The Center for Economic Research
Columbian College of Arts & Sciences

WORKING PAPER SERIES

The Role of Social Connections in the Racial Segregation of US Cities

Andreas Diemer

Tanner Regan

Cheng Keat Tang

Working Paper No. 2025-008

November, 2025

THE CENTER FOR ECONOMIC RESEARCH
Department of Economics
Columbian College of Arts & Sciences
The George Washington University
Washington, DC 20052
https://cer.columbian.gwu.edu/

The Role of Social Connections in the Racial Segregation of US Cities

Andreas Diemer*
Stockholm Univ. (SOFI)

Tanner Regan
George Washington Univ.

Cheng Keat Tang
Nanyang Tech. Univ.

June 10, 2024

Abstract

We study the extent of segregation in the social space of urban America. We measure segregation as the (lack of) actual personal connections between neighbourhoods as opposed to conventional measures that assume the strength of these connections. We distinguish social segregation from geographical definitions of segregation, building and comparing city-level indices of each. We apply our measures to the 75 largest MSAs in the USA. Cities like Miami, Washington DC, and Cincinnati rank higher in social segregation than they do based on the conventional residential isolation, while New Orleans, San Francisco, and Richmond fall in ranks. Conditional on residential segregation, cities with more institutions that foster social cohesion (churches and community associations) are less socially segregated. Looking at within-city variation across neighbourhoods, growing up more socially exposed to non-White neighbourhoods is related to various adulthood outcomes (jailed, income rank, married, and non-migrant) for Black individuals. Social exposure to non-White neighbourhoods is related to worsening adulthood outcomes in neighbourhoods that are majority non-White. Our results suggest that social connections, beyond residential location or other spatial relationships, are important for understanding the effective segregation of race in America.

Keywords: Residential and Social Segregation; Networks; Social connectedness.
JEL Codes: R23; J15.

*Corresponding author. Email: andreas.diemer@sofi.su.se. Funding from NordForsk (IntegrateYouth, 95263) and from FORTE (Interlocking Inequalities, 2016-07099) is gratefully acknowledged.

1 Introduction

Residential segregation by race and ethnicity persists across the United States (US) despite many initiatives aimed at desegregation (Graham, 2018). For instance, in 2020, around 13 percent of US ZIP Codes were mostly non-White with more than 80% racial or ethnic minority residents. Since at least the work of Wilson (1987), social scientists have developed tremendous interest towards measuring the extent and understanding the impacts of racial and ethnic segregation. Many studies have shown that residential segregation is a crucial factor explaining the disparity in different socioeconomic outcomes across neighbourhoods, including educational attainment, earnings, family structure, crime, health, and subjective well-being (e.g., Cutler and Glaeser, 1997; Fryer, 2011; Massey, 2017; Krivo et al., 2009; Ludwig et al., 2012). In addition to the well-known challenges of identifying causal effects of residential segregation,[1] many studies have emphasized the inadequacies associated with existing segregation measures that are constructed based only on the geography of residences. For instance, Massey and Denton (1988) and Graham (2018) have highlighted that studies often proposed different measures to depict residential segregation, with little consensus on which is the most appropriate.

Against this backdrop, a burgeoning stream of literature aims to improve the measurement of residential segregation. One dimension of improvement recognises that individuals are mobile, and so works towards developing segregation indices that integrate the racial composition of locations individuals visit over different times of day (e.g., Wang et al., 2018; Davis et al., 2019; Athey et al., 2021; Abbiasov, 2020; Cook et al., 2022; Magontier et al., 2022). However, these measures are inherently spatial in nature and do not capture actual social connections of, or interactions between, an individual and others. Another dimension of improvement aims to measure social contact directly. As emphasised by Echenique and Fryer (2007): *"The ideal data to estimate residential segregation would contain information on the nature of each household's interactions with other households"*. The authors offer a seminal theoretical contribution to this dimension by constructing an index that is based on people's social interactions. However, their empirical application is limited to a small subset of high school students responding to the Add Health survey. Only recently have studies begun to leverage large-scale network data, such as phone call records or online social media platforms, to study social ties and residential location (e.g., Cornelson, 2017; Büchel et al., 2020; Tóth et al., 2021; Chetty et al., 2022*a*,*b*).

In this paper, we propose a new segregation index that incorporates direct measures of social connections between granular neighbourhoods. The index captures the racial isolation of social connections in US cities, which we simply refer to as 'social isolation'. We

[1] Ananat (2011) outlines the empirical challenges associated with identifying the causal effect of residential segregation. This is primarily driven by the presence of omitted variables and sorting of households.

show that this index can be decomposed into a weighted sum of conventional residential isolation (Gentzkow and Shapiro, 2011) and measures of isolation of social connections across neighbourhoods within the city. We then take this measure to the data using Facebook social connectedness between ZIP Codes (Bailey et al., 2020).[2] We use data from Facebook because it is the world's largest social network with more than 258 million active users, or around 70% of the population, in the United States and Canada. The representativeness of Facebook usage means that social connections can realistically depict actual friendship networks across granular neighbourhoods.[3] Using our social isolation measure, we offer novel estimates of segregation for the 75 largest US Metropolitan Statistical Areas (MSAs), and benchmark our results with conventional residential or spatial measures at different levels of aggregation.

First, we illustrate using the examples of Chicago and Washington DC how social connections can be strongly biased towards neighbourhoods with similar racial compositions compared with spatially defined measures. On this basis, we argue that spatial proxies miss meaningful variation in the actual social connections of a city. Cross-neighbourhood segregation indices that rely on spatial relationships can therefore significantly under-represent actual social segregation.

Building up our indices to the city level, then, we show that residential isolation and social isolation are highly correlated, as is to be expected based on the decomposition we propose. At the same time, we illustrate how meaningful variation exists in relative terms between social and residential measures. For example, cities like Miami, Washington DC, and Cincinnati rank higher in social segregation than they do based on conventional residential isolation, while New Orleans, San Francisco, and Richmond fall in ranks. These changes reflect different propensities of ZIPs with similar residential compositions to interact with members of other racial groups. For instance, a largely White neighbourhood in Washington DC may be no more socially segregated than its own racial composition would suggest, while a comparable neighbourhood in New Orleans is less socially segregated — that is to say more socially exposed to non-White neighbourhoods — than would be implied by its own racial composition.

Next, we characterise cities that are highly socially isolated, even for comparable levels of

[2] Bailey et al. (2020) focuses on explaining how geographical distances and public transit networks influence the establishment of social connections, before measuring how social connectedness across space affect commuting behaviours and providing correlations of geographical concentration of social connectedness and various socio-economic outcomes. Our paper is different as we rely on the social connectedness index (SCI) to quantify between neighbourhoods (or cross-boundary) social interactions, before using these measures to re-evaluate how racially segregated neighbourhoods are.

[3] The use of these data to study patterns of social interactions and consequences of 'economic connectedness' is well established (Chetty et al., 2022*a*,*b*). These papers, however, do not explore interactions of individuals belonging to different racial or ethnic groups, which is the focus of our paper.

residential segregation. We show that, conditional on the latter, the social component of isolation is negatively related to association density and church adherents rate, mirroring a long literature that argues for the functional value of these institutions in fostering social cohesion (Putnam, 2000, 2007; Chetty et al., 2022*a*).

Finally, we consider social segregation at the ZIP Code level and examine its association with a range of socio-economic outcomes of Black residents. Our results suggest that growing up in more socially isolated neighbourhoods is materially correlated with various adulthood outcomes. There is also substantial heterogeneity in this relationship depending on whether one is residing in a residentially segregated neighbourhood. Specifically, social isolation matters most for Black children growing up in homogeneous neighbourhoods, i.e., largely made up of households of the same race. In such neighbourhoods, there is a positive relationship between social isolation and the fraction of children who never move away from their commuting zone or who end up in jail, and a negative relationship with mean income ranks or the probability of being married.

Our paper contributes to a growing literature that explores the use of large-scale information arising from novel sources to advance our understanding of segregation. Our primary contribution is to conceptualise a measure of segregation that allows for interactions of people across spatial units. We show how traditional residential segregation is nested in our index through the (often implicit) assumption of no linkages across different neighbourhoods.

Secondly, using this measure, we leverage comprehensive social networks data to estimate social isolation metrics for the 75 largest US urban areas and discuss changes in city-level rankings of segregation when using our measure. We offer specific examples of cities that are relatively more or less socially segregated than residential measures would suggest. Incorporating new data on social connections into our analysis allows us to create an empirical segregation measure more closely tied to the theoretical ideal. Previous work focused on using new data to improve racial segregation measures based on where and when people spend their time. More generally, many existing analyses are forced to make assumptions about the intensity of social connections within and across geographical boundaries.[4] Instead, we are the first to explicitly account for social connections between people, and to propose a measure of social segregation of US racial minorities informed by the near universe of social interactions across the US that involve the MSAs we consider.

[4]For instance, residential isolation measures assume that social interactions occur only within one's home geographical neighbourhood, and not at all across (Gentzkow and Shapiro, 2011). Experienced isolation measures assume that social interactions take place between people who co-locate in time and space, but still do not measure social interactions between these co-locators directly (Athey et al., 2021).

In addition, we also offer a formal decomposition of our measure that demonstrates how social segregation is related to traditional residential isolation, adjusted for the propensity to socially interact with people from other groups within and across neighbourhood boundaries. Importantly, together with this decomposition, the measure we propose can be used in urban research to separately account for the influence of different dimensions of segregation.

The organization of the rest of our paper is as follows. Section 2 outlines the conceptual framework on how residential segregation has been measured traditionally in the literature, introduces our index of social isolation, and elucidates how this measure is an improvement to conventional measures. Section 3 discusses data and measurement aspects, including the information on social interactions that our index relies on. Section 4 presents descriptive results, emphasising discrepancies between residential and social isolation measures within and across US urban areas. It also discusses city-level features that correlate with such discrepancies, as well as the relationship between neighbourhood-exposure to minorities and various socio-economic outcomes. Section 5 concludes.

2 Conceptual Framework

2.1 Traditional Measures of Residential Segregation

The concept of residential segregation describes how different groups of individuals, typically categorized by race or socio-economic status (e.g., income), are living apart from one another. Residential segregation can be broadly classified into five different groups, namely *evenness* (relative distribution of certain groups across space), *exposure* (how likely different groups are going to be in contact across space), *concentration* (relative amount of space occupied by certain groups), *centralization* (extent certain groups are located in city centers) and *clustering* (extent to which contiguous areas are inhabited by certain groups) (Massey and Denton, 1988). Among the different measures, the two most relevant concepts to our paper are exposure and clustering.

Exposure is particularly relevant as we care about how social networks can affect interactions between different groups that could, in turn, influence socio-economic outcomes. A particularly prominent way of measuring exposure is the isolation index because of its intuitive appeal (White, 1986; Cutler et al., 1999; Echenique and Fryer, 2007; Gentzkow and Shapiro, 2011; Athey et al., 2021). This index can be interpreted as the expected share of a minority group in a unit occupied by a minority person, or the extent to which minorities disproportionately reside in areas where other residents are also minorities. Following Gentzkow and Shapiro (2011) and Athey et al. (2021), the residential isolation index for city c ($RISO_c$) can be expressed as follows:

$$RISO_c = \underbrace{\sum_{i \in c} \left(\frac{x_i}{X_c} \frac{x_i}{t_i} \right)}_{\text{Average minority exposure to minorities}} - \underbrace{\sum_{i \in c} \left(\frac{y_i}{Y_c} \frac{x_i}{t_i} \right)}_{\text{Average majority exposure to minorities}} , \quad (1)$$

where i denotes a neighbourhood that is situated in city c. The terms x_i and y_i represent the minority and majority population counts respectively, and t_i is the total neighbourhood population ($t_i = y_i + x_i$). Hence, $\frac{x_i}{t_i}$ is the share of minority population in neighbourhood i. Under the assumption that individuals are exposed uniformly to residents living in other neighbourhoods, this term can be interpreted as minority exposure. Finally, X_c and Y_c are the sum of the minority and majority group populations for city c respectively, i.e. $X_c \equiv \sum_{i \in c} x_i$ and $Y_c \equiv \sum_{i \in c} y_i$. Hence, $RISO_c$ measures the minority-weighted average exposure of the minority group to minorities subtracted by the majority-weighted average exposure of majority group to minorities. This measure varies from 0 (no isolation) to 1 (complete isolation). The second term, sometimes omitted in traditional applications, adjusts for the effect of overall city composition. With a small minority population, the minority exposure to minorities mechanically tends to be smaller. By subtracting the majority population's exposure to minorities (also referred to as the majority's interaction index[5]), this measure is comparable across cities with different compositions.

We also build on Massey and Denton's notion of clustering in residential segregation. A major concern associated with the exposure measure in Equation (1) is that the degree of segregation depends on how administrative boundaries are drawn — sometimes referred to as the 'grid problem'. Redrawing these boundaries could drastically influence segregation measures. We illustrate this in Appendix Figure A.1 (based on Echenique and Fryer, 2007) where a hypothetical city can move from perfect integration to full segregation depending only on how boundaries are drawn. The fundamental pitfall is that traditional measures assume interactions are confined within neighbourhood boundaries. Hence, once boundaries are drawn, individuals' location within the city does not matter — an issue known as the 'checkerboard problem' (White, 1983). In reality, individuals can communicate, interact, and be influenced by others beyond these boundaries, for instance, if they commute from their residence to different places during the day. Recent research indeed documents that daily movements across the city result in exposure to communities potentially very different from those in one's place of residence (Athey et al., 2021). In the past, measuring cross-boundary relationships has been very difficult in practice, because researchers did not have information on social connections.

[5]More generally, the interaction index measures how the average member of a group is exposed to members not belonging to that same group. In other words, the interaction index is the inverse of the isolation index.

2.2 An Index of Isolation with Cross-Boundary Interactions

To allow for cross-boundary linkages between areas, we incorporate and modify the distance-decay isolation index introduced by Morgan (1983), which can be expressed as follows:

$$SISO_c = \sum_{i \in c} \left(\frac{x_i}{X_c} \sum_j \omega_{ij} \frac{x_j}{t_j} \right) - \sum_{i \in c} \left(\frac{y_i}{Y_c} \sum_j \omega_{ij} \frac{x_j}{t_j} \right). \quad (2)$$

The notable difference from Equation (1) is the inclusion of ω_{ij} weights to account for exposure to minorities between all neighbourhoods i and j. In fact, Equation (1) represents a special case of this new measure, where own-neighbourhood weights are set to one and all others to zero, that is to say $\omega_{ij} = \mathbf{1}(i = j)$, meaning that individuals are only exposed to others who reside in the same area. It is not hard to see that the major downside of the traditional residential segregation measure $RISO_c$ is that it completely discounts cross-boundary interactions, and that we can improve on this measure by allowing for some form of non-zero linkages between areas. The key empirical challenge, thus, is to define weights so as to accurately capture the strength of connections between spatial units.

Massey and Denton (1988) suggest that one can set interaction weights equal to a negative exponential distance decay function:

$$\omega_{ij}^d = \frac{exp(-\delta d_{ij}) t_j}{\sum_k exp(-\delta d_{ik}) t_k}. \quad (3)$$

This assumes that the probability of meeting individuals living in neighbourhood j when living in neighbourhood i decreases as the bilateral distance d_{ij} between the two neighbourhood increases. The interpretation of these weights is also straightforward as exposure to other areas is measured as a constant decay function in the geographical space. By replacing ω_{ij} in equation 2 with ω_{ij}^d, the resulting $SISO_c^d$ measure can be interpreted as capturing spatial isolation, and can be expressed as follows:

$$SISO_c^d = \sum_{i \in c} \left(\frac{x_i}{X_c} \sum_j \omega_{ij}^d \frac{x_j}{t_j} \right) - \sum_{i \in c} \left(\frac{y_i}{Y_c} \sum_j \omega_{ij}^d \frac{x_j}{t_j} \right). \quad (4)$$

There are, however, at least two issues with this spatial isolation measure. First, it assumes that exposure is a smooth function of distance. This may not hold true in practice, as interactions could be influenced by choice of commuting, workplace, interests, social life, as well as by natural and man-made barriers in the urban environment such as rivers or railroads (e.g., Ananat, 2011; Tóth et al., 2021; Mahajan, 2024). For instance, if cross-boundary interactions are largely determined by workplace, spatial isolation will be placing too much weight on nearby residential neighbourhoods that may bear little relevance with the actual residential location of co-workers across the city. Second, this measure requires researchers to determine the decay parameter (δ) and there is traditionally little prior knowledge or consensus on the range of its possible values. Nevertheless, a spatial definition of isolation can still prove useful in characterising dimensions of segregation that matter for outcomes mostly determined by spatial considerations, such as access to jobs, absent any actual social interaction between people across neighbourhoods.

2.3 Incorporating Social Interactions

Evidently, the accuracy of the isolation index depends on how we measure the exposure weights (ω_{ij}). To measure how households interact across neighbourhoods, previous research relied on social networks information from survey data (Echenique and Fryer, 2007) and, more recently, on GPS co-location of mobile devices throughout the course of a day (Athey et al., 2021; Abbiasov, 2020; Cook et al., 2022). The direct measurement approach using GPS, while a significant improvement over just using bilateral distance, also faces several limitations. As noted in Athey et al. (2021), while the authors observe when devices occupy the same geographical space, they cannot observe actual interactions between individuals.[6] Hence, to mitigate these concerns and accurately model cross-boundary linkages (ω_{ij}) that account for existing contacts between people, we propose to use a measure introduced by Bailey et al. (2020), the Social Connectedness Index (SCI), which is based on counts of friendship connections between Facebook users across different ZIP Codes with at least 500 residents. This index builds on the universe of active Facebook users as of March 2020, and captures the (scaled) relative probability of a friendship link between users in two locations. It is expressed as follows:

$$SCI_{ij} = \mu \frac{Connections_{ij}}{Users_i \times Users_j}, \tag{5}$$

where $Connections_{ij}$ is the observed number of Facebook friendships between ZIP Code i and ZIP Code j, $Users_i$ and $Users_j$ are the number of Facebook users in ZIP Codes i and j (i.e., the denominator is the total possible number of Facebook connections across ZIP Codes i and j), and μ is a re-scaling constant for privacy purposes. This measure can be interpreted as the likelihood that a random individual from ZIP Code i is friends with

[6] For instance, consider a hypothetical scenario of a restaurant with two customers and a chef. Their measure assumes that these two customers are as exposed to one another as they are exposed to the cook based on the GPS locations. This is despite the fact that individuals might not know each other and have zero interactions with one another. We also highlight that segregation policies in the early 20th century USA would often operate at highly localised levels. There might be minimal or zero interactions between Black and White patrons even when they co-locate in the same theatres and/or restaurants.

a random individual from ZIP Code j.[7] We assume that real-life cross-boundary social interactions of individuals in ZIP Codes i and j can be accurately proxied by $Connections_{ij}$ and we set interaction weights as follows:

$$\omega_{ij}^{s} = \frac{SCI_{ij} t_j}{\sum_k SCI_{ik} t_k}, \tag{6}$$

where ω_{ij}^{s} captures the proportion of friendships between neighbourhood i and j, out of all friendships involving neighbourhood i. This is an improvement in measuring between-neighbourhood interactions compared to Equation 3 that requires researchers to assume a value for the spatial decay parameter (δ). As mentioned earlier, not only is it hard for researchers to determine an appropriate value for δ, it is unlikely that interactions across space follow a constant exponential decay. We will see evidence of the limitation of such an assumption in Section 4, e.g. in Figure 1b.

Note that despite its merits, this strategy comes with some limitations of its own. Notably, because we do not observe friendships at an individual level, we must assume that residents of a neighbourhood are exposed to residents of other neighbourhoods based on the mean neighbourhood-neighbourhood social connections. Therefore, an unobserved correlation between demographics and social connections across individuals within neighborhood pairs, in turn, can lead to a bias in our measure of isolation. The direction of this bias is uncertain a priori, though, if unobserved connections exhibit homophily the smoothing likely pushes estimates closer to the difference in average minority and majority composition in the city.

Replacing ω_{ij} with ω_{ij}^{s} in Equation 2, we can derive our social isolation measure ($SISO_c^s$) in equation 7. By rearranging terms, we can further decompose $SISO_c^s$ and express it in terms of own-area residential isolation ($T1_c$), own-area social isolation ($T2_c$), and other

[7] There is a special case links within the same ZIP Code (i.e., when $i = j$). In this case the scaled probability of a friendship link is $\mu \frac{Connections_{ii}}{0.5\,Users_i(Users_i - 1)}$. The SCI for these pairs is constructed by doubling friendship connections and not counting self-friendships (i.e., $SCI_{ii} = \mu \frac{2Connections_{ii}}{Users_i \times Users_i}$). Therefore, the scaled probability of a friendship link is $SCI_{ii} \times \frac{Users_i}{Users_i - 1} \approx SCI_{ii}$.

areas social isolation ($T3_c$), as follows:

$$\begin{aligned}
SISO^s_c &= \sum_{i\in c}\left(\frac{x_i}{X_c}\sum_j \omega^s_{ij}\frac{x_j}{t_j}\right) - \sum_{i\in c}\left(\frac{y_i}{Y_c}\sum_j \omega^s_{ij}\frac{x_j}{t_j}\right) \\
&= \sum_{i\in c}\left(\frac{x_i}{X_c}-\frac{y_i}{Y_c}\right)\omega^s_{ii}\frac{x_i}{t_i} + \sum_{i\in c}\left(\frac{x_i}{X_c}-\frac{y_i}{Y_c}\right)\sum_{j\neq i}\omega^s_{ij}\frac{x_j}{t_j} \\
&= \underbrace{\overline{\omega^s_{iic}} RISO_c}_{\substack{\text{Own-area} \\ \text{residential term} \\ (T1_c)}} + \underbrace{\mathrm{cov}_c\left[\left(N_c\frac{x_i}{X_c}-N_c\frac{y_i}{Y_c}\right)\frac{x_i}{t_i}, \omega^s_{ii}\right]}_{\substack{\text{Own-area} \\ \text{social/spatial term} \\ (T2_c)}} + \underbrace{\sum_{i\in c}\left(\frac{x_i}{X_c}-\frac{y_i}{Y_c}\right)\sum_{j\neq i}\omega^s_{ij}\frac{x_j}{t_j}}_{\substack{\text{Other-area} \\ \text{social/spatial term} \\ (T3_c)}}. \quad (7)
\end{aligned}$$

Here, $T1_c$ is the residential isolation of city c scaled by the average social weight assigned to own-area interactions in the city, $\overline{\omega^s_{iic}}$. This expression is akin to traditional measures of segregation based on exposures within the boundaries of one's place-of-residence.

The second term, $T2_c$, is the city-level covariance between own-neighbourhood interaction weights, ω^s_{ii}, and each neighbourhood's excess residential exposure of minorities to other minorities relative to exposure of the majority to minorities.[8] This term can be thought of as capturing the own-area contribution to social isolation, that is, the average propensity to interact within the same area depending on that area's composition. It can take positive and negative values, depending on the nature of this relationship. When positive, local social interactions exacerbate overall social isolation. Intuitively, cities where more residentially isolated neighbourhoods tend to interact more within their own boundaries (i.e., higher covariance) will display higher levels of social isolation overall. Vice-versa, negative values suggest that residents of relatively more homogeneous neighbourhoods seek friendships elsewhere in the city — note that these may or may not be with out-group members. This reduces social isolation for any given value of $T1_c$ and $T3_c$.

Finally, $T3_c$ is the city-level weighted average exposure to minorities residing in different neighbourhoods (scaled by the intensity of interaction with these places) with weights proportional to the difference in relative concentration of minority and majority group members in each city's neighbourhoods. It is constructed analogously to the expression in Equation 2, but explicitly excludes within-neighbourhood exposures from the calculation (effectively imposing $\omega_{ii} = 0$). This last term arguably captures social isolation that depends solely on interactions with other neighbourhoods.

[8]More accurately, the term inside round brackets captures how much a particular Zip Code disproportionately deviates from the city average number of non-White residents, $\frac{N_c}{X_c}$, relative to White, $\frac{N_c}{Y_c}$, where N_c denotes the the number of neighbourhoods in the city.

There are several reasons for decomposing $SISO_c^s$ into its constituents. First, it allows us to understand how our social isolation measure compares with traditional residential isolation as expressed in Equation 1. Specifically, to what extent does residential isolation influence how socially isolated cities are overall? Second, the decomposition is useful to examine how different aspects of social isolation relate to each other. Particularly interesting is a comparison of own- and other-area social terms. For instance, in cities where residents of more residentially isolated neighbourhoods tend to form connections outside neighbourhood boundaries (i.e., $T2_c < 0$), do cross-neighbourhood ties in these cities tend to connect places that are demographically more different or similar to their own neighbourhood (i.e., are $T2_c$ and $T3_c$ positively or negatively correlated)? Third, we also observe in our data that social interactions tend to be strongest within one's own ZIP Code. This decomposition allows us to disentangle and separately account for the influence of residential isolation, own-area social isolation and other areas' social isolation on various city or neighbourhood-level outcomes.

By relying on this decomposition, in short, urban researchers can gain a more nuanced picture of what it means for a city to be segregated. Noticeably, by replacing ω_{ij}^s with ω_{ij}^d, we can also construct and decompose spatial isolation ($SISO_c^d$) into own-neighbourhood residential isolation, own-area spatial isolation, and other areas spatial isolation. While that is not the focus of this paper, we note that this decomposition is also useful for researchers studying the constituents of spatial isolation on other socio-economic outcomes.

3 Data and Measurement

Our definition of neighbourhoods is based on the 2010 US Census Bureau's five-digit Zip Code Tabulation Areas (ZCTA, henceforth also referred to as ZIP Codes, ZIP areas, or simply ZIPs). This is the most granular level of aggregation for which we observe information on social connections. There are 33,120 ZCTAs in the US, each formed by grouping together Census blocks. We only consider ZCTAs within MSAs,[9] and limit our analysis to MSAs with at least 50 ZCTAs. This leaves us with just under 10,000 ZCTAs across 75 MSAs.

We measure connection strength using Facebook's SCI (Bailey et al., 2020), already introduced in Section 2.3, assuming that it accurately proxies real-life cross-boundary social interactions of individuals in ZIP Codes i and j. There are at least two reasons to justify this. First, more than 70 percent of the US and Canadian population uses Facebook, making the SCI a representative measure of social connections. Second, existing literature

[9]MSAs are a subset of Core Based Statistical Areas (CBSA) involving at least one urban area with a population of 50,000 or more.

has shown that social interactions revealed through Facebook connections can materially affect various socio-economic outcomes and decision making.[10] Furthermore, because the measure is based on active users only, connections likely portray cumulative friendships up to the point the index was constructed, rather than old friendships only. With the SCI, thus, we can conveniently measure interactions directly at a granular ZIP Code level without the need of measuring social connections and identifying real-time locations at an individual level.

For each urban ZCTA, we retain information on geographical distance and social connections with other ZIPs regardless of where these other ZIPs are located (they may or may not be in the same city). For computational purposes, however, we calculate social and spatial exposures imposing a restriction on relevant linkages. For social exposures, we retain the top 1,000 paired ZIP Codes by SCI strength. For spatial exposures, we consider the top 1,000 paired ZIP Codes by distance, conditional on these ZIPs being within 100 miles from each other (about 160 km). We retain all available ZIPs if there are fewer than 1,000 ZIPs within 100 miles.[11] In either case, we assume linkages to be absent or negligible beyond these cut-offs, which we show is also true in our data (see Appendix Figure A.2). We estimate the spatial decay parameter δ in Equation (3) as:

$$\ln SCI_{ij} = \alpha_i + \delta \ln d_{ij} + \epsilon_{ij}, \tag{8}$$

where α_i denotes ZIP Code fixed effects. We fit Equation (8) on the set of nearest 1,000 ZIPs for each neighbourhood in the US for which we also observe social connectedness, weighting each ZIP pair by the product of their residents. The parameter δ takes a value of -1.38, the elasticity of friendships for percentage distance increments.[12]

Information on the demographic composition of neighbourhoods comes from time series tables compiled by the National Historical Geographic Information System (NHGIS) IPUMS project (Manson et al., 2021). This source provides 1990, 2000, 2010, and 2020 US Census data standardised for 2010 definitions of geographical units. We distinguish between White (W) and non-White (NW) Americans (including all Hispanic or Latino) in the main analyses, but also separately consider Black and Hispanic or Latino people for selected empirical applications.

[10] See for instance: Bailey, Cao, Kuchler, Stroebel and Wong (2018); Bailey, Cao, Kuchler and Stroebel (2018); Bailey et al. (2019, 2022); Gee et al. (2017); Wilson (2020).

[11] For consistency, we also drop any pair involving at least one ZIP Code for which no data on social connectedness is available – i.e., ZIPs that never appear in the SCI files.

[12] Appendix Table A.1 provides estimation details, and alternative specifications. In particular, column (2) shows that the elasticity obtained by restricting the sample to ZIP Codes in large urban areas only (the focus of our analysis in the rest of the paper) is comparable to that for all US ZIP Codes. Incidentally, the elasticity we obtain for distance using all US ZIP Codes is also comparable in magnitude to that of -1.42 estimated by Bailey et al. (2020) for transit travel time using data on the New York metro area.

4 Results

4.1 Segregation Within and Across Urban Areas

We begin by illustrating the differences between the approaches to measure minority exposure discussed in Section 2. Figure 1 compares social and spatial weights (panels b and c respectively) with the geographical distribution of non-White residents (panel a) for highly segregated ZCTAs in Chicago (subfigure A) and Washington DC (subfigure B). We use the same scale in panels (b) and (c) to facilitate like-for-like comparison.[13] The contrast in the geographical distribution of weights strength between panels (b) and (c) is evident. Linkages measured using social connections are not only much more spatially concentrated than their distance-based counterparts, but they also more closely track the locations of other highly non-White neighbourhoods irrespective of where they are located in the city. The maps in Figure A.3 in Appendix further show that this is also true when considering friendships with other ZIPs across the US, some of which are located several hundreds of kilometres away. This result, likely reflecting homophily in network formation (McPherson et al., 2001), offers powerful visual evidence for the importance of moving beyond purely spatial measures to characterise cross-boundary connections between areas in the study of segregation.[14] These patterns are observed both in Chicago and in Washington DC. Analogous maps drawn for other MSAs are available in the Appendix (Figures A.6-A.8).

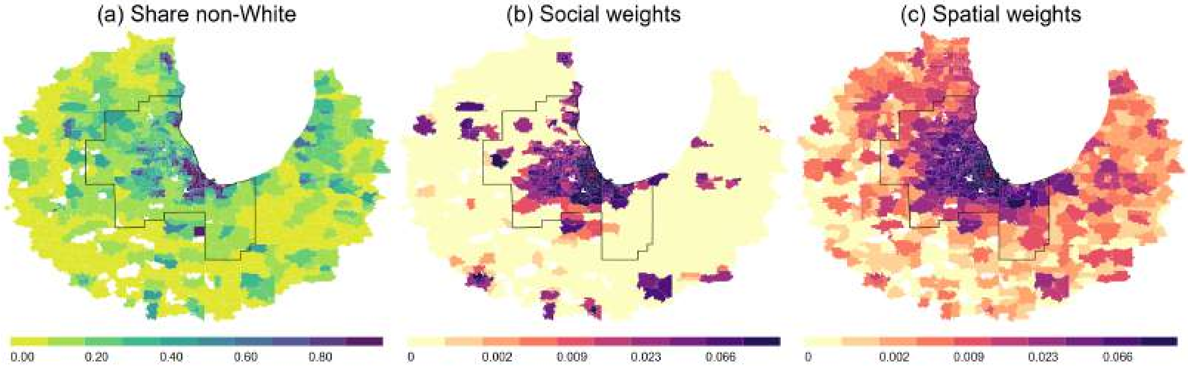
Next, we map ZCTA-level minority exposures for Chicago in Figure 2. Panel (a) shows residential exposure, which is equivalent to the simple share of non-White Americans in the neighbourhood. Panels (b) and (c) show social and spatial exposures respectively, omitting the contribution of own-ZIP compositions to emphasise differences in these two measures.[15] Two features stand out. First, compared to residential exposure, social and spatial measures display a somewhat smoother geographical distribution, which is due to the averaging of minority shares over multiple neighbourhoods. As a result, some ZIPs with very high (low) values of residential exposures have lower (greater) values in the social and spatial counterparts of this measure. This is because we depart from the assumption that individuals are uniformly exposed to co-residents only, and allow for

[13] Possible weight values range from null (no connections) to one (all connections to a single area). Note that, for readability, we restrict the maps to the nearest 1,000 ZIPs to our focal neighbourhood. This captures all spatial linkages by definition, but may mask social ties to some areas outside this range.

[14] The discrepancy between social and spatial characterisations of linkages is even starker when considering weights that are not re-scaled by the relative population counts in each neighbourhood – i.e., if the t_j term and its summation across all neighbourhoods are omitted from Equations (3) and (6). Appendix Figure A.4 gives maps showing these alternative weights for Chicago (panel a is unchanged).

[15] In other words, we map exposures obtained from the inner summation of the third term in Equation 7. Own-ZIP compositions are common to both measures and apportioned with high weights. Thus, including them would partly mask the distinctive traits of considering a social, rather than spatial, definition of neighbourhood linkages.

A. Chicago-Joliet-Naperville, ZCTA no. 60620, 99.48% non-White.



B. Washington-Arlington-Alexandria, ZCTA no. 20743, 98.10% non-White.

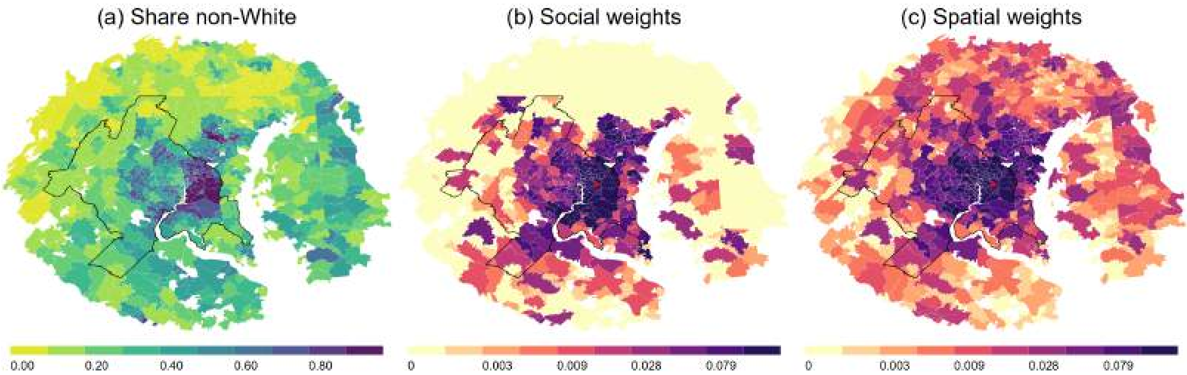


Figure 1 – Maps with MSA race composition (a), social (b) and spatial (c) weights for the least White ZCTA in the city (marked in red on the map). Weights are rescaled by a factor of 100 for legibility. Breaks are defined at each decile of the distribution obtained by pooling both social and spatial weights together. MSA boundaries are in black.

cross-boundary interactions. Chicago's inner city neighbourhoods, for instance, largely non-White, are not as highly exposed to other non-Whites according to these measure as pure residential composition would suggest. Second, spatial exposures in panel (c) show much less geographical variation than social exposures in panel (b), which is a result of the high-degree of smoothing imposed when constructing the former measure. We believe this to reflect a somewhat inaccurate modelling of cross-boundary interactions as purely depending on spatial proximity. These observations also hold for several other urban areas in the US (maps are available in Appendix, Figures A.5-A.9). Next, we consider segregation measures at city-level.

City-level indices of segregation are constructed in line with Equation 2, as the minority-weighted average of ZIP-level exposures (including own-area contributions) minus the majority-weighted average of the same measure. Due to the limitations of using spatial proxies for cross-boundary linkages outlined above, we focus this discussion on residential

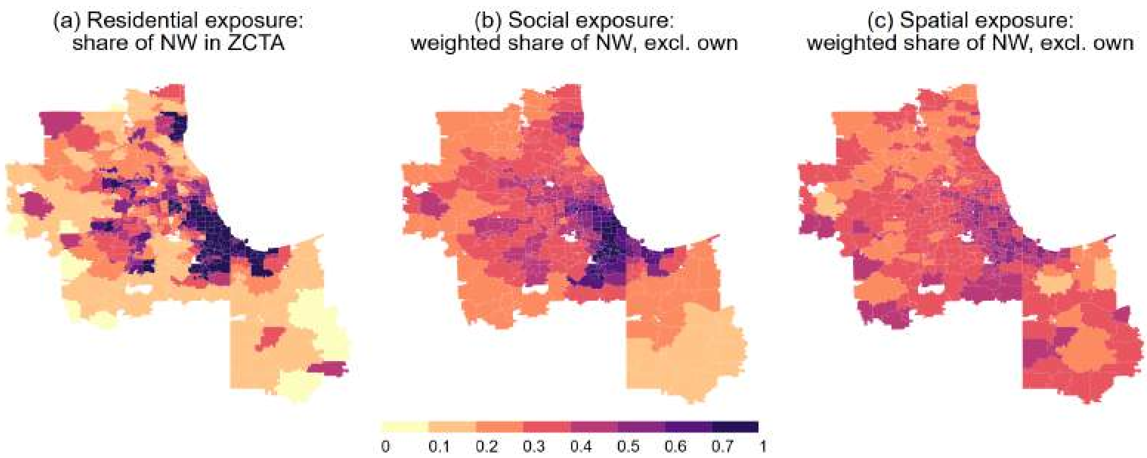


FIGURE 2 – Residential, social, and spatial exposures in Chicago-Joliet-Naperville. The maps show within-MSA variation in minority exposure measures for the city's neighbourhoods, as an example. Exposure values for each ZCTA in panels (b) and (c) are obtained as the linkage-weighted sum of non-White residents proportions across all connected neighbourhoods. Possible exposure values range from null (all White) to one (all non-White). The same choropleth scale was used in all panels to facilitate comparisons across maps. This, however, hides some variation at lower levels of social and spatial exposure.

and social isolation measures (henceforth also just $SISO_c$, omitting the s superscript).[16] The average MSA has a social isolation score of 0.04. By contrast, mean residential isolation is 0.10. If we only consider the largest 75 MSAs, those with at least 50 ZCTAs, mean values for social and residential isolation are 0.08 and 0.18 respectively. Appendix Figures A.13 and A.14 map residential and social isolation to illustrate spatial variation across the US.

According to these measures, American cities tend to be more residentially than socially segregated. In fact, social isolation is systematically lower than residential in all the largest MSAs, as illustrated in panel A of Appendix Figure A.10.[17] Similar to what is documented in Echenique and Fryer (2007), residential and social isolation are strongly correlated with

[16] We also run into two empirical issues when estimating the spatial measure that are specific to our application. First, by considering the nearest 1,000 ZCTAs to construct spatial exposures in each neighbourhood, we pick up areas in suburbs, even outside city boundaries, which are predominantly White. As a result, the scaling of minority exposure to other minorities by the majority's exposure to minorities using city-wide averages mechanically centers the index around low values for most urban areas, particularly the smallest ones. Second, the high-degree of smoothing involved in the spatial isolation measure exacerbates this issue, notably because we are forced to use relatively aggregate data, which increases measurement error. To a lesser extent, these two concerns also exist for the social isolation index.

[17] Athey et al. (2021) discuss a similar result when comparing their measures of experienced and residential isolation for a sample of large cities. This is not entirely surprising, considered that the rank correlation of our measure with theirs is 0.84 (see also Figure A.11 in Appendix). Figure A.10 also shows in panel B a comparison of social isolation with its spatial counterpart. Despite their positive correlation, there is more variation between these two, which appears to be driven in part by city size. This however is also in line with our concern about downward bias in the spatial measure, which is particularly acute in smaller urban areas.

one another. For all 75 largest MSAs the Pearson correlation between the two measures is 0.97. We interpret this to indicate that social ties mimic, or perhaps even consolidate, existing patterns of residential segregation, rather than reducing it. This is also true for index constituents. The first term of the decomposition in Equation 7, which isolates the role of residential composition, has a correlation coefficient of 0.83 with the third term, which captures the contribution of social interactions with other areas to overall isolation. Interestingly, this third term is negatively correlated with the second one for own-area interactions. To interpret this result, it is useful to first note that the average value of $T2_c$ in the cities we consider is negative. This negative value indicates that, in the average city, neighbourhoods with more residentially isolated non-Whites tend to display a lower share of friendships within their own neighbourhood boundaries (their friendship networks are geographically broader). At the same time, the negative correlation between $T2_c$ and $T3_c$ suggests that, across cities, this tendency for neighbourhoods with more residentially isolated non-Whites to interact outside of their neighbourhood does not translate in more diverse friendships outside the local area. A reduction in overall city social isolation from this term tends to be counter-weighted by an increase in social isolation stemming from interactions with other places.

Despite its positive correlation with residential measures, there is also meaningful variation in social isolation, conditional on residential segregation. Appendix Table A.2 lists index values for all MSAs, along with standardized scores of each index that allow to compare cities in relative terms. Places like St. Louis MO, Duluth MN, Las Vegas NV or Detroit MI, for instance, are relatively more socially than residentially segregated compared to the average urban area. The opposite is true for Jackson MS, Memphis TN, Richmond VA, and Providence RI. Some places rank very differently in the two measures too. Scranton PA is only 55th in terms of residential isolation, but 31th for social isolation. Washington DC and Miami climb from 27th to 19th and from 16th to 10th in the same comparison. By contrast, Charlotte NC is 41st for residential isolation, despite ranking only 59th in the social measure. New Orleans and San Francisco respectively fall from the 25th and 42nd (residential) to the 33rd and 51st places (social). A visual comparison of ranks for all largest MSAs is available in Appendix Figure A.15.

Generalizing the ZIP-level results from the previous section, Figure 3 compares residential and social segregation across selected large cities with at least 50 ZCTAs. Each graph plots using blue hollow markers the average NW social exposure of every ZIP against that ZIP's share of NW residents (i.e., residential exposure, on the x-axis). A population-weighted local polynomial fit is also overlaid (solid blue line), along with an estimate for the linear association between the two variables (also population-weighted, reported in each graph with blue text). For reference, we also plot in red the 45° line marking the relationship between residential exposure of each ZIP and its composition (which by definition are the same), and in green the population-weighted local polynomial fit between average spatial exposure and residential composition. We consider three types of cities: panels A and B on the top plot cities that rank higher in terms of social isolation relative to residential, C and D in the middle consider cities that rank about the same, E and F at the bottom those that rank higher in residential rather than social isolation.

Confirming city-level patterns, social and residential exposures are strongly correlated.[18] A closer examination, however, also reveals some interesting discrepancies. Firstly, cities that are relatively more socially than residentially isolated (top panels) also display the steepest profile for this correlation, whereas those that rank higher in terms of residential relative to social segregation show weaker associations (bottom panels). In Washington DC for example (panel A) ZIPs that are twenty percent non-White are socially exposed to about thirty percent of non-White individuals living elsewhere. In New Orleans instead (panel E) a ZIP with corresponding non-White composition is socially exposed to nearly forty percent non-White people from other areas. On the other end, an 80 percent non-White ZIP in DC shows average social exposure scores of nearly 0.7, in contrast to an average score of nearly 0.5 in New Orleans. Estimates for the linear association between own-ZIP composition and social exposure confirm this pattern (0.48 vs. 0.33 in DC and New Orleans, respectively). Chicago, a city that is about as residentially as socially segregated in terms of ranks, lies somewhere in between these two cases. In short, cities that are less socially segregated than residentially segregated, are places where White ZIPs are more exposed to non-Whites and non-White ZIPs are more exposed to Whites.

Second, this pattern holds true regardless of the overall composition of the city (although levels change somewhat). Panels A, C, and E (left-hand side) describe cities whose residents are about as likely to be White than non-White, whereas panels B, D, and F (right-hand side) focus on more diverse cities that are about two-thirds non-White. In all cases, the linear association is strongest for MSAs in the top panels, and weakest for those at the bottom. In sum, these six cases illustrate how in two cities with similar overall rates of non-White residents a largely White neighbourhood may be no more socially segregated than its own composition would suggest in one case (e.g., Washington DC), whereas in the other (e.g., New Orleans) a comparable neighbourhood is less socially segregated, that is to say more socially exposed to non-White individuals, than the local residential composition would suggest. Importantly, this distinction bears out on average at MSA-level too, where we observe cities like DC or Miami ranking relatively higher on social segregation than residential, and vice versa for New Orleans or San Francisco.

[18]The polynomial fit for spatial exposure, by contrast, tends to be flat, suggesting that own ZIP composition plays a relatively minor role in determining this measure. Another interpretation, which we emphasize in light of our findings on social exposure, is that spatial linkages are a poor proxy for how people actually interact with each other.

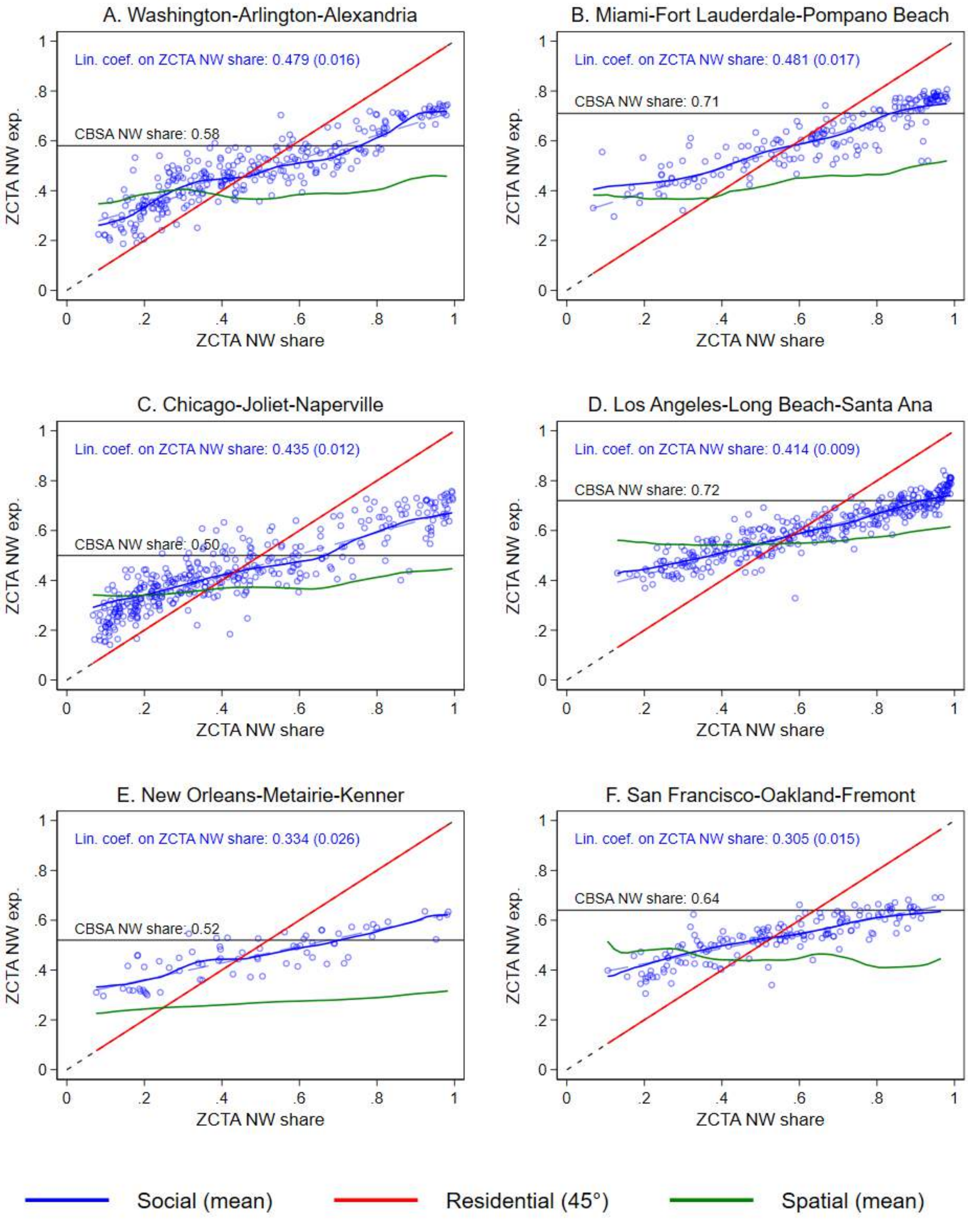


FIGURE 3 – Scatter plots comparing residential, social, and spatial segregation across selected cities with at least 50 ZCTAs. For each city, ZCTA exposures are plotted against each ZCTA's composition (share of non-White residents). Fitted lines for social and spatial exposures are local polynomials weighted by ZCTA size. For social exposure, markers also show individual ZCTAs. A population-weighted linear fit is also overlaid in this case (light-blue dashed line), along with a point estimate for the slope. Panels A and B consider cities ranking higher in terms of social relative to residential isolation, C and D consider cities that rank about the same, E and F those that rank lower in social isolation.

Next, we describe what features correlate with levels of social segregation across US cities, as a way to make sense of cross-sectional differences in social segregation, and specifically also of differences between social and residential measures.

4.2 Correlates of Social Isolation in US Cities

We characterise socially segregated cities by examining features that are associated with greater social isolation. However, social isolation is in part mechanically related to its residential counterpart. Hence, as illustrated in Equation 7, we decompose our social isolation measures to study the role of interactions with other areas in a city, separately from the effects arising from residential composition alone, or from interactions within the same neighbourhood. This exercise is useful to identify dimensions along which residential and social isolation may have divergent relationships. To this end, we estimate the following two models:

$$Y_c = \alpha + \beta SISO_c + \epsilon_c, \tag{9a}$$

$$Y_c = \alpha + \beta_1 T1_c + \beta_2 T2_c + \beta_3 T3_c + \upsilon_c, \tag{9b}$$

where Y_c is a characteristic of interest for city c, $SISO_c$ is the social isolation index, and $T1_c$, $T2_c$ and $T3_c$ are respectively the own-area residential, own-area social, and other-area social components of $SISO_c$, as defined in Equation 7. Of key interest here is the coefficient β_3, which captures variation in social isolation due to differences in how residents of a neighbourhood interact with residents of other neighbourhoods. This can be compared to either β_1, which captures a residential composition effect, or with β from Equation 9a, which is the overall effect of all three components of $SISO_c$.[19]

We consider a set of urban characteristics Y_c typically examined by the segregation literature (e.g., Cutler et al., 1999; Athey et al., 2021) and standardize them (in terms of z-scores) so that coefficients can be easily compared across models. Figure 4 summarises these results graphically. Markers show coefficients β_{1-3}, obtained from estimating the model in 9b. For reference, we also overlay β from Equation 9a using the blue vertical reference line (solid lines denote statistical significance at the 95 percent level). Magnitudes are scaled by the sample standard deviation of $SISO_c$, or of each component, for ease of comparison and interpretation.[20] It is important to emphasise that our findings are descriptive correlations.

What emerges from this figure is that the headline association of $SISO_c$ with urban fea-

[19] In addition, note that $T2_c$ is proportional to the number of ZIP Codes in a city, N_c, which means this term also controls for a scaling effect arising from larger cities with more neighbourhoods.

[20] Appendix Table A.3 reports raw coefficients.

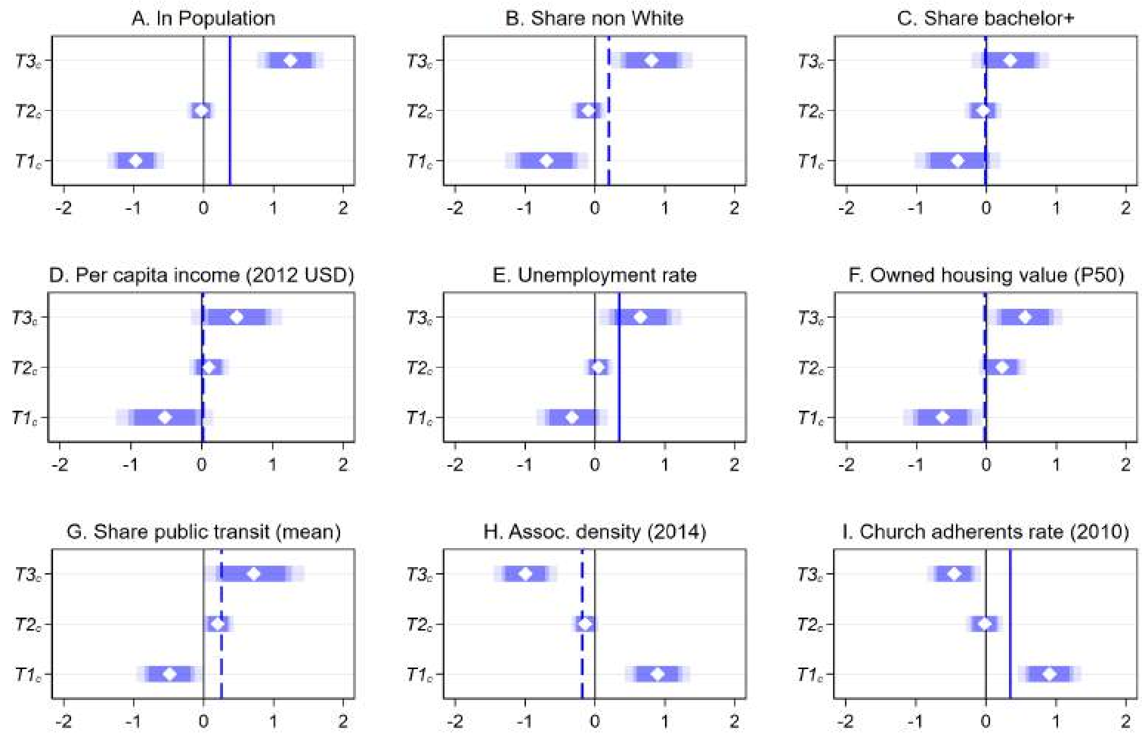


FIGURE 4 – Coefficient plots for correlates of MSA social isolation. For each characteristic, markers denote results from fitting the model in Equation 9b. Robust confidence intervals at the 90, 95, and 99 percent levels are displayed in progressively lighter shades. The blue reference line shows the overall effect of $SISO_c$, estimated using the model in Equation 9a. Solid lines denote statistical significance of the coefficient at the 95 percent level. Both the dependent and independent variables are expressed in terms of z-scores.

tures (blue reference line) often masks substantial differences in the underlying effect of own-area components ($T1_c$ and $T2_c$) versus interactions with other areas ($T3_c$). In line with Cutler et al. (1999), for instance, social isolation is greater in larger cities, but unrelated to racial composition (share of non-White residents) or to public transit use for commuting to work (as a proxy for transportation costs and urban mobility). This lack of association, however, is in part due to $T1_c$ and $T3_c$ having opposite effects. Greater urban diversity is actually positively associated with other-area interactions, but negatively when it comes to own-area composition. The same is true for public transit networks. The latter, in particular, suggests that transport networks potentially consolidate existing racial homophily in friendship ties, similar to what Wang et al. (2018) documented for residential segregation. Similar considerations apply to income per capita, housing costs (median value of owned housing), and unemployment. In addition, we consider associational density (Rupasingha et al., 2006) and adherence rate to churches of all faiths (ARDA, 2010) as proxies for social capital. This is often posited to be negatively related to segregation (e.g. Putnam, 2000, 2007; Athey et al., 2021).[21] Again, other-area interac-

[21] More specifically, Putnam (2007) describes diversity as generally decreasing social capital in local

tions and own-area composition have divergent estimates, with only the former displaying the expected negative association. In other words, for comparable levels of residential segregation, greater participation to associations and local religious communities appears to limit the in-group exposure of non-White residents, reducing social isolation at least in relative terms. This is consistent with recent evidence by Chetty et al. (2022*b*), who show that participation to religious organizations can mitigate socio-economic friending bias. This relationship, as well as the role of public transport, may warrant future investigation in dedicated studies.

Appendix Figure A.16 reports results from an alternative version of the specifications in Equations 9a and 9b, where in addition to either $SISO_c$ or each component we also control for all the other urban characteristics considered in this analysis (excluding the dependent variable), as many of the measures we consider co-vary with each other. Results are sensitive to this change in specification, except for coefficients on church adherence, which are remarkably robust. With Appendix Figure A.17, we confirm that population size is a key determinant of these changes, as including this control is by itself enough to substantially alter findings. Results for church adherence, however, remain unchanged.[22] Finally, we show in Appendix Table A.4 that our findings for social isolation, notably with respect to its other-area interactions constituent ($T3_c$), are robust to controlling for the spatial isolation index.[23] Finally, Appendix Figure A.18 takes a closer look at the association with social capital, by considering correlations of our isolation index with ZCTA-level measures of connectedness (degree of interaction low-income people have with high-income), cohesiveness (degree to which social networks are fragmented into cliques), and civic engagement (rates of volunteering and participation in community organizations) defined in Chetty et al. (2022*a*,*b*). Reassuringly, with the exception of economic connectedness (EC), results align with our findings on associational density and church adherence.

communities, fostering social isolation. Organizational activity and religious involvement, however, stand out as notable exceptions. Community resources like religious institutions, sport clubs, and civic associations, if anything, are found to display a positive association with diversity as measured by ZIP Code composition. The author concludes that religious institutions, in particular, can play an important role in building shared identities that cut across ethnic and racial boundaries.

[22] In untabulated results, we also controlled for log population density instead of population. This influenced some outcomes (notably public transit, income, and owned housing value, which become insignificant throughout), but once again had no effect on coefficients for church adherence.

[23] Associational density and church adherence, in particular, remain negatively associated with social isolation of non-local friendship networks, whereas spatial isolation plays a more limited role (displaying weaker, noisier, or even statistically insignificant associations).

4.3 Associations with Outcomes in Adulthood at ZIP Level

So far we considered features of the urban environment that are associated with greater social isolation at MSA level, but our data also allow us to examine variation in isolation at a more granular scale. Here, we zoom-in at the neighbourhood level to examine how living in a ZIP Code with greater or lower social exposure to non-Whites relates with a series of socio-economic outcomes. We focus on outcomes specific to the Black population from Chetty et al. (2018)'s Opportunity Atlas[24,25] and construct ZIP-level measures of social exposure that are specific to this group.[26] We also contrast the effect of social exposure to its spatial counterpart, holding residential exposure fixed, to test which exposure measure is more relevant. We estimate the following two related models:

$$Y_i = \beta SOC\_EXP_i + \gamma SPT\_EXP_i + \sum_{d=1}^{10} \delta_d r_{id} + \mu_c + \epsilon_i, \tag{10a}$$

$$Y_i = \sum_{d=1}^{10} [\beta_d SOC\_EXP_i \times r_{id} + \delta_d r_{id}] + \gamma SPT\_EXP_i + \mu_c + \upsilon_i, \tag{10b}$$

where $SOC\_EXP_i \equiv \sum_{j \neq i} \omega^s_{ij} \frac{x_j}{t_j}$ is the social exposure of ZIP i residents to Black residents in other neighbourhoods, $SPT\_EXP_i \equiv \sum_{j \neq i} \omega^d_{ij} \frac{x_j}{t_j}$ is its spatial analogue, and d denotes intervals of residential exposure RES_EXP_i (the neighbourhood's own proportion of Black residents) in ten percentage point increments. The variable $r_{id} = [RES\_EXP_i \in d]$ indicates whether ZIP Code i falls in the interval d in terms of residential exposure. We allow the intercept of this model to vary across cities by including MSA fixed-effects (μ_c). In Equation 10b, we depart from the baseline model in 10a by estimating a separate effect of SOC_EXP_i for every residential exposure interval, to study how β varies depending on local neighbourhood composition. Non-linearity is informative with respect to patterns of complementarity or substitution between residential and social exposure. Being socially exposed to other Black Americans outside one's neighbourhood may matter differently depending on that neighbourhood's own demographic composition.

We consider the following outcomes Y_i for Black children born around 1980 and growing up in neighbourhood i (until the age of 23), measured conditionally on parental income falling at the bottom quartile of the national income distribution:[27]

- `jail`: fraction of children who are in jail in 2010;

[24] Available at: https://opportunityinsights.org/data/.

[25] A comparable analysis for the group of Hispanic or Latino Americans, the largest ethnic minority in the US, is available in the online Appendix, Figures A.22 and A.23, and Tables A.6 and A.8.

[26] Specifically, we re-define the term x_j (first introduced in Equation 1) to be the share of Black residents in each neighbourhood, rather than more generally any non-White racial or ethnic minority.

[27] We refer to the original paper for details on the definition and construction of each variable.

- `kfr`: mean income rank in 2015 in the US distribution of their birth-cohort;
- `married`: fraction of children who are married in 2015;
- `staycz`: fraction of children who live in one of their childhood commuting zones.

We visualise results for both models in Figure 5.[28] We emphasise that these results are purely descriptive and unlikely to be causal. One notable issue with our specification is that outcomes are observed prior to the snapshot of social interactions on Facebook, which could lead to reverse causality.[29] Similarly, outcomes are assigned to each ZIP based on where the children grew up, but social connections in that neighbourhood may have looked different during the 1980-2015 period compared to what we observed in 2020. Therefore, a necessary underlying assumption to our estimates is that social interactions are fixed in time, or at least very slow-changing. Relatedly, this also hinges on the stability of residential demographics across neighbourhoods. The validity of this second assumption is easier to check. Most neighbourhoods changed very little between 1990 and 2020. The median ZCTA experienced an increase in the share of Black residents of 0.006, with the middle 50 percent of observations falling between -0.002 and 0.032. This is also true at city level. The median MSA experienced an increase in the share of Black residents of 0.014, with the middle 50 percent of observations falling between 0.000 and 0.025. Appendix Figure A.19 visualises this high degree of persistence in composition. We also compare two maps of Chicago in 1990 and 2020 to stress this point (Appendix Figure A.20).[30]

Estimates of β and γ from Equation 10a are in blue and red lines respectively, with a solid pattern for coefficients that are statistically significant at the 95 percent level, and dashed otherwise (refer to Appendix Table A.7 for details on the point estimates). Results suggest that growing up in neighbourhoods with high social exposure to Black residents in other areas is associated on average with lower likelihood of being jailed, married, or living in a different labour market in adulthood. No correlation is detected with respect to income ranks. This holds true controlling for levels of residential segregation (flexibly) and spatial exposure. The latter, in particular, does not appear to be better than social exposure at explaining the outcomes we consider: while we do not observe any association with being jailed or living in the childhood commuting zone, the association

[28] For reference, Appendix Figure A.21 shows conditional averages of each outcome by levels of residential exposure (in increments of ten percentage points), absorbing MSA fixed effects. Essentially, these represent flexible estimates for the effect of residential exposure in discrete intervals (r_d), imposing β and γ to be zero. We largely confirm the adverse socio-economic effects of residential isolation previously documented in much of the existing research on segregation, with a clear negative gradient emerging with respect to income, marriage, and mobility across commuting zones (less so for jail).

[29] This concern is also highlighted by Chetty et al. (2022*a*), who perform a similar analysis using their measure of economic connectedness that is based on the same Facebook data we rely on.

[30] We consider 1990 composition rather than 1980, which would be closer to the children's birth year, because outcomes are measured for those who lived in the same neighbourhood until the age of 23. The 1990 composition, then, is arguably more relevant.

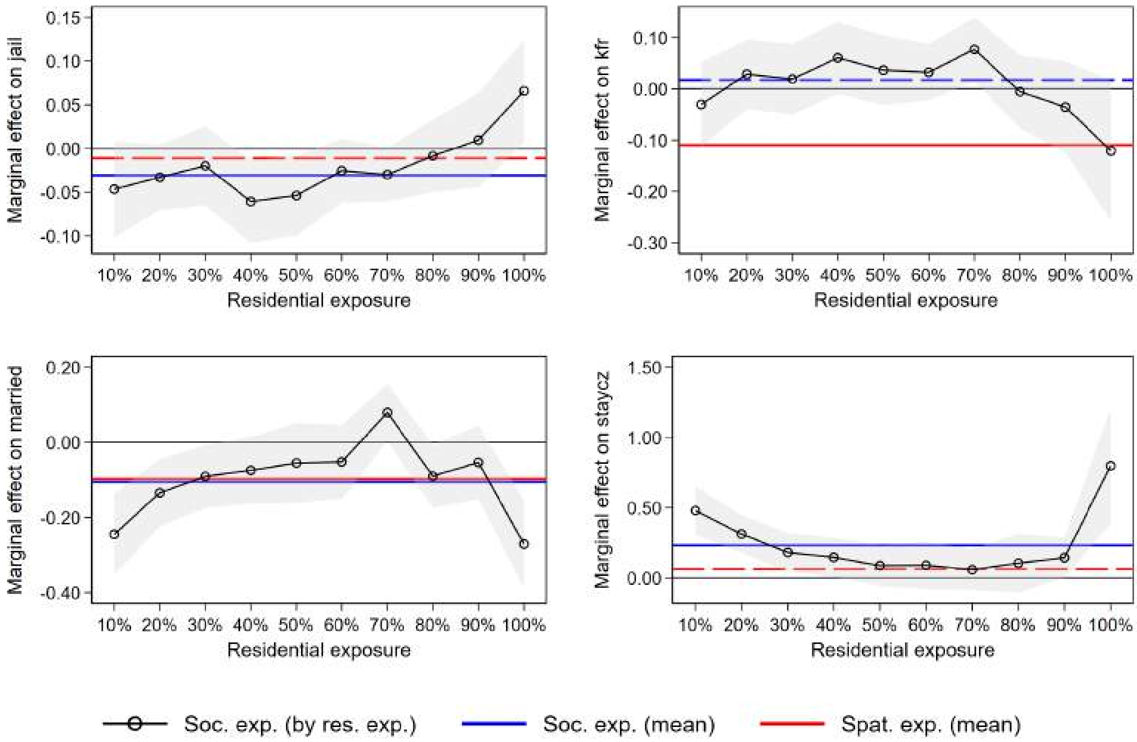


FIGURE 5 – Black connected markers gives marginal effects of social exposure to Black population β_d by levels of residential exposure to Black population from the estimation of Equation 10b. The shaded area denotes 95 percent confidence intervals. Lines in blue and red denote the average marginal effect for social exposure, β, and spatial exposure, γ, to Black population from the estimation of Equation 10a. Lines are solid if the coefficients are statistically significant at the 95% level or dashed otherwise. Outcomes and exposure measures are always specific to the Black population. All models absorb MSA fixed-effects. Standard errors are clustered at the MSA level.

with marriage is comparable both qualitatively and quantitatively to that detected on SOC_EXP_i. Interestingly, however, Black Americans who grew up in neighbourhoods located in larger spatial clusters of predominantly Black ZIPs (higher SPT_EXP_i) appear to rank lower in terms of income compared to other cohort peers. This result is consistent with poor access to employment opportunities, as emphasised by the spatial mismatch literature (Kain, 1968; Coulson et al., 2001). Any interpretation in this direction however is purely speculative as we do not actually study the location of jobs in the city.

We also acknowledge the possibility of non-linearities in the effects of social exposure. Estimates for β_d by levels of residential exposure (from Equation 10b) are summarised with black connected markers, along with 95 percent confidence intervals in grey. Confirming our hypothesis, we detect significant heterogeneity in our previous findings, depending on own-neighbourhood composition. Black children growing up in ZCTAs with higher social exposure to other Black neighbourhoods may be less likely to be jailed on average, but if their own ZCTA is itself homogeneous, they are actually more likely to end up in prison as adults. They are also less likely to be married, particularly if they reside in very homogeneous neighbourhoods (irrespective of whether it is predominantly Black or White/other race or ethnicity). Similarly, Black children growing up in homogeneous neighbourhoods are more likely to live in the same commuting zone in adulthood if they are socially exposed to predominantly Black areas outside their own ZIP. This does not seem to be true for children in more evenly mixed neighbourhoods.

All in all, one important takeaway from this analysis is that social exposure is a particularly relevant variable to consider when people reside in neighbourhoods that are relatively homogeneous in terms of racial or ethnic composition. In the literature, social connections have been shown to be important in determining adult outcomes, even explaining away the correlation with racial composition (Chetty et al., 2022*a*). Here, we point out an important interaction between the two: that social connections matter differentially depending on racial composition.[31]

5 Conclusions

Residential segregation of racial and ethnic minorities is a pressing social concern whose causes and consequences continue to stir a flurry of debates in both academic and policy circles. Our paper considers this question from a new angle, by studying the extent of segregation of US urban residents in the social space. By leveraging novel and granular data on the universe of online friendships between US neighbourhoods, we propose a new measure of segregation defined as the lack of personal social connections between individuals belonging to different racial or ethnic groups. We refer to this as social isolation. In so doing, we depart from a key assumption plaguing much of existing research on this subject: that individuals only interact with people in their own residential neighbourhood. We discuss and show evidence of why this distinction matters.

The social lens we adopt allows us to uncover new interesting facts about US segregation. First, we confirm that residential and social isolation measures tend to be highly correlated, suggesting that segregation in residential terms also persists in the social domain. Second, we show that despite this correlation social isolation tends to be lower than its residential counterpart. Still, there is also substantial discrepancy in these measures in relative terms: many cities are more socially than residentially isolated, and vice-versa, depending on variation across urban areas in the propensity of ZIPs with similar residential compositions to form ties with members of other racial groups. Third, we examine features of the urban environment that correlate with social isolation. We demonstrate that

[31]In Appendix Table A.9, we also show that our findings on social exposure are essentially unchanged if we replace the control for spatial exposure with one for the Economic Connectedness variable defined in Chetty et al. (2022*a*). This further underscores the relevance of our social isolation measure, and the novelty of this analysis relative to the one carried out in Chetty et al. (2022*a*).

the headline indicator hides substantial heterogeneity in own-area residential composition and other-area social interaction components. This underscores the importance of taking into account neighbourhood linkages when studying segregation. Public transport use and participation in local community life emerge as key variables that warrant future study. Finally, we examine how growing up in more socially isolated neighbourhoods correlates with the economic outcomes of Black individuals, reporting substantial heterogeneity in this relationship across neighbourhoods. Specifically, social isolation matters the most for individuals who live in neighbourhoods whose composition is relatively homogeneous to begin with.

There are, however, some important limitations to our analysis. Although we have extensive information on social connections across ZIP Codes, we do not observe these for individuals and are hence unable to construct measures of social segregation at this level. Our data of social networks is also limited to a snapshot observed in 2020. Without more granular data and lacking variation over time, our analysis is necessarily descriptive and explorative in scope. Nevertheless, the wealth of data on social connections that is increasingly becoming available to researchers offers new opportunities to conceptualise and measure segregation. Our study aims to provide a better understanding of what it means to be a segregated minority in today's society, with the hope of laying the groundwork for more research exploring this important subject from the angle of social interactions.

References

Abbiasov, T. (2020), 'Do Urban Parks Promote Racial Diversity? Evidence from New York City', pp. 1–35.

Ananat, E. O. (2011), 'The wrong side(s) of the tracks: The causal effects of racial segregation on urban poverty and inequality', *American Economic Journal: Applied Economics* **3**(2), 34–66.

ARDA (2010), 'Churches and Church Membership in the United States. Collected by the Association of Statisticians of American Religious Bodies (ASARB) and distributed by the Association of Religion Data Archives.'.

Athey, S., Ferguson, B., Gentzkow, M. and Schmidt, T. (2021), 'Estimating experienced racial segregation in US cities using large-scale GPS data', *Proceedings of the National Academy of Sciences* **118**(46), e2026160118.

Bailey, M., Cao, R., Kuchler, T. and Stroebel, J. (2018), 'The economic effects of social networks: Evidence from the housing market', *Journal of Political Economy* **126**(6), 2224–2276.

Bailey, M., Cao, R., Kuchler, T., Stroebel, J. and Wong, A. (2018), 'Social connectedness: Measurement, determinants, and effects', *Journal of Economic Perspectives* **32**(3), 259–80.

Bailey, M., Dávila, E., Kuchler, T. and Stroebel, J. (2019), 'House price beliefs and mortgage leverage choice', *The Review of Economic Studies* **86**(6), 2403–2452.

Bailey, M., Farrell, P., Kuchler, T. and Stroebel, J. (2020), 'Social connectedness in urban areas', *Journal of Urban Economics* **118**, 103264.

Bailey, M., Johnston, D., Kuchler, T., Stroebel, J. and Wong, A. (2022), 'Peer Effects in Product Adoption', *American Economic Journal: Applied Economics* **14**(3), 488–526.

Büchel, K., Ehrlich, M. V., Puga, D. and Viladecans-Marsal, E. (2020), 'Calling from the outside: The role of networks in residential mobility', *Journal of Urban Economics* **119**, 103277.

Cattaneo, M. D., Crump, R. K., Farrell, M. H. and Feng, Y. (2024), 'On Binscatter', *American Economic Review* **114**(5), 1488–1514.

Chetty, R., Friedman, J. N., Hendren, N., Jones, M. R. and Porter, S. R. (2018), The opportunity atlas: Mapping the childhood roots of social mobility, Working Paper 25147, National Bureau of Economic Research.

Chetty, R., Jackson, M. O., Kuchler, T., Stroebel, J., Hendren, N., Fluegge, R. B., Gong, S., Gonzalez, F., Grondin, A., Jacob, M., Johnston, D., Koenen, M., Laguna-Muggenburg, E., Mudekereza, F., Rutter, T., Thor, N., Townsend, W., Zhang, R., Bailey, M., Barberá, P., Bhole, M. and Wernerfelt, N. (2022*a*), 'Social capital I: measurement and associations with economic mobility', *Nature 2022* pp. 1–14.

Chetty, R., Jackson, M. O., Kuchler, T., Stroebel, J., Hendren, N., Fluegge, R. B., Gong, S., Gonzalez, F., Grondin, A., Jacob, M., Johnston, D., Koenen, M., Laguna-Muggenburg, E., Mudekereza, F., Rutter, T., Thor, N., Townsend, W., Zhang, R., Bailey, M., Barberá, P., Bhole, M. and Wernerfelt, N. (2022*b*), 'Social capital II: determinants of economic connectedness', *Nature 2022* pp. 1–13.

Cook, C., Currier, L. and Glaeser, E. L. (2022), Urban Mobility and the Experienced Isolation of Students and Adults, Working Paper 29645, National Bureau of Economic Research.

Cornelson, K. (2017), Social Interactions and Racial Inequality, PhD thesis, University of Toronto.

Coulson, N. E., Laing, D. and Wang, P. (2001), 'Spatial Mismatch in Search Equilibrium', *Journal of Labor Economics* **19**(4), 949–972.

Cutler, D. M. and Glaeser, E. L. (1997), 'Are Ghettos Good or Bad?', *The Quarterly Journal of Economics* **112**(3), 827–872.

Cutler, D. M., Glaeser, E. L. and Vigdor, J. L. (1999), 'The rise and decline of the American Ghetto', *Journal of Political Economy* **107**(3), 455–506.

Davis, D. R., Dingel, J. I., Monras, J. and Morales, E. (2019), 'How Segregated Is Urban Consumption?', *Journal of Political Economy* **127**(4), 1684–1738.

Echenique, F. and Fryer, R. G. (2007), 'A Measure of Segregation Based on Social Interactions', *The Quarterly Journal of Economics* **122**(2), 441–485.

Fryer, R. G. (2011), The Importance of Segregation, Discrimination, Peer Dynamics, and Identity in Explaining Trends in the Racial Achievement Gap, Vol. 1, North-Holland, chapter 21, pp. 1165–1191.

Gee, L. K., Jones, J. and Burke, M. (2017), 'Social networks and labor markets: How strong ties relate to job finding on facebook's social network', *Journal of Labor Economics* **35**(2), 485–518.

Gentzkow, M. and Shapiro, J. M. (2011), 'Ideological segregation online and offline', *The Quarterly Journal of Economics* **126**(4), 1799–1839.

Graham, B. S. (2018), 'Identifying and Estimating Neighborhood Effects', *Journal of Economic Literature* **56**(2), 450–500.

Kain, J. F. (1968), 'Housing Segregation, Negro Employment, and Metropolitan Decentralization', *The Quarterly Journal of Economics* **82**(2), 175–197.

Krivo, L. J., Peterson, R. D. and Kuhl, D. C. (2009), 'Segregation, racial structure, and neighborhood violent crime', *American Journal of Sociology* **114**(6), 1765–1802.

Ludwig, J., Duncan, G. J., Gennetian, L. A., Katz, L. F., Kessler, R. C., Kling, J. R. and Sanbonmatsu, L. (2012), 'Neighborhood effects on the long-term well-being of low-income adults', *Science* **337**(6101), 1505–1510.

Magontier, P., von Ehrlich, M. and Schläpfer, M. (2022), The Fragility of Urban Social Networks: Mobility as a City Glue, Working Paper 38, Universitaet Bern, Departement Volkswirtschaft - CRED.

Mahajan, A. (2024), 'Highways and segregation', *Journal of Urban Economics* **141**, 103574.

Manson, S., Schroeder, J., Riper, D. V., Kugler, T. and Ruggles, S. (2021), *IPUMS National Historical Geographic Information System: Version 16.0 [dataset]*, IPUMS, Minneapolis, MN.

Massey, D. S. (2017), 'Why death haunts black lives', *Proceedings of the National Academy of Sciences* **114**(5), 800–802.

Massey, D. S. and Denton, N. A. (1988), 'The Dimensions of Residential Segregation', *Social Forces* **67**(2), 281–315.

McPherson, M., Smith-Lovin, L. and Cook, J. M. (2001), 'Birds of a Feather: Homophily in Social Networks', *Annual Review of Sociology* **27**(1), 415–444.

Morgan, B. S. (1983), 'A Distance-Decay Based Interaction Index to Measure Residential Segregation', *Area* **15**(3), 211–217.

Putnam, R. D. (2000), *Bowling Alone: The Collapse and Revival of American Community*, Simon & Schuster, New York, NY.

Putnam, R. D. (2007), 'E Pluribus Unum: Diversity and Community in the Twenty-first Century The 2006 Johan Skytte Prize Lecture', *Scandinavian Political Studies* **30**(2), 137–174.

Rupasingha, A., Goetz, S. J. and Freshwater, D. (2006), 'The production of social capital in US counties', *Journal of Socio-Economics* **35**(1), 83–101.

Tóth, G., Wachs, J., Clemente, R. D., Jakobi, Á., Ságvári, B., Kertész, J. and Lengyel, B. (2021), 'Inequality is rising where social network segregation interacts with urban topology', *Nature Communications* **12**(1), 1–9.

Wang, Q., Phillips, N. E., Small, M. L. and Sampson, R. J. (2018), 'Urban mobility and neighborhood isolation in America's 50 largest cities', *Proceedings of the National Academy of Sciences* **115**(30), 7735–7740.

White, M. J. (1983), 'The Measurement of Spatial Segregation', *American Journal of Sociology* **88**(5), 1008–1018.

White, M. J. (1986), 'Segregation and diversity measures in population distribution.', *Population index* **52**(2), 198–221.

Wilson, R. (2020), 'The Impact of Social Networks on EITC Claiming Behavior', *The Review of Economics and Statistics* pp. 1–45.

Wilson, W. J. (1987), *The Truly Disadvantaged: The Inner City, the Underclass, and Public Policy*, University of Chicago Press, Chicago, IL.

Online Appendix to "The Role of Social Connections in the Racial Segregation of US Cities "

Andreas Diemer, Tanner Regan, Cheng Keat Tang

A Additional Figures and Tables

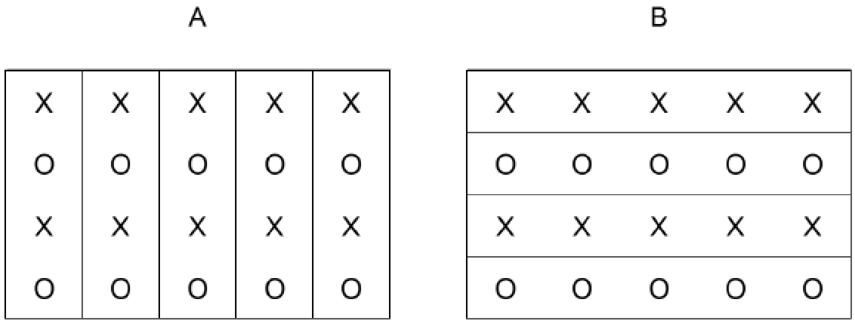


FIGURE A.1 – Residential segregation in a hypothetical city with different boundaries drawn (Echenique and Fryer, 2007). In the scenario in panel A, boundaries are drawn so that there is perfect integration between groups based on both the dissimilarity and isolation indices. In panel B, instead, boundaries are redrawn so that the groups are fully segregated. Note that the location of each group in the city is unchanged in either scenario.

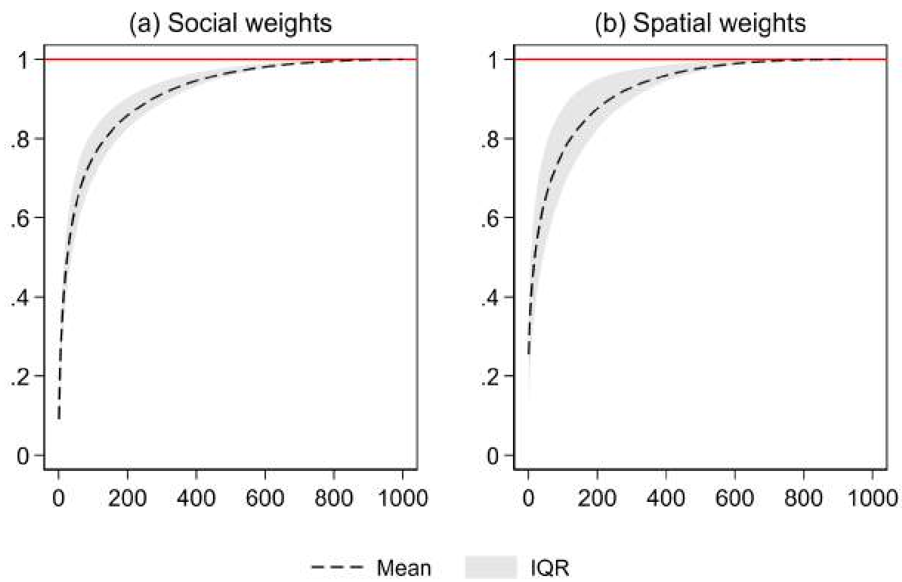


FIGURE A.2 – The figure shows averages and inter-quartile ranges (IQR) of the cumulative sum of social and spatial weights over their ranking. Observations are restricted to linkages for 9,747 neighbourhoods in 75 urban areas (MSAs) encompassing at least 50 ZCTAs. For each of these zips, the top 1,000 paired observations by link strenght are retained (irrespective of whether they involve another zip outside an urban area). For spatial weights, we also condition on pairs being within 100 miles from each other (approximately 160 km). In panel (a), N=9,701,159; in panel (b), N=7,310,173. In both instances, linkages ranking 500 or lower contribute very little relative to other higher-ranked linkages (1-500).

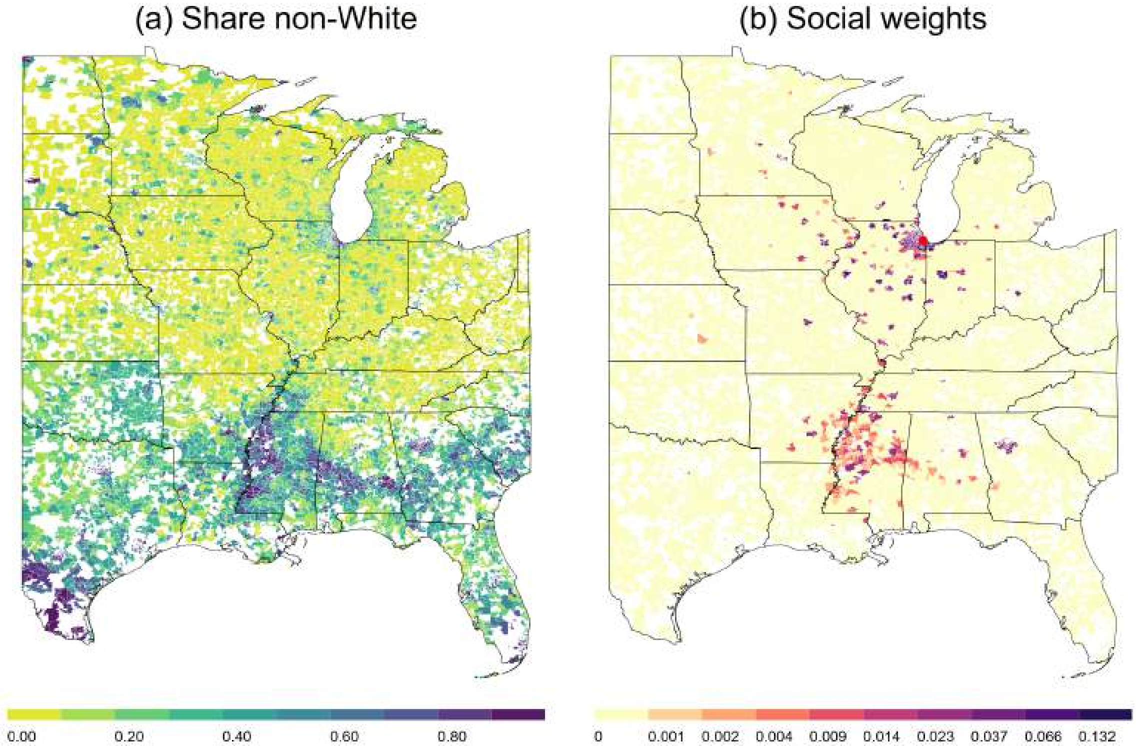


FIGURE A.3 – Race composition in ZCTAs across the US (panel a) and weights for ZCTA no. 60620 (99.48% non-White), denoted with a red dot on the map (panel b). Social weights in panel (b) are multiplied by 100 for legibility. Breaks are defined at each decile of the distribution obtained by pooling both social and spatial weights together.

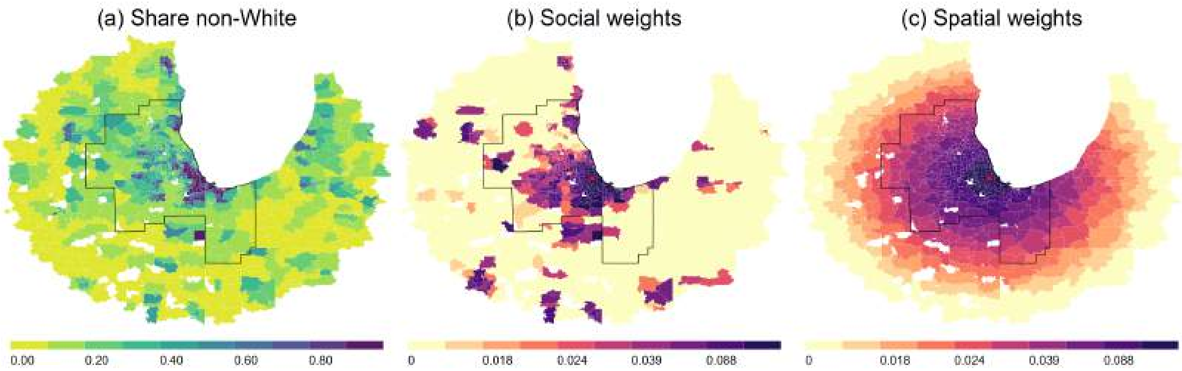


FIGURE A.4 – Maps with MSA race composition (a), social (b) and spatial (c) weights for the least White ZCTA in Chicago-Joliet-Naperville, ZCTA no. 60620, 99.48% non-White. (marked in red on the map). Weights are not adjusted for relative population size, and are rescaled by a factor of 100 for legibility. Breaks are defined at each decile of the distribution obtained by pooling both social and spatial weights together. MSA boundaries are in black.

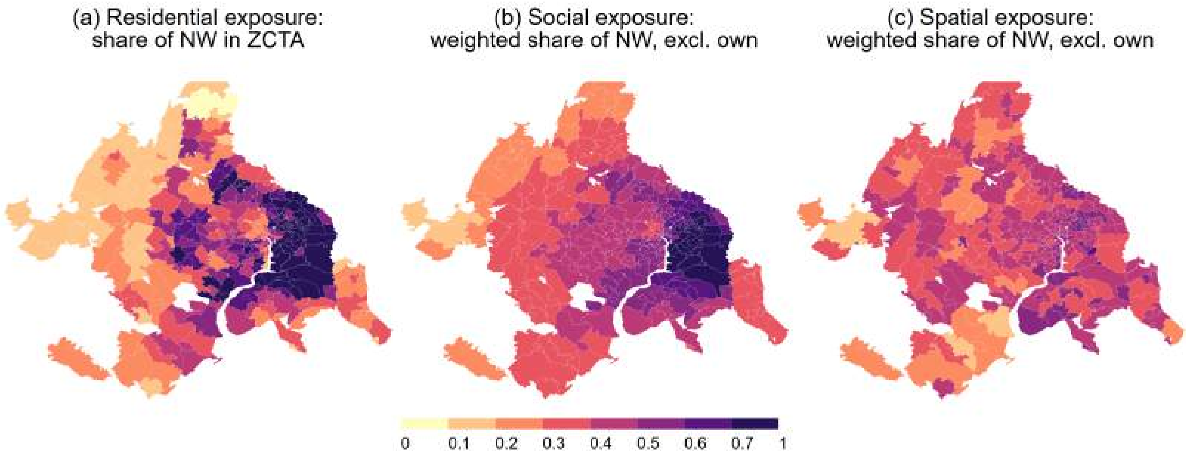


FIGURE A.5 – Residential, social, and spatial exposures in Washington-Arlington-Alexandria. The maps show within-MSA variation in minority exposure measures for the city's neighbourhoods. For reference, MSA-level segregation is as follows: $RISO$=0.205, $SISO^s$=0.106, $zRISO$=0.288, $zSISO^s$=0.545, $zSISO^s - zRISO$=0.257.

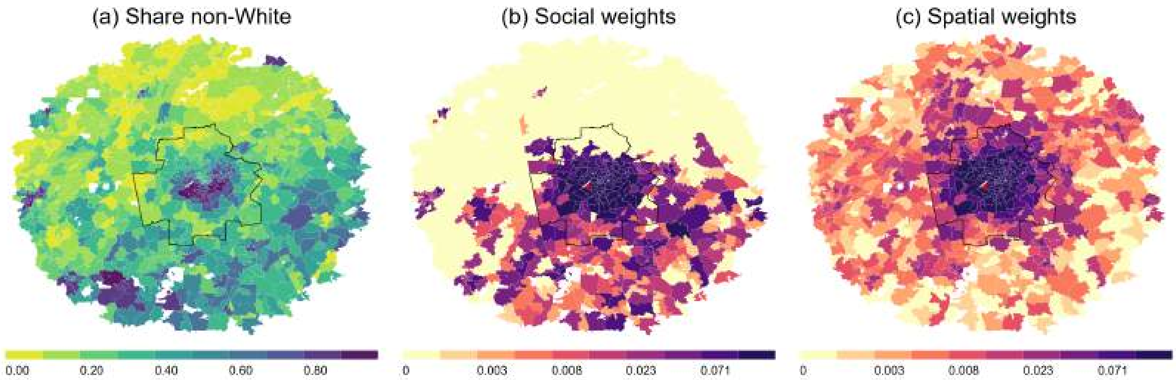


FIGURE A.6 – Maps with MSA race composition (a), social (b) and spatial (c) weights for the least White ZCTA in Atlanta-Sandy Springs-Marietta, ZCTA no. 30331, 98.70% non-White. (marked in red on the map). Weights are not adjusted for relative population size, and are rescaled by a factor of 100 for legibility. Breaks are defined at each decile of the distribution obtained by pooling both social and spatial weights together. MSA boundaries are in black.

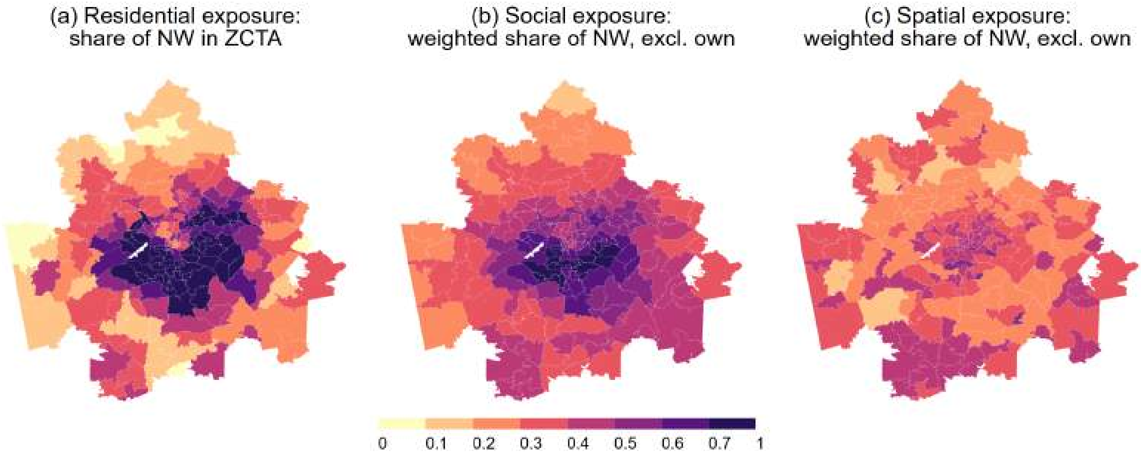


FIGURE A.7 – Residential, social, and spatial exposures in Atlanta-Sandy Springs-Marietta. The maps show within-MSA variation in minority exposure measures for the city's neighbourhoods. For reference, MSA-level segregation is as follows: $RISO$=0.234, $SISO^s$=0.117, $zRISO$=0.691, $zSISO^s$=0.809, $zSISO^s - zRISO$=0.117.

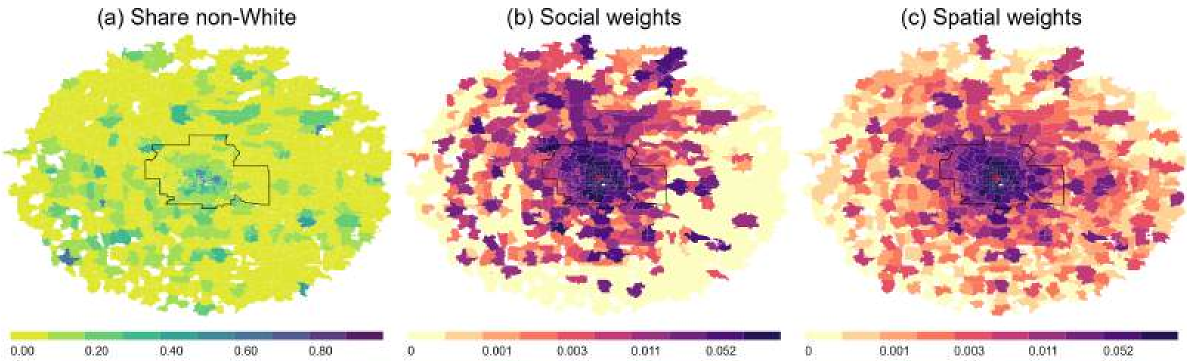


FIGURE A.8 – Maps with MSA race composition (a), social (b) and spatial (c) weights for the least White ZCTA in Minneapolis-St. Paul-Bloomington, ZCTA no. 55411, 83.23% non-White. (marked in red on the map). Weights are not adjusted for relative population size, and are rescaled by a factor of 100 for legibility. Breaks are defined at each decile of the distribution obtained by pooling both social and spatial weights together. MSA boundaries are in black.

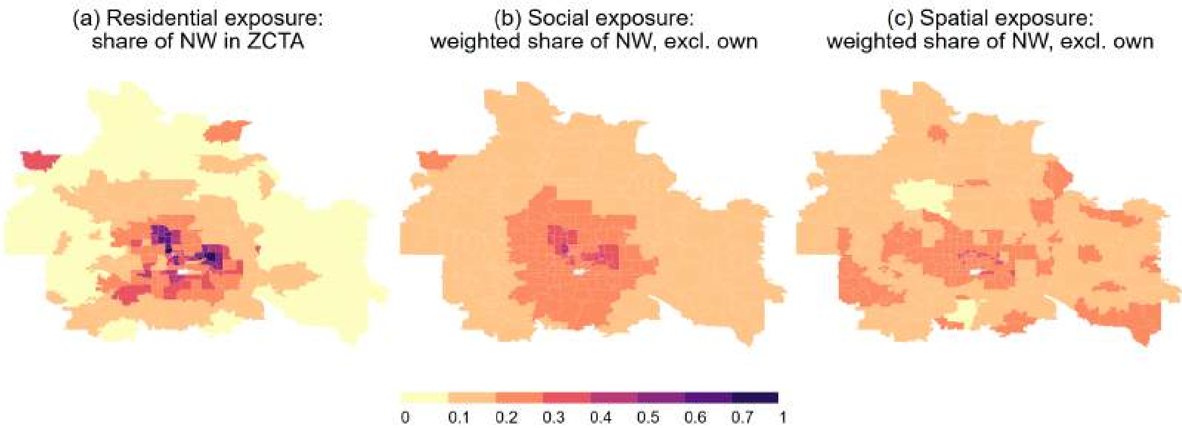


FIGURE A.9 – Residential, social, and spatial exposures in Minneapolis-St. Paul-Bloomington. The maps show within-MSA variation in minority exposure measures for the city's neighbourhoods. For reference, MSA-level segregation is as follows: $RISO$=0.146, $SISO^s$=0.055, $zRISO$=-0.532, $zSISO^s$=-0.638, $zSISO^s - zRISO$=-0.107.

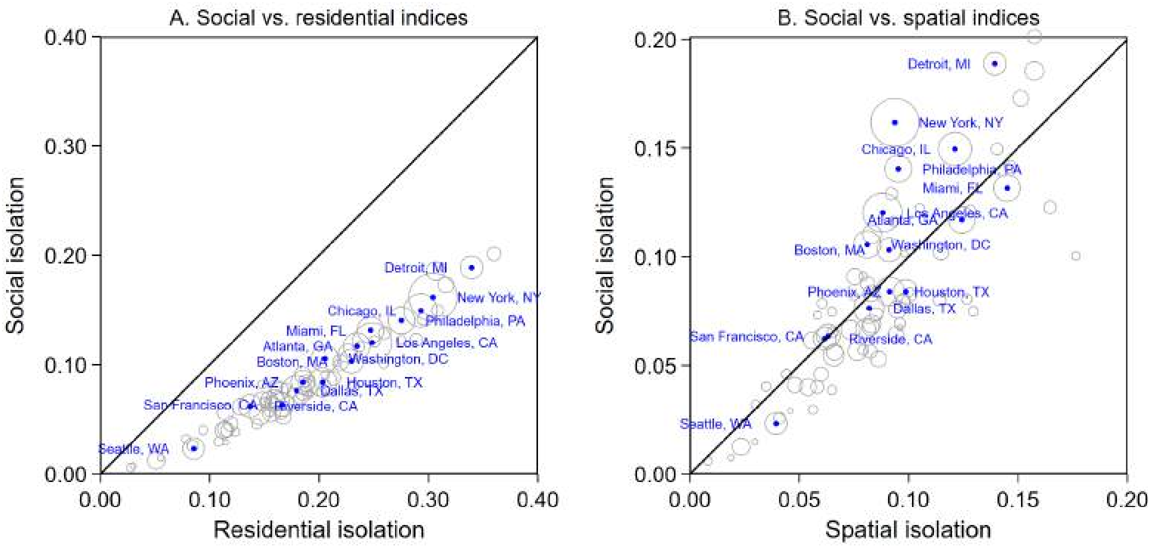


Figure A.10 – Scatter plots comparing social isolation with its residential or spatial counterparts, in panels A and B respectively. Only MSAs with at least 50 ZCTAs were retained (N=75). Marker sizes in grey are proportional to each city's population. The largest 15 cities in the sample are also marked and labeled in blue. Panel A highlights that social segregation is lower than residential segregation for all cities in our sample, and that residential and social isolation are strongly correlated with one another. However, there are also some noteworthy discrepancies (see also Appendix Table A.2, which lists index values for all these MSAs). St. Louis MO, for instance, one of the most segregated places in the US according to our measure, is also a lot more socially than residentially segregated. The same is true for Duluth MN, despite the fact that the area actually ranks amongst the least segregated in our sample. Las Vegas NV, Detroit MI, Atlanta GA, and Washington DC follow a similar pattern. By contrast, some urban areas like Jackson MS, Memphis TN, Richmond VA, and Providence RI, owe much of their segregation to residential, rather than social isolation. Similar considerations apply to the MSAs of New Orleans LA and San Francisco CA. To describe the geography of these discrepancies, Figure A.12 in Appendix maps the difference in z-scores of residential and social isolation. Panel B shows that despite their positive correlation, there is more variation in the isolation index depending on whether we use weights based on geographical distance or actual social interactions. Larger cities appear to display larger values in both indices, and larger values of social relative to spatial isolation.

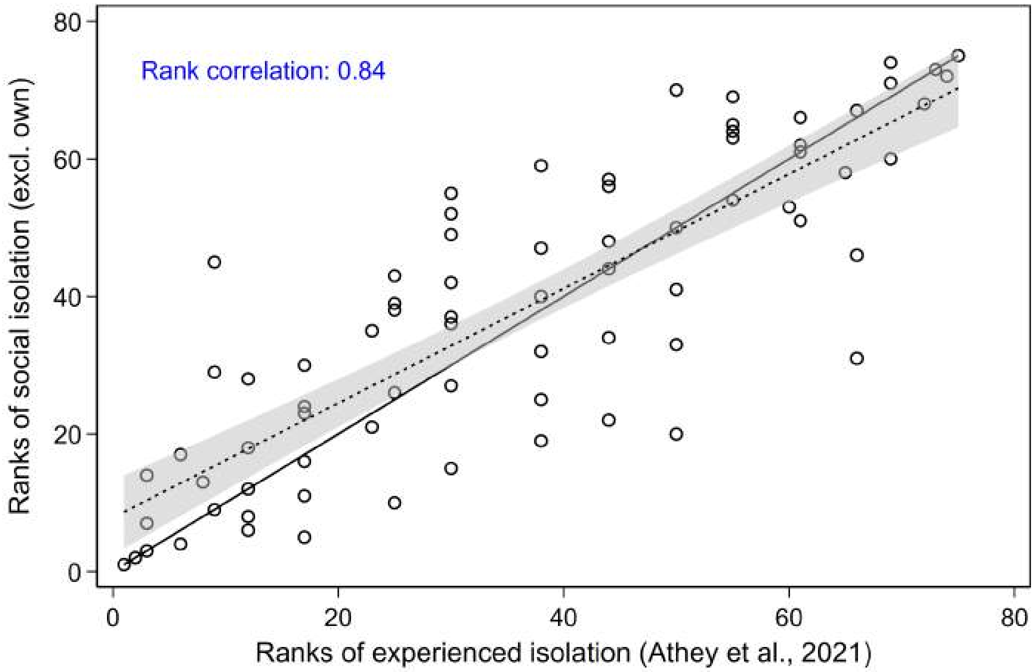


FIGURE A.11 – Spearman's rank correlation of social isolation with Athey et al. (2021)'s experienced isolation. The comparison is restricted to cities included in our sample only (i.e., MSAs with at least 50 ZCTAs within their boundaries). The 45 degree line is in solid black. The dotted line gives the linear fit, along with a 95 percent confidence band.

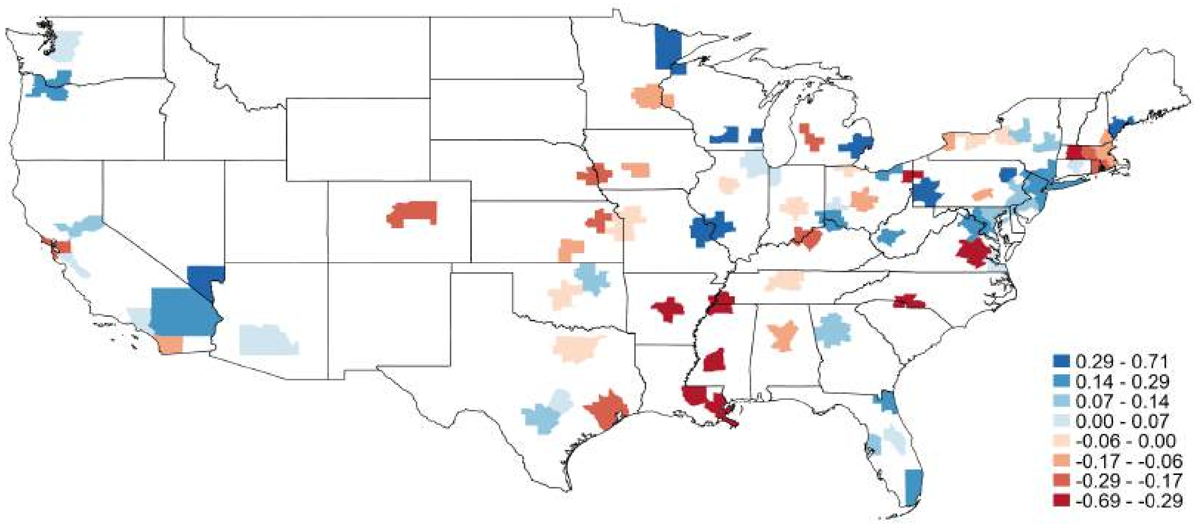


FIGURE A.12 – Differences in z-scores of social and residential isolation. Positive values entail that the MSA is relatively more socially than residentially segregated compared to the average city in our sample. Only MSAs with at least 50 ZCTAs are retained.

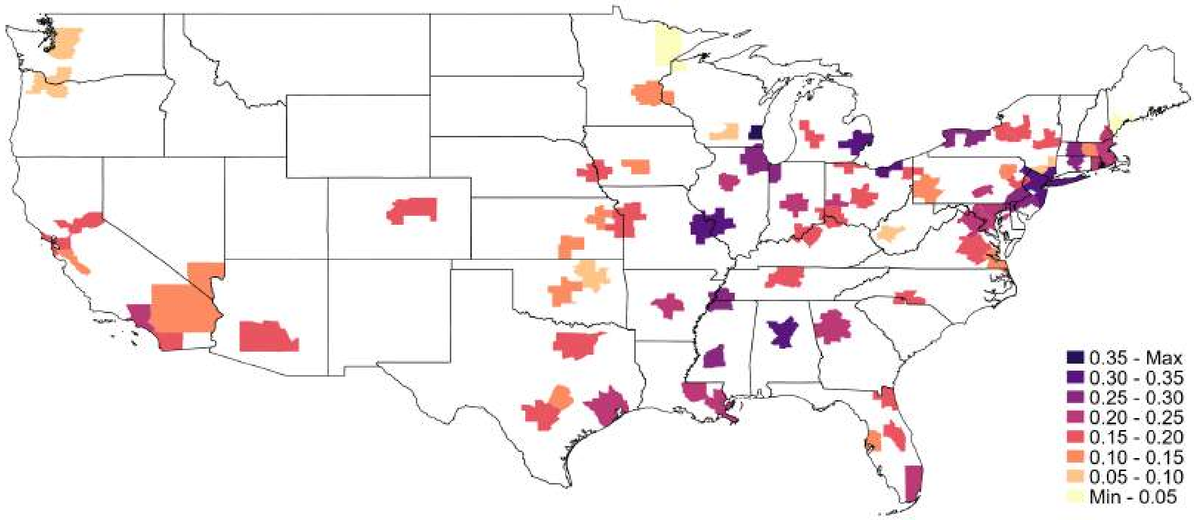


FIGURE A.13 – Residential isolation in the largest US cities. Legend categories are consistent with those in Figure A.14 to facilitate comparison.

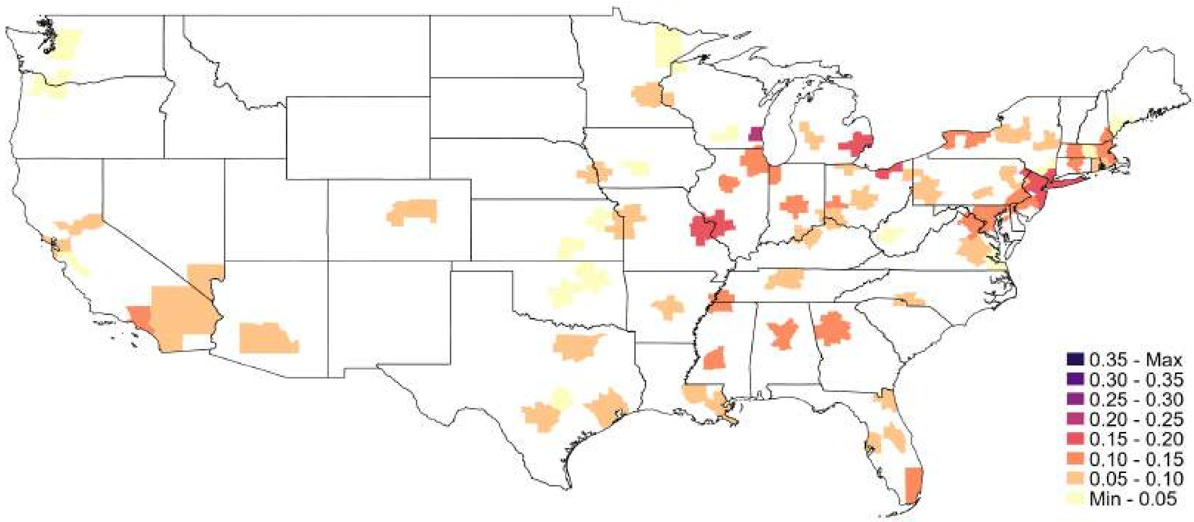


FIGURE A.14 – Social isolation in largest US cities, excluding own-area interactions. Legend categories are consistent with those used in Figure A.13 to facilitate comparison.

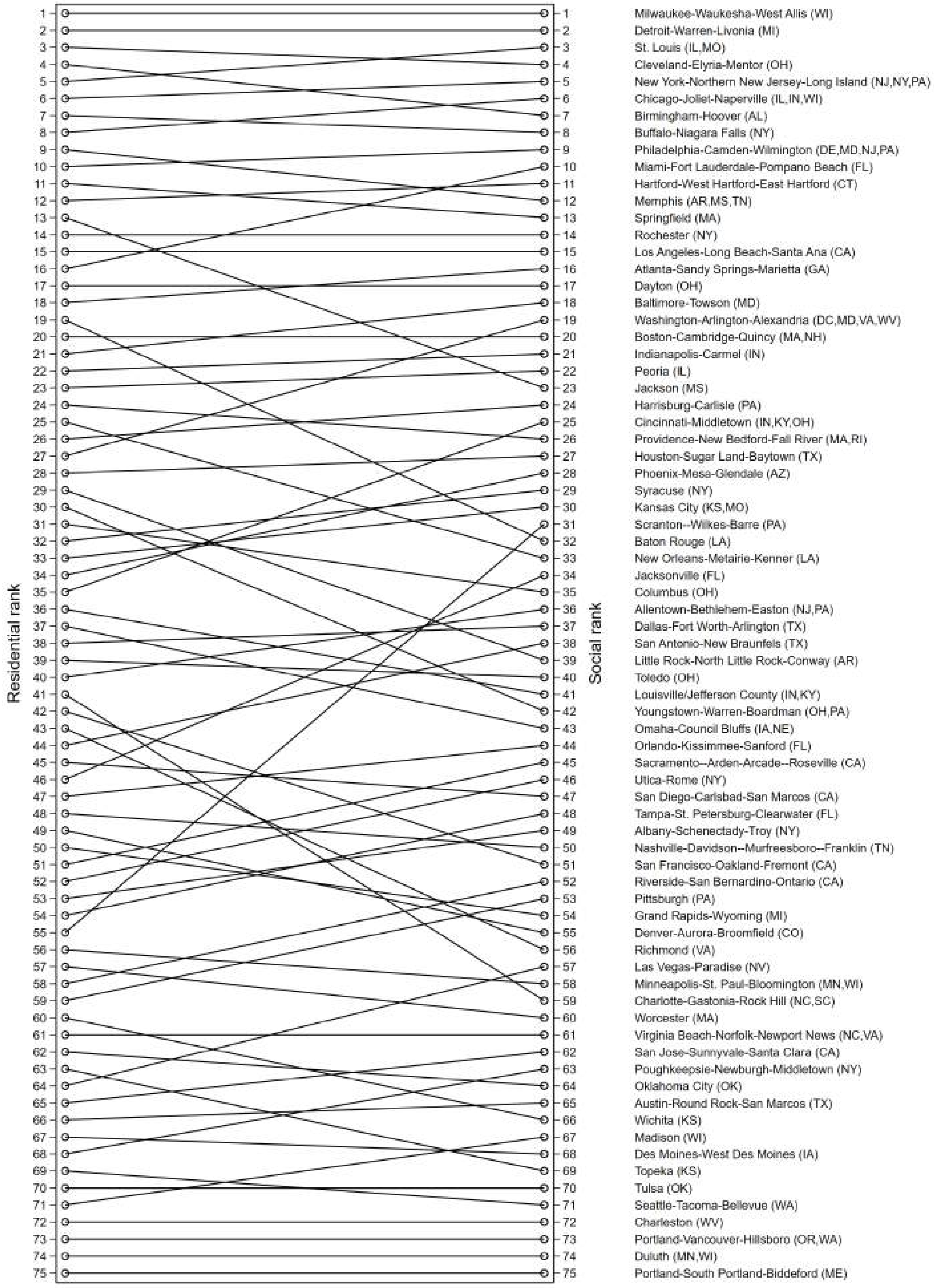


FIGURE A.15 – Comparison of MSA ranks defined in terms of residential segregation ($RISO_c$, on the left-hand side) with ranks defined in terms of social segregation ($SISO_c$, on the right-hand). MSAs are labeled according to their social segregation rank, so the figure is best read right-to-left.

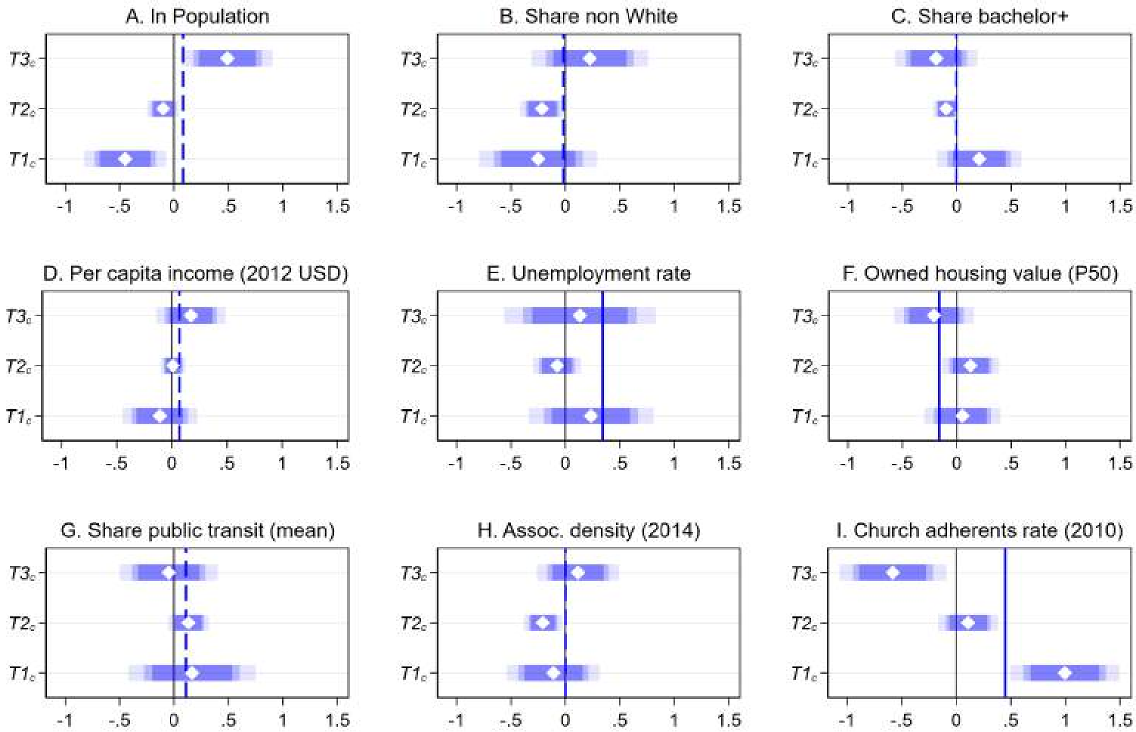


Figure A.16 – Coefficient plots for correlates of MSA social isolation. For each characteristic, markers denote results from fitting the model in Equation 9b, additionally controlling for all other characteristics. Robust confidence intervals at the 90, 95, and 99 percent levels are displayed in progressively lighter shades. The blue reference line shows the overall effect of $SISO_c$, estimated using the model in Equation 9a, again controlling for all other characteristics. Solid lines denote statistical significance of the coefficient at the 95 percent level. Both the dependent and independent variables are expressed in terms of z-scores.

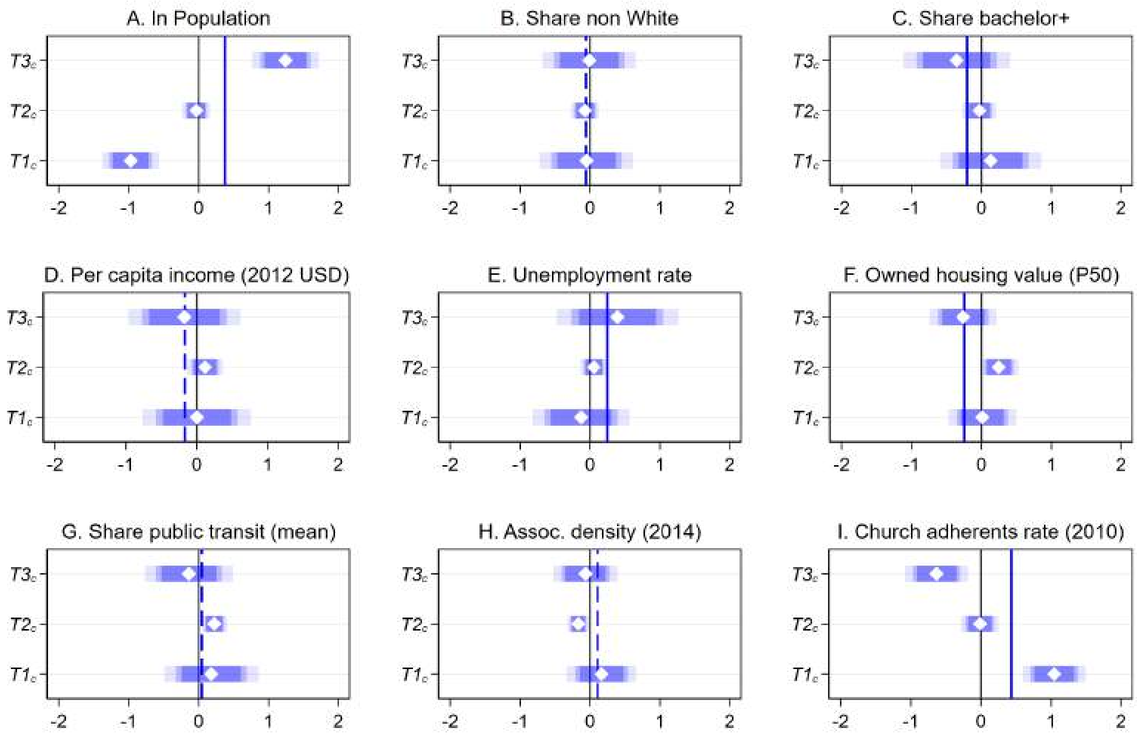


FIGURE A.17 – Coefficient plots for correlates of MSA social isolation. For each characteristic, markers denote results from fitting the model in Equation 9b, additionally controlling for the log of population size (except for panel A). Robust confidence intervals at the 90, 95, and 99 percent levels are displayed in progressively lighter shades. The blue reference line shows the overall effect of $SISO_c$, estimated using the model in Equation 9a, again controlling for the log of population. Solid lines denote statistical significance of the coefficient at the 95 percent level. Both the dependent and independent variables are expressed in terms of z-scores.

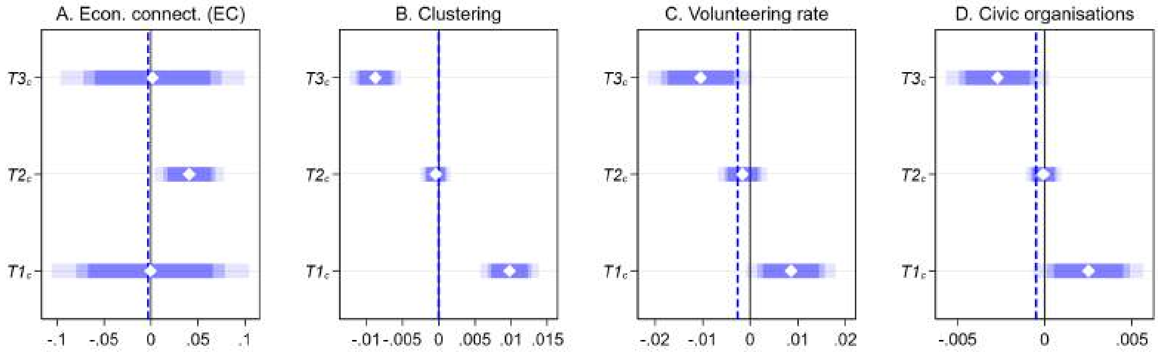


FIGURE A.18 – Coefficient plots for the association of isolation indices and its constituents (MSA-level) with social capital measures in Chetty et al. (2022*a*,*b*). For each characteristic, markers denote results from fitting the model in Equation 9b. Outcome data varies at the ZCTA level, but social isolation and its constituents are measured at the MSA level. Standard errors are clustered by MSA. Confidence intervals at the 90, 95, and 99 percent levels are displayed in progressively lighter shades. The blue reference line shows the overall effect of $SISO_c$, estimated using the model in Equation 9a. Solid lines denote statistical significance of the coefficient at the 95 percent level. Independent variables are expressed in terms of z-scores.

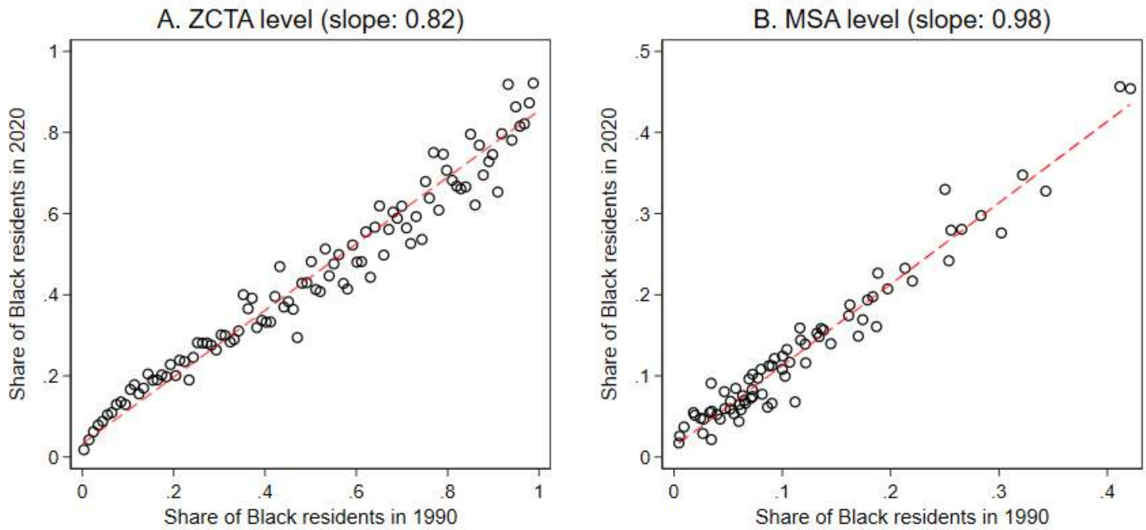


FIGURE A.19 – Persistence in the demographic composition of ZCTAs (panel A) and MSAs (panel B) in the largest 75 US cities (MSAs with 50 or more ZCTAs). Each binned scatter plot shows the average proportion of Black residents in 2020 (y axis) for each partition of the same variable, measured in 1990 (Cattaneo et al., 2024). A linear fit is overlaid in red. Panel A uses 100 evenly-spaced bins, panel B uses 75 bins (corresponding to the number of MSAs, i.e., to a conventional scatter plot). Source: NHGIS time series tables standardised for 2010 definitions of geographical units (Manson et al., 2021).

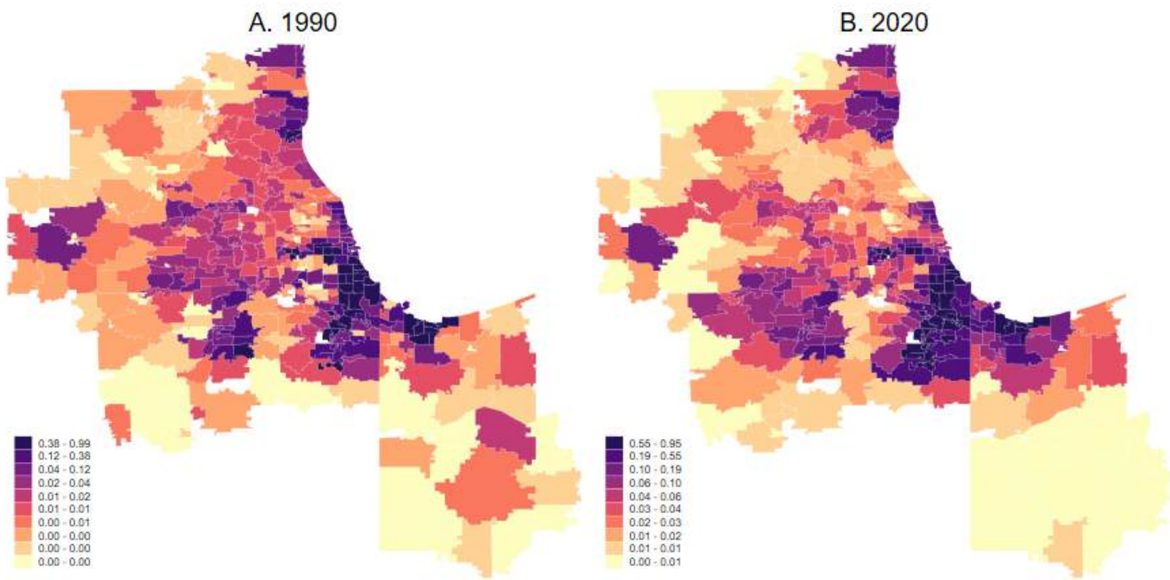


FIGURE A.20 – Persistence in the demographic composition of ZCTAs in Chicago. Each map plots the share of Black residents in each ZCTA, with darker polygons denoting higher values. Legend scales are allowed to vary between each map, to emphasize the similar spatial demographic distribution, despite overall levels changing slightly over time. The share of Black residents in the MSA was 0.19 in 1990 and 0.16 in 2020.

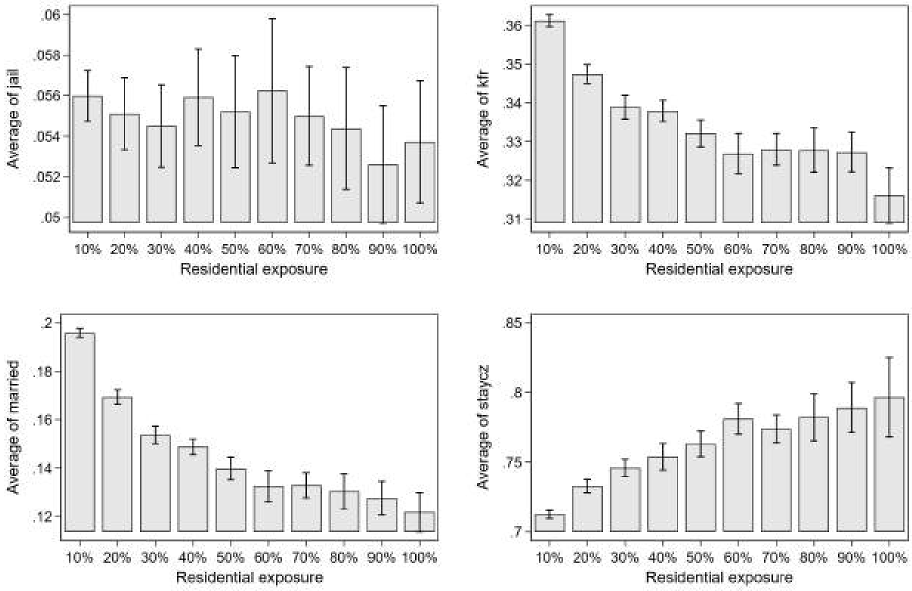


FIGURE A.21 – Conditional averages of outcomes for Black adults by levels of residential exposure (bins of 10 percentage points width) of the neighbourhood they grew up in. Averages obtained by regressing each outcome on the set of residential exposure dummies, absorbing MSA fixed-effects. Vertical bars denote 90 percent confidence intervals. Detailed regression results are summarised in Appendix Table A.5.

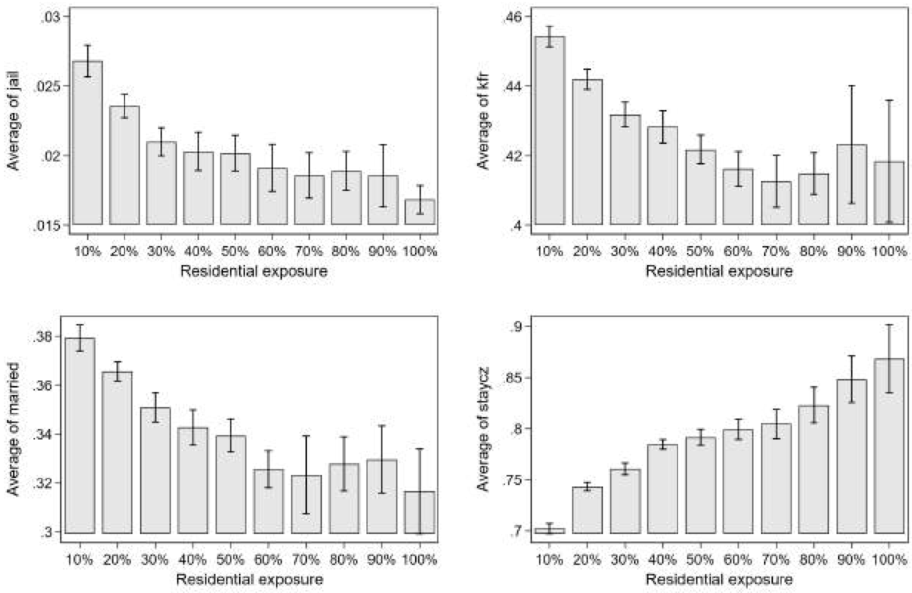


FIGURE A.22 – Conditional averages of outcomes for Hispanic/Latino adults by levels of residential exposure (bins of 10 percentage points width) of the neighbourhood they grew up in. Averages obtained by regressing each outcome on the set of residential exposure dummies, absorbing MSA fixed-effects. Vertical bars denote 90 percent confidence intervals. Detailed regression results are summarised in Appendix Table A.6.

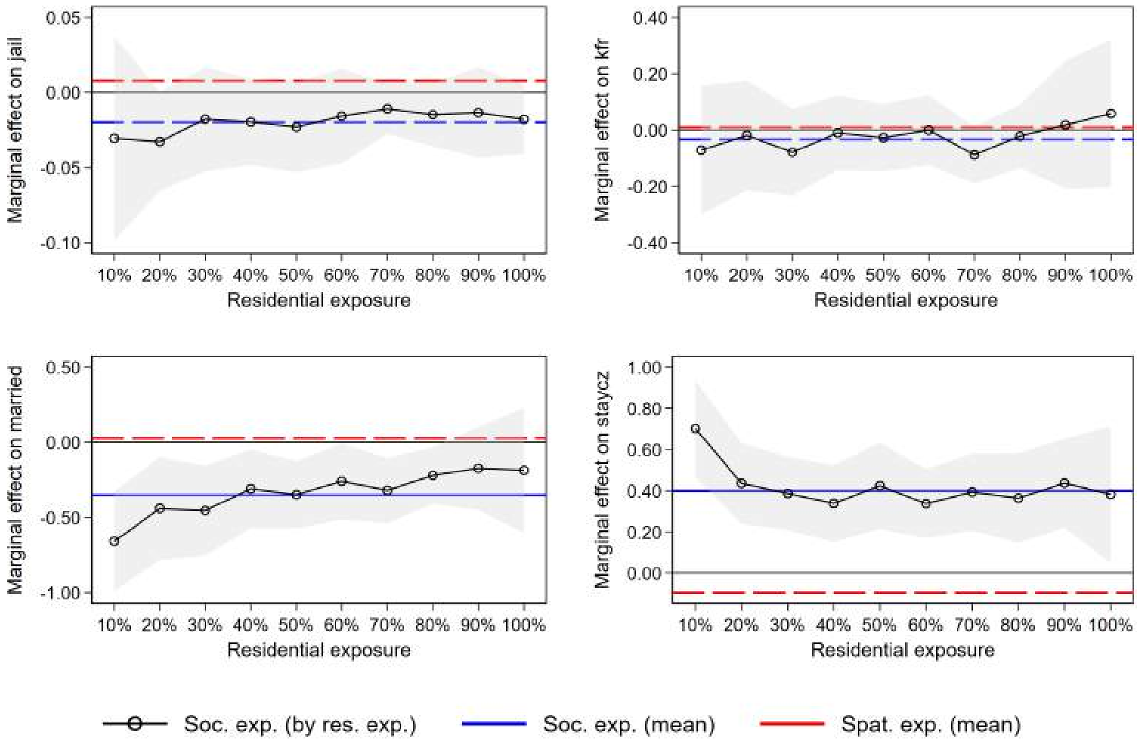


FIGURE A.23 – Black connected markers gives marginal effects of social exposure β_d by levels of residential exposure, obtained by estimating Equation 10b. The shaded area denotes 95 percent confidence intervals. For reference, added lines in blue and red give the average marginal effect for social exposure, β, and spatial exposure, γ, obtained by estimating Equation 10a. Lines are solid if the coefficient is significant at the 95 percent level, dashed otherwise. Outcomes and exposure measures are always specific to the Hispanic/Latino population. All models absorb MSA fixed-effects. Standard errors are clustered at the MSA level.

TABLE A.1 – Estimates of the spatial decay parameter of friendship

	(1)	(2)	(3)	(4)
ln Distance	-1.381^a	-1.286^a		
	(0.00545)	(0.00550)		
Distance (km)			-0.0171^a	-0.0383^a
			(0.0000743)	(0.000182)
Distance^2 (km)				0.0000874^a
				(0.000000629)
Constant	15.38^a	14.74^a	11.15^a	12.00^a
	(0.0240)	(0.0232)	(0.00802)	(0.0108)
Within R^2	0.52	0.60	0.45	0.52
Observations	17,926,190	7,310,173	17,926,190	17,926,190
Absorbed ZCTA FEs	26,142	9,747	26,142	26,142

Notes: The table shows estimates of friendship decay, obtained by regressing the log of social connectedness onto measures of physical distance between ZIP Code pairs. For the purpose of this estimation, the sample includes the top 1,000 paired ZIP Code by distance, for all US ZIPs (irrespective of where they are located). An exception is the estimate in column (2), which restricts the sample to ZIP Code in MSAs encompassing at least 50 ZCTAs. All regressions absorb ZCTA fixed effects and are weighted by the product of ZIP Code populations in each pair. Robust standard errors clustered at ZCTA level. Sig. lev.: $^a p < 0.01$; $^b p < 0.05$; $^c p < 0.1$.

Table A.2 – Residential and social segregation

MSA name	RISO	Rank	SISO	Rank	zRISO	zSISO	zSISO-zRISO
St. Louis (IL,MO)	0.306	5	0.186	3	1.684	2.397	0.713
Scranton–Wilkes-Barre (PA)	0.148	55	0.082	31	–0.507	–0.006	0.501
Duluth (MN,WI)	0.030	74	0.008	74	–2.129	–1.726	0.403
Portland-South Portland-Biddeford (ME)	0.028	75	0.006	75	–2.160	–1.763	0.397
Las Vegas-Paradise (NV)	0.115	64	0.056	57	–0.963	–0.600	0.363
Milwaukee-Waukesha-West Allis (WI)	0.360	1	0.201	1	2.422	2.761	0.339
Detroit-Warren-Livonia (MI)	0.339	2	0.189	2	2.138	2.475	0.337
Madison (WI)	0.078	71	0.032	67	–1.461	–1.160	0.301
Pittsburgh (PA)	0.129	59	0.061	53	–0.770	–0.478	0.292
Cleveland-Elyria-Mentor (OH)	0.316	3	0.173	4	1.816	2.106	0.289
Miami-Fort Lauderdale-Pompano Beach (FL)	0.247	16	0.132	10	0.864	1.147	0.283
Poughkeepsie-Newburgh-Middletown (NY)	0.094	68	0.040	63	–1.248	–0.970	0.278
Jacksonville (FL)	0.160	46	0.080	34	–0.331	–0.054	0.276
Washington-Arlington-Alexandria (DC,MD,VA,WV)	0.205	27	0.106	19	0.288	0.545	0.257
Portland-Vancouver-Hillsboro (OR,WA)	0.051	73	0.013	73	–1.839	–1.612	0.228
Charleston (WV)	0.055	72	0.015	72	–1.783	–1.557	0.226
Riverside-San Bernardino-Ontario (CA)	0.137	58	0.062	52	–0.656	–0.458	0.198
Cincinnati-Middletown (IN,KY,OH)	0.185	35	0.091	25	0.007	0.202	0.195
New York-Northern New Jersey-Long Island (NJ,NY,PA)	0.304	6	0.162	5	1.652	1.846	0.194
Tampa-St. Petersburg-Clearwater (FL)	0.148	54	0.066	48	–0.502	–0.364	0.138
San Antonio-New Braunfels (TX)	0.164	44	0.076	38	–0.279	–0.145	0.133
Allentown-Bethlehem-Easton (NJ,PA)	0.170	40	0.079	36	–0.205	–0.083	0.122
Atlanta-Sandy Springs-Marietta (GA)	0.234	18	0.117	16	0.691	0.809	0.117
Philadelphia-Camden-Wilmington (DE,MD,NJ,PA)	0.275	10	0.141	9	1.254	1.353	0.099
Utica-Rome (NY)	0.152	52	0.067	46	–0.442	–0.345	0.097
Sacramento–Arden-Arcade–Roseville (CA)	0.154	51	0.068	45	–0.416	–0.333	0.083
Albany-Schenectady-Troy (NY)	0.150	53	0.065	49	–0.471	–0.390	0.082
Baltimore-Towson (MD)	0.226	21	0.110	18	0.578	0.655	0.077
Tulsa (OK)	0.084	70	0.026	70	–1.378	–1.305	0.073
Chicago-Joliet-Naperville (IL,IN,WI)	0.293	8	0.150	6	1.497	1.564	0.067
Dayton (OH)	0.237	17	0.116	17	0.728	0.792	0.064
Virginia Beach-Norfolk-Newport News (NC,VA)	0.119	61	0.046	61	–0.899	–0.839	0.060
Hartford-West Hartford-East Hartford (CT)	0.260	12	0.129	11	1.042	1.086	0.044
Phoenix-Mesa-Glendale (AZ)	0.185	34	0.084	28	0.013	0.043	0.030
Orlando-Kissimmee-Sanford (FL)	0.160	47	0.069	44	–0.333	–0.303	0.030
Austin-Round Rock-San Marcos (TX)	0.112	66	0.040	65	–0.994	–0.974	0.020
San Jose-Sunnyvale-Santa Clara (CA)	0.114	65	0.041	62	–0.968	–0.953	0.016
Seattle-Tacoma-Bellevue (WA)	0.085	69	0.023	71	–1.366	–1.362	0.004
Los Angeles-Long Beach-Santa Ana (CA)	0.248	15	0.120	15	0.884	0.885	0.000
Rochester (NY)	0.251	14	0.121	14	0.912	0.905	–0.008
Peoria (IL)	0.217	23	0.101	22	0.455	0.446	–0.008
Syracuse (NY)	0.189	32	0.084	29	0.058	0.041	–0.017
Indianapolis-Carmel (IN)	0.220	22	0.102	21	0.486	0.463	–0.023
Oklahoma City (OK)	0.116	62	0.040	64	–0.940	–0.973	–0.033
Kansas City (KS,MO)	0.189	33	0.083	30	0.057	0.022	–0.035
Nashville-Davidson–Murfreesboro–Franklin (TN)	0.157	48	0.064	50	–0.372	–0.411	–0.039
Toledo (OH)	0.175	39	0.075	40	–0.128	–0.171	–0.043
Dallas-Fort Worth-Arlington (TX)	0.179	38	0.076	37	–0.070	–0.132	–0.062
San Diego-Carlsbad-San Marcos (CA)	0.163	45	0.067	47	–0.289	–0.354	–0.066
Harrisburg-Carlisle (PA)	0.205	26	0.091	24	0.290	0.205	–0.085
Minneapolis-St. Paul-Bloomington (MN,WI)	0.146	56	0.055	58	–0.532	–0.638	–0.107
Boston-Cambridge-Quincy (MA,NH)	0.230	20	0.103	20	0.624	0.488	–0.136
Buffalo-Niagara Falls (NY)	0.295	7	0.142	8	1.526	1.381	–0.145
Birmingham-Hoover (AL)	0.308	4	0.150	7	1.711	1.563	–0.148
Wichita (KS)	0.123	60	0.039	66	–0.843	–1.006	–0.163
Columbus (OH)	0.191	31	0.079	35	0.086	–0.077	–0.163
Des Moines-West Des Moines (IA)	0.108	67	0.030	68	–1.050	–1.215	–0.166
Grand Rapids-Wyoming (MI)	0.155	50	0.057	54	–0.408	–0.579	–0.171
San Francisco-Oakland-Fremont (CA)	0.166	42	0.064	51	–0.252	–0.431	–0.178
Louisville/Jefferson County (IN,KY)	0.184	36	0.074	41	0.002	–0.198	–0.200
Denver-Aurora-Broomfield (CO)	0.156	49	0.057	55	–0.387	–0.593	–0.206
Houston-Sugar Land-Baytown (TX)	0.203	28	0.084	27	0.258	0.044	–0.214
Omaha-Council Bluffs (IA,NE)	0.182	37	0.069	43	–0.036	–0.297	–0.261
Topeka (KS)	0.115	63	0.029	69	–0.962	–1.226	–0.264
Worcester (MA)	0.143	57	0.046	60	–0.570	–0.838	–0.268
Providence-New Bedford-Fall River (MA,RI)	0.213	24	0.088	26	0.400	0.126	–0.274
Youngstown-Warren-Boardman (OH,PA)	0.191	30	0.073	42	0.096	–0.210	–0.306
Springfield (MA)	0.275	11	0.122	13	1.252	0.935	–0.316
New Orleans-Metairie-Kenner (LA)	0.206	25	0.080	33	0.293	–0.044	–0.337
Richmond (VA)	0.166	43	0.056	56	–0.257	–0.595	–0.338
Little Rock-North Little Rock-Conway (AR)	0.202	29	0.075	39	0.244	–0.165	–0.409
Charlotte-Gastonia-Rock Hill (NC,SC)	0.167	41	0.053	59	–0.237	–0.676	–0.439
Memphis (AR,MS,TN)	0.289	9	0.123	12	1.438	0.941	–0.497
Jackson (MS)	0.259	13	0.100	23	1.030	0.424	–0.606
Baton Rouge (LA)	0.232	19	0.081	32	0.657	–0.037	–0.694

Notes: The table reports values of residential (RISO) and social (SISO) isolation indices respectively, along with associated ranks. It also gives z-scores of these values (zRISO and zSISO), along with their difference. MSAs are listed in descending order of this variable. Top entries are relatively more socially than residentially segregated, and vice-versa at the bottom.

Table A.3 – Association of urban features with social isolation

	A. ln Population					B. Share non White					C. Share bachelor+				
	(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)
Social isolation	8.740^a					4.690^c					-0.358				
	(2.604)					(2.398)					(2.296)				
Other area social term ($T3_c$)		31.50^a	11.17^a				20.50^a	6.508^b				8.707	0.347		
		(4.579)	(2.814)				(5.585)	(2.674)				(5.284)	(2.495)		
Own area social term ($T2_c$)		-5.222		-38.78^a			-14.71		-35.38^b			-7.399		-13.84	
		(12.04)		(13.31)			(15.05)		(14.23)			(15.93)		(16.66)	
Own area residential term ($T1_c$)		-171.6^a			12.03		-122.0^a			-0.926		-72.41^c			-20.83
		(27.07)			(20.98)		(39.57)			(19.81)		(41.03)			(19.83)
Observations	75	75	75	75	75	75	75	75	75	75	75	75	75	75	75
R^2	0.142	0.490	0.194	0.058	0.005	0.041	0.232	0.066	0.049	0.000	0.000	0.057	0.000	0.007	0.014
	D. Per capita income (2012 USD)					E. Unemployment rate					F. Owned housing value (P50)				
	(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)
Social isolation	0.464					8.109^a					-0.589				
	(2.246)					(2.613)					(2.508)				
Other area social term ($T3_c$)		12.53^b	0.854				16.51^a	9.262^a				14.08^a	-0.651		
		(6.163)	(2.464)				(5.657)	(2.795)				(5.141)	(2.679)		
Own area social term ($T2_c$)		16.29		5.613			7.924		-13.33			36.73^c		25.52	
		(17.00)		(17.97)			(12.52)		(10.45)			(20.20)		(21.59)	
Own area residential term ($T1_c$)		-91.80^b			-21.26		-57.76^c			37.05^b		-110.2^a			-33.41
		(45.89)			(19.91)		(34.27)			(18.44)		(37.58)			(20.22)
Observations	75	75	75	75	75	75	75	75	75	75	75	75	75	75	75
R^2	0.000	0.083	0.001	0.001	0.015	0.123	0.165	0.134	0.007	0.044	0.001	0.140	0.001	0.025	0.036
	G. Share public transit (mean)					H. Assoc. density (2014)					I. Church adherents rate (2010)				
	(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)
Social isolation	5.988					-4.174^c					8.100^a				
	(3.923)					(2.399)					(2.738)				
Other area social term ($T3_c$)		18.20^b	6.546				-25.21^a	-5.342^b				-11.47^a	7.791^b		
		(6.896)	(4.319)				(4.323)	(2.600)				(3.688)	(2.986)		
Own area social term ($T2_c$)		32.77^b		11.87			-22.51^c		1.995			-2.713		-1.561	
		(13.79)		(12.82)			(11.66)		(13.01)			(16.18)		(15.30)	
Own area residential term ($T1_c$)		-85.85^a			15.36		158.0^a			14.69		159.9^a			93.66^a
		(31.65)			(25.18)		(31.20)			(21.14)		(30.28)			(20.04)
Observations	75	75	75	75	75	75	75	75	75	75	75	75	75	75	75
R^2	0.067	0.158	0.067	0.005	0.008	0.032	0.283	0.044	0.000	0.007	0.122	0.342	0.095	0.000	0.282

Notes: The table displays coefficients obtained from regressing urban characteristics on social isolation and its components. Urban characteristics are expressed in z-scores to facilitate comparison of magnitudes. Models in column (1) consider social isolation alone ($SISO_c$). Models in (2) break-down social isolation into the terms defined in Equation 7. Models in (3-5) show the independent effect associated with each component separately. Robust standard errors in parentheses. Sig. lev.: $^a p < 0.01$; $^b p < 0.05$; $^c p < 0.1$.

Table A.4 – Association of urban features with social and spatial isolation

	A. ln Population				B. Share non White				C. Share bachelor+			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Social isolation	8.740^{a}		15.14^{a}		4.690^{c}		-2.051		-0.358		6.919^{c}	
	(2.604)		(4.363)		(2.398)		(4.064)		(2.296)		(3.532)	
Spatial isolation		5.783^{c}	-9.298^{c}	0.456		7.755^{a}	9.799^{b}	19.94^{a}		-3.683	-10.58^{b}	-8.394
		(3.127)	(4.677)	(4.636)		(2.896)	(4.877)	(4.983)		(2.769)	(4.544)	(5.529)
Other area social term ($T3_c$)				31.35^{a}				13.82^{b}				11.52^{b}
				(4.862)				(5.420)				(5.144)
Own area social term ($T2_c$)				-5.169				-12.42				-8.362
				(12.14)				(15.40)				(15.17)
Own area residential term ($T1_c$)				-173.2^{a}				-188.6^{a}				-44.38
				(33.87)				(28.47)				(47.26)
Observations	75	75	75	75	75	75	75	75	75	75	75	75
R^2	0.142	0.043	0.177	0.490	0.041	0.077	0.080	0.362	0.000	0.017	0.046	0.080

	D. Per capita income (2012 USD)				E. Unemployment rate				F. Owned housing value (P50)			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Social isolation	0.464		14.55^{a}		8.109^{a}		8.759^{b}		-0.589		11.93^{a}	
	(2.246)		(3.609)		(2.613)		(4.094)		(2.508)		(3.895)	
Spatial isolation		-5.979^{b}	-20.48^{a}	-17.82^{a}		7.782^{b}	-0.945	3.495		-6.310^{b}	-18.20^{a}	-12.47^{b}
		(2.441)	(4.118)	(4.995)		(2.993)	(4.382)	(4.780)		(2.520)	(4.537)	(5.020)
Other area social term ($T3_c$)				18.50^{a}				15.34^{a}				18.26^{a}
				(5.536)				(5.718)				(5.553)
Own area social term ($T2_c$)				14.24				8.325				35.30^{c}
				(14.35)				(12.82)				(18.19)
Own area residential term ($T1_c$)				-32.29				-69.43^{c}				-68.51^{c}
				(47.47)				(37.95)				(38.41)
Observations	75	75	75	75	75	75	75	75	75	75	75	75
R^2	0.000	0.046	0.170	0.187	0.123	0.078	0.123	0.169	0.001	0.051	0.135	0.191

	G. Share public transit (mean)				H. Assoc. density (2014)				I. Church adherents rate (2010)			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Social isolation	5.988		20.74^{b}		-4.174^{c}		-5.890		8.100^{a}		0.687	
	(3.923)		(8.849)		(2.399)		(3.649)		(2.738)		(3.738)	
Spatial isolation		-0.777	-21.44^{b}	-18.37^{c}		-3.374	2.494	-9.537^{b}		11.46^{a}	10.77^{b}	1.897
		(1.831)	(8.557)	(9.955)		(3.356)	(5.028)	(4.666)		(3.380)	(5.012)	(4.345)
Other area social term ($T3_c$)				24.35^{a}				-22.01^{a}				-12.10^{a}
				(8.878)				(4.333)				(3.802)
Own area social term ($T2_c$)				30.67^{b}				-23.61^{c}				-2.496
				(11.85)				(12.76)				(16.43)
Own area residential term ($T1_c$)				-24.50				189.9^{a}				153.6^{a}
				(41.14)				(37.34)				(34.83)
Observations	75	75	75	75	75	75	75	75	75	75	75	75
R^2	0.067	0.001	0.253	0.269	0.032	0.015	0.035	0.312	0.122	0.169	0.169	0.343

Notes: The table displays coefficients obtained from regressing urban characteristics on social isolation and its components, controlling for spatial isolation in columns (3) and (4). Urban characteristics are expressed in z-scores to facilitate comparison of magnitudes. Models in column (1) consider social isolation alone ($SISO_c^s$), replicating the same column in Table A.3. Models in (2) consider spatial isolation alone ($SISO_c^d$). Robust standard errors in parentheses. Sig. lev.: $^a p < 0.01$; $^b p < 0.05$; $^c p < 0.1$.

Table A.5 – Average outcomes for Black residents by residential exposure

	(1) Jail	(2) Inc. rank	(3) Married	(4) Stay CZ
Resid. exp. dummy=20	-0.000877	-0.0138^a	-0.0266^a	0.0203^a
	(0.00168)	(0.00229)	(0.00266)	(0.00380)
Resid. exp. dummy=30	-0.00149	-0.0224^a	-0.0423^a	0.0334^a
	(0.00182)	(0.00268)	(0.00303)	(0.00485)
Resid. exp. dummy=40	-0.0000655	-0.0233^a	-0.0471^a	0.0414^a
	(0.00195)	(0.00229)	(0.00265)	(0.00666)
Resid. exp. dummy=50	-0.000777	-0.0291^a	-0.0561^a	0.0507^a
	(0.00214)	(0.00270)	(0.00350)	(0.00664)
Resid. exp. dummy=60	0.000253	-0.0344^a	-0.0634^a	0.0687^a
	(0.00270)	(0.00365)	(0.00441)	(0.00802)
Resid. exp. dummy=70	-0.000992	-0.0333^a	-0.0630^a	0.0613^a
	(0.00197)	(0.00294)	(0.00379)	(0.00739)
Resid. exp. dummy=80	-0.00161	-0.0335^a	-0.0656^a	0.0697^a
	(0.00224)	(0.00384)	(0.00497)	(0.0110)
Resid. exp. dummy=90	-0.00339	-0.0340^a	-0.0684^a	0.0767^a
	(0.00220)	(0.00374)	(0.00469)	(0.0118)
Resid. exp. dummy=100	-0.00226	-0.0452^a	-0.0742^a	0.0843^a
	(0.00215)	(0.00488)	(0.00541)	(0.0182)
Constant (baseline)	0.0560^a	0.361^a	0.196^a	0.712^a
	(0.000759)	(0.000979)	(0.00110)	(0.00178)
Within R^2	0.00	0.04	0.06	0.04
Observations	5,527	5,937	5,937	5,919
Absorbed MSA FEs	73	74	74	74

Notes: Conditional averages of outcomes for Black adults by levels of residential exposure (bins of 10 percentage points width) of the neighbourhood they grew up in. Averages obtained by regressing each outcome on the set of residential exposure dummies, absorbing MSA fixed-effects. Sample restricted to observations in MSAs containing at least 50 ZCTAs. Outcomes, measured for the Black group only: fraction incarcerated in 2010; mean percentile rank in the national distribution of household income in 2014-15; fraction of children who are married in 2015; fraction of children living in the same CZ in adulthood. Each regression includes all Zip Codes for which outcomes are available. Robust standard errors clustered at MSA level. Sig. lev.: $^a p < 0.01$; $^b p < 0.05$; $^c p < 0.1$.

Table A.6 – Average outcomes for Hispanics/Latino residents by residential exposure

	(1) Jail	(2) Inc. rank	(3) Married	(4) Stay CZ
Resid. exp. dummy=20	-0.00324^{a}	-0.0123^{a}	-0.0137^{a}	0.0409^{a}
	(0.00105)	(0.00305)	(0.00472)	(0.00443)
Resid. exp. dummy=30	-0.00581^{a}	-0.0224^{a}	-0.0284^{a}	0.0585^{a}
	(0.00117)	(0.00348)	(0.00631)	(0.00561)
Resid. exp. dummy=40	-0.00649^{a}	-0.0260^{a}	-0.0366^{a}	0.0824^{a}
	(0.00141)	(0.00393)	(0.00693)	(0.00464)
Resid. exp. dummy=50	-0.00662^{a}	-0.0325^{a}	-0.0400^{a}	0.0893^{a}
	(0.00128)	(0.00390)	(0.00676)	(0.00727)
Resid. exp. dummy=60	-0.00769^{a}	-0.0380^{a}	-0.0537^{a}	0.0971^{a}
	(0.00157)	(0.00429)	(0.00713)	(0.00843)
Resid. exp. dummy=70	-0.00822^{a}	-0.0416^{a}	-0.0560^{a}	0.102^{a}
	(0.00149)	(0.00539)	(0.0121)	(0.0111)
Resid. exp. dummy=80	-0.00790^{a}	-0.0394^{a}	-0.0515^{a}	0.121^{a}
	(0.00140)	(0.00487)	(0.00919)	(0.0128)
Resid. exp. dummy=90	-0.00824^{a}	-0.0310^{a}	-0.0497^{a}	0.146^{a}
	(0.00182)	(0.0110)	(0.0101)	(0.0155)
Resid. exp. dummy=100	-0.00995^{a}	-0.0359^{a}	-0.0627^{a}	0.166^{a}
	(0.00122)	(0.0113)	(0.0117)	(0.0215)
Constant (baseline)	0.0268^{a}	0.454^{a}	0.379^{a}	0.702^{a}
	(0.000693)	(0.00181)	(0.00326)	(0.00303)
Within R^2	0.01	0.02	0.02	0.06
Observations	5,953	6,361	6,361	6,345
Absorbed MSA FEs	72	73	73	73

Notes: Conditional averages of outcomes for Hispanic/Latino adults by levels of residential exposure (bins of 10 percentage points width) of the neighbourhood they grew up in. Averages obtained by regressing each outcome on the set of residential exposure dummies, absorbing MSA fixed-effects. Sample restricted to observations in MSAs containing at least 50 ZCTAs. Outcomes, measured for the Hispanic/Latino group only: fraction incarcerated in 2010; mean percentile rank in the national distribution of household income in 2014-15; fraction of children who are married in 2015; fraction of children living in the same CZ in adulthood. Each regression includes all zip-codes for which outcomes are available. Robust standard errors clustered at MSA level. Sig. lev.: $^{a}p < 0.01$; $^{b}p < 0.05$; $^{c}p < 0.1$.

Table A.7 – Outcomes for Black residents: social and spatial exposures

	(1) Jail	(2) Inc. rank	(3) Married	(4) Stay CZ
Social exp. (Black)	-0.0307^b	0.0173	-0.106^a	0.232^a
	(0.0143)	(0.0274)	(0.0381)	(0.0521)
Spatial exp. (Black)	-0.0111	-0.110^a	-0.0981^a	0.0625
	(0.0160)	(0.0282)	(0.0318)	(0.0588)
Within R^2	0.00	0.04	0.07	0.05
Observations	5,527	5,937	5,937	5,919
Absorbed MSA FEs	73	74	74	74

Notes: Sample restricted to observations in MSAs containing at least 50 ZCTAs. All models absorb MSA fixed effects and dummies for levels of residential exposure (in 10pp increments). Exposure measures and outcomes are specific to the Black group. The following outcomes are considered: fraction incarcerated in 2010; mean percentile rank in the national distribution of household income in 2014-15; fraction of children who are married in 2015; fraction of children living in the same CZ in adulthood. Each regression includes all Zip Codes for which outcomes are available. Robust standard errors clustered at MSA level in parentheses. Sig. lev.: $^a p < 0.01$; $^b p < 0.05$; $^c p < 0.1$.

Table A.8 – Outcomes for Hispanics/Latino residents: social and spatial exposures

	(1) Jail	(2) Inc. rank	(3) Married	(4) Stay CZ
Social exp. (Hisp.)	-0.0198^c	-0.0326	-0.352^b	0.399^a
	(0.0117)	(0.0654)	(0.136)	(0.0829)
Spatial exp. (Hisp.)	0.00794	0.00963	0.0276	-0.0947^c
	(0.00863)	(0.0521)	(0.0544)	(0.0505)
Within R^2	0.01	0.02	0.03	0.08
Observations	5,953	6,361	6,361	6,345
Absorbed MSA FEs	72	73	73	73

Notes: Sample restricted to observations in MSAs containing at least 50 ZCTAs. All models absorb MSA fixed effects and dummies for levels of residential exposure (in 10pp increments). Exposure measures and outcomes are specific to the Hispanic/Latino group. The following outcomes are considered: fraction incarcerated in 2010; mean percentile rank in the national distribution of household income in 2014-15; fraction of children who are married in 2015; fraction of children living in the same CZ in adulthood. Each regression includes all zip-codes for which outcomes are available. Robust standard errors clustered at MSA level. Sig. lev.: $^a p < 0.01$; $^b p < 0.05$; $^c p < 0.1$.

Table A.9 – Outcomes for Black residents with an additional control for Economic Connectedness (EC) as defined in Chetty et al. (2022a)

	Jail			Inc. rank			Married			Stay CZ		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Social exp. (Black)	-0.0363^{b}	-0.0539^{a}	-0.0625^{a}	0.0325	0.00385	0.0973^{a}	-0.101^{b}	-0.124^{a}	-0.0428	0.199^{a}	0.216^{a}	0.132^{b}
	(0.0168)	(0.0127)	(0.0170)	(0.0333)	(0.0216)	(0.0246)	(0.0426)	(0.0227)	(0.0313)	(0.0583)	(0.0402)	(0.0572)
Spatial exp. (Black)	-0.00143		0.0157	-0.132^{a}		-0.173^{a}	-0.114^{a}		-0.150^{a}	0.113^{c}		0.155^{b}
	(0.0197)		(0.0178)	(0.0357)		(0.0268)	(0.0406)		(0.0353)	(0.0585)		(0.0652)
EC (Chetty et al., 2022)		-0.0309^{a}	-0.0312^{a}		0.0732^{a}	0.0769^{a}		0.0654^{a}	0.0686^{a}		-0.0752^{a}	-0.0785^{a}
		(0.00508)	(0.00517)		(0.00600)	(0.00622)		(0.00883)	(0.00913)		(0.0137)	(0.0141)
Within R^2	0.00	0.02	0.02	0.04	0.08	0.09	0.07	0.08	0.09	0.05	0.06	0.07
Observations	5,261	5,261	5,261	5,641	5,641	5,641	5,641	5,641	5,641	5,623	5,623	5,623
Absorbed MSA FEs	73	73	73	74	74	74	74	74	74	74	74	74

Notes: Sample restricted to observations in MSAs containing at least 50 ZCTAs. Each regression further restricts the sample to Zip Codes for which each outcome as well as Economic Connectedness (EC), social and spatial exposures are observed. All models absorb MSA fixed effects and dummies for levels of residential exposure (in 10pp increments). Exposure measures and outcomes are specific to the Black group. The following outcomes are considered: fraction incarcerated in 2010; mean percentile rank in the national distribution of household income in 2014-15; fraction of children who are married in 2015; fraction of children living in the same CZ in adulthood. Robust standard errors clustered at MSA level in parentheses. Sig. lev.: $^{a}p < 0.01$; $^{b}p < 0.05$; $^{c}p < 0.1$.