VERY PRELIMINARY EXTENDED ABSTRACT - DO NOT CIRCULATE

Rethinking Segregation: The Role of Social Connections in Racial Segregation

Andreas Diemer
SOFI, LSE

Tanner Regan
LBS, CEP

Cheng Keat Tang
NTU

February 22, 2022

Abstract

We study the extent of segregation in the social space of American cities. We measure segregation as the (lack of) actual personal connections between groups as opposed to conventional measures based on the spatial relationship between residents. We conceptualise how social segregation in American cities compares to geographical segregation and why the differentiation is important. Using data on the universe of Facebook friendships between urban zip codes, we then create city-level indices of social segregation and compare these with their geographical counterparts. We explore how various features of the urban space correlate with the discrepancies we observe. We also consider outcomes related to social exposure across neighbourhoods within-city, holding residential exposure fixed.

1 Introduction

Despite policy makers' efforts to desegregate neighborhoods, and the benefits associated with such programs (Guryan, 2004; Reber, 2010), preferences to live with other households of the same race and socio-economic status and sorting of households, and collective action by majority group that increase the cost of minority races to integrate continue to drive the persistence of residential segregation across United States (Shertzer and Walsh, 2019).

Social scientists developed tremendous interest towards understanding the extent and the impacts of ethnic segregation because many believed that it is a crucial factor explaining the disparity in different socioeconomic outcomes across neighbourhoods. This includes education (Hoxby, 1994; Cutler and Glaeser, 1997; Echenique and Fryer, 2007), wages (Ananat, 2011), crime (Kling et al., 2005), subjective well-being (Ludwig et al., 2012), consumption (Davis et al., 2019) and provision of public goods (Alesina et al., 1999).

The general consensus in the literature is that residential segregation exacerbates the black-white gaps in education outcomes, labor market outcomes and incarceration rates, along with many other outcomes.

Interest towards residential segregation led to a plethora of measures developed to improve our understanding towards the segregation and its impact on the society. Notwithstanding the difficulty associated with identifying the causal effects of residential segregation on socioeconomic outcomes[1], many studies have illustrated the inadequacies associated with existing geographical-based residential segregation measures. For instance, Massey and Denton (1988) highlighted that studies often proposed different measures to depict residential segregation, with little consensus towards what is the most appropriate measure. Furthermore, Echenique and Fryer (2007) pointed out that geographical-based measures depend arbitrarily on how cities are partitioned (e.g Census Blocks) and assume that individuals do not communicate with others beyond the boundaries. They illustrate that depending on how these geographical boundaries drawn, neighborhoods can become more or less segregated.

Our paper is related to the burgeoning stream of literature that improves the measurement of residential segregation with additional information on social networks and individuals' real-time location. These papers are primarily concerned that conventional measures of residential/geographical segregation assume individuals only communicate with others within the boundaries of the neighborhood of residence. The fact that individuals move across neighborhood boundaries across the day (e.g to work, to school) and communicate with others beyond residence, and could interact with others via social media platforms (e.g Facebook, Twitter, Instagram etc.) meant that traditional geographical based measures could be inaccurately depicting residential segregation.

Pioneering work by Echenique and Fryer Jr (2007) improves on traditional segregation measure by constructing individual segregation measures based on social interactions from friendship connections. They apply their strategy to construct school segregation for students from grade 7 to 12 in the Addhealth database. While this comprehensive database provides information not only on academic achievement and social behaviours, but also detailed information on their friendship networks, the survey is only conducted to a subset of the population. To construct residential segregation across US, Echenique and Fryer Jr (2007) are forced to assume that individuals residing in blocks within 1km from one other know each other. They report immaterial differences between the results from traditional dissimilarity index and their spectral segregation index.

[1] Ananat (2011) provides a good overview outlining the empirical challenges associated with identifying the causal effect of residential segregation. This is primarily driven by the presence of omitted variables and sorting of households. Existing literature typically relies on historical or natural geographical boundaries to identify exogenous variation in racial segregation. Some examples include (Hoxby, 1994; Ananat, 2011)

Recently, Athey et al. (2020) redefine residential segregation based on the concept of experienced isolation, which is determined by how individuals spend their time across the day and how they interact with others. Specifically, they rely on Global Positioning System (GPS) data of more than 17 million users that reveal their locations across the day based on cell phone apps. They are able to construct measures of experienced segregation for each individual based on whereabouts of an individual throughout the day relative to the locations of others. Without knowing whether individuals actually know one another, Athey et al. (2020) are forced to make the assumption that individuals interact with the people that co-locate in these places at the same time. Their findings reveal that traditional geographical segregation measures overestimate how segregated individuals are. This is because individuals commute to less segregated areas during the day and are exposed to people of different races. Experienced isolation is considerably lower in areas with higher population density and better public transit connectivity.

In similar fashion, we propose a new method of measuring residential segregation that incorporates social connections between neighbourhoods. Specifically, we rely on social connections from Facebook between granular geographical areas (at zipcode levels) across United States and integrate them into our measures of residential segregation to more accurately depict how areas can be connected with one another. We rely on social connections on Facebook because it is the world's largest social network with more than 258 million active users across the United States and Canada. This constitutes around 70% of the total population. We argue that the representative-ness of Facebook usage meant that social connections could realistically depict actual friendship networks across US. Incorporating the social connectivity index (SCI) allow us to more accurately and conveniently improve our existing measures of residential segregation that depend on how administrative boundaries are drawn within the city (e.g census tracts or blocs). We are no longer assuming that individuals are only interacting with others within the boundaries as SCI provides detailed zipcode-to-zipcode level of social connections across US. We then benchmark our measure of residential segregation with conventional measures and explain what is driving the disparity in the two measures across space.

An important question is why should we care about social connections. The growing influence of social media in our lives has drawn considerable attention from academics. Existing literature has shown that social connections materially influence economic decision making. For instance, Bailey et al. (2018a) document that peers on social network can influence home purchasing decisions. An individual is more likely to buy a bigger house and pay more for his/her purchase if peers on social network experience house price appreciation. Peers can also encourage homeowners to become more optimistic and undertake more debts (Bailey et al., 2019a). Bailey et al. (2019b) further show that peers can influence smaller decisions such as selecting the brand of cellphone while Gee et al. (2017) also show the social connections can improve the probability of finding a job. The recent outbreak of the pandemic further illustrated the importance of social media in shaping human behaviors. Specifically, Bailey et al. (2020) show that individuals with more friends in areas affected by COVID-19 are more likely to stay at home, reducing their visits to restaurants, bars and other recreational venues. Similar results are echoed by Charoenwong et al. (2020) who show that areas that are more socially connected to China and Italy has 50% higher compliance to mobility restrictions. Tóth et al. (2021) show that fragmentation of social networks (density of social ties within vs. across groups) closely interacts with various features of the urban space. Hence, we argue that even when spatial frictions (e.g commuting cost) are high, the ease of communication with friends via social media platforms such as Facebook could materially affect the extent individuals are isolated from others.

2 Methodology

Residential segregation is formally defined as how different groups of individuals, categorized by race or socio-economic status (e.g income), are living apart from one another. In their review, Massey and Denton (1988) classified various residential segregation measures into 5 different groups, namely Evenness (*distribution of certain groups across space*), Exposure (*how likely different groups are going to be in contact across space*), Concentration (*relative amount of space occupied by certain groups*), Centralization (*extent certain groups are located in city centers*) and Clustering (*extent to which contiguous areas are inhabited by certain groups*). The concept of exposure is particularly relevant in our study as we care about how social networks can affect exposure between groups that could influence socio-economic outcomes.

Isolation index is widely adopted in the literature as an exposure measure (Cutler et al., 1999; Echenique and Fryer, 2007; Gentzkow and Shapiro, 2011; Athey et al., 2020) to measure the extent to which minority groups are exposed to other members of the same group because of its intuitive appeal. Specifically, the residential isolation index for city c ($RISO_c$) can be expressed as follows[2]:

$$RISO_c = \underbrace{\sum_{i \in c} \left(\frac{x_i}{X_c} \frac{x_i}{t_i} \right)}_{\text{Minority group exposure to minorities}} - \underbrace{\sum_{i \in c} \left(\frac{y_i}{Y_c} \frac{x_i}{t_i} \right)}_{\text{Majority group exposure to minorities}} \tag{1}$$

where i denotes zipcode and c denotes city. y_i represents majority population and x_i

[2]Interaction index measures how the average member in the minority group is exposed to the majority. In other words, the interaction index is the inverse of isolation index.

represents minority population. t_i denotes total population in zipcode i such that $t_i = y_i + x_i$. X_c and Y_c denotes the sum of the minority and majority group population for city c respectively, i.e. $X_c \equiv \sum_{i\in c} x_i$ and $Y_c \equiv \sum_{i\in c} x_i$. Hence, $RISO_c$ varies from 0 to 1 and measures the average exposure of minority group to minorities minus the average exposure of majority group to minorities in city c.

A major concern associated with conventional measures of residential segregation (in equation 1), as highlighted by Echenique and Fryer (2007), is that the degree of segregation depends on how administrative boundaries are drawn. Redrawing these boundaries could drastically influence segregation measures. The fundamental pitfall is that we are assuming interactions between members are confined within neighbourhood boundaries.

To allow for cross-boundary interactions between areas, Massey and Denton (1988) suggest that we can set interaction weights equal to an exponential distance decay function to compute spatial isolation ($SPISO_c$). Massey and Denton (1988) motivate the spatial weights in the clustering index by arguing that"as one moves away from the home area, the actual likelihood of minority interaction with majority members decays rapidly... The probability of meeting a member of another group decreases as a function of distance". The interpretation is also straightforward as exposure to other areas is now measured as a constant decay function in geographical space. Spatial isolation index ($SPISO_c$) can be expressed as follows:

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

where the notable difference from equation 1 is the inclusion of ω_{ij} to account for cross-boundary interactions. Here, we allow the interaction between zipcode i and j (ω_{ij}) to decay exponentially with distance from one another (i.e. $\omega_{ij} = \frac{exp(-d_{ij})t_j}{\sum_{k\in c} exp(-d_{ik})t_k}$). There are, however, at least two issues with the spatial isolation measure ($SPISO_c$). First, it assumes that exposure is a smooth function in distance. Second, it requires a decay parameter that must be assumed or estimated. Consider the simple case that an individual who stays relative far from his/her work place and gets exposed to communities at a distance from his/her residence. In this context, such a distance decay function might not accurately capture "actual" cross-boundaries exposure.

To more accurately measure cross-boundary interactions (ω_{ij}), researchers measure interactions based on social networks from survey data (Echenique and Fryer, 2007), and more recently, based on GPS locations throughout the course of a day (Athey et al., 2020; Abbiasov, 2020). The distance decay approach is obviously an over-simplification, for instance if cross-boundary interactions come from the workplace, such a measure will place too much weight on nearby residential neighbourhoods that commuters pass through to reach their place of work. The direct measurement approach using GPS, while a significant improvement, faces several limitations. As noted in Athey et al. (2020), while they can observe when devices occupy the same geographic space, they cannot observe actual interactions between individuals. 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. Therefore, it is of paramount importance to measure social interactions between groups to accurately measure residential segregation across space. In this paper, we propose an alternative method to compute cross boundary interactions based on actual social connections between zipcodes from Facebook to mitigate pitfalls of previous cross-boundary interaction measures and we explain how we compute this measure in subsequent sections.

To measure cross-boundary social interactions, we rely on the Social Connectedness Index (SCI) for all US zipcode pairs (Bailey et al., 2018b). It can be interpreted as the probability that a random individual from zipcode i is friends with a random individual from zipcode j and can be expressed as follows:

$$SCI_{i,j} = c * \frac{FB\_Connections_{i,j}}{users_i \times users_j} \tag{3}$$

Where $FB\_Connections_{i,j}$ is the observed number of facebook friendships between zipcode i and zipcode j, $users_i$ and $users_j$ are the number of facebook users in zipcodes i and j (i.e. the denominator is the total possible number of facebook connections across zipcodes i and j), and c is a constant scaling. Therefore, SCI is a measure of the (scaled) 'relative probability of a friendship link' (Bailey et al., 2018b).[3]

In this paper, we focus on improving existing isolation measure, as illustrated in equation 1, by allowing cross-boundaries exposure between areas. Here, we express our measure of social isolation ($SOCISO_c$) as follows:

[3]There is a special case for zipcode links to itself, i.e. when $i = j$. In this case the scaled probability of a friendship link is $c * \frac{FB\_Connections_{i,i}}{0.5*users*(users-1)}$. The SCI for these pairs is constructed by doubling friendship connections and not counting self-friendships, i.e. $SCI_{ii} = c * \frac{2*FB\_Connections_{i,i}}{users_i \times users_i}$. Therefore the scaled probability of a friendship link is $SCI_{ii} * \frac{users_i}{users_i - 1} \approx SCI_{ii}$.

$$
\begin{aligned}
SOCISO_c &= \sum_{i \in c} \left(\frac{x_i}{X_c} \sum_{j \in c} \omega_{ij} \frac{x_j}{t_j} \right) - \sum_{i \in c} \left(\frac{y_i}{Y_c} \sum_{j \in c} \omega_{ij} \frac{x_j}{t_j} \right) \\
&= \sum_{i \in c} \left(\frac{x_i}{X_c} - \frac{y_i}{Y_c} \right) \omega_{ii} \frac{x_i}{t_i} + \sum_{i \in c} \left(\frac{x_i}{X_c} - \frac{y_i}{Y_c} \right) \sum_{j \neq i} \omega_{ij} \frac{x_j}{t_j} \\
&= \underbrace{\overline{\omega_{ii}}_c RISO_c + N_c * cov(\left(\frac{x_i}{X_c} - \frac{y_i}{Y_c} \right) \frac{x_i}{t_i}, \omega_{ii})}_{\text{Own zip contribution to ISO}} + \underbrace{\sum_{i \in c} \left(\frac{x_i}{X_c} - \frac{y_i}{Y_c} \right) \sum_{j \neq i} \omega_{ij} \frac{x_j}{t_j}}_{\text{Other zip contribution to ISO}}
\end{aligned} \tag{4}
$$

Evidently, the accuracy of the isolation index depends on how we measure the exposure weights (ω_{ij}). As explained 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". We propose using SCI (denoted in equation 3), which captures the social connectivity between zipcodes, to compute ω_{ij}. Specifically, we allow for cross-boundary interactions by setting the weights to the relative propensity of friendship (i.e. $\omega_{ij} = \frac{SCI_{ij} t_j}{\sum_{k \in c} SCI_{ik} t_k}$). In short, we are measuring the exposure to minority for the average minority and majority individual living in zipcode i to all other j zipcodes based on exposure weights ω_{ij} computed from the SCI.

We assume that the cross-boundary interactions of individuals in zipcode i with zipcode j can be measured by the share of the friends living in zipcode j from the SCI. There are at least two reasons to justify this. First, as mentioned earlier, more than 70% of the US and Canadian population uses Facebook, making the SCI a representative measure of social connections. Furthermore, existing literature has shown that social interactions on Facebook can materially affect various socio-economic outcomes and decision making. With the SCI, we can conveniently measure interactions directly at a granular zipcode level without the need of measuring social connections and identifying real-time locations at an individual level.

It is also not hard to relate this to traditional isolation measure (equation 1) proposed by Massey and Denton (1988) and Gentzkow and Shapiro (2011) to illustrate how our measure ($SOCISO_c$) improves on the existing literature. In short, traditional isolation measures set the own-zipcode weights to one and all others to zero (i.e. $\omega_{ij} = \mathbf{1}(i = j)$, meaning that individuals are only exposed to others who reside in the same area). The interpretation is intuitive: isolation measures the minority share of exposure for the average minority individual subtracted by the minority share of exposure for the average majority individual, where exposure is measured as living in the same neighbourhood. The major downside of this measure is that cross-boundary exposure is completely discounted and we improve on this measure by allowing for cross-boundary exposure based on social interactions via Facebook.

3 (Very) preliminary findings

Figure 1 shows within MSA variation showing the components of our indices for Chicago, as an example. Panel (a) gives the difference in each ZCTA's share of MSA's non-white and white. Panel (b) gives the residential exposure measure (simply the share of non-white in each ZCTA). Panel (c) gives the social exposure, obtained as the weighted share of non-white (excluding own ZCTA shares), with weights proportional to the intensity of social interaction between areas. The same scale was imposed to panels (b) and (c), making the two easier to compare visually.

Figure 2 presents the estimated residential and social isolation measures across different Metropolitan Statistical Areas (MSAs). An important fact that emerges from this map is that many areas are not as isolated as previously depicted by traditional residential isolation measures that do not allow between areas interaction when we refer to social isolation. Another observation is that the disparity in residential and social isolation is non-trivial in some areas.

Figure 3 gives a scatter plot of residential isolation against social isolation. Based on the 45 degree line, we can identify MSAs of which social isolation is greater than residential isolation (and vice versa). Panel (a) compares absolute values, panel (b) compares z-scores, which allows to interpret results relative to each measures' average.

4 Intended contribution

This research contributes to the existing literature with an improved measure of residential segregation across US. We rely on a social connections on Facebook - a representative social medium used by many Americans - to accurately measure residential segregation at granular geographical levels. With reliable data on how areas are connected to one another, we no longer need to assume that individuals communicate with one another when they are living near to one another, or when individuals who co-locate in certain areas actually know each other.

There are, however, some limitations to our research. In particular, although we have extensive zipcode-to-zipcode social connections across US, we do not observe social connections at an individual level. Hence, we are unable to construct individual measure of social segregation.

References

Abbiasov, T. (2020). Do Urban Parks Promote Racial Diversity? Evidence from New York City. pages 1–35.

Alesina, A., Baqir, R., and Easterly, W. (1999). Public Goods and Ethnic Divisions. *The Quarterly Journal of Economics*, 114(4):1243–1284.

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.

Athey, S., Ferguson, B., Gentzkow, M., and Schmidt, T. (2020). Experienced Segregation. *National Bureau of Economic Research*, pages 1–54.

Bailey, M., Cao, R., Kuchler, T., and Stroebel, J. (2018a). 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. (2018b). Social connectedness: Measurement, determinants, and effects. *Journal of Economic Perspectives*, 32(3):259–80.

Bailey, M., Dávila, E., Kuchler, T., and Stroebel, J. (2019a). House price beliefs and mortgage leverage choice. *The Review of Economic Studies*, 86(6):2403–2452.

Bailey, M., Johnston, D., Koenen, M., Kuchler, T., Russel, D., and Stroebel, J. (2020). Social distancing during a pandemic: The role of friends. *NBER Working Paper*, (w28234).

Bailey, M., Johnston, D. M., Kuchler, T., Stroebel, J., and Wong, A. (2019b). Peer effects in product adoption. Technical report, National Bureau of Economic Research.

Charoenwong, B., Kwan, A., and Pursiainen, V. (2020). Social connections with covid-19–affected areas increase compliance with mobility restrictions. *Science advances*, 6(47):eabc3054.

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

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

Davis, D. R., Dingel, J. I., Monras, J., and Morales, E. (2019). How segregated is urban consumption? *Journal of Political Economy*, 127(4):1684–1738.

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

Echenique, F. and Fryer Jr, R. G. (2007). A measure of segregation based on social interactions. *The Quarterly Journal of Economics*, 122(2):441–485.

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.

Guryan, J. (2004). Desegregation and black dropout rates. *American Economic Review*, 94(4):919–943.

Hoxby, C. M. (1994). Do private schools provide competition for public schools? Technical report, National Bureau of Economic Research.

Kling, J. R., Ludwig, J., and Katz, L. F. (2005). Neighborhood effects on crime for female and male youth: Evidence from a randomized housing voucher experiment. *The Quarterly Journal of Economics*, 120(1):87–130.

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.

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

Reber, S. J. (2010). School desegregation and educational attainment for blacks. *Journal of Human resources*, 45(4):893–914.

Shertzer, A. and Walsh, R. P. (2019). Racial sorting and the emergence of segregation in american cities. *Review of Economics and Statistics*, 101(3):415–427.

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 2021 12:1*, 12(1):1–9.

Figure 1: Components of the isolation indices for Chicago

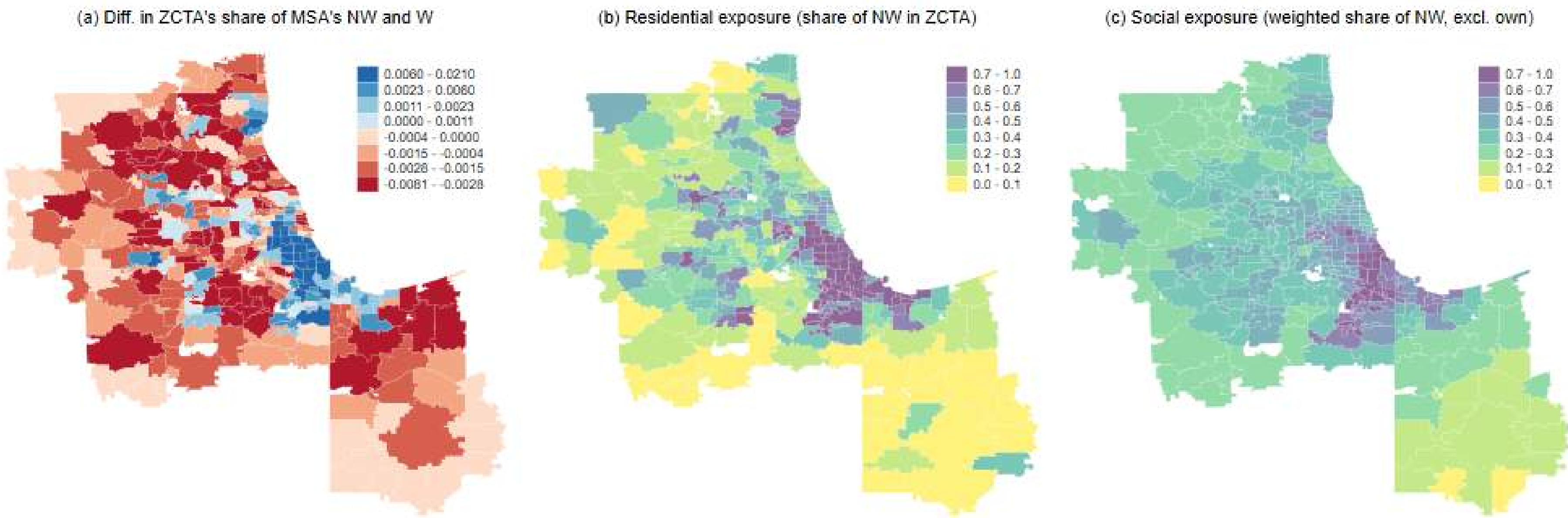


Figure 2: Residential and social isolation across MSAs

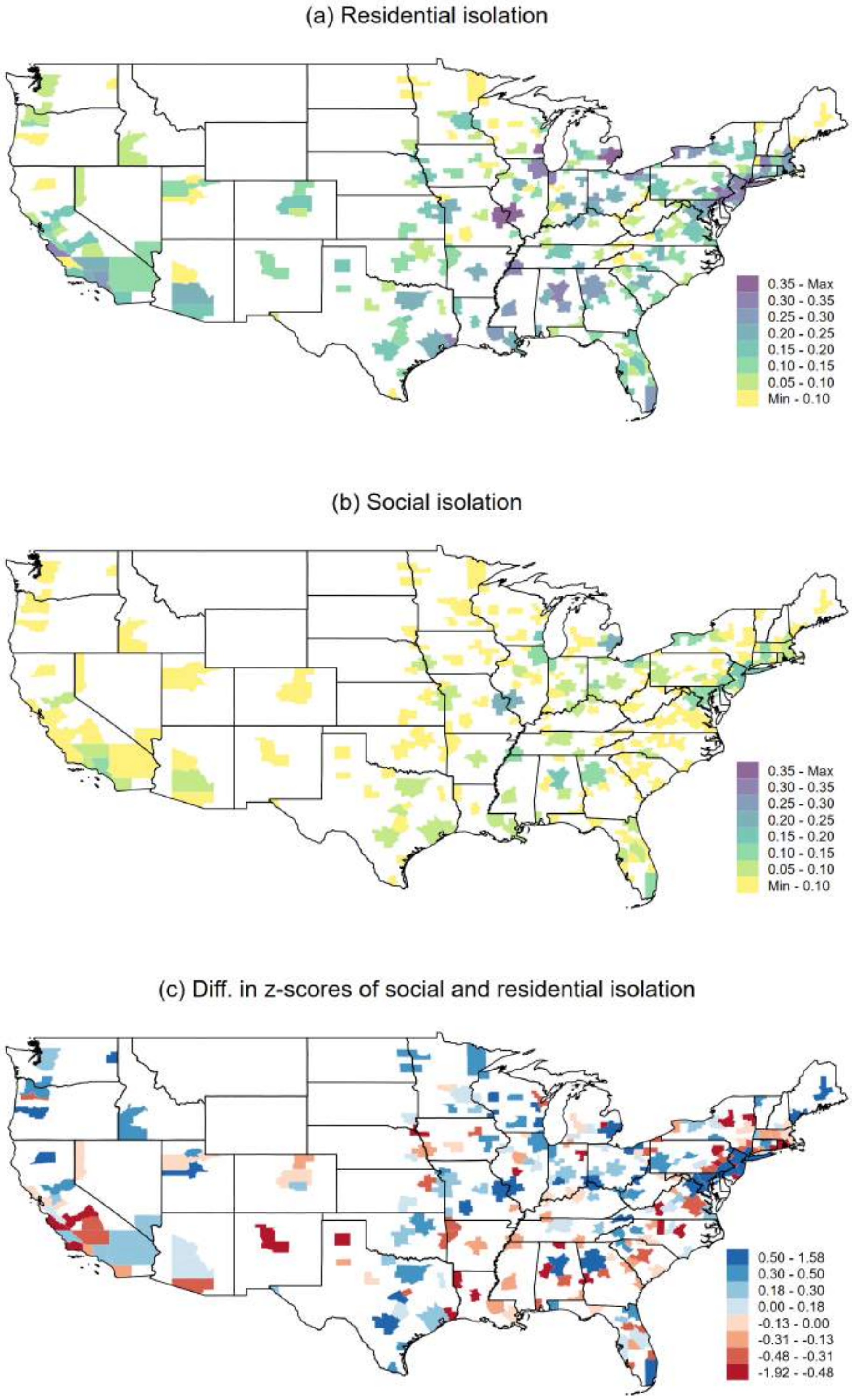
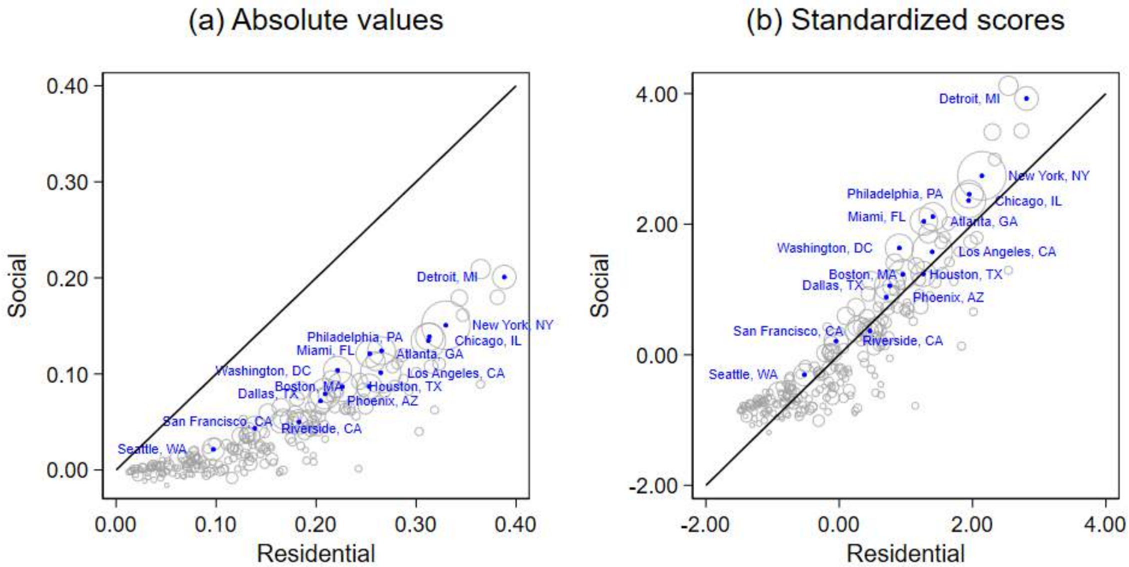


Figure 3: Scatter plots of segregation indices



Note: Only MSAs with at least 20 ZCTAs were retained.