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.*

February 11, 2026

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. We also decompose social segregation into residential, within-neighbourhood, and cross-neighbourhood interaction components, revealing that while social ties mirror residential patterns, geographically broader networks can partially offset residential isolation. Examining heterogeneity across racial minority groups, we find variation in social isolation with Black individuals being relatively more socially segregated than Hispanic and Asian minorities. Our results suggest that social connections, beyond residential geography, are important for understanding racial segregation in America.

Keywords: Residential and social segregation; social networks; social connectedness.
JEL Codes: R23; J15.

[*]Corresponding author. Swedish Institute for Social Research (SOFI), Stockholm University, and Institute for Futures Studies (IFFS). Email: andreas.diemer@sofi.su.se. Funding from NordForsk (95263), VR (2019-02168, 2022-05235) and the ERC (101142895) 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 (e.g., Ananat, 2011), 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 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 in the city. We then take this measure to the data using Facebook's Social Connectedness Index (SCI) for ZIP Code Tabulation Areas (Bailey et al., 2020).[1] 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.[2] 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. This is to be expected based on the decomposition we propose that allows us to separately examine the contribution of residential, within-neighbourhood, and cross-neighbourhood interactions to overall isolation. We show that social ties largely mirror residential patterns, but also discuss how geographically broader networks can still partially offset residential segregation. This explains why, on average, our index of social isolation is lower than its residential counterpart in absolute levels. 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

[1] 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 affects 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 SCI to quantify within and between neighbourhoods (or cross-boundary) social interactions, before using these measures to re-evaluate how racially segregated neighbourhoods are.

[2] 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.

be implied by its own racial composition.

Next, we characterise cities that are highly socially isolated, even for comparable levels of 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 examine how our findings vary across different racial and ethnic minority groups, uncovering substantial heterogeneity in social isolation and its underlying components. In particular, we find that compared to racial minorities in general: Black individuals are more socially segregated in most cities, Asians are less socially segregated in most cities, and Hispanics are sometimes more and sometimes less socially segregated. Further, we show that this is not simply driven by residential isolation. Blacks are more socially relative to residentially isolated from the white majority than are Hispanics or Asians, even conditional on the overall size of the racial minority in the city.

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 based on how individuals actually interact socially rather than where they reside. Because social connections are not constrained by administrative boundaries, this naturally implies allowing for linkages across spatial units based on social ties — similar in spirit but distinct from segregation measures that account for spatial clustering of minority groups. We show how traditional residential segregation is nested in our index through the often implicit assumption of no linkages across neighbourhoods, and how considering a purely spatial definition of linkages could be misleading. Second, using this measure, we leverage comprehensive social networks data for granular neighbourhoods to estimate social isolation metrics in 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. This is especially true for applications that focus on how actual interpersonal contacts determine economic and social outcomes — as opposed to more loosely defined 'contextual factors' based on where people live. In such instances, measuring segregation using social interactions can improve our understanding of the specific ways in which contacts between different groups influence outcomes (Echenique and Fryer, 2007). Recent work focused on using new data to improve racial segregation measures based on where and when people spend their time. Generally, however, many existing analyses are forced to make assumptions about the intensity of social connections within and across geographical boundaries.[3] Instead, we are the first to explicitly account for social connections between people on a large scale, proposing 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.

In addition, we 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, our measure 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, 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 within and across US urban areas. It also discusses city-level features that correlate with such discrepancies, and how social isolation varies across minority groups. 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. There are multiple ways to measure segregation, each tied to different conceptual aspects of spatial separation between groups (Massey and Denton, 1988). The two most prominent indices in the empirical literature are the dissimilarity index and the isolation index. The former index measures *evenness* — whether groups are distributed uniformly across neighbourhoods — while the latter measures *exposure* — how likely different groups are going to be in contact across space. Traditionally, much of the segregation literature emphasized dissimilarity, partly because it provides an intuitive summary of how far a city deviates from perfect integration, and partly due to its mathematical simplicity (White, 1983). However, more recent work has increasingly adopted isolation and other exposure-based measures (e.g., Cutler et al., 1999; Echenique and Fryer, 2007; Gentzkow

[3]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).

and Shapiro, 2011; Athey et al., 2021; Cook et al., 2022; Chetty et al., 2022*a*,*b*).

We follow this trend toward exposure-based measurement for several reasons. First, because exposure emphasises inter-group contact (Cutler et al., 1999), it aligns conceptually with mechanisms through which segregation affects many economic outcomes, such as peer effects, information flows, and labour market networks (Topa and Zenou, 2015). Thus, 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. Second, isolation naturally handles many neighbourhoods and decomposes cleanly, as we will demonstrate in Equation 7, allowing us to explicitly account for interactions both within and across urban areas. Finally, the isolation index we propose has a structure directly analogous to gravity models widely used in quantitative spatial economics, thus lending itself to interpretation through the lens of established theory. Just as market access in trade models or commuting accessibility in urban models aggregates opportunities across locations weighted by bilateral frictions (e.g., Ahlfeldt et al., 2015; Donaldson and Hornbeck, 2016; Redding, 2023), social isolation as we define it aggregates exposure across neighbourhoods weighted by connection strength and composition.

To fix concepts, we start with a simple residential isolation index ($RISO_c$), which can generally 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. In particular, following Gentzkow and Shapiro (2011) and Athey et al. (2021), the residential isolation of city c can be expressed as:

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

where i denotes a neighbourhood 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 of their neighbourhood, 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 is mechanically smaller. Subtracting the majority population's exposure to minorities makes this measure comparable across cities with different compositions.

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.[4] Indeed, the correct shape or scale of neighbourhood boundaries is highly uncertain and likely tied to specific research questions (Mast and Barca, 2026). The fundamental pitfall is that traditional measures assume interactions are confined within certain 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. Such movements can 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, because researchers did not have information on social connections. The increasing availability of large-scale social network data, however, makes it possible to overcome this limitation. In what follows, we extend the conventional isolation index to incorporate cross-boundary interactions in general, and actual social connections across neighbourhood boundaries in particular.[5]

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) as follows:

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

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.

[4] 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.

[5] By incorporating cross-neighbourhood social ties into isolation, our approach also relates to Massey and Denton (1988)'s notion of *clustering* — the extent to which certain groups reside in contiguous areas. We effectively extend spatial clustering to include social clustering, capturing how groups interact not just within nearby areas but across the broader metropolitan social network. This positions our contribution in continuity with the frontier of segregation research while extending it to incorporate actual social ties.

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:

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

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 neighbourhoods 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:

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

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; Whaley, 2024; Mast and Barca, 2026). 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.

These challenges notwithstanding, a purely spatial definition of isolation such as the one in (4) can be useful to characterise segregation in applications where outcomes are mostly determined by geographical processes, such as access to jobs and services, or accessibility more in general. By contrast, research that builds conceptually on the existence of actual interpersonal contacts, such as the study of job search, referral networks, or knowledge spillovers, can benefit from a measure that is grounded in social interactions.

2.3 Incorporating Social Interactions

In applications where the strength of interpersonal contacts matters more than spatial proximity, the accuracy of the isolation index depends on how exposure weights (ω_{ij}) are defined. 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; Couture et al., 2025). 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 Code Tabulation Areas with at least 500 residents. This index builds on the universe of active non-institutional Facebook users as of March 2020, and captures the (scaled) relative probability of a friendship link between users in two locations. We discuss the underlying data, its construction and representativeness more in detail in Section 3. Here, we focus on the mathematical definition of the index and how we incorporate it into our analysis. The SCI 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 ZIPs 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 i is friends with a random individual from ZIP 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

[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.

[7] For within-ZIP ties ($i = j$), 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., $SCI_{ii} = \mu \frac{2 Connections_{ii}}{Users_i \times Users_i}$). Thus, the scaled probability of a friendship link is $SCI_{ii} \times \frac{Users_i}{Users_i - 1} \approx SCI_{ii}$.

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 or race-group 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 neighbourhood pairs, in turn, may lead to a bias in our measure of isolation. The direction of this potential bias is uncertain a priori, though. If unobserved connections across racial groups and within ZIP Code pairs exhibit homophily, then the index will be biased towards lower segregation. If instead these connections are relatively strong across racial groups, then the index will be biased towards higher segregation. In other words, to interpret our results at face-value, we must assume that micro level social connections, within ZIP Code pairs, are independent of racial identity.[8] Alternatively, we can also interpret our measure in a broader sense as the 'social context' that every resident experiences in the their local area, either directly, or indirectly through their neighbours — an interpretation closer to the concept of 'experienced segregation' (Athey et al., 2021).

Replacing ω_{ij} with ω^s_{ij} in Equation (2), we can derive our social isolation measure ($SISO^s_c$) in Equation (7). By rearranging terms, we can further decompose $SISO^s_c$ and express it in terms of own-area residential isolation ($T1_c$), own-area social isolation ($T2_c$), and other areas social isolation ($T3_c$), as follows:

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

Here, $T1_c$ is the residential isolation of city c scaled by the average social weight assigned

[8]While studying variation within-ZIP Code pairs is outside the scope of this paper due to data limitations, we note it as a promising research avenue which is technically possible with access to proprietary data.

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.[9] 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.

There are several reasons for decomposing $SISO^s_c$ 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.

[9] 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.

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

3.1 Sources and Definitions

Our definition of neighbourhoods is based on the 2010 US Census Bureau's five-digit ZIP Code Tabulation Areas (ZCTA, which we also refer to as ZIP Codes, ZIP areas, or 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,[10] 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. The index builds on a de-identified snapshot of the universe of active, non-institutional Facebook users as of March 2020. Users are active if they interacted with the platform at least once over the previous 30 days. They are non-institutional in the sense that they do not represent businesses, organizations, or other non-human entities.[11] Location is assigned based on a combination of profile information, user activity, device and connection history (e.g., IP addresses). To preserve confidentiality, the index is restricted to ZIPs with at least 500 residents. In our application, this also mitigates empirical concerns associated with estimates of segregation for small units (Rathelot, 2012; D'Haultfœuille and Rathelot, 2017). We discuss representativeness issues below.

Information on the demographic composition of neighbourhoods comes from time series tables compiled by the National Historical Geographic Information System (NHGIS)

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

[11]These entities are permitted on the platform in the form of Pages rather than personal profiles under Facebook's terms of service. Facebook also emphasises that accounts are regularly screened to remove duplicate and false users, including misclassified profiles (e.g., institutional pages registered as individuals) and bots. As a result, the global incidence of false accounts among active monthly users was estimated to be low, at about 5 percent, and even lower within the US, as many such accounts reportedly locate abroad (Meta Platforms, Inc., 2021).

IPUMS project (Manson et al., 2021). This source provides 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 consider Black, Hispanic, and Asian minorities separately in Section 4.3.

3.2 Restrictions to Social and Spatial Interactions

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 no more than 160 miles from each other (about 250 km). We retain all available ZIPs if there are fewer than 1,000 ZIPs within this threshold.[12] In either case, we assume linkages to be absent or negligible beyond these cut-offs. Our data supports this assumption: the weights implied by the relative strength of social connections or by the decay of these connections over geographical distance are essentially null already past the 500^{th} connection (see Appendix Figure A.2). This restriction, in effect, matters only for computational convenience and is inconsequential for our results.

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.[13]

[12]For consistency, we also drop any pair involving at least one ZIP Code for which no data on social connectedness is available — ZIPs that never appear in the SCI files. The spatial radius of 160 miles may seem large. However, note that it only applies to areas for which the nearest 1,000 ZIPs extend beyond this threshold. Conversely, we see value in retaining, whenever possible, the same number of ZIPs for social and spatial exposures. In either case, this ensures the scope of possible exposure is defined consistently in terms of ranks, and is wide enough to capture all relevant connections (particularly in the social case). This also helps us cover the full extent of the urban area even when considering focal ZIPs in fringe MSA locations. At the same time, note that weakly connected or far away ZIPs get assigned small weights, which reduces their relative influence in the analysis.

[13]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.

3.3 Representativeness of Facebook Friendships

An important assumption underlying our social isolation measure is that it builds on social interactions data that accurately reflect real life connections across the entire US geography. In other words, how representative are Facebook friendships of the general patterns of interactions in the population?

Several factors support the validity of this representativeness assumption.[14] First, about 70 percent of the US and Canadian population uses Facebook (Meta Platforms, Inc., 2021), making the SCI among the most comprehensive measures of social connections available at this geographic scale. Facebook's user base is also evenly distributed geographically across both rural and urban places, with comparable penetration rates in all counties (Auxier and Anderson, 2021; Chetty et al., 2022*a*), which is especially reassuring for our analysis.

Second, Facebook user demographics broadly mirror the US population across most dimensions relevant to segregation patterns. According to nationally representative survey data, usage rates differ only slightly between White, Black, or Hispanic adults (with point estimates at 67, 74, and 72 percent respectively), and show minimal variation across income and educational attainment (Auxier and Anderson, 2021). Chetty et al. (2022*a*) confirm this result by comparing key characteristics in their sample of over 70 million Facebook users to estimates of population demographics from the American Community Survey (ACS). Two groups show modest under-representation. Men use Facebook at somewhat lower rates than women (61 compared to 77 percent), and usage declines among the elderly, with only 50 percent of those aged 65 and older active on the platform compared to 70–84 percent for younger cohorts (Auxier and Anderson, 2021). Since residential segregation patterns have historically varied across age groups, our social isolation measures may be somewhat less representative of social connections for older Americans. However, because our analysis aggregates connections at the neighbourhood level rather than examining individual networks, and because the overall penetration rate remains substantial even for these groups, we expect any resulting bias to be limited.

Third, existing research demonstrates that social interactions revealed through Facebook proxy meaningful social relationships. These ties have been shown to materially affect labour market outcomes, housing decisions, and other behaviours (e.g., Bailey, Cao, Kuchler, Stroebel and Wong, 2018; Bailey, Cao, Kuchler and Stroebel, 2018; Bailey et al., 2019, 2022; Gee et al., 2017; Wilson, 2020; Diemer, 2026), providing external validation that the

[14] We emphasize that the SCI measures the *stock* of friendship connections circa 2020. So, our analysis should be thought as representative of social connections over the period 2006 (when Facebook first became broadly accessible to the public) to 2020. Segregation of the social space may differ today, especially post-COVID. While we are unable to study how the social network has changed since 2020, we encourage future research to explore this interesting question.

SCI captures consequential social interactions rather than superficial online connections. Indeed, several studies have shown that interactions over Facebook are good predictors of real life friendships and friendship strength (Gilbert and Karahalios, 2009; Jones et al., 2013). According to Hampton et al. (2011), over 93 percent of Facebook connections link to someone the average user has met in person at least once. They reflect relationships with high school or college peers (31 percent of all connections), immediate or extended family members (20), co-workers (10), and neighbours or acquaintances (9).

Facebook networks naturally include ties of varying strength — from close family to distant acquaintances. While we do not observe the strength of each tie (nor proxies thereof, such as messaging or reactions activity), several considerations suggest this variation does not undermine our aggregate measure. First, the sociological literature recognizes that 'weak ties' often facilitate important flows of information and opportunities (Granovetter, 1973), so there is no conceptual justification to ignore such connections. Second, especially at the neighbourhood level, there is no reason to expect systematic differences in strong- and weak-tie networks. Third, even if such differences existed, they would not bias our segregation measures unless they varied systematically across racial groups, neighbourhoods, or cities. Absent such systematic variation, our measure accurately captures the full social context through which norms, information, and opportunities diffuse, which is precisely what matters for understanding the effective extent of social segregation.[15]

In sum, the SCI allows us to 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.

4 Results

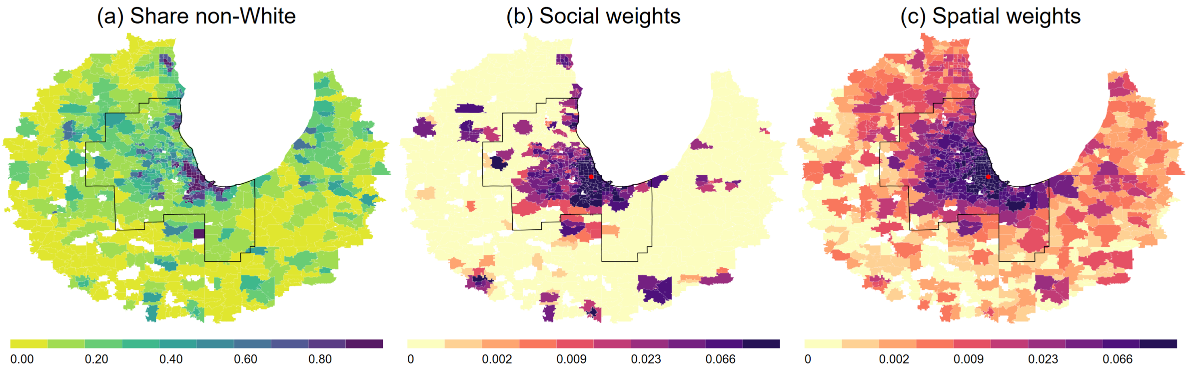
4.1 Segregation Within and Across Urban Areas

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.[16]

[15]Using the same SCI data we rely on in this paper, moreover, Diemer (2026) demonstrates that the average connectedness of each ZCTA — interpreted as a proxy for general sociability — only weakly correlates with local area demographics, ruling out that certain groups would disproportionately contribute to our measure by being more prone to form friendships.

[16]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.

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



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

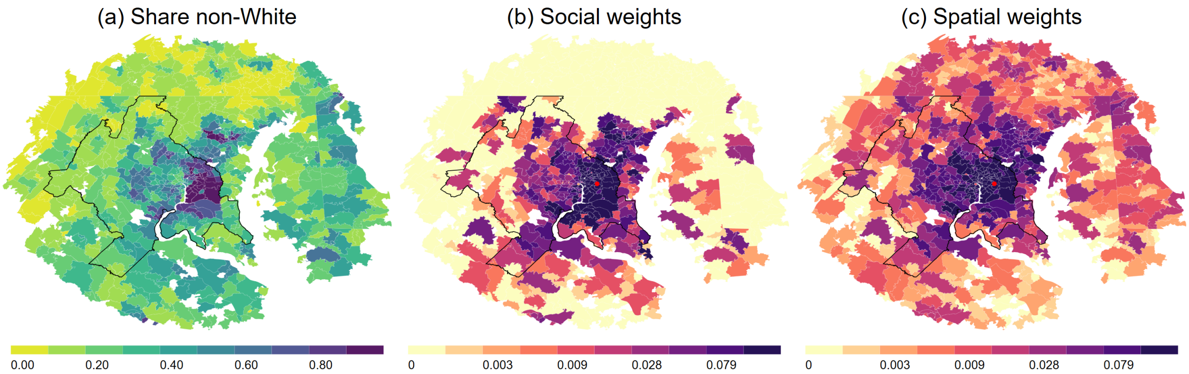


Figure 1 – Maps with MSA race composition (a), social (b) and spatial (c) weights for the least White ZCTA in the city (marked in red on the map). Weights are rescaled by a factor of 100 for legibility. Breaks are defined at each decile of the distribution obtained by pooling both social and spatial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

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.[17] These patterns are observed

[17]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).

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.[18] 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 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. Acknowledging the limitations of using spatial proxies for cross-boundary linkages, we focus this discussion on residential and social isolation measures (henceforth also just $SISO_c$, omitting the s superscript).[19] 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

[18]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.

[19]When estimating the spatial measure we also run into two empirical issues 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 centres 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 concerns also apply to social isolation.

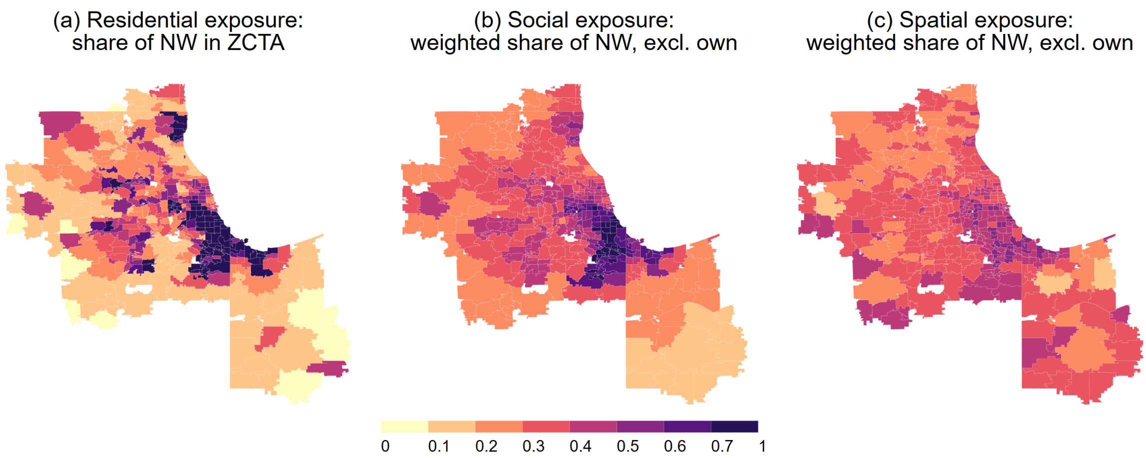


FIGURE 2 – Residential, social, and spatial exposures in Chicago-Joliet-Naperville. The maps show within-MSA variation in minority exposure measures for the city's neighbourhoods, as an example. Exposure values for each ZCTA in panels (b) and (c) are obtained as the linkage-weighted sum of non-White 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 A.14 map residential and social isolation to illustrate 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.[20] Similar to what is documented in Echenique and Fryer (2007), residential and social isolation are strongly correlated with 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 closely mirror existing patterns of residential segregation, partially preserving racial boundaries established by residential sorting and potentially even offsetting integration gains from casting a spatially broader network of connections.

To illustrate this point, it is useful to examine correlations between index components. 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. This strong positive correlation indicates that cities with higher residential isolation also exhibit higher cross-area social isolation — social boundaries mirror residential ones. Interestingly, this third

[20] 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 consistent with the high degree of smoothing imposed with the spatial measure, which is particularly acute in smaller urban areas.

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. This geographic broadening mechanically lowers the contribution of $RISO_c$ to social isolation in the first component (through $\overline{\omega^s_{iic}}$), which helps explain why cities are on average more residentially than socially segregated. 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 form connections outside their neighbourhood does not translate into more diverse non-local friendships. Cities where $T2_c$ is more negative (broader local networks) tend to have higher $T3_c$ (greater cross-area social isolation), offsetting the reduction in overall isolation. It is in this sense that social ties partially preserve residential segregation patterns: while friendships extend beyond immediate neighbourhoods, cross-area social connections do not proportionally increase interracial exposure. Put differently, geographically broader social networks act as partial substitutes for residential proximity in maintaining local connections, but this substitution tends not to extend to forming more racially integrated ties across neighbourhoods.

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.[21] 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 eighty 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. This complex neighbourhood-level exposure heterogeneity reflects the partial substitution pattern discussed above at city-level, where residents of more isolated areas tend to interact more intensely outside their neighbourhood boundaries, but not necessarily in a more racially integrated way.

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

[21] 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.

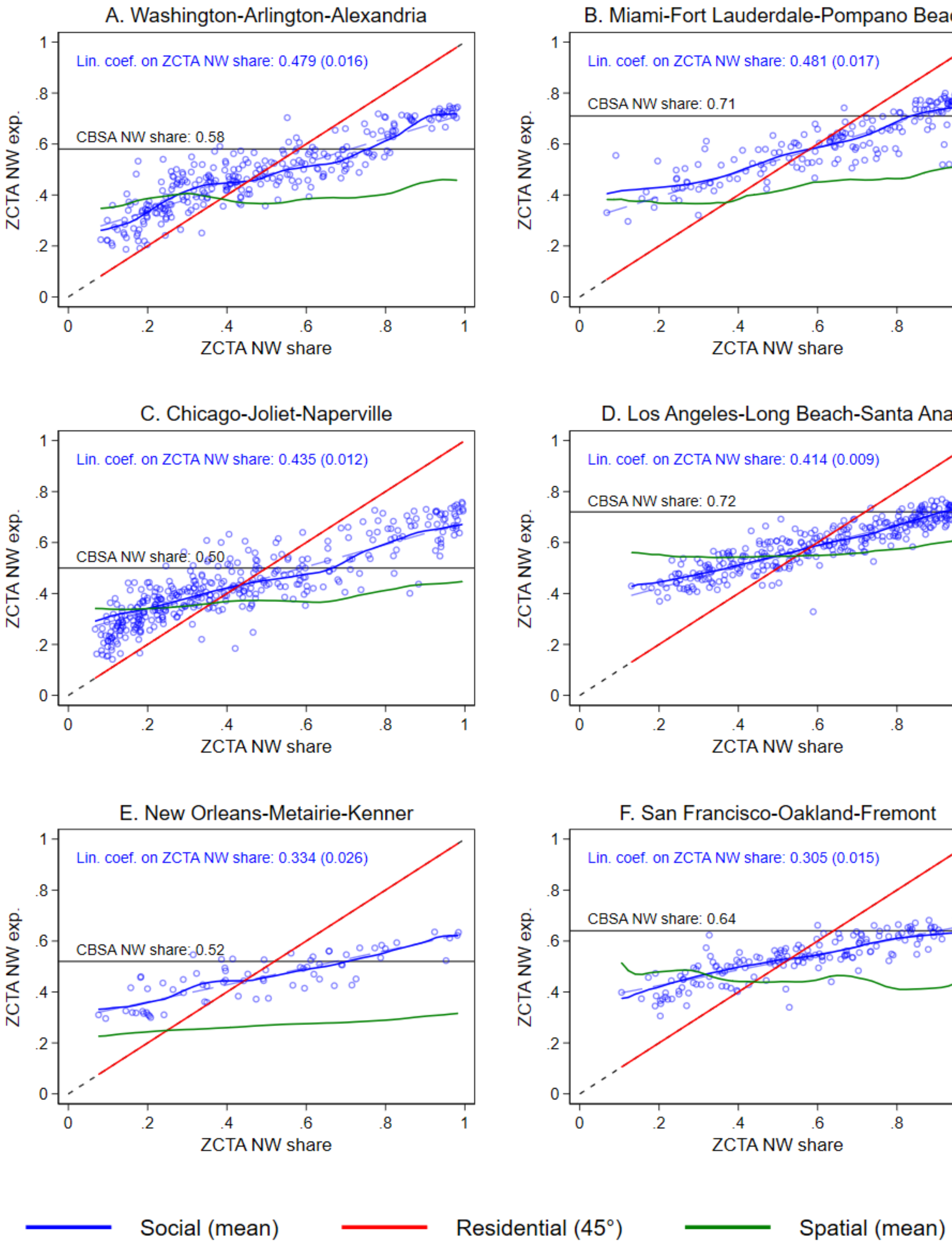


FIGURE 3 – Scatter plots comparing residential, social, and spatial segregation across selected cities with at least 50 ZCTAs. For each city, ZCTA exposures are plotted against each ZCTA's composition (share of non-White residents). Fitted lines for social and spatial exposures are local polynomials 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.

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.

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 $SISO_c$.[22]

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

[22]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.

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 (raw coefficients are in Appendix Table A.3). It is important to emphasise that our findings are descriptive correlations.

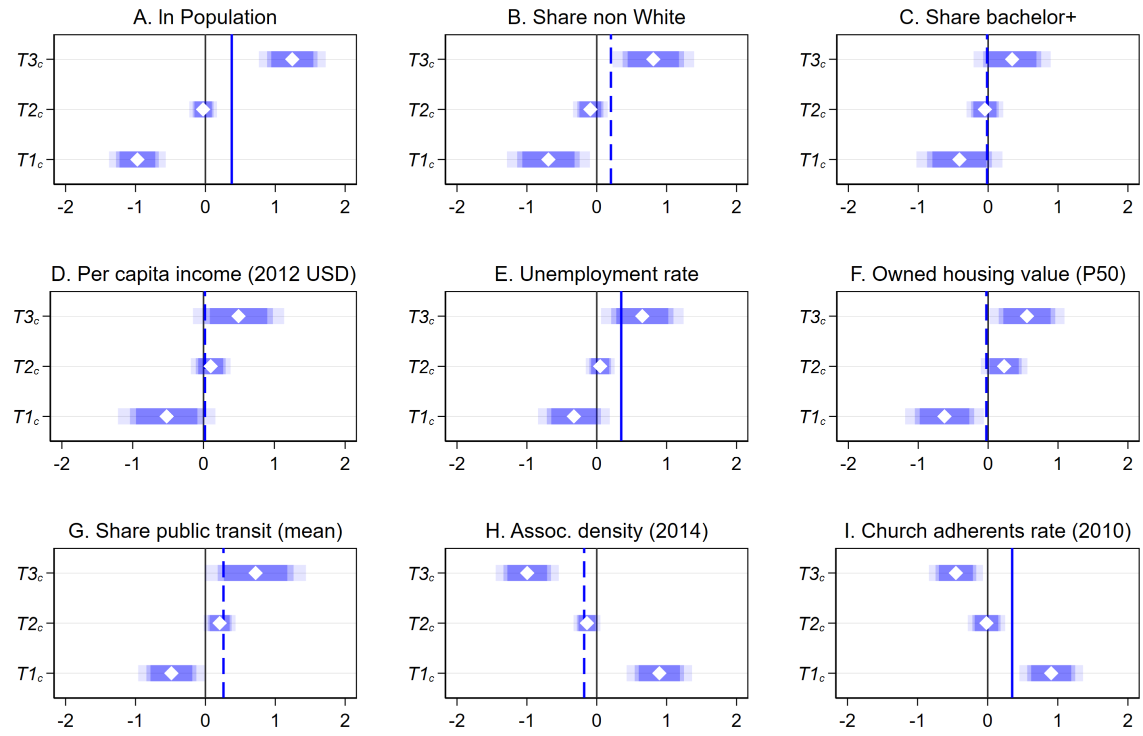


FIGURE 4 – Coefficient plots for correlates of MSA social isolation. For each characteristic, markers denote results from fitting the model in Equation (9b). Robust confidence intervals at the 90, 95, and 99 percent levels are displayed in progressively lighter shades. The blue reference line shows the 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.

What emerges from this figure is that the headline association of $SISO_c$ with urban features (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 and Glaeser (1997) and Cutler et al. (1999), for instance, social isolation is greater in larger cities and in cities with higher unemployment, 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. To a lesser extent, similar considerations apply to education, income, and housing costs (median value of owned housing). These findings are consistent with an emerging literature on urban revival and gentrification driven by college graduates (Couture and Handbury, 2020, 2023), often interacting with a racial dimension (Baum-Snow and Hartley, 2020).

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, but this relationship is complex and can be very heterogeneous (e.g. Putnam, 2000, 2007; Athey et al., 2021).[23] Again, other-area interactions and own-area composition have divergent estimates, with the former displaying a 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, and with a sustained trend of racial integration in US congregations (Dougherty et al., 2020). The patterns we detect suggest that civic and religious institutions facilitate cross-neighbourhood social ties that reduce isolation at city level. However, the mechanisms underlying this association — particularly whether institutions draw from diverse neighbourhoods and whether they mirror, bridge, or even reinforce racial boundaries in the local environment — are complex and likely context-dependent (e.g., see Emerson and Kim, 2003; Dougherty and Huyser, 2008; Dougherty and Emerson, 2018, on congregations).[24] This relationship, as well as the role of public transport, may warrant future investigation in dedicated studies.

Appendix Figure A.17 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.18, we confirm that population size is a key determinant of these changes, as including this control is by itself

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

[24]We tentatively explore the relevance of neighbourhood context in this relationship using ZIP-level data in Appendix Figure A.16. We uncover patterns consistent with bridging, rather than boundary reinforcement, but also point to possible selection bias in the types of institutions driving this association.

enough to substantially alter findings. Results for church adherence, however, remain unchanged.[25] We also 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.[26] Finally, Appendix Figure A.19 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. (2022a,b). Reassuringly, with the exception of economic connectedness (EC), results align with our findings on associational density and church adherence.

4.3 Social Isolation by Racial or Ethnic Group

So far, we grouped together minorities belonging to different racial or ethnic groups into one unique non-White category, which we contrasted to Whites. However, this simplification could mask important heterogeneity in social isolation across individuals belonging to different minorities. To explore variation across groups, we construct equivalent $SISO_c$ measures to that defined in Equation (7), including the decomposition, where we let x_i and X_c be the group-specific population counts in each neighbourhood i and city c respectively.[27] We consider three different groups: Black, Hispanic, and Asian residents.

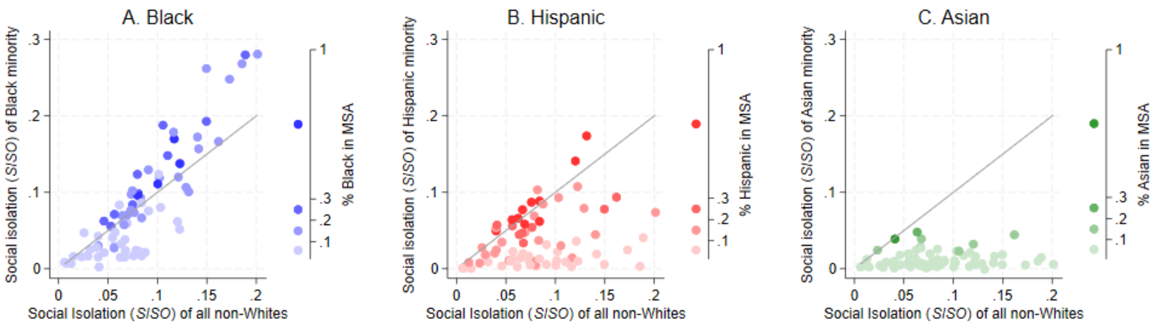


FIGURE 5 – Scatter plots of $SISO_c$ measured for specific racial or ethnic minority groups against the equivalent $SISO_c$ measured for all minorities combined. The grey line denotes equivalence between the two measures. Markers are shaded proportionally to the share of the MSA population that belongs to the particular minority considered in each panel.

[25] 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.

[26] 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).

[27] All other terms are unchanged. Specifically, y_i and Y_c are counts of Whites, and t_i is the total population in each neighbourhood.

Figure 5 compares group-specific social isolation (from the White majority) against the social isolation of all non-White minorities in each city. A few results stand out. First, in many cities Black residents are more socially isolated than the average non-White minority (panel A). By contrast, Hispanic and Asian residents tend to be less socially isolated — particularly the latter — suggesting that aggregate results reflected largely the isolation of Black residents. Second, minorities are more socially isolated in cities where they represents a larger proportion of the overall population, mirroring results from the residential segregation literature (e.g., Cutler et al., 1999; Logan et al., 2004). This suggests that larger minority groups at the city level is associated with more homophilistic friending preferences or greater constraints to cross-racial friendships. Third, Hispanic Americans show the largest variation in group-specific social isolation relative to non-Whites in general, likely due to the inherent heterogeneity of this ethnic group, which includes both Black and White people, among other smaller minorities. Analogous scatter plots for each index subcomponent are available in the Appendix (Figures A.20, A.21, and A.22), each pointing to very similar findings.[28]

Next, we examine how the decomposition of social isolation in each sub-term varies by minority group. In particular, we compare the first and the third index-components ($T1_c$ and $T3_c$), respectively capturing own-area residential isolation and other-area social isolation, since the second component ($T2_c$) for own-area social isolation contributes little to the overall index in relative terms (see Appendix Figure A.23, which summarises the relative incidence of each component by group). The scatter plot in Figure 6 shows the ratio of $T3_c$ over $T1_c$ in each city for different racial and ethnic groups against that city's group-specific population share (note that markers are MSA-group specific). This plot is informative about how differences in social and residential isolation vary by demographic group, and what role the relative size of this group in each city plays in these differences.

Several notable patterns emerge. First, other-area social isolation ($T3_c$) plays a more important role in overall social isolation than own-area residential isolation ($T1_c$) for nearly all cities and demographic groups, as revealed by ratios mostly greater than one.[29] Second, we observe variation in the $T3_c$ to $T1_c$ ratio between different groups in cities with comparable demographic composition. In particular, relative to own-area residential segregation, other-area social isolation is especially pronounced among Black Americans.

[28]Something worth highlighting from Appendix Figure A.21, which cannot be evinced from the one reproduced here, is the striking homogeneity of the own-area contribution to social isolation ($T2_c$) among Asian Americans. This term is narrowly clustered around zero, with little cross-city variation. This suggests that for this group local social interactions tend to be uncorrelated with the composition of one's residential area.

[29]Note that this is not to say that residential isolation per se is lower than social isolation overall. In fact, as discussed in Section 4.1, it is the opposite ($RISO_c > SISO_c$). Instead this results suggests that generally $T3_c > T1_c$. Both statements can be true for a relatively low ω_{ii} and relatively negative $T2_c$.

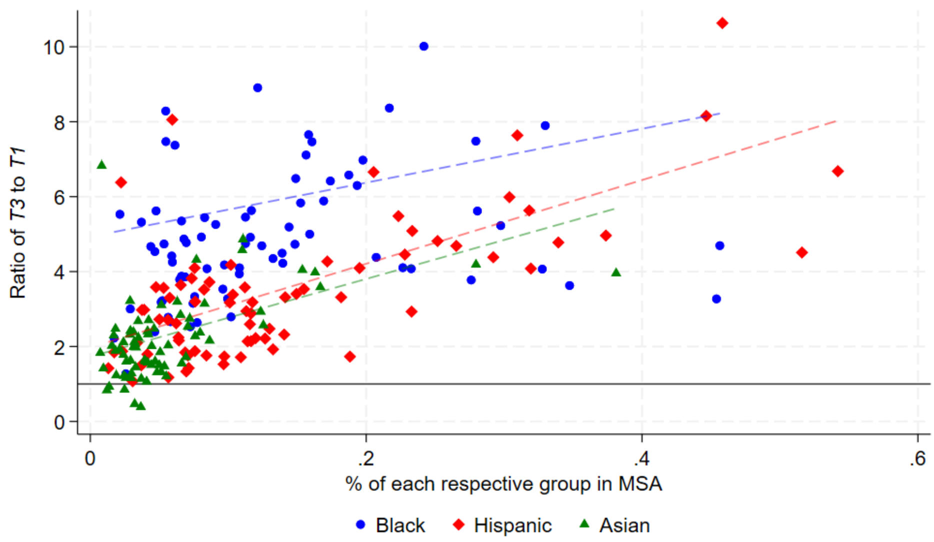


FIGURE 6 – Contribution of cross-boundary social isolation relative to residential isolation by racial or ethnic group. The figure plots the ratio of $T3_c$ over $T1_c$ for specific racial or ethnic groups against the share of each respective group in the MSA population. Each marker represents an MSA-group combination. The population-weighted linear fit for each group is overlaid using dashed lines.

This pattern demonstrates that the differences in Figure 5 are not simply due to differences in residential isolation across groups, but other-area social connections are particularly important for Black isolation. Third, pooling across all racial and ethnic groups, there is an overall positive relationship between city-level composition and the relevance of social vs. residential isolation. As Black, Hispanic, and Asian minorities become more concentrated across cities, the ratio of $T3_c$ to $T1_c$ increases, with the other-area social component becoming up to one-order of magnitude larger than the own-area residential one in MSAs with minority shares between 30 and 50 percent of the urban population. The strength of this association is largely comparable across minority groups but appears to taper off at very high levels of concentration. In short, we conclude that city-wide demographics mediate the relative importance of social vs. residential isolation across all cities and groups, but that group-specific considerations also apply for comparable levels of demographic concentration.

Finally, we turn to other factors that are associated with greater social isolation of particular minorities across cities, replicating the analysis in Section 4.2 for our group-specific $SISO_c$ measures and decompositions. We report results graphically in Appendix Figure A.24. Results for Black Americans are by and large comparable to previous estimates. Some heterogeneity emerges for Hispanic and Asian minorities. Both groups show a positive relationship with the overall non-White share in the city and the median value of owned housing, and a negative one with associational density. This suggests these groups are more isolated in more diverse cities with less affordable housing, but less isolated in cities with more vibrant community life. In all these instances, differences are driven primarily by the third component, that is, by social isolation with respect to cross-boundary interactions. Contrary to the Black group, Hispanic and Asian minorities also display no association between social isolation and church adherence overall, although we confirm the negative and statistically significant relationship with the third component for all three minority groups, pointing once again at the role of community organisations. For Asians in particular, we also observe positive headline associations with educational attainment, income, and public transit, suggesting that this group is more socially isolated in more affluent, educated, and well-connected urban areas. Like before, however, we caution from interpreting any of these descriptive relationships in a causal sense.

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 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 uncover substantial heterogeneity in social isolation and its underlying components across different racial and ethnic groups, documenting how city-wide demographics matter for this relationship.

Our findings have implications for the design and evaluation of policies aimed at reducing segregation. Conventional residential segregation measures capture the *potential* for intergroup contact based on spatial proximity, while our social segregation measures capture *realized* interaction patterns, including connections that extend beyond neighbourhood boundaries. The systematic gap between these two measures indicates that residential proximity alone does not fully determine social interaction. This has several policy implications. First, while residential measures remain essential for understanding proximity-based outcomes such as access to services, school quality, and commuting patterns, they overstate the social isolation individuals actually experience. Second, policies aimed solely at residential desegregation may achieve spatial mixing without fostering intergroup contact, particularly when social boundaries remain strong across neighbourhoods. Third, our findings suggest that policymakers might consider targeting the social interaction margin more directly rather than focusing exclusively on residential sorting. Interventions that facilitate cross-group connections — such as improved public transit linking diverse neighbourhoods, support for civic organizations that draw from multiple areas, or programs that foster bridging social capital — could prove more cost-effective and less disruptive than policies that physically relocate households. Our finding that institutional density is associated with lower social isolation provides suggestive evidence for this approach. Finally, the two measures offer complementary guidance for policy design: residential measures identify *where* intervention is needed and matter most when proximity-based outcomes are the concern, while social measures reveal *whether* interventions foster actual intergroup contact when interaction-related outcomes, such as labour market networks, information flows, or social cohesion, are the policy goal.

While our analysis reveals systematic patterns in how social ties matter for segregation, several important limitations qualify both our descriptive findings and the policy insights that rest upon them. 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.

Ahlfeldt, G. M., Redding, S. J., Sturm, D. M. and Wolf, N. (2015), 'The Economics of Density: Evidence From the Berlin Wall', *Econometrica* **83**(6), 2127–2189.

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.

Auxier, B. and Anderson, M. (2021), Social Media Use in 2021, Technical report, Pew Research Center.

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.

Baum-Snow, N. and Hartley, D. (2020), 'Accounting for central neighborhood change, 1980–2010', *Journal of Urban Economics* **117**, 103228.

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.

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.

Couture, V., Dingel, J. I., Green, A. E. and Handbury, J. (2025), Demographic Preferences and Income Segregation, Working Paper 33386, National Bureau of Economic Research.

Couture, V. and Handbury, J. (2020), 'Urban revival in America', *Journal of Urban Economics* **119**, 103267.

Couture, V. and Handbury, J. (2023), 'Neighborhood Change, Gentrification, and the Urbanization of College Graduates', *Journal of Economic Perspectives* **37**(2), 29–52.

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.

D'Haultfœuille, X. and Rathelot, R. (2017), 'Measuring segregation on small units: A partial identification analysis', *Quantitative Economics* **8**(1), 39–73.

Diemer, A. (2026), 'Spatial diffusion of local economic shocks in social networks: Evidence from the US fracking boom', *Journal of Labor Economics* **44**(1), 00–00.

Donaldson, D. and Hornbeck, R. (2016), 'Railroads and American Economic Growth: A "Market Access" Approach', *The Quarterly Journal of Economics* **131**(2), 799–858.

Dougherty, K. D., Chaves, M. and Emerson, M. O. (2020), 'Racial Diversity in U.S. Congregations, 1998–2019', *Journal for the Scientific Study of Religion* **59**(4), 651–662.

Dougherty, K. D. and Emerson, M. O. (2018), 'The Changing Complexion of American Congregations', *Journal for the Scientific Study of Religion* **57**(1), 24–38.

Dougherty, K. D. and Huyser, K. R. (2008), 'Racially Diverse Congregations: Organizational Identity and the Accommodation of Differences', *Journal for the Scientific Study of Religion* **47**(1), 23–44.

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

Emerson, M. O. and Kim, K. C. (2003), 'Multiracial Congregations: An Analysis of Their Development and a Typology', *Journal for the Scientific Study of Religion* **42**(2), 217–227.

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.

Gilbert, E. and Karahalios, K. (2009), Predicting tie strength with social media, *in* 'Conference on Human Factors in Computing Systems - Proceedings', ACM Press, New York, New York, USA, pp. 211–220.

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

Granovetter, M. (1973), 'The Strength of Weak Ties', *The American Journal of Sociology* **78**(6), 1360–1380.

Hampton, K. N., Goulet, L. S., Rainie, L. and Purcell, K. (2011), Social Networking Sites and Our Lives, Technical report, Pew Research Center, Washington, DC.

Jones, J. J., Settle, J. E., Bond, R. M., Fariss, C. J., Marlow, C. and Fowler, J. H. (2013), 'Inferring Tie Strength from Online Directed Behavior', *PLOS ONE* **8**(1), 1–6.

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.

Logan, J. R., Stults, B. J. and Farley, R. (2004), 'Segregation of minorities in the metropolis: Two decades of change', *Demography* **41**(1), 1–22.

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.

Mast, E. and Barca, A. (2026), Off the Beaten Tract: Constructing a New Neighborhood Geography Using Revealed Preference. Unpublished manuscript.

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.

Meta Platforms, Inc. (2021), Form 10-k: Annual report for the fiscal year ended december 31, 2020, Technical report, U.S. Securities and Exchange Commission.

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.

Rathelot, R. (2012), 'Measuring Segregation When Units are Small: A Parametric Approach', *Journal of Business & Economic Statistics* **30**(4), 546–553.

Redding, S. J. (2023), 'Quantitative Urban Models: From Theory to Data', *Journal of Economic Perspectives* **37**(2), 75–98.

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.

Topa, G. and Zenou, Y. (2015), Neighborhood and Network Effects, *in* G. Duranton, J. V. Henderson and W. C. Strange, eds, 'Handbook of Regional and Urban Economics', Vol. 5, Elsevier, Amsterdam, NL, pp. 561–624.

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.

Whaley, K. (2024), 'Residential Segregation at Physical Neighborhood Boundaries', *Journal of Economics, Race, and Policy* **7**(3), 141–153.

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

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

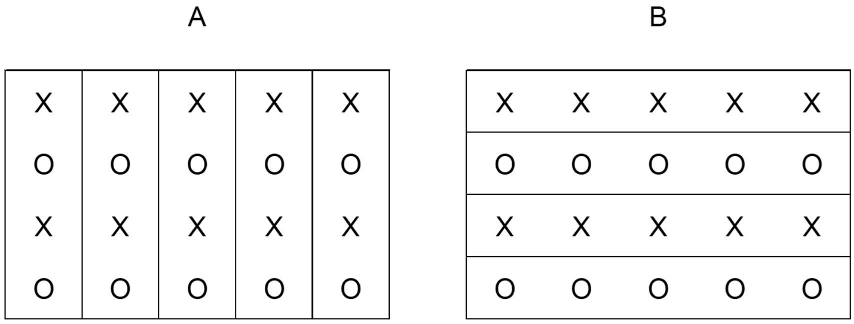


Figure A.1 – Residential segregation in a hypothetical city with different boundaries drawn (Echenique and Fryer, 2007). In the scenario in panel A, boundaries are drawn so that there is perfect integration between groups based on both the dissimilarity and isolation indices. In panel B, instead, boundaries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

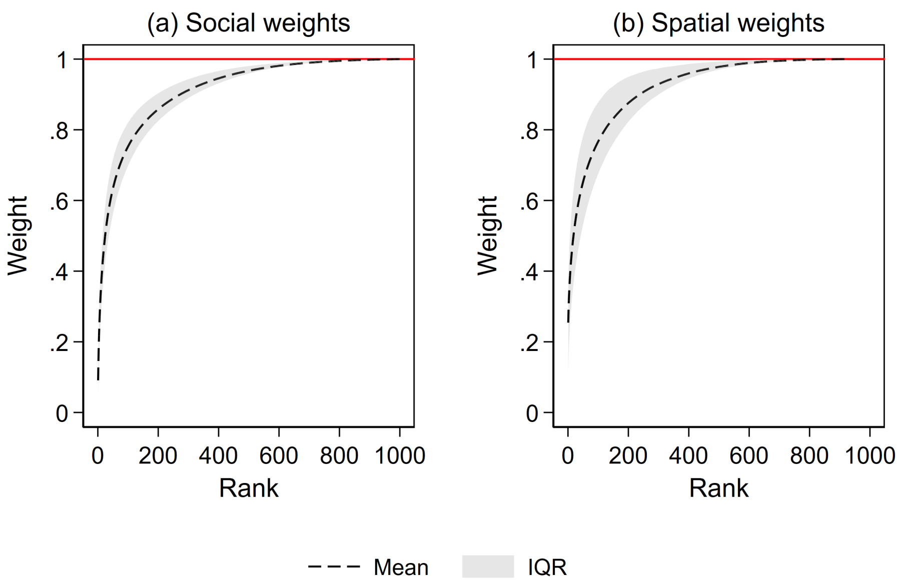


Figure A.2 – The figure shows averages and inter-quartile ranges (IQR) of the cumulative sum of social and spatial weights over their ranking. Observations are restricted to linkages for 9,747 neighbourhoods in 75 urban areas (MSAs) encompassing at least 50 ZCTAs. For each of these ZIPs, the top 1,000 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 160 miles from each other (approximately 260 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).

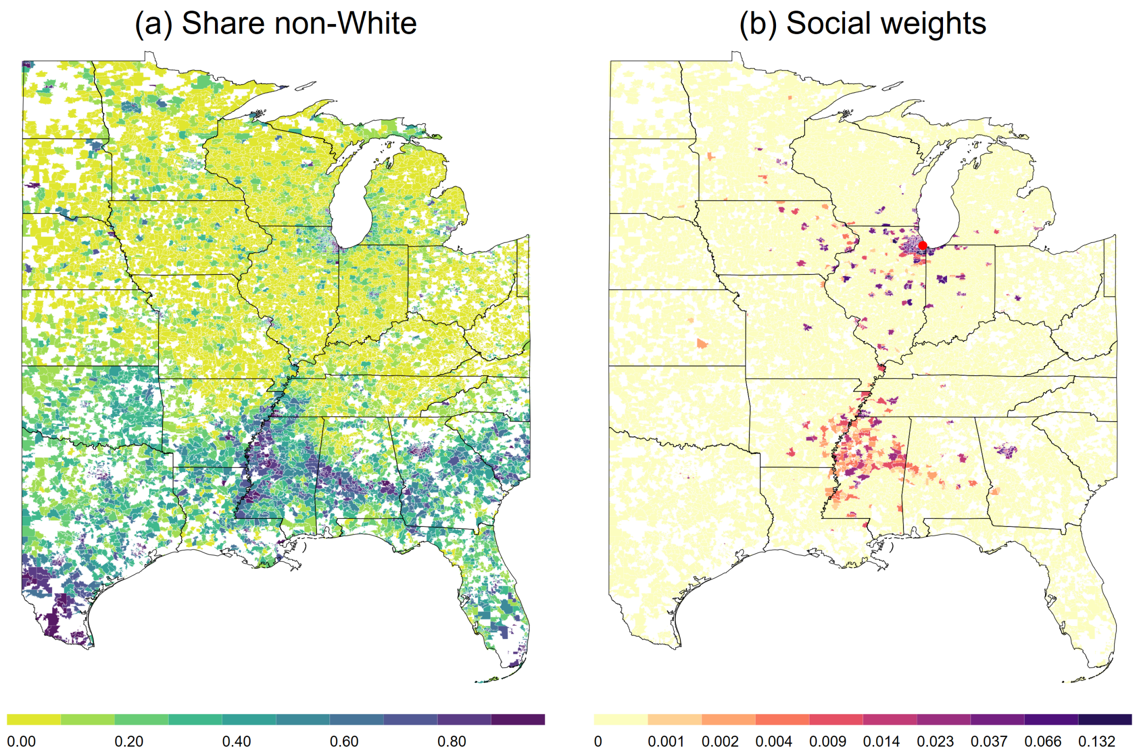


FIGURE A.3 – Race composition in ZCTAs across the US (panel a) and weights for ZCTA no. 60620 (99.48% non-White), denoted with a red dot on the map (panel b). Social weights in panel (b) are multiplied by 100 for legibility. Breaks are defined at each decile of the distribution obtained by pooling both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

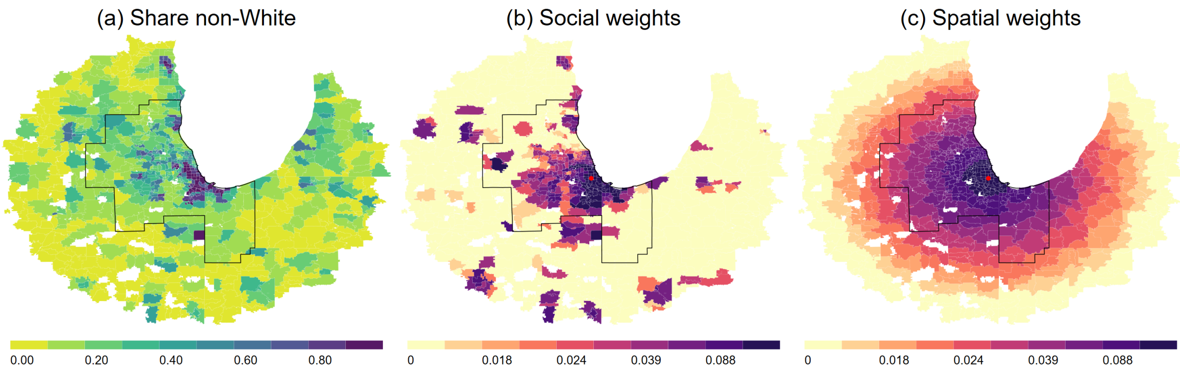


FIGURE A.4 – Maps with MSA race composition (a), social (b) and spatial (c) weights for the least White ZCTA in Chicago-Joliet-Naperville, ZCTA no. 60620, 99.48% non-White. (marked in red on the map). Weights are not adjusted for relative population size, and are rescaled by a factor of 100 for legibility.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

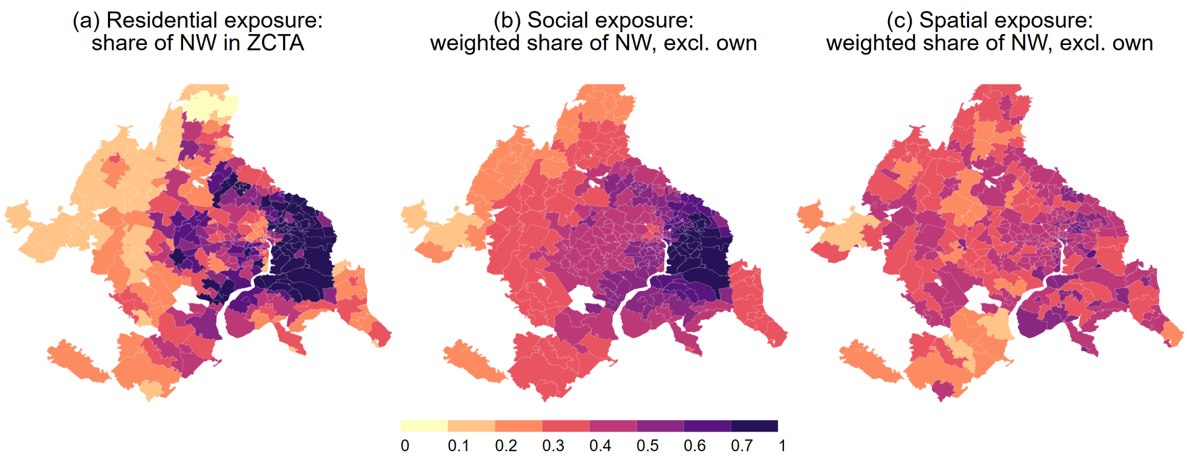


FIGURE A.5 – Residential, social, and spatial exposures in Washington-Arlington-Alexandria. The maps show within-MSA variation in minority exposure measures for the city's neighbourhoods. For reference, MSA-level segregation is as follows: $RISO$=0.205, $SISO^s$=0.106, $zRISO$=0.288, $zSISO^s$=0.545,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

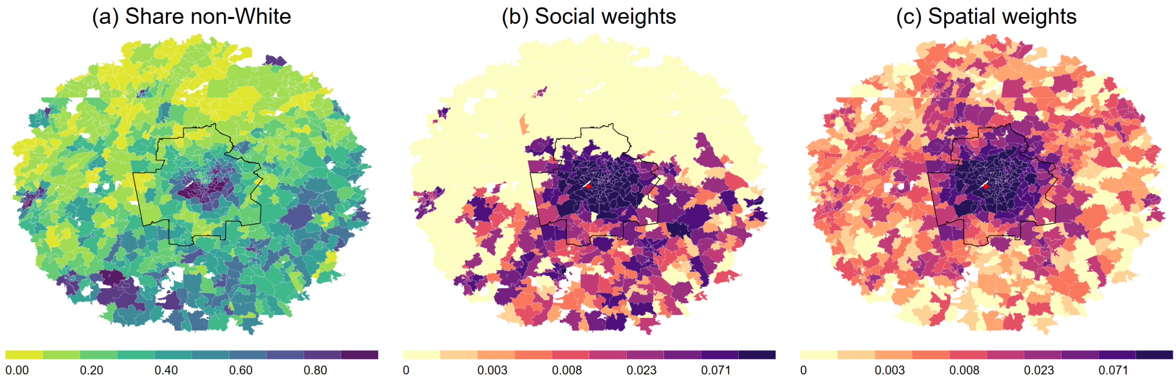


FIGURE A.6 – Maps with MSA race composition (a), social (b) and spatial (c) weights for the least White ZCTA in Atlanta-Sandy Springs-Marietta, ZCTA no. 30331, 98.70% non-White. (marked in red on the map). Weights are not adjusted for relative population size, and are rescaled by a factor of 100 for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

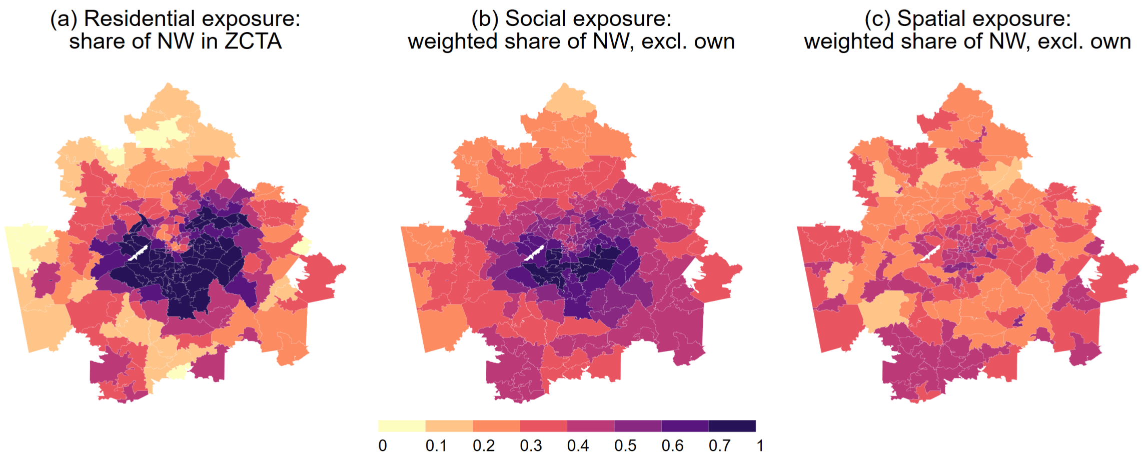


FIGURE A.7 – Residential, social, and spatial exposures in Atlanta-Sandy Springs-Marietta. The maps show within-MSA variation in minority exposure measures for the city's neighbourhoods. For reference, MSA-level segregation is as follows: $RISO$=0.234, $SISO^s$=0.117, $zRISO$=0.691, $zSISO^s$=0.809,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

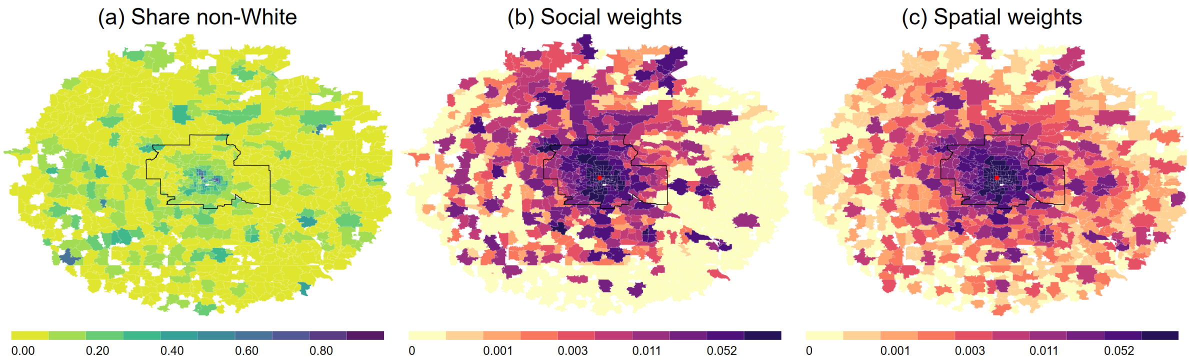


FIGURE A.8 – Maps with MSA race composition (a), social (b) and spatial (c) weights for the least White ZCTA in Minneapolis-St. Paul-Bloomington, ZCTA no. 55411, 83.23% non-White. (marked in red on the map). Weights are not adjusted for relative population size, and are rescaled by a factor of 100 for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

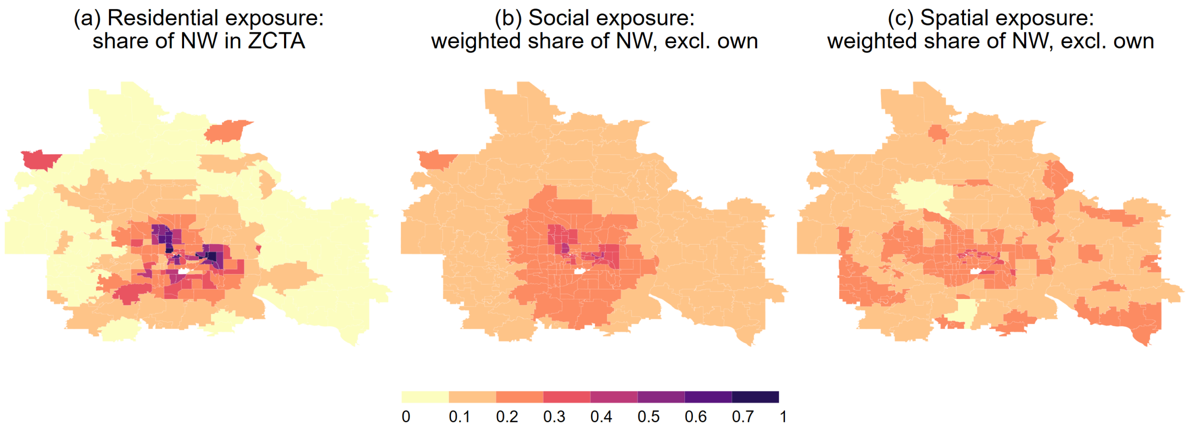


FIGURE A.9 – Residential, social, and spatial exposures in Minneapolis-St. Paul-Bloomington. The maps show within-MSA variation in minority exposure measures for the city's neighbourhoods. For reference, MSA-level segregation is as follows: $RISO$=0.146, $SISO^s$=0.055, $zRISO$=-0.532, $zSISO^s$=-0.638,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

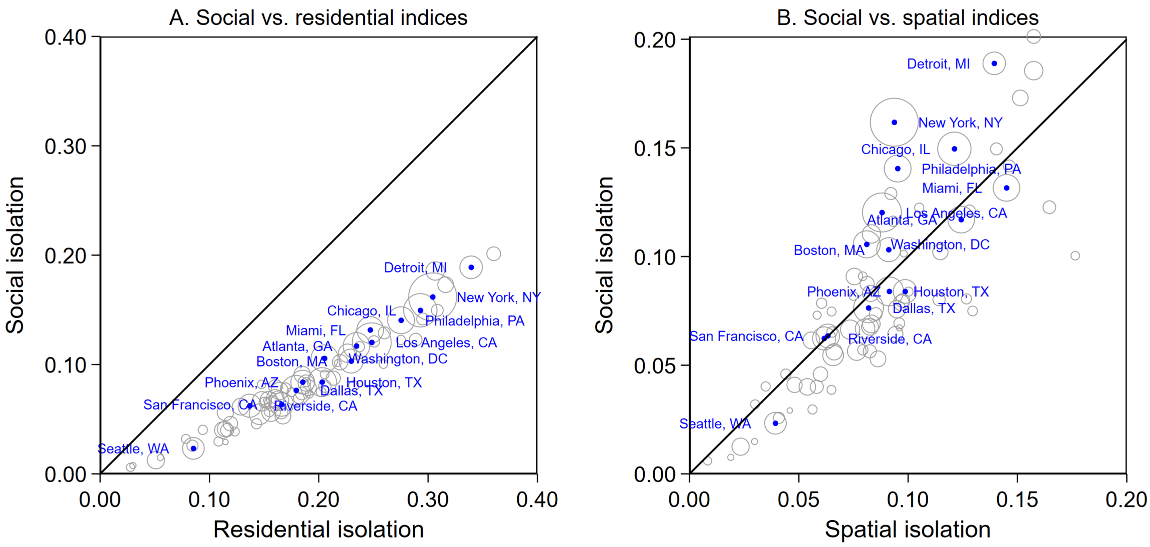


FIGURE A.10 – Scatter plots comparing social isolation with its residential or spatial counterparts, in panels A and B respectively. Only MSAs with at least 50 ZCTAs were retained (N=75). Marker sizes in grey are proportional to each city's population. The largest 15 cities in the sample are also marked 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.

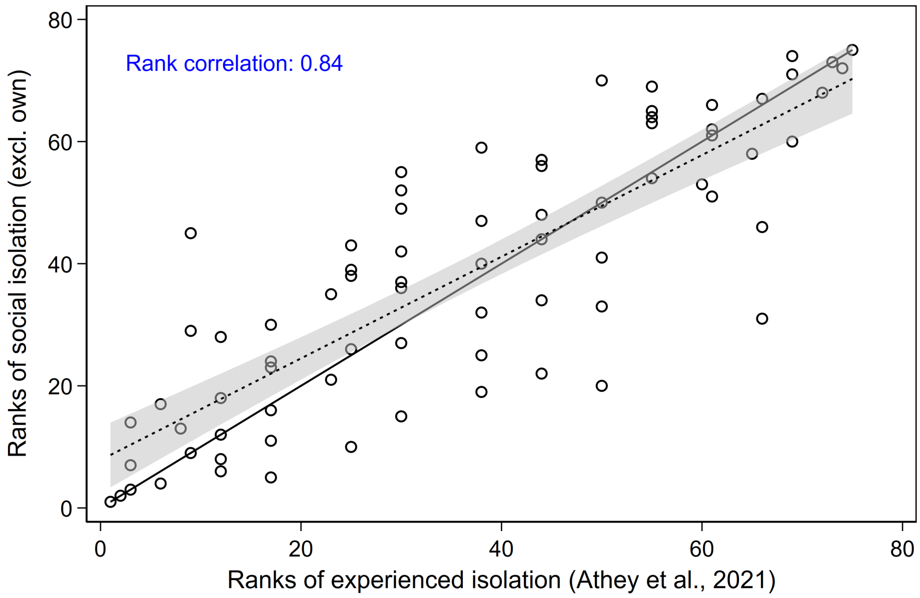


Figure A.11 – Spearman's rank correlation of social isolation with Athey et al. (2021)'s experienced isolation. The comparison is restricted to cities included in our sample only (i.e., MSAs with at least 50 ZCTAs within their boundaries). The 45 degree line is in solid black. The dotted line gives the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

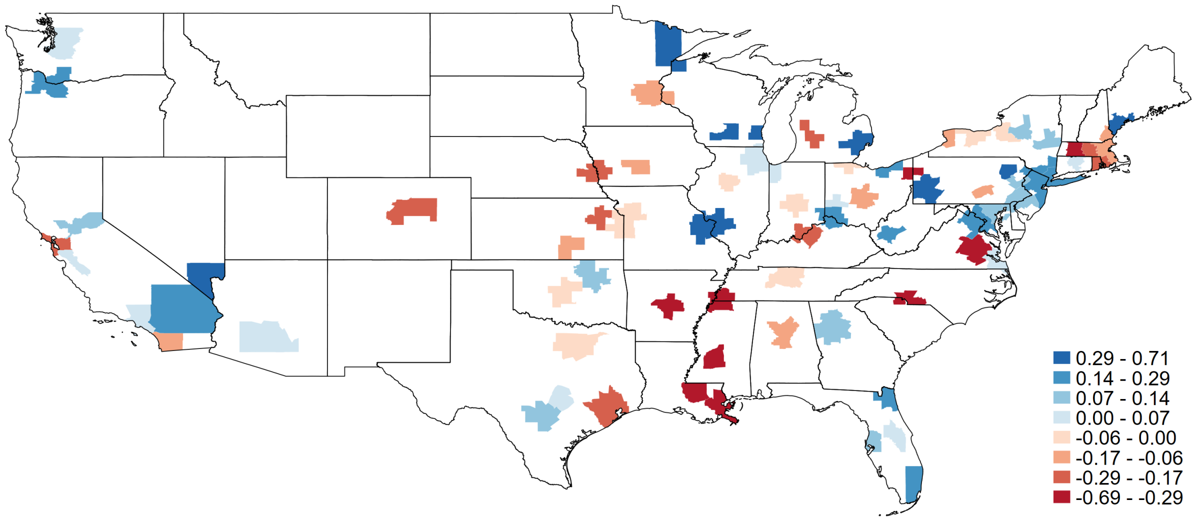


Figure A.12 – Differences in z-scores of social and residential isolation. Positive values entail that the MSA is relatively more socially than residentially segregated compared to the average city in our sample. Only MSAs with at least 50 ZCTAs are retained.

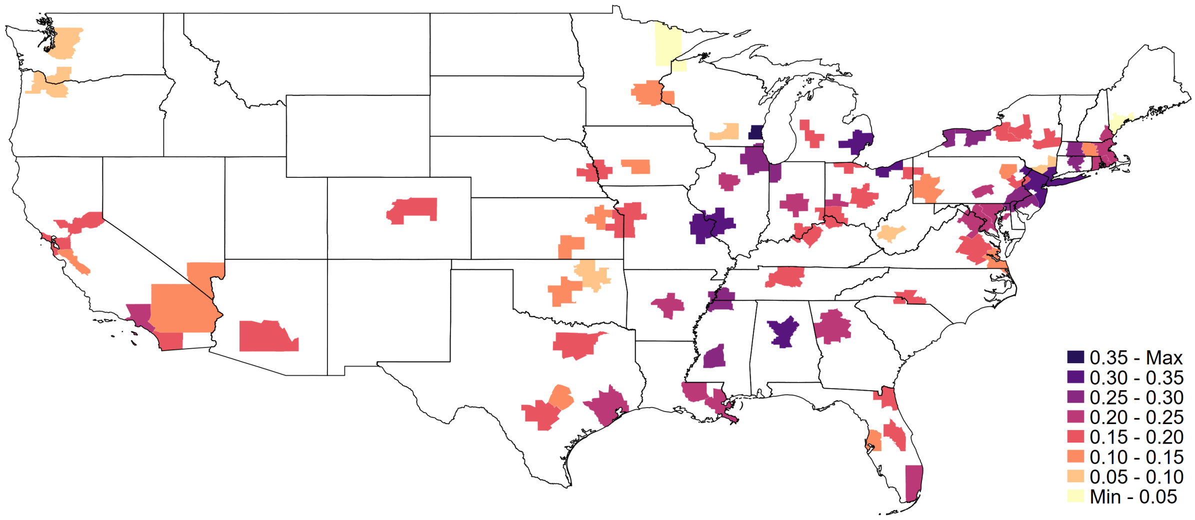


FIGURE A.13 – Residential isolation in the largest US cities. Legend categories are consistent with those in Figure A.14 to facilitate comparison.

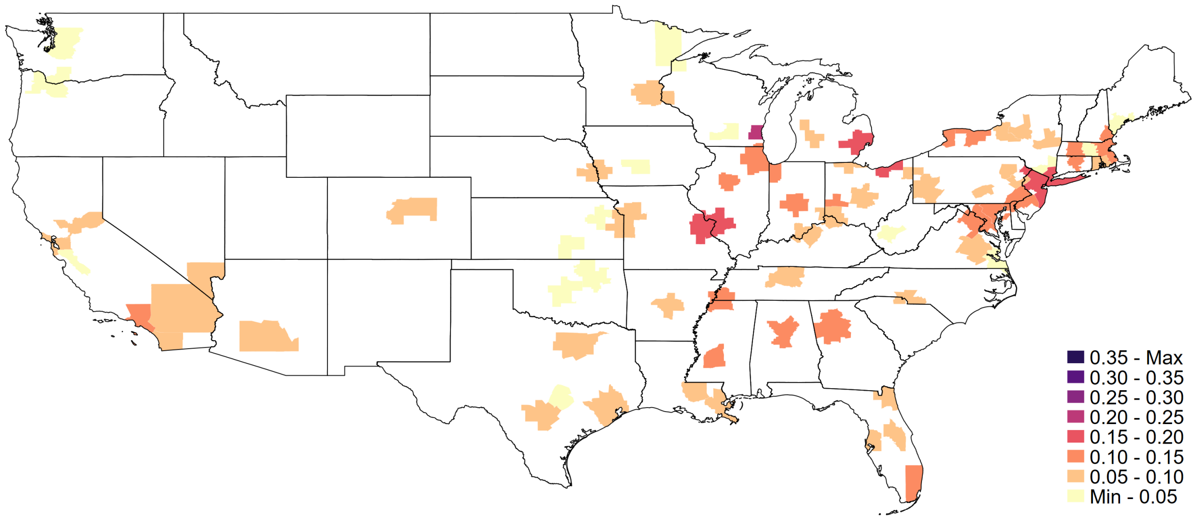


FIGURE A.14 – Social isolation in largest US cities, excluding own-area interactions. Legend categories are consistent with those used in Figure A.13 to facilitate comparison.

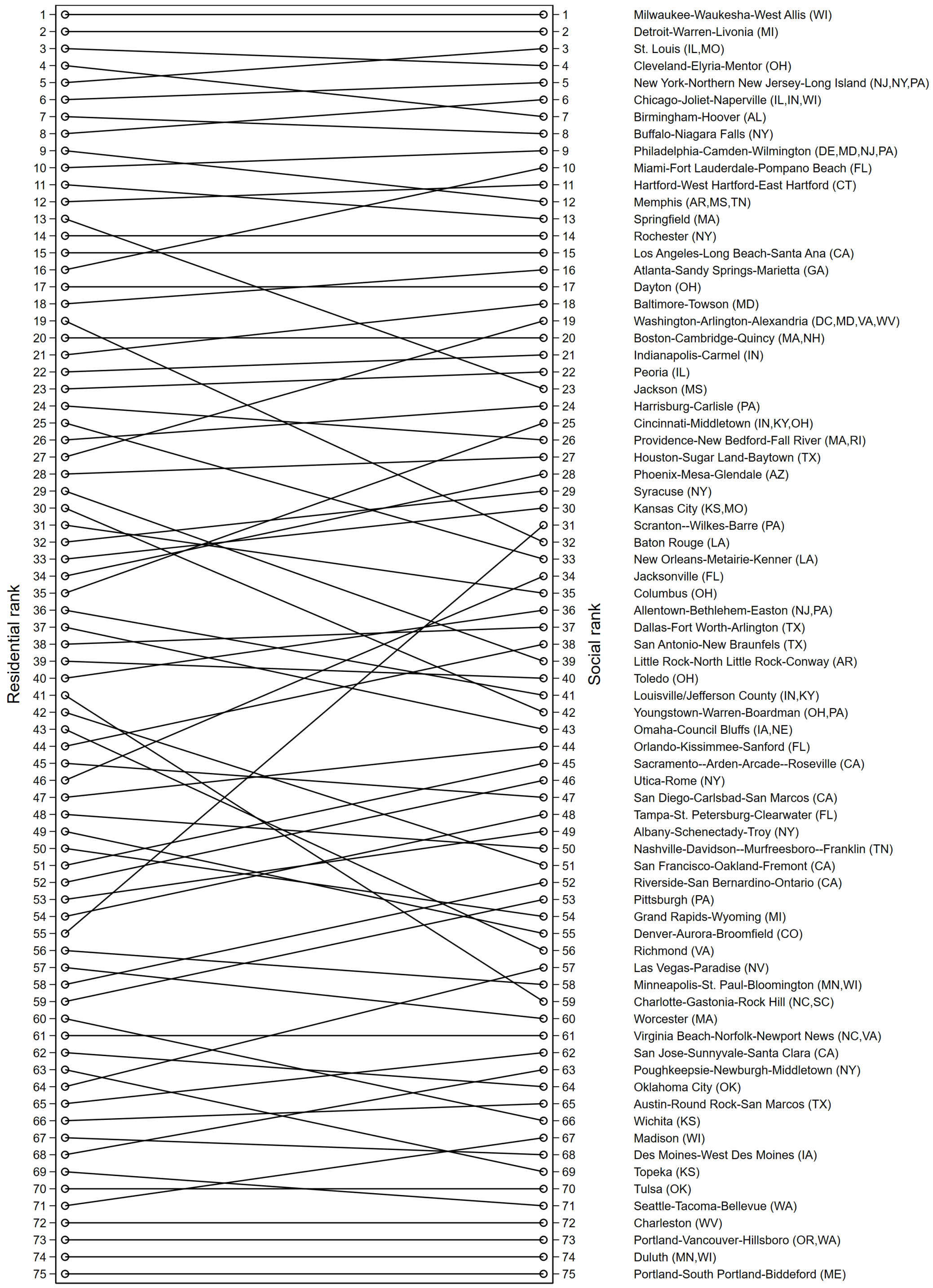


Figure A.15 – Comparison of MSA ranks defined in terms of residential segregation ($RISO_c$, on the left-hand side) with ranks defined in terms of social segregation ($SISO_c$, on the right-hand). MSAs are labeled according to their social segregation rank, so the figure is best read right-to-left.

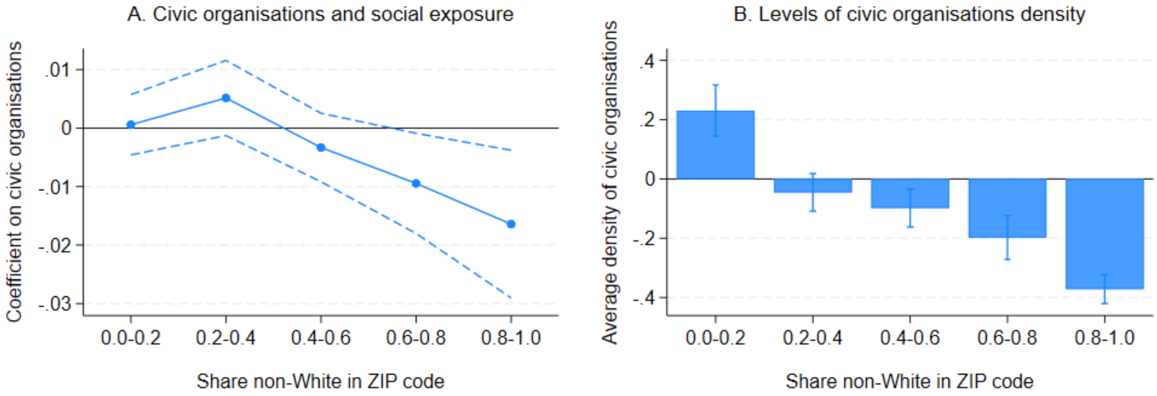


FIGURE A.16 – ZCTA-level association of civic engagement with social exposure (panel A) and average civic engagement (panel B) by share of non-White ZCTA residents. Civic engagement is obtained from Chetty et al. (2022*a*,*b*) and expressed in terms of z-scores. It is defined as the rate of volunteering and participation in community organizations. Panel A plots coefficients of a regression of other-area social exposure to non-Whites on civic engagement, interacted with ZCTA composition dummies and separately controlling for main effects of these dummies. Panel B plots the average levels of civic engagement for each ZCTA composition category. Standard errors at the 95% level, clustered by MSA, are reported using band and whiskers. Panel A shows a weakly positive (albeit not statistically significant) association between organizational density and cross-area social exposure to non-Whites in predominantly White ZIPs, whereas the association reverses and becomes negative and statistically significant in mostly non-White areas. This pattern is consistent with a 'bridging' interpretation of community organisations in both moderately White and largely non-White areas — though especially so in the latter case. At the same time, Panel B reveals that organizational density is substantially lower in mostly non-White ZIPs, raising the possibility that these organizations are selected to serve specific bridging functions rather than representing typical institutions.

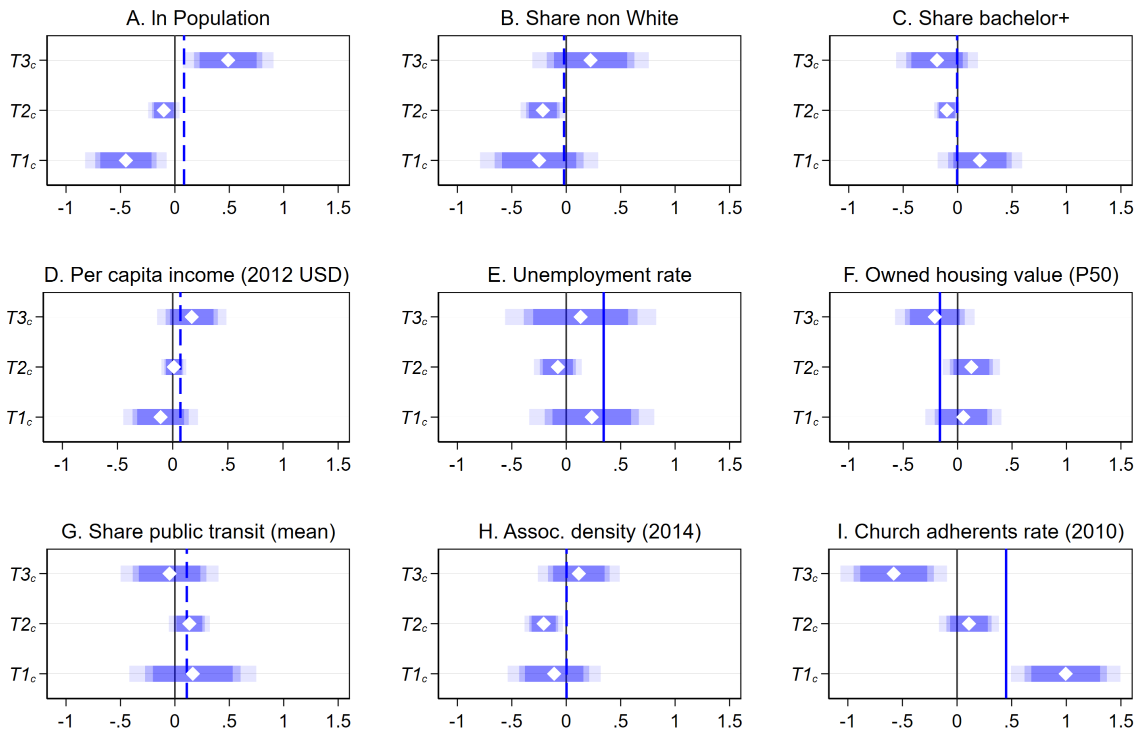


Figure A.17 – Coefficient plots for correlates of MSA social isolation. For each characteristic, markers denote results from fitting the model in Equation (9b), additionally controlling for all other characteristics. Robust confidence intervals at the 90, 95, and 99 percent levels are displayed in progressively 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.

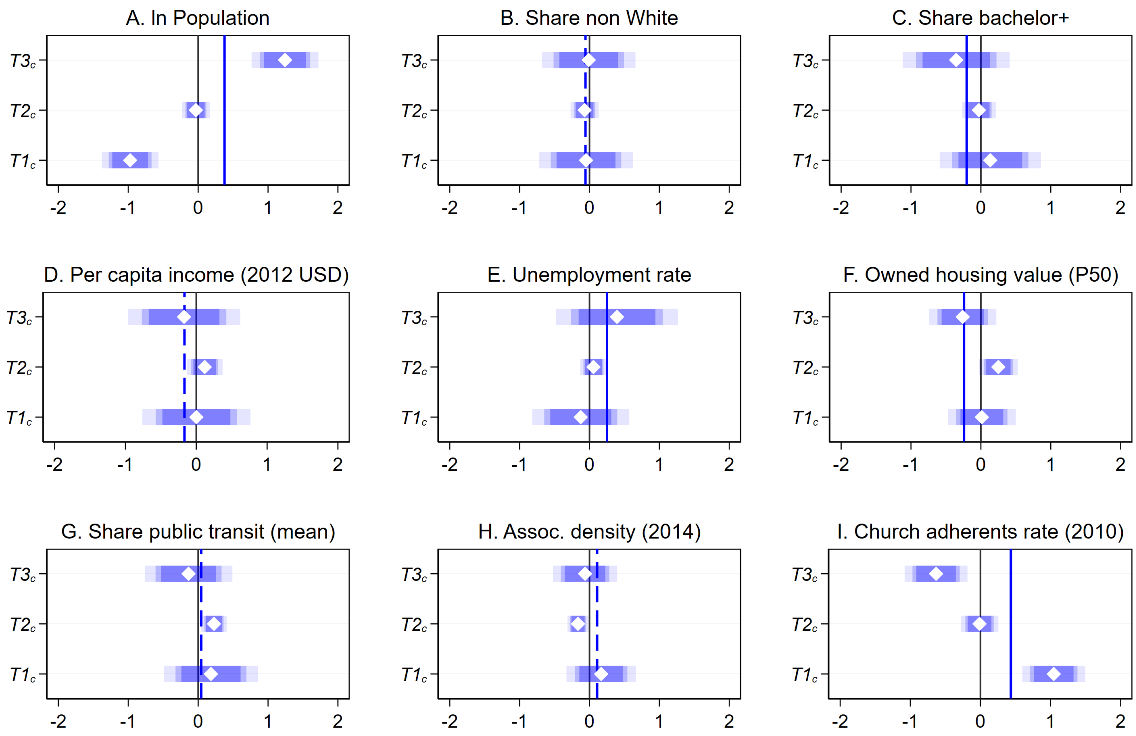


FIGURE A.18 – Coefficient plots for correlates of MSA social isolation. For each characteristic, markers denote results from fitting the model in Equation (9b), additionally controlling for the log of population size (except for panel A). Robust confidence intervals at the 90, 95, and 99 percent levels 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.

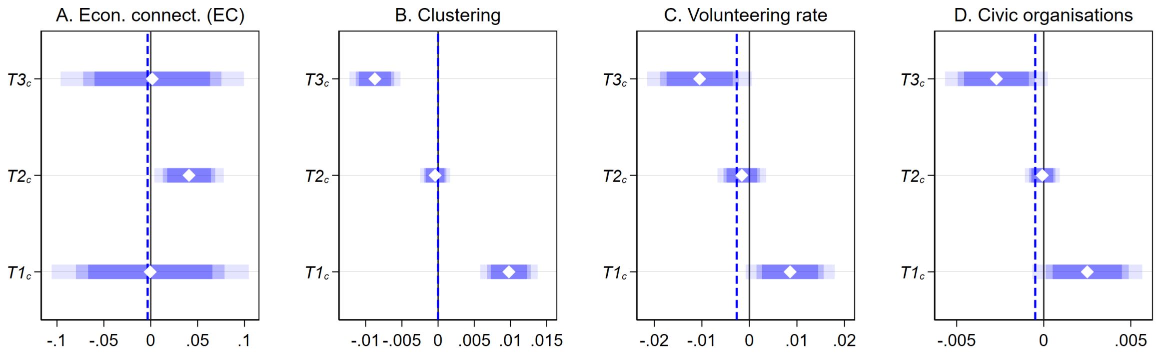


Figure A.19 – Coefficient plots for the association of isolation indices and its constituents (MSA-level) with social capital measures in Chetty et al. (2022*a*,*b*). For each characteristic, markers denote results from fitting the model in Equation (9b). Outcome data varies at the ZCTA level, but social 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.

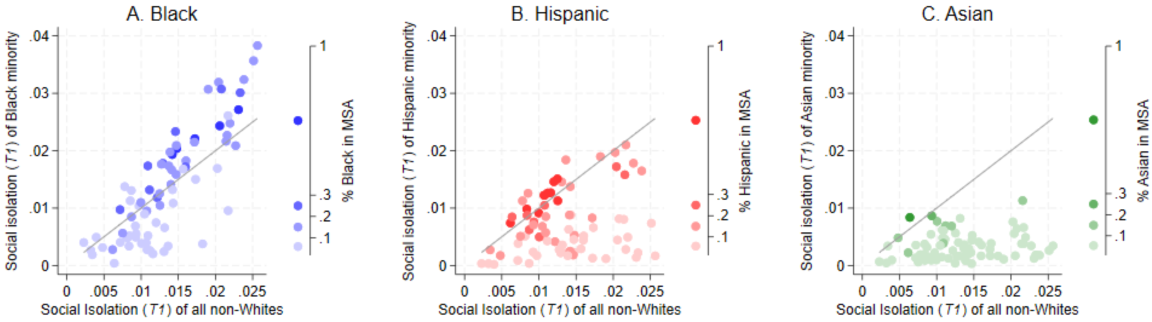


FIGURE A.20 – Scatter plots of component $T1_c$ of $SISO_c$ measured for specific racial or ethnic minority groups against the equivalent $T1_c$ measured for all minorities combined. The grey line denotes equivalence between the two measures. Markers are shaded proportionally to the share of the MSA population that belongs to the particular minority considered in each panel.

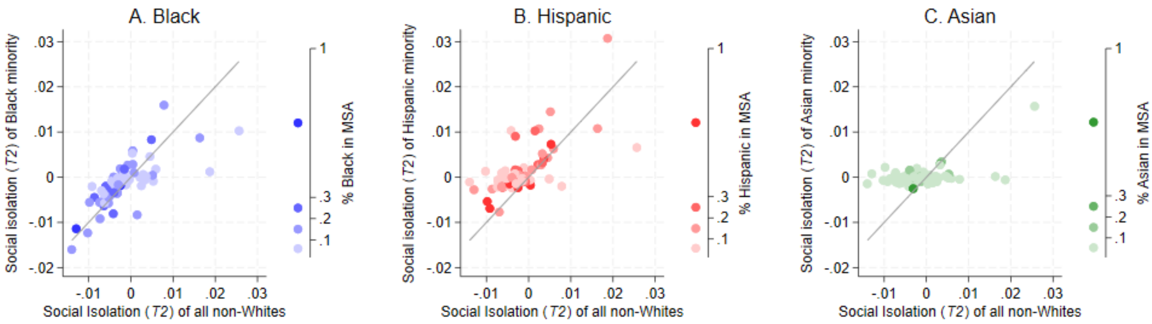


FIGURE A.21 – Scatter plots of component $T2_c$ of $SISO_c$ measured for specific racial or ethnic minority groups against the equivalent $T2_c$ measured for all minorities combined. The grey line denotes equivalence between the two measures. Markers are shaded proportionally to the share of the MSA population that belongs to the particular minority considered in each panel.

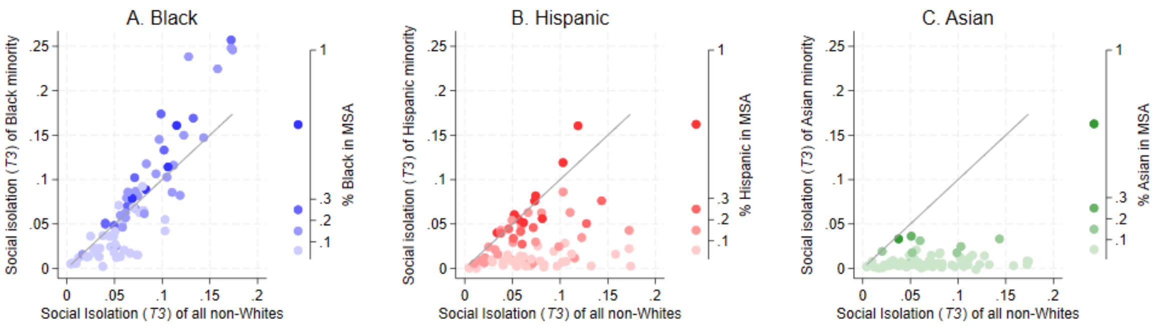


FIGURE A.22 – Scatter plots of component $T3_c$ of $SISO_c$ measured for specific racial or ethnic minority groups against the equivalent $T3_c$ measured for all minorities combined. The grey line denotes equivalence between the two measures. Markers are shaded proportionally to the share of the MSA population that belongs to the particular minority considered in each panel.

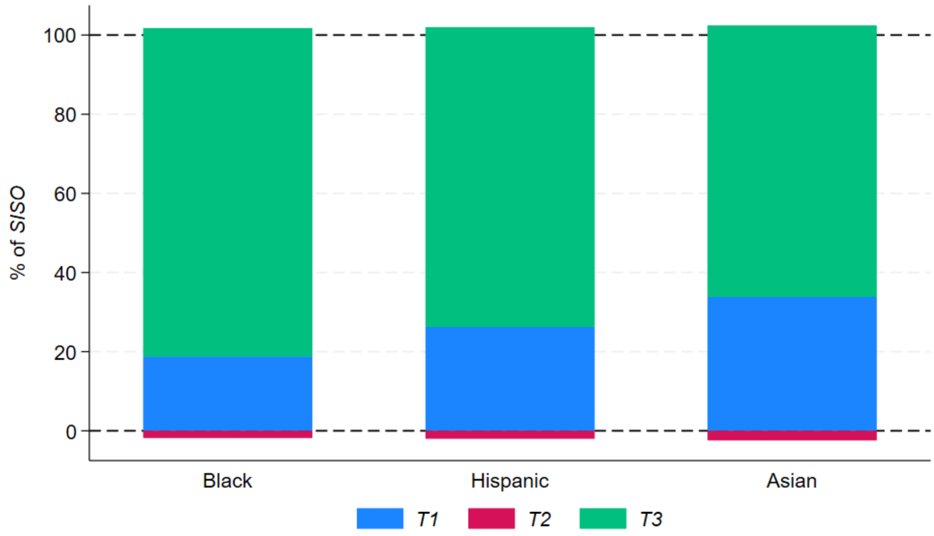


FIGURE A.23 – Average contribution across cities of each component to the overall index (i.e., the average ratio of $T1_c$, $T2_c$ and $T3_c$ on $SISO_c$) by racial or ethnic minority group.

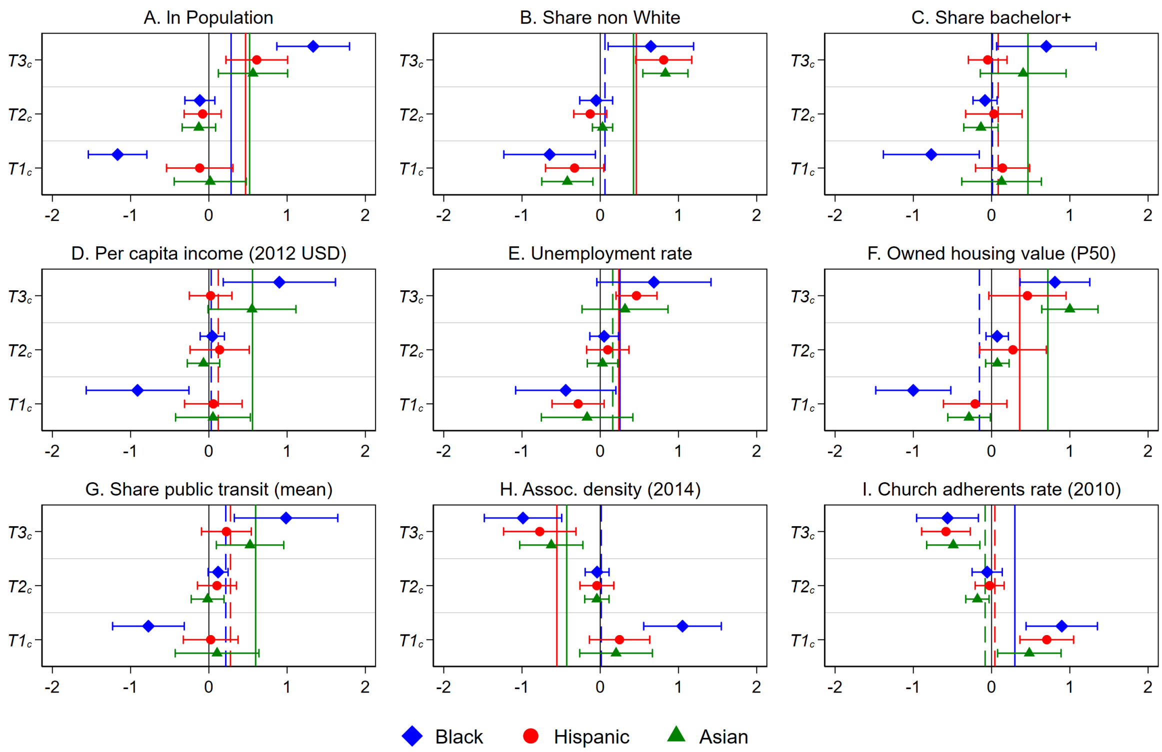


FIGURE A.24 – Coefficient plots for correlates of MSA social isolation by minority group. For each characteristic, markers denote results from fitting the model in Equation (9b). Robust confidence intervals at 95 percent levels are displayed using whiskers. The vertical reference lines shows the overall effect of $SISO_c$ for each group, 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.

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.: $^a p < 0.01$; $^b p < 0.05$; $^c p < 0.1$.