may be a greater supply of people in the same economic class available to interact with). We test this hypothesis via a null network model, in which which we preserve network nodes (individuals and their ES values) but randomize edges[56,57]. We randomly assign interactions between pairs of people, weighting the likelihood of interaction between people of similar ES higher according to a constant homophily function. Specifically, the probability of interaction ($p_{i,j}$) between two individuals of ES_i and ES_j is weighted by their similarity in ES, defined as the complement of the normalized Euclidean distance in ES: $p_{i,j} \propto Similarity(ES_i, ES_j) = 1 - \frac{|ES_i - ES_j|}{max(ES) - min(ES)}$. We choose 75 interactions per person such that the mean number of interactions per person is 150, which corresponds to Dunbar's number[58]. We find that under this null model, there is no positive association between interaction segregation and population size; in fact, larger cities are less segregated on average, as there is an increase in supply of diverse individuals in economic standing in larger cities. These findings are also robust to a variety of null model specifications (Supplementary Figure S10).

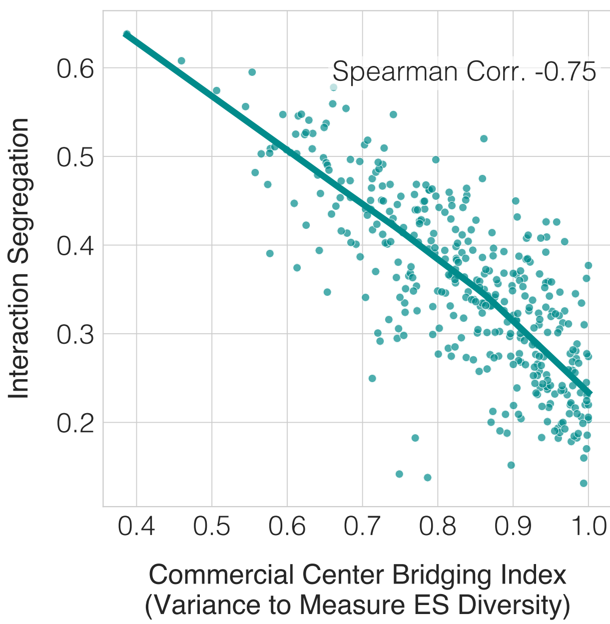
(b) *Between Activity Homophily*: i.e. it is not the differentiation of individual venues that drives segregation, but rather that in large cities individuals choose different categories of activities which results in segregation (e.g. in small cities, there are less country clubs so everybody visits restaurants to socialize, whereas in large cities high-ES individuals segregate by spending a higher proportion of time in exclusive venues such as country clubs). We test this hypothesis via a configuration model[56,57], a prominent null network model in which node degree is preserved. Specifically, by applying a configuration model to reconfigure network edges for each leisure category separately, we preserve network nodes (individuals and their ES values) as well as the number of interactions they had in each category of POI (node degree), but randomize the specific venue in which each interaction occurred. For instance, if an individual interacts with 5 people inside of restaurants and 100 people inside of a fitness center, they will be randomly assigned to interact with 5 people from all of those who visited restaurants, and 100 people from all of those who visited fitness centers. This null model preserves between-activity homophily which results from activity choices (e.g. whether to visit a country club or restaurant), but erases within-activity homophily (e.g. individuals who visit any restaurant are equally likely to interact). We find that under this null model, there is no positive association between interaction segregation and population size; in fact, there is minimal segregation across all cities as variation between activity categories is insufficient to retain segregation. This is further supported by Supplementary Table S3, which shows relatively small differences in ES between participants in different categories of leisure activity (e.g. the lowest ES activity, limited service restaurants has a median visitor ES of $1,352, the highest ES activity, golf courses and country clubs has a median visitor ES of $1,648.4).



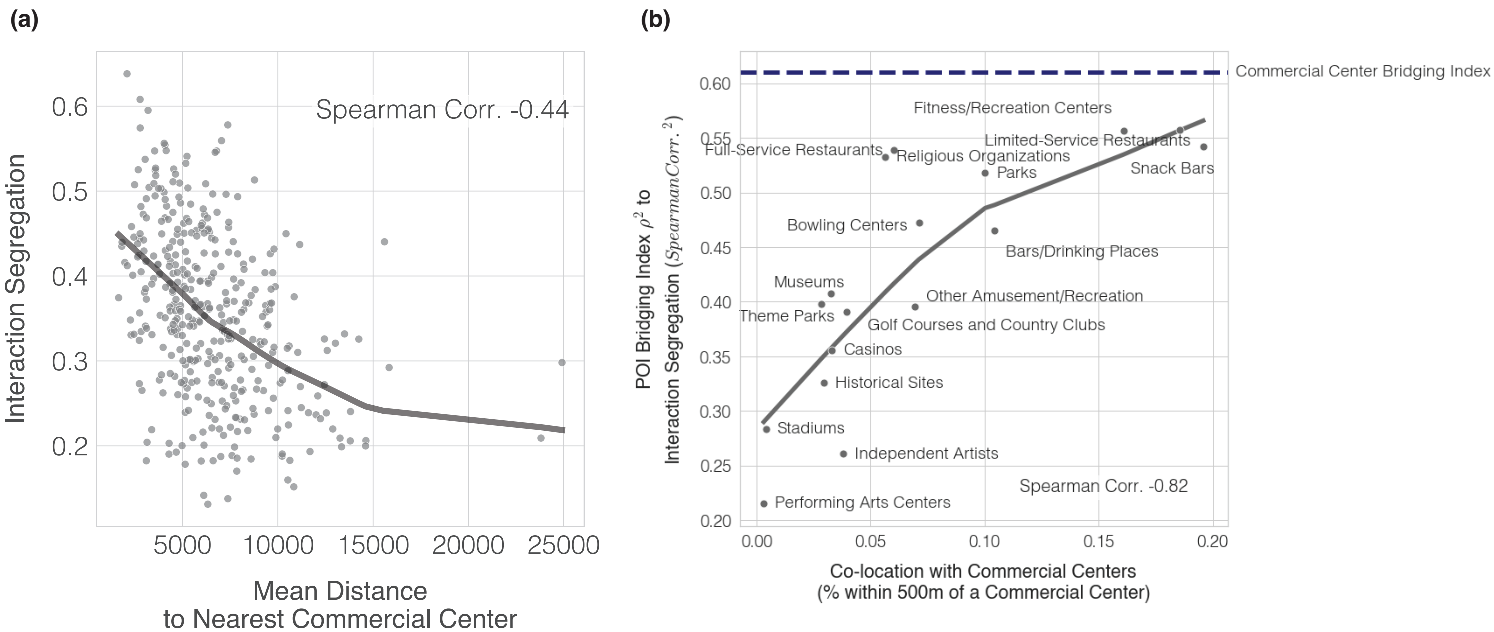
Supplementary Figure S10: Baseline Homophily null model results are robust to varying null model specifications. We re-run the analysis in Supplementary Figure S9a under a variety of null model specifications, and find that *in all cases there is no evidence to suggest that the Constant Homophily hypothesis explains the high segregation observed in large cities*. **(a-c)** We first consider varying the extent of homophily by adding a constant parameter H to the homophilous weighting of edges, to exponentially increase/decrease the extent of homophily in our null model for the probability of individuals i and j interacting: $p_{i,j} \propto Similarity(ES_i, ES_j)^H$. We find that regardless of if we **(a)** decrease homophily (H=0.5) or **(b-c)** increase homophily mildly (H=2) or strongly (H=25), there is no positive association between population size and segregation in our simulations. In fact, larger cities are less segregated on average, as there is an increase in supply of diverse individuals in economic standing in larger cities. We also consider alternative null model specifications such as **(d)** a softmax homophily function $p_{i,j} \propto \frac{e^{Similarity(ES_i,ES_j)}}{\sum_{k=1}^{N} e^{Similarity(ES_i,ES_k)}}$, **(e)** applying the original null model to percentile-scored values economic standing **(f)** applying the original null model to percentile-scored values economic standing, and calculating Interaction Segregation using percentile-scored values economic standing. This suggests that the high segregation in large cities is due to a change in resident behavior, facilitated by the built environment of large cities, and not an artifact the economic standing distribution in large cities.



Supplementary Figure S11: Understanding why CCBI explains Interaction Segregation. We show via an ablation study that commercial center locations, in addition to number of centers and residential segregation, contributes to the explanatory power of CCBI. As illustrated in Extended Data Figure 6, CCBI captures three factors of built environment: (1) locations of commercial centers (2) number of centers and (3) residential segregation. In this analysis, we aim to disentangle how these three factors contribute to the ability of CCBI to explain Interaction Segregation (as measured by ρ^2, the squared Spearman correlation with Interaction Segregation). We find that number of commercial centers (orange, $\rho^2 = 0.436$) and residential segregation (blue, $\rho^2 = 0.437$) are each correlated with Interaction Segregation. To measure the combined explanatory power of these two factors within CCBI, independent of commercial center locations, we conduct an ablation study in which we calculate Commercial Center Bridging Index for each MSA, using the actual home location data and number of commercial centers for each MSA, but randomize commercial center locations (light purple, $\rho^2 = 0.486$). For each MSA, we estimate this value over 1000 random trials. We find that calculating CCBI using randomized commercial center locations is a significantly weaker predictor (p=0.0008<0.01, Fisher's z-test) compared to CCBI values computed using actual commercial center locations (dark purple, $\rho^2 = 0.604$). This demonstrates that commercial center locations contributes to the explanatory power of CCBI, i.e. CCBI explains Interaction Segregation because it captures the extent to which the locations of commercial centers in different cities facilitate the interaction of diverse individuals.

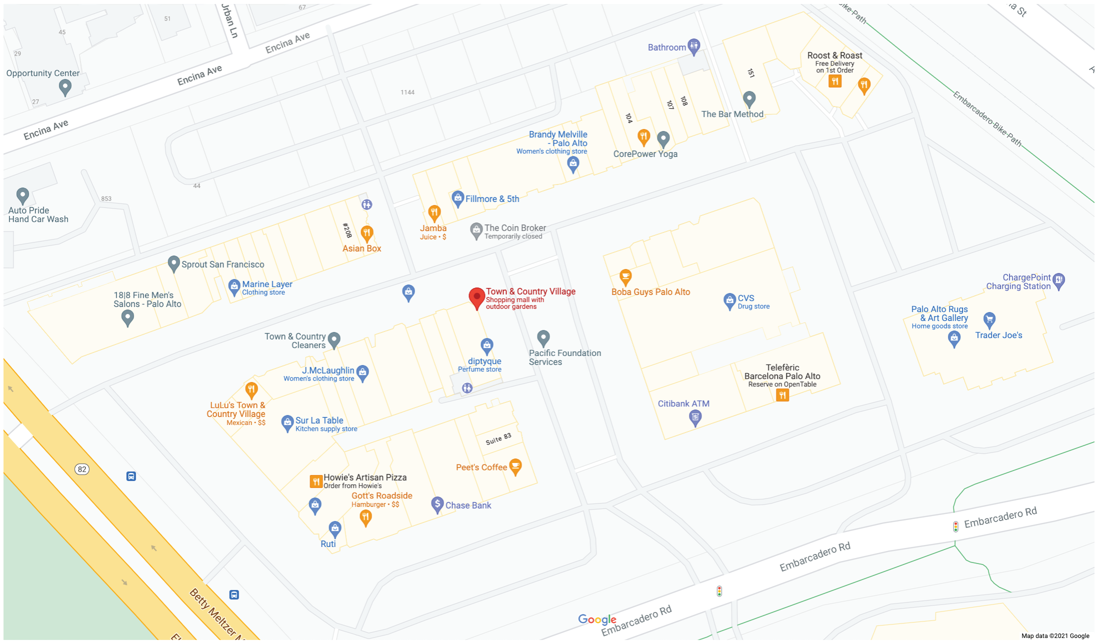


Supplementary Figure S12: Robustness of CCBI to definition of ES diversity. We calculate a version of Commercial Center Bridging Index which uses variance in economic standing to operationalize economic diversity: $Commercial\ Center\ Bridging\ Index\ (CCBI) = \frac{\sum_{i=1}^{K} |\mathcal{C}_i| \cdot Var(\mathcal{C}_i)}{|\mathcal{V}_{MSA}| \cdot Var(\mathcal{V}_{MSA})}$. This variant of CCBI explains Interaction Segregation comparably (Spearman Corr. -0.75 vs. -0.78) to our primary measure of CCBI which uses Gini Index to operationalize economic standing diversity for each commercial center. Thus, we find that the ability of Commercial Center Bridging Index to explain Interaction Segregation is robust to the definition of ES diversity.

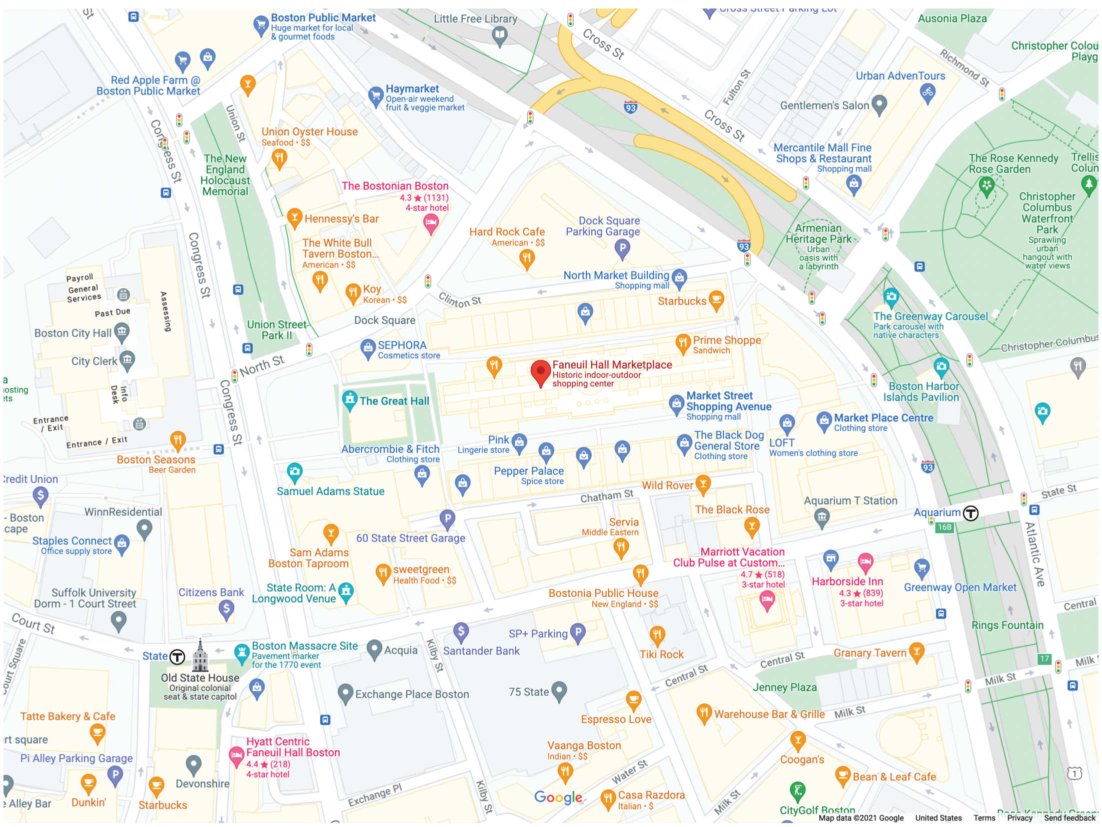


Supplementary Figure S13: We consider alternative mechanisms through which built environment may mitigate Interaction Segregation. (a) Following the inverse relationship between POI localization and segregation established in Extended Data Figure 2a, we consider whether the de-localization of commercial centers alone can provide an alternative to CCBI. We compute mean distance to nearest commercial center for each MSA, which is the same measure in Extended Data Figure 2a but calculated for commercial centers. We find that while commercial center localization is inversely correlated with Interaction Segregation (Spearman Corr -0.44), this correlation is significantly less than the correlation between CCBI and Interaction Segregation (Spearman Corr -0.78). This suggests that commercial center bridging, as quantified by CCBI may be a more promising direction to investigate as a potential mitigator of segregation. **(b)** We also consider whether fine-grained POIs may function as bridges between diverse individuals. For each of the fine-grained leisure POI categories in Figure 1e, we calculate a POI Bridging Index across all MSAs (using the same procedure to calculate CCBI as shown in Extended Data Figure S12, except using fine-grained POI locations instead of commercial center locations). For instance, to calculate the Restaurant Bridging Index for an MSA with K restaurants, we cluster all homes by the nearest restaurant location, and then calculate: $Restaurant\ Bridging\ Index = \frac{\sum_{i=1}^{K} |\mathcal{R}_i| \cdot Gini\ Index(\mathcal{R}_i)}{|\mathcal{V}_{MSA}| \cdot Gini\ Index(\mathcal{V}_{MSA})}$. After calculating the bridging index for all fine-grained POI categories and for each of the 382 MSAs, we then measure the correlation between each bridging index and Interaction Segregation across all MSAs (as measured by ρ^2, the squared Spearman correlation). We find that CCBI provides a stronger correlation ($\rho^2 = 0.604$, horizontal line), than all other bridging indices which are plotted as points on the scatter-plot in **(b)**. Further, we find that POI categories which are often located inside or near commercial centers (co-location, X-axis) have bridging indices which are stronger predictors of Interaction Segregation (e.g. for fitness/recreation centers, snack bars etc.). The high correlation between (Spearman Corr -0.82, $p < 0.001$) between co-location of POIs and bridging index predictive ability demonstrates asymptotic convergence between all other predictive bridging index metrics and CCBI. This further suggests that commercial center bridging should be the primary metric of interest for mitigators of segregation, because other bridging indexes which leverage fine-grained POI locations are at best proxies for CCBI. Supplementary Figures S14-S16 illustrate the frequent co-location between commercial centers and other fine-grained POIs.

a) Town & County Village, Palo Alto, California

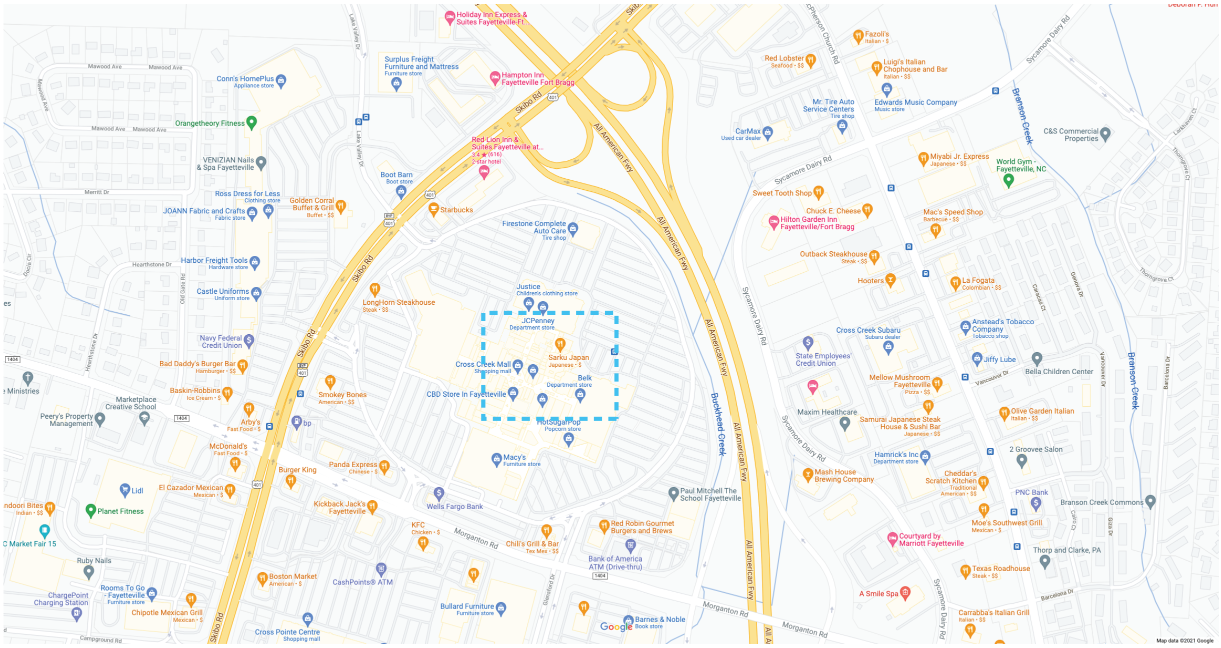


b) Faneuil Hall Marketplace, Boston, MA



Supplementary Figure S14: Examples of commercial centers in coastal cities of (a) San Francisco Bay Area and (b) Boston, MA. Commercial centers frequently contain a diverse assortment of POIs including restaurants, fitness centers/gyms, grocery stores, etc. and are also frequently hubs around which other POIs are located nearby.

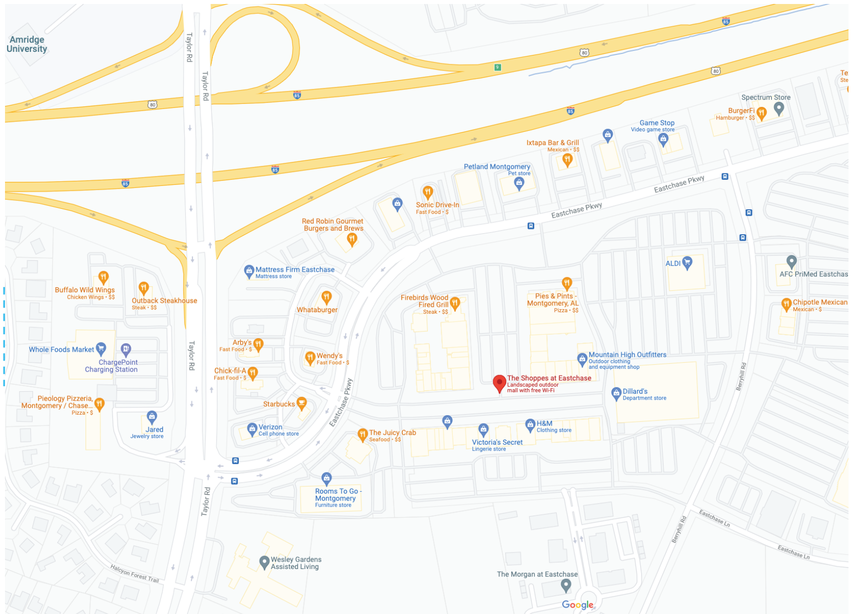
a) Cross Creek Mall and Surrounding Area, Fayetteville, NC



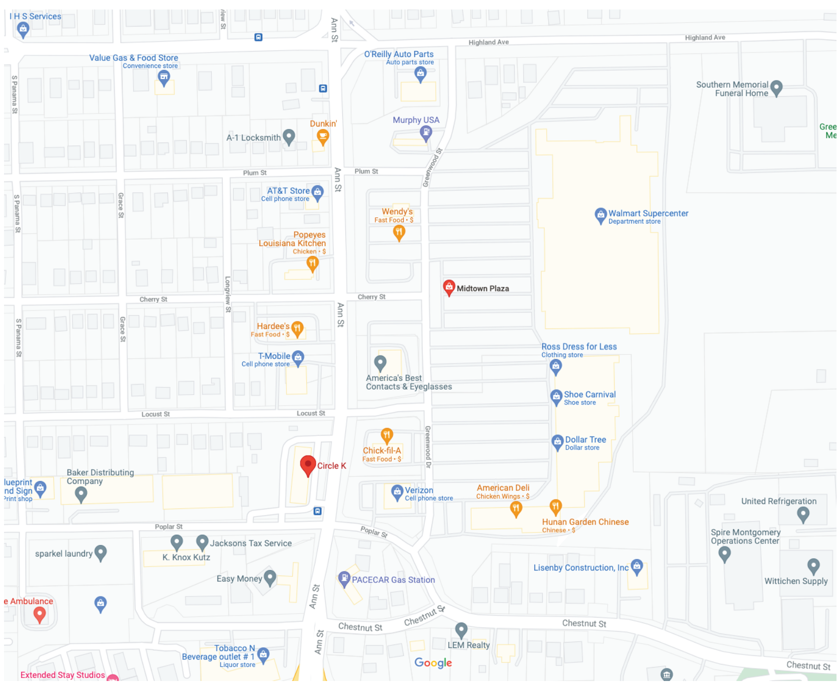
b) [Zoomed In] Cross Creek Mall, Fayetteville, NC

Supplementary Figure S15: Example of a major commercial center in Fayetteville, NC (a) zoomed-out view of commercial center and surrounding co-located POIs (b) zoomed-in view of the center core and businesses contained inside. We find that in Fayetteville, a city with a high commercial center bridging, large commercial centers contain a variety of POIs which cater to diverse individuals of both high and low-ES.

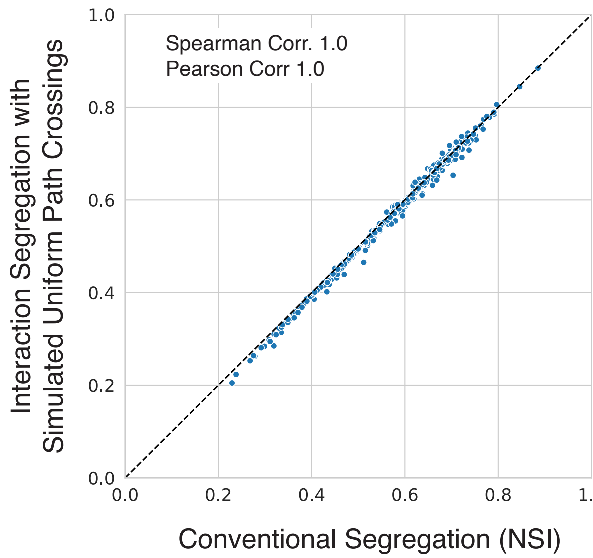
a) The Shoppes at Eastchase, Montgomery, AL



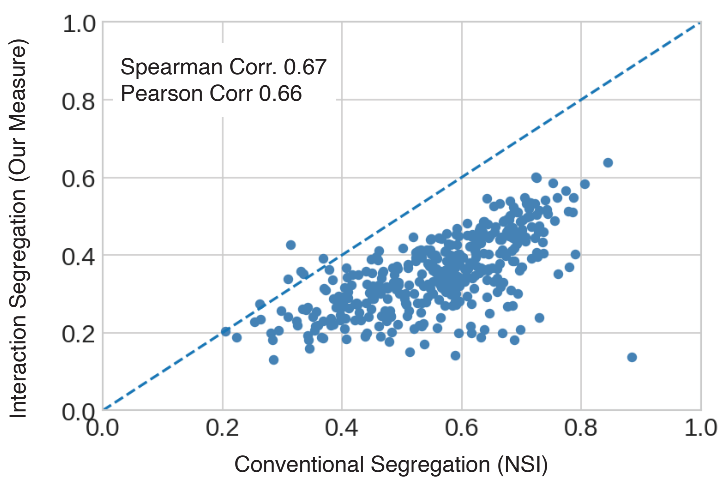
b) Midtown Plaza, Montgomery, AL



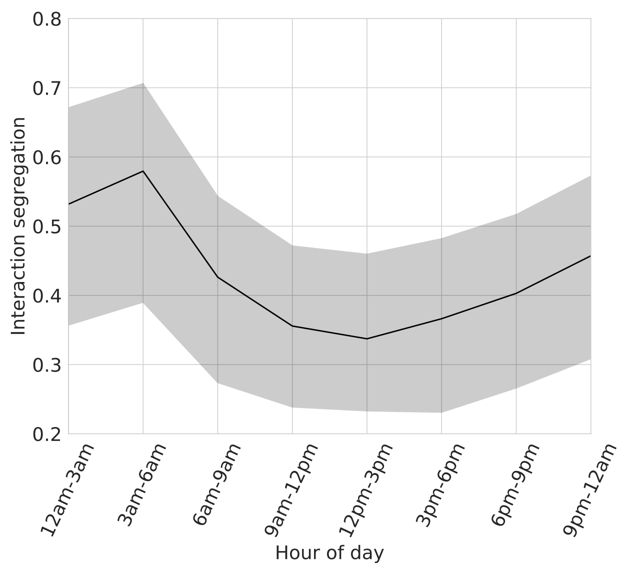
Supplementary Figure S16: Examples two commercial centers in Montgomery, AL which have visitors of predominantly (a) high economic standing (b) low economic standing. We find that in Montgomery, AL a city with a low commercial center bridging, smaller commercial centers exist which contain POIs which cater to a narrow band of individuals in a specific economic stratum. For instance, we find that the nearby grocery store **(a)** is a Whole Foods Market in the high-ES commercial center, in contrast to the **(b)** Walmart Supercenter in the low-ES commercial center.



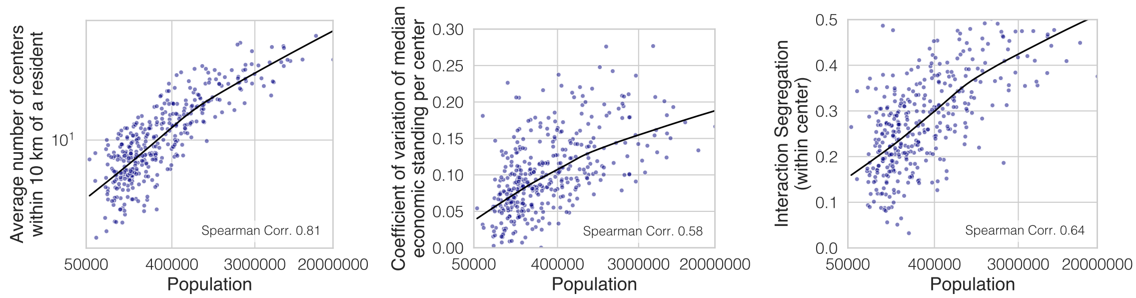
Supplementary Figure S17: Our model with simulated uniform within-tract crossings is equivalent to the conventional neighbourhood sorting index (NSI). Each point is a Metropolitan Statistical Area (MSA). The y-axis shows the Interaction Segregation estimate from the mixed model with a simulated path crossing between every person in a tract (in our dataset). The x-axis shows the correlation between a person's ES and the average ES of people in their tract, which is the neighbourhood sorting index (NSI). As these measures are equivalent, Spearman Corr = 1.0 and Pearson Corr. = 1.0.



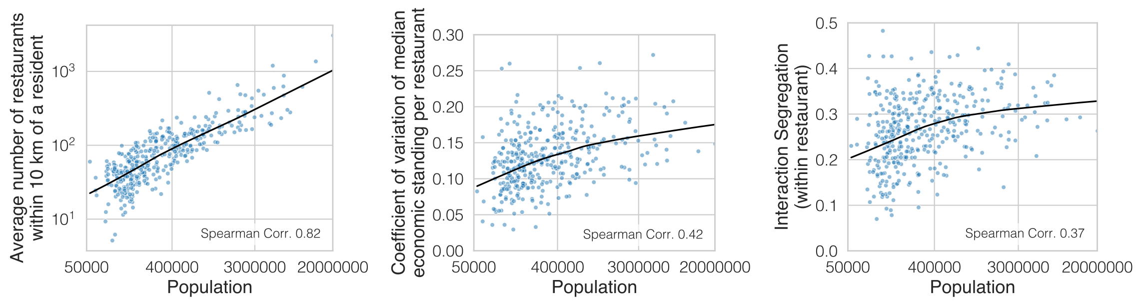
Supplementary Figure S18: Our segregation measure versus a conventional residential segregation, neighborhood sorting index (NSI). Each point is a Metropolitan Statistical Area (MSA). Regardless of whether we compare the numerical segregation values (Pearson Corr. 0.67) or the MSA ranking (Spearman 0.66), only moderate correlation indicates that our measure is different in kind from residential segregaiton as measured conventionally by NSI.



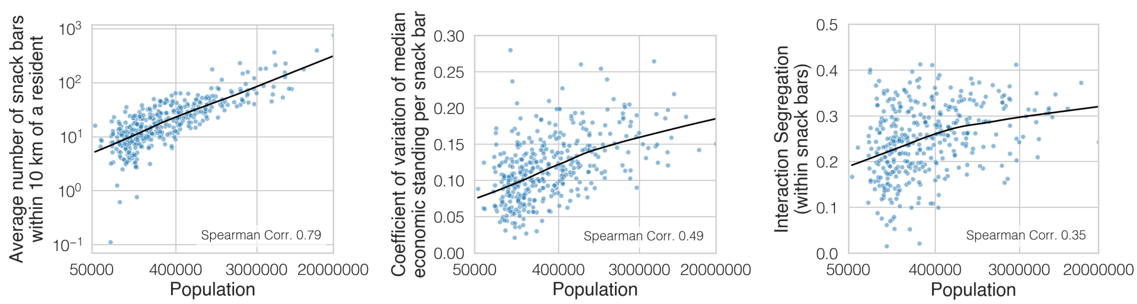
Supplementary Figure S19: Segregation decomposed by time. As described in Methods M3, our fine-grained interaction network allows us to decompose our overall Interaction Segregation into estimates of segregation during different hours of the day, by filtering for interactions that occurred within a specific hour. In Supplementary Figure S19, we partition estimates of segregation by 3 hour windows to illustrate how segregation varies throughout the day (see Supplementary Information). We observe that segregation increases by 61% between the afternoon and early morning hours. Segregation is lowest during commute and work hours, indicating higher levels of interaction with people of different SES while at work or otherwise away from home. Segregation is higher during nighttime hours. This is driven by individuals returning to their home neighborhoods, which are more homogeneous in economic standing, as we well as mechanically a result of ES being is defined by rent value, such that people who live in the same household will have the same ES (and thus will be highly segregated).



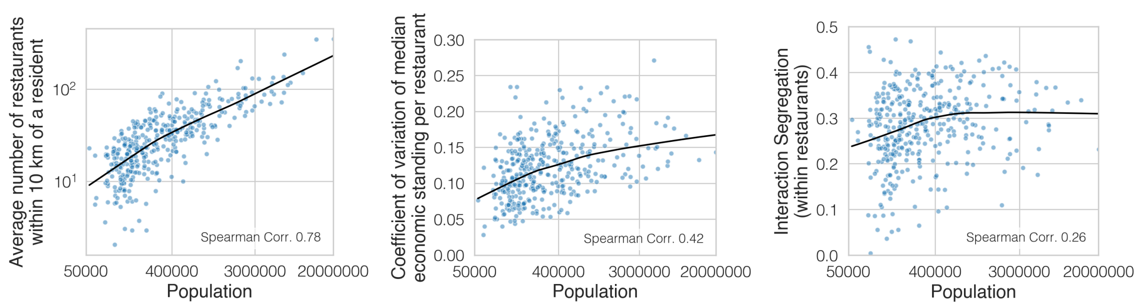
(a) Commercial centers



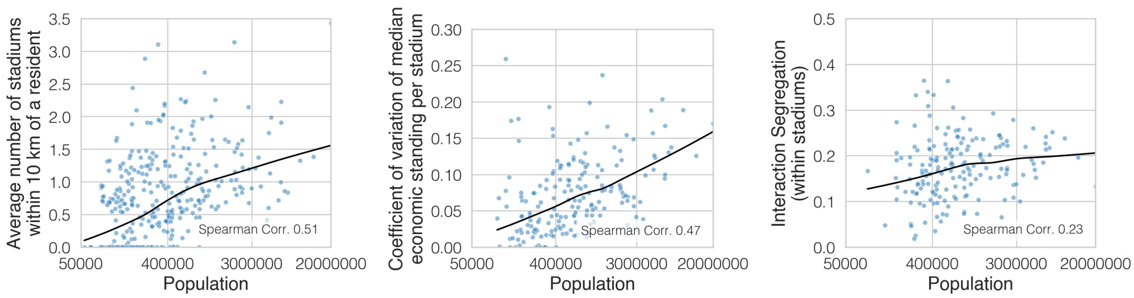
(b) Full-Service Restaurants



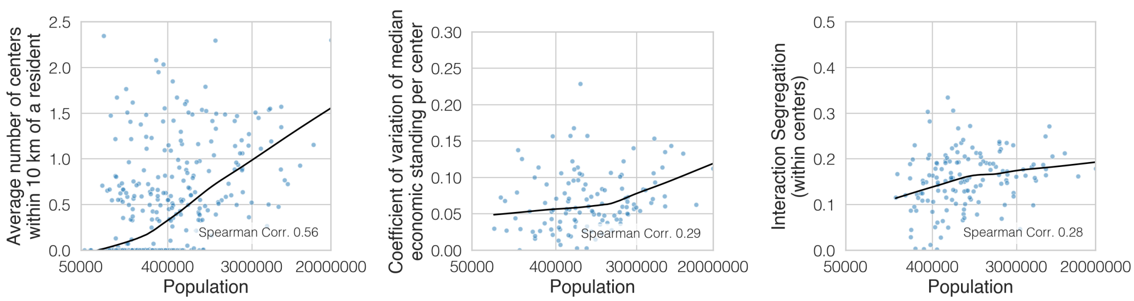
(c) Snack Bars



(d) Limited-Service Restaurants



(e) Stadiums



(f) Performing Arts Centers

Supplementary Figure S20: Across many activities, POIs in large cities are more differentiated and consequently more segregated. This figure shows that the trend towards more options, increased differentiation, and consequently higher segregation is consistent across many prominent POI categories. Here we find similar results for the 5 most frequently visited fine-grained Safegraph place features, in addition to commercial centers. The analyses for full-service restaurants and commercial centers correspond to Figures 2c-e and Extended Data Figure 4, and we additionally show the same trend for snack bars, limited-service restaurants, stadiums, and performing arts centers (ranked 2-5 after full-service restaurants in terms of most frequently visited POIs among Safegraph places). Across the board, large, densely populated metropolitan areas are associated with increased options and economic differentiation of POIs, which may facilitate higher self-segregation.

	Dependent variable:				Interaction Segregation	
	(1)	(2)	(3)	(4)	(5)	(6)
Intercept	0.355***	0.355***	0.355***	0.356***	0.355***	0.355***
	(0.005)	(0.004)	(0.004)	(0.004)	(0.003)	(0.003)
Population Density	0.039***		0.024***	0.022***	0.017***	0.017***
	(0.005)		(0.004)	(0.004)	(0.004)	(0.004)
Gini Index (Estimated Rent)		0.064***	0.058***	0.059***	0.049***	0.050***
		(0.004)	(0.004)	(0.004)	(0.003)	(0.003)
Political Alignment (% Democrat in 2016 Election)				0.009*		0.006
				(0.005)		(0.004)
Racial Demographics (% non-Hispanic White)				-0.005		0.003
				(0.004)		(0.003)
Mean ES (Estimated Rent)				-0.009*		-0.003
				(0.005)		(0.004)
Walkability (Walkscore)					0.002	0.001
					(0.003)	(0.004)
Commutability (% Commute to Work)					-0.012***	-0.013***
					(0.004)	(0.004)
Conventional Segregation (NSI)					0.048***	0.047***
					(0.003)	(0.003)
Observations	382	382	382	376	382	376
R^2	0.151	0.419	0.475	0.490	0.682	0.680
Adjusted R^2	0.149	0.417	0.472	0.483	0.678	0.673

*p<0.1; **p<0.05; ***p<0.01

Supplementary Table S7: Population density is significantly associated with Interaction Segregation, after controlling for MSA income inequality (Gini Index), political alignment (% Democrat in 2016 election), racial demographics (% non-Hispanic White), mean ES, walkability (Walkscore[53]), commutability (% of residents commuting to work), and residential segregation (NSI). This table is from an analogous regression to the regression shown in Extended Data Figure 1, using population density instead of population size (we look at each separately due to co-linearity between population size and density). Here we show the coefficients (after normalizing via z-scoring to have mean 0 and variance 1) from the primary specifications estimating the effect of population density on Interaction Segregation across all MSAs. Columns (1-5) are models specified with different subsets of covariates; Column 6 shows model specification with all covariates. Differences between sample size in models is due to missing data for several covariates in a small number of MSAs (Walkscores were not available for all MSAs). (*p $<$ 0.1; **p $<$ 0.05; *** p $<$ 0.01).

Supplementary Table S8: Interaction Segregation and related variables (i.e. # Interactions, Mean ES, NSI, Gini Index, Population Size, and Commercial Center Bridging Index (CCBI) by MSA

MSA	Interaction Segregation	# Interactions	Mean ES	NSI	Gini	Pop. Size	CCBI
Abilene, TX	0.44	561,896.00	1,245.47	0.63	0.21	170,516.00	0.79
Akron, OH	0.55	2,211,810.00	1,338.21	0.71	0.27	704,367.00	0.65
Albany, GA	0.40	355,999.00	1,077.89	0.47	0.26	151,293.00	0.79
Albany, OR	0.27	153,057.00	1,425.95	0.40	0.11	124,977.00	0.95
Albany-Schenectady-Troy, NY	0.40	2,058,079.00	1,651.16	0.62	0.18	882,130.00	0.81
Albuquerque, NM	0.35	1,979,325.00	1,316.79	0.59	0.19	912,897.00	0.78
Alexandria, LA	0.37	267,228.00	1,011.98	0.52	0.23	153,604.00	0.88
Allentown-Bethlehem-Easton, PA-NJ	0.44	2,629,311.00	1,647.92	0.62	0.19	838,081.00	0.79
Altoona, PA	0.14	199,365.00	799.18	0.59	0.08	123,175.00	0.83
Amarillo, TX	0.45	2,027,531.00	1,358.46	0.62	0.26	264,955.00	0.82
Ames, IA	0.22	178,625.00	1,255.36	0.37	0.19	97,260.00	0.97
Anchorage, AK	0.37	1,189,861.00	1,921.88	0.65	0.17	400,647.00	0.84
Ann Arbor, MI	0.42	984,188.00	2,087.55	0.69	0.19	369,208.00	0.76
Anniston-Oxford-Jacksonville, AL	0.27	498,621.00	949.79	0.43	0.19	114,664.00	0.96
Appleton, WI	0.28	798,727.00	1,201.77	0.60	0.10	236,058.00	0.94
Asheville, NC	0.36	1,449,906.00	1,634.66	0.44	0.18	455,255.00	0.88
Athens-Clarke County, GA	0.26	431,582.00	1,427.27	0.46	0.19	208,997.00	0.83
Atlanta-Sandy Springs-Roswell, GA	0.50	41,054,246.00	1,805.49	0.70	0.22	5,874,249.00	0.62
Atlantic City-Hammonton, NJ	0.55	788,439.00	1,993.59	0.79	0.27	266,328.00	0.59
Auburn-Opelika, AL	0.35	415,339.00	1,409.80	0.44	0.19	161,641.00	0.86
Augusta-Richmond County, GA-SC	0.39	1,698,887.00	1,257.25	0.57	0.21	600,006.00	0.79
Austin-Round Rock, TX	0.53	13,378,670.00	1,954.85	0.70	0.19	2,115,230.00	0.66
Bakersfield, CA	0.29	2,168,162.00	1,399.40	0.69	0.18	888,988.00	0.81
Baltimore-Columbia-Towson, MD	0.61	15,120,132.00	1,898.23	0.80	0.17	2,798,587.00	0.60
Bangor, ME	0.21	56,124.00	1,264.12	0.65	0.17	151,190.00	0.91
Barnstable Town, MA	0.32	556,457.00	2,483.81	0.43	0.18	213,482.00	0.90
Baton Rouge, LA	0.45	3,449,982.00	1,374.01	0.70	0.14	831,182.00	0.81
Battle Creek, MI	0.37	329,581.00	1,013.54	0.66	0.13	134,358.00	0.84
Bay City, MI	0.23	204,294.00	904.87	0.48	0.11	104,189.00	0.97
Beaumont-Port Arthur, TX	0.41	1,935,665.00	1,276.93	0.63	0.16	412,616.00	0.83
Beckley, WV	0.22	61,971.00	988.00	0.41	0.19	118,639.00	0.94
Bellingham, WA	0.18	300,129.00	1,946.31	0.35	0.16	221,650.00	0.95
Bend-Redmond, OR	0.38	257,079.00	1,983.21	0.53	0.17	186,807.00	0.92
Billings, MT	0.36	316,679.00	1,370.83	0.56	0.18	170,740.00	0.91
Binghamton, NY	0.21	325,247.00	1,125.73	0.56	0.16	241,609.00	0.90
Birmingham-Hoover, AL	0.56	7,522,699.00	1,392.29	0.73	0.26	1,149,685.00	0.63
Bismarck, ND	0.18	341,245.00	1,397.18	0.48	0.13	132,418.00	0.95
Blacksburg-Christiansburg-Radford, VA	0.33	260,509.00	1,340.74	0.57	0.17	182,692.00	0.82
Bloomington, IL	0.34	599,796.00	1,245.47	0.62	0.18	188,754.00	0.89
Bloomington, IN	0.35	405,186.00	1,517.20	0.47	0.22	167,513.00	0.91
Bloomsburg-Berwick, PA	0.19	80,784.00	994.00	0.64	0.14	83,924.00	0.94
Boise City, ID	0.41	1,299,521.00	1,584.28	0.60	0.16	710,080.00	0.81
Boston-Cambridge-Newton, MA-NH	0.38	22,163,903.00	2,655.52	0.63	0.16	4,844,597.00	0.70
Boulder, CO	0.36	744,121.00	2,570.91	0.57	0.21	324,073.00	0.84
Bowling Green, KY	0.43	541,630.00	1,227.13	0.59	0.22	174,962.00	0.93
Bremerton-Silverdale, WA	0.40	636,416.00	2,044.31	0.60	0.16	266,550.00	0.77
Bridgeport-Stamford-Norwalk, CT	0.51	3,009,778.00	2,840.03	0.79	0.32	943,457.00	0.53
Brownsville-Harlingen, TX	0.33	1,204,081.00	1,106.57	0.49	0.17	423,181.00	0.88
Brunswick, GA	0.55	345,233.00	1,939.59	0.65	0.33	117,728.00	0.56
Buffalo-Cheektowaga-Niagara Falls, NY	0.41	3,152,570.00	1,285.13	0.68	0.20	1,129,660.00	0.69
Burlington, NC	0.30	489,581.00	1,237.01	0.40	0.21	163,529.00	0.89
Burlington-South Burlington, VT	0.43	140,807.00	1,990.17	0.33	0.16	218,881.00	0.86
California-Lexington Park, MD	0.19	442,775.00	1,739.01	0.35	0.12	112,413.00	0.98
Canton-Massillon, OH	0.41	1,037,327.00	1,235.25	0.38	0.29	399,418.00	0.81
Cape Coral-Fort Myers, FL	0.39	6,067,007.00	1,990.09	0.57	0.27	739,506.00	0.74
Cape Girardeau, MO-IL	0.24	174,129.00	1,013.11	0.36	0.18	96,873.00	0.99
Carbondale-Marion, IL	0.21	182,895.00	855.65	0.37	0.13	125,065.00	0.99
Carson City, NV	0.33	124,126.00	1,700.73	0.59	0.18	54,608.00	0.98
Casper, WY	0.18	103,682.00	1,377.58	0.30	0.21	79,556.00	0.97
Cedar Rapids, IA	0.33	1,010,446.00	1,200.65	0.47	0.14	270,594.00	0.96
Chambersburg-Waynesboro, PA	0.26	158,476.00	1,080.40	0.48	0.10	154,487.00	0.96
Champaign-Urbana, IL	0.32	799,317.00	1,229.14	0.67	0.19	239,877.00	0.88
Charleston, WV	0.25	365,352.00	988.62	0.44	0.21	214,398.00	0.93
Charleston-North Charleston, SC	0.47	4,062,901.00	2,014.52	0.69	0.24	775,089.00	0.66
Charlotte-Concord-Gastonia, NC-SC	0.50	12,750,805.00	1,699.80	0.68	0.24	2,524,863.00	0.64
Charlottesville, VA	0.31	510,779.00	1,840.79	0.50	0.21	233,586.00	0.95
Chattanooga, TN-GA	0.46	2,432,138.00	1,376.30	0.62	0.18	556,081.00	0.86
Cheyenne, WY	0.33	234,335.00	1,450.51	0.67	0.15	98,460.00	0.96
Chicago-Naperville-Elgin, IL-IN-WI	0.44	61,552,971.00	1,943.77	0.66	0.21	9,520,784.00	0.68
Chico, CA	0.29	324,613.00	1,772.67	0.55	0.17	229,207.00	0.92
Cincinnati, OH-KY-IN	0.47	10,110,144.00	1,533.13	0.66	0.22	2,179,858.00	0.76
Clarksville, TN-KY	0.30	989,270.00	1,100.48	0.56	0.16	285,691.00	0.83
Cleveland, TN	0.23	421,419.00	1,072.17	0.33	0.15	122,082.00	0.92
Cleveland-Elyria, OH	0.54	6,830,481.00	1,385.15	0.68	0.25	2,058,549.00	0.63
Coeur d'Alene, ID	0.13	243,473.00	1,680.34	0.32	0.15	157,485.00	0.97
College Station-Bryan, TX	0.40	1,243,139.00	1,430.02	0.59	0.18	258,825.00	0.87
Colorado Springs, CO	0.42	2,666,493.00	1,758.07	0.64	0.16	725,438.00	0.72
Columbia, MO	0.36	425,486.00	1,194.35	0.50	0.20	178,523.00	0.86
Columbia, SC	0.42	3,047,549.00	1,390.00	0.59	0.21	825,110.00	0.81
Columbus, GA-AL	0.47	724,780.00	1,143.38	0.72	0.24	303,436.00	0.74
Columbus, IN	0.41	250,666.00	1,411.07	0.55	0.22	82,429.00	0.94
Columbus, OH	0.55	9,849,191.00	1,623.39	0.71	0.23	2,082,475.00	0.69

Continued on next page

Supplementary Table S8: (cont'd) Interaction Segregation and related variables (i.e. # Interactions, Mean ES, NSI, Gini Index, Population Size, and Commercial Center Bridging Index (CCBI) by MSA

MSA	Interaction Segregation	# Interactions	Mean ES	NSI	Gini	Pop. Size	CCBI
Corpus Christi, TX	0.50	2,288,424.00	1,487.67	0.72	0.16	453,684.00	0.79
Corvallis, OR	0.24	104,179.00	1,936.53	0.48	0.15	91,567.00	0.97
Crestview-Fort Walton Beach-Destin, FL	0.43	1,868,711.00	1,852.07	0.53	0.24	271,959.00	0.76
Cumberland, MD-WV	0.26	30,347.00	936.53	0.51	0.14	98,566.00	0.98
Dallas-Fort Worth-Arlington, TX	0.51	48,228,424.00	1,996.27	0.73	0.22	7,407,944.00	0.62
Dalton, GA	0.14	235,701.00	876.66	0.89	0.14	143,872.00	0.77
Danville, IL	0.24	70,716.00	693.62	0.73	0.05	77,776.00	0.96
Daphne-Fairhope-Foley, AL	0.34	1,066,095.00	1,602.47	0.41	0.16	212,619.00	0.96
Davenport-Moline-Rock Island, IA-IL	0.45	945,921.00	1,287.37	0.70	0.22	381,854.00	0.76
Dayton, OH	0.49	2,617,342.00	1,278.37	0.70	0.25	803,713.00	0.73
Decatur, AL	0.29	403,755.00	981.61	0.42	0.12	151,888.00	0.94
Decatur, IL	0.35	233,031.00	919.63	0.76	0.15	105,533.00	0.86
Deltona-Daytona Beach-Ormond Beach, FL	0.33	5,092,950.00	1,686.87	0.46	0.18	648,117.00	0.82
Denver-Aurora-Lakewood, CO	0.35	11,589,449.00	2,312.40	0.67	0.17	2,892,979.00	0.77
Des Moines-West Des Moines, IA	0.46	1,966,610.00	1,577.32	0.53	0.22	645,100.00	0.76
Detroit-Warren-Dearborn, MI	0.57	15,495,989.00	1,554.43	0.77	0.26	4,321,704.00	0.49
Dothan, AL	0.33	466,788.00	1,142.03	0.45	0.23	147,923.00	0.89
Dover, DE	0.26	462,113.00	1,471.35	0.34	0.13	176,445.00	0.93
Dubuque, IA	0.36	201,954.00	1,313.10	0.57	0.18	97,009.00	0.86
Duluth, MN-WI	0.44	458,229.00	1,373.15	0.67	0.21	278,659.00	0.74
Durham-Chapel Hill, NC	0.44	2,019,082.00	1,713.50	0.57	0.19	566,491.00	0.83
East Stroudsburg, PA	0.24	452,426.00	1,526.37	0.39	0.12	168,089.00	0.95
Eau Claire, WI	0.18	377,072.00	1,114.77	0.41	0.11	167,436.00	0.95
El Centro, CA	0.18	271,797.00	1,340.14	0.69	0.12	181,574.00	0.96
El Paso, TX	0.32	1,667,796.00	1,157.84	0.60	0.17	845,145.00	0.76
Elizabethtown-Fort Knox, KY	0.19	245,608.00	1,241.61	0.24	0.22	150,531.00	0.91
Elkhart-Goshen, IN	0.37	567,181.00	1,186.01	0.41	0.17	204,310.00	0.89
Elmira, NY	0.36	121,786.00	1,173.73	0.58	0.20	84,874.00	0.78
Enid, OK	0.38	320,003.00	1,102.25	0.59	0.24	61,492.00	0.89
Erie, PA	0.34	571,583.00	1,129.58	0.61	0.21	273,892.00	0.85
Eugene, OR	0.29	514,167.00	1,600.88	0.49	0.15	375,617.00	0.93
Evansville, IN-KY	0.47	751,625.00	1,240.07	0.60	0.26	314,960.00	0.80
Fairbanks, AK	0.16	60,754.00	1,619.53	0.36	0.12	99,725.00	0.99
Fargo, ND-MN	0.30	816,028.00	1,333.79	0.52	0.12	241,619.00	0.92
Farmington, NM	0.28	67,876.00	1,265.57	0.52	0.17	126,902.00	0.98
Fayetteville, NC	0.27	1,253,869.00	1,082.27	0.48	0.19	385,380.00	0.90
Fayetteville-Springdale-Rogers, AR-MO	0.43	1,854,285.00	1,371.30	0.60	0.18	538,412.00	0.91
Flagstaff, AZ	0.31	184,120.00	2,012.92	0.42	0.15	141,107.00	0.96
Flint, MI	0.50	981,799.00	1,063.23	0.72	0.22	407,673.00	0.58
Florence, SC	0.37	312,804.00	1,335.06	0.49	0.21	205,546.00	0.93
Florence-Muscle Shoals, AL	0.22	331,068.00	912.14	0.50	0.14	147,100.00	0.97
Fond du Lac, WI	0.19	220,305.00	873.95	0.47	0.09	102,371.00	0.92
Fort Collins, CO	0.28	926,669.00	1,916.57	0.43	0.13	343,993.00	0.91
Fort Smith, AR-OK	0.32	638,594.00	887.85	0.60	0.16	281,990.00	0.93
Fort Wayne, IN	0.50	1,475,260.00	1,342.98	0.69	0.25	434,001.00	0.66
Fresno, CA	0.35	2,137,796.00	1,471.73	0.69	0.18	986,542.00	0.73
Gadsden, AL	0.30	465,122.00	894.00	0.59	0.19	102,937.00	0.89
Gainesville, FL	0.35	1,165,180.00	1,593.03	0.56	0.22	284,685.00	0.82
Gainesville, GA	0.34	758,361.00	1,756.64	0.32	0.22	199,439.00	0.85
Gettysburg, PA	0.22	205,160.00	1,410.01	0.34	0.10	102,367.00	0.99
Glens Falls, NY	0.27	131,519.00	1,486.70	0.59	0.21	125,917.00	0.84
Goldsboro, NC	0.27	275,443.00	1,088.76	0.28	0.18	123,257.00	0.87
Grand Forks, ND-MN	0.33	184,078.00	1,228.24	0.57	0.21	102,277.00	0.98
Grand Island, NE	0.25	236,404.00	1,138.90	0.40	0.14	84,862.00	1.00
Grand Junction, CO	0.32	248,844.00	1,410.36	0.59	0.16	151,406.00	0.88
Grand Rapids-Wyoming, MI	0.40	2,808,054.00	1,540.98	0.65	0.16	1,060,326.00	0.71
Grants Pass, OR	0.19	84,482.00	1,601.76	0.38	0.14	86,653.00	1.00
Great Falls, MT	0.27	156,642.00	1,210.84	0.51	0.13	81,604.00	1.00
Greeley, CO	0.44	749,425.00	1,912.90	0.57	0.14	305,274.00	0.79
Green Bay, WI	0.44	1,141,954.00	1,416.76	0.71	0.21	319,786.00	0.86
Greensboro-High Point, NC	0.48	2,269,305.00	1,226.69	0.67	0.26	763,486.00	0.69
Greenville, NC	0.36	702,454.00	1,259.95	0.34	0.22	178,617.00	0.94
Greenville-Anderson-Mauldin, SC	0.46	2,888,574.00	1,383.12	0.56	0.21	895,422.00	0.82
Gulfport-Biloxi-Pascagoula, MS	0.35	1,349,098.00	1,223.72	0.49	0.18	394,322.00	0.92
Hagerstown-Martinsburg, MD-WV	0.32	599,583.00	1,349.39	0.54	0.16	265,295.00	0.96
Hammond, LA	0.31	354,092.00	1,182.56	0.38	0.14	132,322.00	0.94
Hanford-Corcoran, CA	0.22	253,730.00	1,343.94	0.62	0.15	149,696.00	0.96
Harrisburg-Carlisle, PA	0.39	1,555,132.00	1,467.19	0.54	0.19	571,101.00	0.78
Harrisonburg, VA	0.26	290,643.00	1,278.25	0.42	0.16	134,220.00	0.95
Hartford-West Hartford-East Hartford, CT	0.43	2,241,050.00	1,710.35	0.66	0.18	1,206,719.00	0.77
Hattiesburg, MS	0.37	311,274.00	1,156.01	0.70	0.16	148,719.00	0.83
Hickory-Lenoir-Morganton, NC	0.41	640,647.00	1,244.75	0.47	0.18	367,004.00	0.91
Hilton Head Island-Bluffton-Beaufort, SC	0.39	732,993.00	2,115.09	0.53	0.21	214,890.00	0.82
Hinesville, GA	0.20	138,926.00	1,134.33	0.29	0.13	80,518.00	0.97
Homosassa Springs, FL	0.30	528,334.00	1,417.05	0.46	0.21	145,512.00	0.94
Hot Springs, AR	0.34	274,972.00	1,162.56	0.39	0.23	98,444.00	0.89
Houma-Thibodaux, LA	0.29	511,209.00	1,159.51	0.65	0.12	209,893.00	0.83
Houston-The Woodlands-Sugar Land, TX	0.47	63,151,024.00	1,866.72	0.72	0.22	6,905,695.00	0.66
Huntington-Ashland, WV-KY-OH	0.34	860,612.00	1,069.95	0.49	0.18	355,582.00	0.87
Huntsville, AL	0.45	1,623,341.00	1,313.85	0.66	0.20	455,741.00	0.81
Idaho Falls, ID	0.25	212,821.00	1,219.23	0.54	0.14	145,792.00	0.95
Indianapolis-Carmel-Anderson, IN	0.52	10,182,520.00	1,466.08	0.68	0.24	2,026,723.00	0.64

Continued on next page

Supplementary Table S8: (cont'd) Interaction Segregation and related variables (i.e. # Interactions, Mean ES, NSI, Gini Index, Population Size, and Commercial Center Bridging Index (CCBI) by MSA

MSA	Interaction Segregation	# Interactions	Mean ES	NSI	Gini	Pop. Size	CCBI
Iowa City, IA	0.36	473,387.00	1,499.89	0.44	0.20	171,470.00	0.97
Ithaca, NY	0.23	130,902.00	1,607.18	0.45	0.14	102,678.00	0.95
Jackson, MI	0.40	294,167.00	1,049.19	0.67	0.15	158,690.00	0.96
Jackson, MS	0.58	1,725,331.00	1,409.20	0.72	0.24	581,552.00	0.68
Jackson, TN	0.40	324,665.00	1,075.77	0.74	0.16	129,186.00	0.82
Jacksonville, FL	0.49	10,861,594.00	1,742.93	0.61	0.24	1,504,841.00	0.64
Jacksonville, NC	0.30	544,185.00	1,162.30	0.45	0.17	194,838.00	0.90
Janesville-Beloit, WI	0.25	328,187.00	1,002.54	0.65	0.10	162,320.00	0.76
Jefferson City, MO	0.24	349,139.00	1,010.56	0.45	0.16	151,298.00	0.96
Johnson City, TN	0.33	357,946.00	1,075.54	0.55	0.17	201,844.00	0.84
Johnstown, PA	0.20	232,779.00	741.42	0.62	0.11	133,054.00	0.98
Jonesboro, AR	0.36	308,755.00	1,106.28	0.64	0.17	131,158.00	0.94
Joplin, MO	0.21	356,202.00	864.53	0.41	0.12	178,330.00	0.97
Kahului-Wailuku-Lahaina, HI	0.22	246,346.00	3,043.60	0.37	0.16	166,491.00	0.97
Kalamazoo-Portage, MI	0.37	910,106.00	1,412.34	0.58	0.14	338,347.00	0.88
Kankakee, IL	0.30	374,176.00	1,249.33	0.70	0.10	110,544.00	0.81
Kansas City, MO-KS	0.54	8,835,941.00	1,541.78	0.75	0.25	2,127,259.00	0.64
Kennewick-Richland, WA	0.37	373,182.00	1,575.53	0.61	0.15	290,570.00	0.88
Killeen-Temple, TX	0.37	1,587,760.00	1,116.71	0.58	0.16	443,653.00	0.87
Kingsport-Bristol-Bristol, TN-VA	0.29	373,431.00	1,097.78	0.46	0.17	306,253.00	0.93
Kingston, NY	0.31	290,633.00	1,750.29	0.47	0.13	178,723.00	0.94
Knoxville, TN	0.43	2,704,521.00	1,467.11	0.60	0.23	875,797.00	0.76
Kokomo, IN	0.31	288,583.00	944.72	0.55	0.16	82,311.00	0.88
La Crosse-Onalaska, WI-MN	0.17	200,045.00	1,144.24	0.54	0.09	136,778.00	0.97
Lafayette, LA	0.39	1,650,845.00	1,126.34	0.67	0.16	490,107.00	0.76
Lafayette-West Lafayette, IN	0.33	768,405.00	1,226.26	0.57	0.16	220,337.00	0.88
Lake Charles, LA	0.27	207,541.00	1,286.99	0.62	0.13	209,256.00	0.88
Lake Havasu City-Kingman, AZ	0.30	321,719.00	1,255.77	0.59	0.18	207,114.00	0.83
Lakeland-Winter Haven, FL	0.31	4,246,971.00	1,435.22	0.43	0.18	685,830.00	0.90
Lancaster, PA	0.28	1,111,168.00	1,304.06	0.54	0.11	541,054.00	0.86
Lansing-East Lansing, MI	0.42	1,146,004.00	1,249.29	0.65	0.19	480,353.00	0.80
Laredo, TX	0.55	682,291.00	1,308.82	0.76	0.18	273,982.00	0.72
Las Cruces, NM	0.27	340,688.00	1,194.16	0.44	0.18	216,186.00	0.97
Las Vegas-Henderson-Paradise, NV	0.27	10,258,483.00	1,703.98	0.51	0.20	2,183,310.00	0.82
Lawrence, KS	0.19	423,826.00	1,417.50	0.47	0.20	120,629.00	0.94
Lawton, OK	0.30	337,308.00	933.32	0.58	0.21	127,589.00	0.88
Lebanon, PA	0.43	276,347.00	1,155.40	0.70	0.13	139,566.00	0.88
Lewiston, ID-WA	0.20	43,087.00	1,309.23	0.23	0.13	62,881.00	0.99
Lewiston-Auburn, ME	0.21	110,078.00	1,111.21	0.53	0.07	107,569.00	0.98
Lexington-Fayette, KY	0.42	2,391,314.00	1,321.56	0.62	0.21	512,732.00	0.85
Lima, OH	0.35	228,949.00	983.20	0.60	0.16	103,069.00	0.95
Lincoln, NE	0.35	1,808,658.00	1,395.51	0.57	0.16	331,179.00	0.87
Little Rock-North Little Rock-Conway, AR	0.51	3,070,717.00	1,192.95	0.72	0.20	737,991.00	0.77
Logan, UT-ID	0.20	154,902.00	1,210.21	0.60	0.10	138,052.00	0.92
Longview, TX	0.35	262,120.00	1,243.45	0.61	0.18	218,594.00	0.86
Longview, WA	0.38	173,605.00	1,459.50	0.52	0.18	106,900.00	1.00
Los Angeles-Long Beach-Anaheim, CA	0.44	110,526,499.00	2,970.24	0.75	0.20	13,298,709.00	0.66
Louisville/Jefferson County, KY-IN	0.51	4,567,106.00	1,436.51	0.67	0.23	1,292,809.00	0.71
Lubbock, TX	0.45	1,549,243.00	1,381.12	0.55	0.22	316,588.00	0.83
Lynchburg, VA	0.31	496,432.00	1,201.66	0.58	0.18	261,954.00	0.84
Macon-Bibb County, GA	0.46	616,989.00	1,232.02	0.58	0.26	229,081.00	0.81
Madera, CA	0.30	249,720.00	1,396.67	0.47	0.14	155,904.00	0.96
Madison, WI	0.37	1,737,217.00	1,628.86	0.60	0.17	654,577.00	0.89
Manchester-Nashua, NH	0.46	929,901.00	2,027.43	0.69	0.18	413,157.00	0.88
Manhattan, KS	0.31	268,899.00	1,285.52	0.38	0.19	97,954.00	0.99
Mankato-North Mankato, MN	0.26	274,393.00	1,407.12	0.35	0.13	100,945.00	0.97
Mansfield, OH	0.30	203,318.00	872.54	0.50	0.12	120,543.00	0.88
McAllen-Edinburg-Mission, TX	0.36	2,672,266.00	1,165.08	0.45	0.22	858,323.00	0.87
Medford, OR	0.32	250,898.00	1,584.27	0.48	0.14	216,761.00	0.93
Memphis, TN-MS-AR	0.56	5,217,305.00	1,409.82	0.74	0.27	1,347,576.00	0.58
Merced, CA	0.24	471,172.00	1,489.37	0.57	0.13	271,340.00	0.92
Miami-Fort Lauderdale-West Palm Beach, FL	0.44	147,998,127.00	2,642.33	0.67	0.28	6,149,687.00	0.70
Michigan City-La Porte, IN	0.31	262,677.00	1,150.07	0.58	0.16	109,911.00	0.93
Midland, MI	0.35	156,234.00	1,161.45	0.63	0.18	83,245.00	0.96
Midland, TX	0.33	821,156.00	2,759.87	0.61	0.19	170,948.00	0.91
Milwaukee-Waukesha-West Allis, WI	0.60	5,452,737.00	1,428.80	0.77	0.24	1,575,151.00	0.63
Minneapolis-St. Paul-Bloomington, MN-WI	0.41	17,181,042.00	1,970.38	0.56	0.18	3,592,669.00	0.78
Missoula, MT	0.20	144,440.00	1,486.18	0.49	0.17	117,863.00	0.96
Mobile, AL	0.28	1,700,477.00	1,102.96	0.52	0.16	414,515.00	0.85
Modesto, CA	0.20	1,396,841.00	1,673.04	0.53	0.12	545,267.00	0.91
Monroe, LA	0.39	420,225.00	1,057.93	0.56	0.25	178,211.00	0.81
Monroe, MI	0.20	298,001.00	1,123.01	0.45	0.08	149,592.00	0.99
Montgomery, AL	0.47	933,055.00	1,116.61	0.73	0.18	374,042.00	0.71
Morgantown, WV	0.28	144,020.00	1,375.31	0.53	0.22	139,739.00	0.99
Morristown, TN	0.35	127,639.00	1,123.38	0.35	0.16	117,843.00	0.95
Mount Vernon-Anacortes, WA	0.29	164,737.00	1,848.65	0.46	0.13	126,026.00	0.99
Muncie, IN	0.34	325,604.00	951.96	0.58	0.19	115,389.00	0.91
Muskegon, MI	0.32	338,931.00	1,059.59	0.50	0.15	173,656.00	0.86
Myrtle Beach-Conway-North Myrtle Beach, SC-NC	0.24	1,937,451.00	1,614.84	0.27	0.22	463,386.00	0.87
Napa, CA	0.19	813,681.00	3,152.52	0.45	0.18	140,386.00	0.95
Naples-Immokalee-Marco Island, FL	0.45	3,239,165.00	3,865.02	0.70	0.36	372,345.00	0.77
Nashville-Davidson–Murfreesboro–Franklin, TN	0.50	10,766,763.00	1,845.97	0.74	0.22	1,900,584.00	0.62

Continued on next page

Supplementary Table S8: (cont'd) Interaction Segregation and related variables (i.e. # Interactions, Mean ES, NSI, Gini Index, Population Size, and Commercial Center Bridging Index (CCBI) by MSA

MSA	Interaction Segregation	# Interactions	Mean ES	NSI	Gini	Pop. Size	CCBI
New Bern, NC	0.32	309,000.00	1,270.00	0.47	0.19	125,010.00	0.89
New Haven-Milford, CT	0.37	2,382,587.00	1,669.24	0.53	0.19	857,794.00	0.76
New Orleans-Metairie, LA	0.40	6,489,654.00	1,556.45	0.68	0.16	1,270,465.00	0.81
New York-Newark-Jersey City, NY-NJ-PA	0.40	168,755,438.00	2,597.55	0.69	0.21	19,998,951.00	0.57
Niles-Benton Harbor, MI	0.51	301,883.00	1,248.76	0.78	0.19	154,362.00	0.80
North Port-Sarasota-Bradenton, FL	0.40	6,601,216.00	2,284.53	0.61	0.27	805,139.00	0.73
Norwich-New London, CT	0.27	286,633.00	1,478.91	0.60	0.15	267,826.00	0.87
Ocala, FL	0.30	1,669,292.00	1,370.81	0.39	0.21	353,717.00	0.90
Odessa, TX	0.28	687,081.00	2,187.43	0.54	0.13	157,173.00	0.88
Ogden-Clearfield, UT	0.38	1,513,318.00	1,609.07	0.63	0.15	664,589.00	0.84
Oklahoma City, OK	0.46	11,029,080.00	1,417.61	0.63	0.27	1,383,249.00	0.74
Olympia-Tumwater, WA	0.19	644,378.00	1,873.77	0.37	0.14	280,289.00	0.98
Omaha-Council Bluffs, NE-IA	0.49	3,944,289.00	1,583.10	0.64	0.22	932,217.00	0.69
Orlando-Kissimmee-Sanford, FL	0.42	25,094,242.00	1,870.76	0.57	0.21	2,512,917.00	0.72
Oshkosh-Neenah, WI	0.36	624,252.00	1,196.65	0.58	0.20	170,375.00	0.90
Owensboro, KY	0.36	203,069.00	1,064.78	0.48	0.21	118,543.00	0.95
Oxnard-Thousand Oaks-Ventura, CA	0.34	3,449,664.00	3,029.11	0.67	0.17	850,802.00	0.81
Palm Bay-Melbourne-Titusville, FL	0.43	4,331,998.00	1,826.84	0.52	0.20	588,265.00	0.77
Panama City, FL	0.28	1,089,950.00	2,071.06	0.40	0.20	200,168.00	0.93
Parkersburg-Vienna, WV	0.25	140,256.00	1,096.22	0.37	0.18	90,873.00	0.95
Pensacola-Ferry Pass-Brent, FL	0.42	3,672,000.00	1,405.47	0.55	0.21	487,327.00	0.78
Peoria, IL	0.46	914,882.00	1,178.67	0.63	0.22	371,810.00	0.80
Philadelphia-Camden-Wilmington, PA-NJ-DE-MD	0.53	24,822,104.00	1,802.57	0.74	0.20	6,078,451.00	0.61
Phoenix-Mesa-Scottsdale, AZ	0.48	17,152,709.00	1,746.59	0.75	0.19	4,761,694.00	0.61
Pine Bluff, AR	0.15	171,555.00	759.22	0.52	0.10	90,923.00	0.95
Pittsburgh, PA	0.47	7,756,479.00	1,348.23	0.70	0.24	2,330,283.00	0.72
Pittsfield, MA	0.37	128,317.00	1,477.61	0.52	0.15	126,485.00	0.86
Pocatello, ID	0.28	116,547.00	1,158.04	0.60	0.18	85,641.00	0.97
Port St. Lucie, FL	0.44	4,896,670.00	2,082.29	0.62	0.23	473,192.00	0.72
Portland-South Portland, ME	0.30	493,445.00	1,896.31	0.52	0.16	532,280.00	0.80
Portland-Vancouver-Hillsboro, OR-WA	0.34	6,614,427.00	2,016.66	0.58	0.16	2,456,462.00	0.86
Prescott, AZ	0.38	384,045.00	1,627.89	0.60	0.17	228,055.00	0.88
Providence-Warwick, RI-MA	0.40	5,763,343.00	1,744.07	0.63	0.16	1,617,057.00	0.79
Provo-Orem, UT	0.32	1,219,235.00	1,546.11	0.66	0.13	617,751.00	0.84
Pueblo, CO	0.36	443,458.00	1,270.67	0.65	0.18	166,426.00	0.85
Punta Gorda, FL	0.34	1,184,701.00	1,829.41	0.60	0.21	181,537.00	0.84
Racine, WI	0.42	548,195.00	1,383.27	0.63	0.16	195,949.00	0.79
Raleigh, NC	0.46	8,986,021.00	1,697.93	0.62	0.16	1,334,342.00	0.79
Rapid City, SD	0.32	250,301.00	1,441.12	0.41	0.20	146,869.00	0.88
Reading, PA	0.49	1,036,022.00	1,344.18	0.68	0.18	417,524.00	0.71
Redding, CA	0.23	258,289.00	1,573.52	0.43	0.19	179,539.00	0.99
Reno, NV	0.47	1,550,552.00	1,953.99	0.63	0.18	461,336.00	0.70
Richmond, VA	0.49	4,524,531.00	1,626.59	0.71	0.20	1,292,911.00	0.68
Riverside-San Bernardino-Ontario, CA	0.43	19,908,134.00	2,103.34	0.70	0.17	4,570,427.00	0.68
Roanoke, VA	0.38	811,899.00	1,291.22	0.58	0.20	313,488.00	0.80
Rochester, MN	0.33	648,812.00	1,540.01	0.58	0.16	217,828.00	0.94
Rochester, NY	0.45	2,870,914.00	1,536.09	0.66	0.19	1,071,589.00	0.77
Rockford, IL	0.40	946,786.00	1,225.98	0.67	0.18	338,252.00	0.72
Rocky Mount, NC	0.33	366,999.00	977.64	0.55	0.16	146,769.00	0.81
Rome, GA	0.36	264,276.00	1,078.32	0.39	0.16	97,427.00	0.99
Sacramento–Roseville–Arden-Arcade, CA	0.39	7,101,248.00	2,048.69	0.66	0.15	2,320,381.00	0.70
Saginaw, MI	0.37	355,021.00	927.43	0.70	0.18	191,996.00	0.77
Salem, OR	0.27	670,775.00	1,603.26	0.45	0.12	424,968.00	0.94
Salinas, CA	0.36	952,169.00	2,642.69	0.65	0.17	435,477.00	0.77
Salisbury, MD-DE	0.48	875,922.00	1,462.24	0.68	0.14	404,067.00	0.78
Salt Lake City, UT	0.33	3,468,862.00	1,763.34	0.60	0.15	1,205,238.00	0.77
San Angelo, TX	0.34	380,590.00	1,321.06	0.66	0.14	119,200.00	0.83
San Antonio-New Braunfels, TX	0.53	14,354,046.00	1,596.19	0.70	0.20	2,474,274.00	0.64
San Diego-Carlsbad, CA	0.42	13,807,983.00	2,854.26	0.73	0.20	3,325,468.00	0.71
San Francisco-Oakland-Hayward, CA	0.41	37,492,367.00	3,925.90	0.71	0.21	4,710,693.00	0.68
San Jose-Sunnyvale-Santa Clara, CA	0.37	8,012,471.00	3,766.32	0.78	0.16	1,993,582.00	0.68
San Luis Obispo-Paso Robles-Arroyo Grande, CA	0.22	656,686.00	2,601.39	0.47	0.13	282,838.00	0.94
Santa Cruz-Watsonville, CA	0.27	611,337.00	3,306.01	0.52	0.13	275,105.00	0.85
Santa Fe, NM	0.45	182,572.00	2,075.86	0.61	0.23	149,617.00	0.91
Santa Maria-Santa Barbara, CA	0.52	1,124,975.00	3,039.10	0.75	0.28	445,606.00	0.59
Santa Rosa, CA	0.22	1,144,462.00	2,793.94	0.54	0.11	503,246.00	0.92
Savannah, GA	0.39	1,601,410.00	1,599.94	0.57	0.19	386,337.00	0.83
Scranton–Wilkes-Barre–Hazleton, PA	0.30	1,064,769.00	1,077.93	0.62	0.16	555,645.00	0.90
Seattle-Tacoma-Bellevue, WA	0.44	18,136,495.00	2,474.85	0.64	0.20	3,884,469.00	0.73
Sebastian-Vero Beach, FL	0.52	1,148,601.00	2,259.56	0.71	0.31	154,314.00	0.78
Sebring, FL	0.26	352,266.00	1,291.88	0.31	0.21	104,060.00	0.98
Sheboygan, WI	0.35	220,669.00	1,204.87	0.58	0.10	115,235.00	0.86
Sherman-Denison, TX	0.39	648,078.00	1,321.20	0.46	0.14	131,214.00	0.95
Shreveport-Bossier City, LA	0.50	1,110,198.00	1,247.36	0.65	0.26	439,631.00	0.79
Sierra Vista-Douglas, AZ	0.21	153,651.00	992.99	0.70	0.14	124,990.00	0.93
Sioux City, IA-NE-SD	0.24	383,415.00	1,137.14	0.41	0.14	168,218.00	0.97
Sioux Falls, SD	0.27	733,324.00	1,239.36	0.68	0.13	260,521.00	0.98
South Bend-Mishawaka, IN-MI	0.48	1,021,877.00	1,289.97	0.67	0.26	321,447.00	0.86
Spartanburg, SC	0.39	913,944.00	1,295.96	0.46	0.23	334,130.00	0.93
Spokane-Spokane Valley, WA	0.34	944,039.00	1,472.98	0.58	0.16	563,958.00	0.88
Springfield, IL	0.40	797,160.00	1,152.17	0.79	0.17	209,175.00	0.87
Springfield, MA	0.45	1,422,143.00	1,632.08	0.69	0.16	629,506.00	0.82

Continued on next page

Supplementary Table S8: (cont'd) Interaction Segregation and related variables (i.e. # Interactions, Mean ES, NSI, Gini Index, Population Size, and Commercial Center Bridging Index (CCBI) by MSA

MSA	Interaction Segregation	# Interactions	Mean ES	NSI	Gini	Pop. Size	CCBI
Springfield, MO	0.39	1,300,346.00	1,100.41	0.59	0.23	462,300.00	0.85
Springfield, OH	0.39	361,791.00	895.90	0.61	0.18	134,649.00	0.89
St. Cloud, MN	0.23	518,337.00	1,273.69	0.40	0.14	198,106.00	0.95
St. George, UT	0.27	243,300.00	1,715.91	0.35	0.17	165,859.00	0.94
St. Joseph, MO-KS	0.27	258,535.00	916.56	0.69	0.14	126,598.00	0.88
St. Louis, MO-IL	0.51	11,016,511.00	1,413.67	0.72	0.25	2,805,850.00	0.62
State College, PA	0.35	215,274.00	1,600.27	0.54	0.18	162,250.00	0.89
Staunton-Waynesboro, VA	0.24	282,626.00	1,324.76	0.28	0.17	121,984.00	0.94
Stockton-Lodi, CA	0.42	2,116,634.00	1,865.80	0.73	0.15	742,516.00	0.72
Sumter, SC	0.35	237,013.00	1,050.69	0.46	0.21	106,514.00	0.92
Syracuse, NY	0.37	1,932,194.00	1,483.28	0.63	0.19	651,048.00	0.82
Tallahassee, FL	0.45	1,973,991.00	1,428.85	0.71	0.22	383,467.00	0.75
Tampa-St. Petersburg-Clearwater, FL	0.45	24,564,393.00	1,805.89	0.62	0.22	3,091,225.00	0.73
Terre Haute, IN	0.32	369,258.00	930.48	0.48	0.18	170,022.00	0.94
Texarkana, TX-AR	0.30	240,478.00	956.50	0.61	0.12	150,254.00	0.89
The Villages, FL	0.39	786,309.00	1,711.00	0.69	0.12	124,933.00	0.90
Toledo, OH	0.53	1,775,752.00	1,217.56	0.69	0.25	603,830.00	0.63
Topeka, KS	0.43	1,205,672.00	1,084.47	0.64	0.20	233,153.00	0.93
Trenton, NJ	0.64	1,240,113.00	2,005.70	0.85	0.21	368,602.00	0.54
Tucson, AZ	0.39	2,564,383.00	1,362.93	0.73	0.17	1,027,502.00	0.72
Tulsa, OK	0.49	5,223,272.00	1,242.38	0.72	0.20	991,610.00	0.77
Tuscaloosa, AL	0.38	1,468,839.00	1,332.43	0.60	0.18	242,700.00	0.91
Twin Falls, ID	0.27	182,971.00	1,187.85	0.41	0.13	109,037.00	0.98
Tyler, TX	0.27	593,804.00	1,432.30	0.41	0.19	227,460.00	0.87
Urban Honolulu, HI	0.33	1,368,021.00	2,616.52	0.62	0.19	986,429.00	0.85
Utica-Rome, NY	0.31	374,846.00	1,123.98	0.66	0.17	292,336.00	0.80
Valdosta, GA	0.32	380,832.00	1,136.00	0.52	0.23	145,403.00	0.94
Vallejo-Fairfield, CA	0.18	1,878,258.00	2,372.07	0.67	0.11	443,877.00	0.85
Victoria, TX	0.37	374,597.00	1,529.97	0.55	0.20	99,651.00	0.94
Vineland-Bridgeton, NJ	0.37	354,594.00	1,371.71	0.63	0.09	151,748.00	0.90
Virginia Beach-Norfolk-Newport News, VA-NC	0.45	6,944,774.00	1,666.43	0.62	0.20	1,724,876.00	0.75
Visalia-Porterville, CA	0.25	854,866.00	1,309.53	0.48	0.16	463,097.00	0.95
Waco, TX	0.41	1,245,450.00	1,334.52	0.55	0.20	268,550.00	0.89
Walla Walla, WA	0.22	53,561.00	1,448.66	0.39	0.13	64,675.00	1.00
Warner Robins, GA	0.40	600,913.00	1,271.38	0.54	0.20	191,227.00	0.86
Washington-Arlington-Alexandria, DC-VA-MD-WV	0.46	127,482,444.00	2,461.96	0.70	0.18	6,200,001.00	0.71
Waterloo-Cedar Falls, IA	0.38	326,152.00	1,111.17	0.62	0.18	169,553.00	0.84
Watertown-Fort Drum, NY	0.24	113,127.00	1,273.60	0.32	0.15	113,063.00	0.97
Wausau, WI	0.24	299,543.00	1,069.93	0.49	0.12	135,415.00	0.93
Weirton-Steubenville, WV-OH	0.24	235,799.00	823.09	0.52	0.12	118,181.00	0.94
Wenatchee, WA	0.21	126,600.00	1,748.88	0.31	0.14	118,646.00	1.00
Wheeling, WV-OH	0.26	102,043.00	1,100.65	0.53	0.15	141,228.00	0.87
Wichita Falls, TX	0.41	729,999.00	1,253.35	0.53	0.26	151,180.00	0.93
Wichita, KS	0.46	3,213,430.00	1,206.35	0.64	0.23	644,949.00	0.78
Williamsport, PA	0.23	148,520.00	1,025.19	0.38	0.14	113,930.00	0.99
Wilmington, NC	0.46	1,684,931.00	1,859.12	0.58	0.26	289,425.00	0.77
Winchester, VA-WV	0.33	296,124.00	1,560.81	0.50	0.16	138,107.00	0.95
Winston-Salem, NC	0.44	1,872,668.00	1,256.11	0.57	0.22	666,746.00	0.79
Worcester, MA-CT	0.48	2,650,313.00	1,777.27	0.71	0.17	942,303.00	0.74
Yakima, WA	0.33	222,165.00	1,236.55	0.54	0.15	250,377.00	0.88
York-Hanover, PA	0.41	1,303,401.00	1,430.52	0.56	0.18	445,722.00	0.85
Youngstown-Warren-Boardman, OH-PA	0.34	1,336,389.00	951.78	0.64	0.19	541,875.00	0.71
Yuba City, CA	0.29	309,905.00	1,574.77	0.57	0.13	173,213.00	0.89
Yuma, AZ	0.22	271,294.00	1,079.08	0.55	0.17	209,756.00	1.00