Using Population Mobility Data to Measure Black-White Residential Segregation in the COVID-19 Pandemic

Yongjun Zhang[a,1]

[a]Department of Sociology and Institute for Advanced Computational Science, State University of New York at Stony Brook, 100 Nicolls Road, Stony Brook, New York 11794; ORCiD: https://orcid.org/0000-0002-8265-925X

June 23, 2021

Racial and ethnic residential segregation has long been the central focus of stratification and inequality research, and it is a linchpin of racial stratification in the U.S. Sociologists and demographers have developed a series of spacial or aspatial measures to capture distinct aspects of segregation. Although the recent development of segregation measures, for instance, spacial exposure, accounts for spacial proximity among different groups, it is static and ignores the social connectedness dimension. This article uses population mobility across communities to correct the potential bias in spacial segregation measures. As population mobility is highly racially segregated, we modify the conventional spatial isolation index by adding an extra layer of social connectedness between communities to create a socially and spatially weighted segregation measure. We then use this spatial and social segregation measure to quantify the level of blacks' isolation with whites in the local neighboring communities.

residential segregation | population mobility | big data | COVID19

Spatial segregation of racial groups has received considerable attention from social scientists, journalists, and policymakers in the U.S., as it relates to a variety of negative outcomes, such as poverty (1, 2), school resegregation (3, 4), and crime (5, 6). Residential segregation serves as a fundamental mechanism of socioeconomic stratification, and residents of poor and geographically isolated communities have limited contacts with and access to the mainstream society with opportunities and resources (7–9). Thus, an appropriate measure of racial and ethnic residential segregation is vital to our society, especially to those minority groups living in disadvantaged neighborhoods.

For over 60 years, social scientists have developed various aspatial and spatial measures to capture the degree of residential segregation and its impact on education, housing, health care, and labor market outcomes (10–17). These measures account for distinct spatial variation in unevenness, exposure, clustering, concentration, and centralization with indices like dissimilarity index, information theory index, and spatial exposure (7, 12, 18). Of these five dimensions, only unevenness and exposure are widely used in empirical studies by social scientists (19). These measures rely on the partition of a city or county into small geographic units such as Census Tract or Block and they are often aspatial and static. A more sophisticated segregation measure is to incorporate spatial relations between two units when defining segregation indices (15, 20). In this research, the primary segregation measure of interest focuses on the exposure dimension, which captures the extent to which members in a group interact with members of different groups in a given space. Typically, scholars use the distance between two geographic units to account for spatial proximity. But this assumes that different social groups living in proximate geographic areas have a greater likelihood to interact with each other. This underlying assumption is not necessarily true, given that two proximate geographic units might not be tightly connected due to various reasons (e.g., rivers, highways).

To address these shortcomings, Echenique and Fryer took a social network approach to develop a measure of segregation based on social interactions (16). The rationale is that an individual is more segregated when interacting more with other segregated agents in a community. They also highlight that the measure of segregation should disaggregate to the individual-level. Yet, their measure receives less attention due to lack of large-scale social interaction data across different social groups. A few notable exceptions are Wang, Candipan, and their collegue's seminal work on using geocoded twitter users' everyday mobility data to measure neighborhood isolation in America's 50 largest cities (9, 17). Their research shows that urban residents of poor minority communities appear to travel far and widely across communities instead of limiting themselves in their neighboring communities. Following their research, we argue that spatial proximity and social interaction are two distinct dimensions to measure residential segregation, even though they are correlated. For instance, a person lives in Stony Brook might never go to East Setauket in Long Island, even though these two areas are geographically proximate. Thus, to better measure residential segregation, we need to account for both spatial proximity and social interactions.

Significance Statement

Spatial or residential segregation among different racial groups has been a longstanding issue in the U.S. This article develops a new approach to correct potential biases in measures of residential segregation between whites and Blacks using population mobility data across different communities. We modify the conventional spatial isolation measure by adding a social interaction layer to account for both spatial proximity and social connectedness when defining residential segregation. Social isolation is distinct from spatial isolation, even though social interaction is a function of spatial proximity. Our approach can be extended to other segregation measures and provide a new perspective to assess racial segregation in the U.S.

Y.Z. designed research, analyzed data, and wrote the paper.

Y.Z. declares no competing interests in the research.

[1]To whom correspondence should be addressed. E-mail: yongjun.zhang@stonybrook.edu

In this article, we focus on black-white residential segregation because historically African Americans are the minority group that experiences the hypersegregation (21). We modify the conventional spatial isolation index by adding another social interaction layer to account for social connections across different communities. Using novel big data on cross-census group block population mobility in New York State, we first examine the population mobility in 2019-2020 and then measure the local racial segregation for any individuals in a given census block group (CBG), creating a spatially and socially weighted measure of cross-racial exposure for CBGs. These spatial and population mobility data afford scholar an opportunity to create a more comprehensive and fine-tuned measure of racial residential segregation in the U.S.

Measuring Social Connectedness across Communities

Racial segregation is not only solely about where people reside, but also about where people travel over the course of everyday activities (17). To capture the strength of social connection, it requires large-scale and representative data on social connectedness between individuals or communities.

We define social connectedness as the extent to which a community is densely connected with its neighboring communities. Following Echenique and Fryer's work, a community is highly connected with another communities if residents within two communities interact with each other densely. We can use different social relations, such as phone call records, online friendship, and traffic visits, to quantify the strength between two communities. For instance, Bailey et al. used online friendship links on Facebook to measure social connectedness across different geographic units (22). Candipan et al. repurposed geocoded Twitter user mobility data to develop the segregated mobility index to assess the isolation among urban communities.

In recent years, especially in the pandemic, to combat COVID-19, big tech firms such as Google, Facebook, and SafeGraph have shared large-scale population mobility data with academic scholars at small geographic units such as Census Block Groups (CBG). Even though these are not designed for segregation research, we can repurpose these data and examine the social interaction dimension of racial segregation (23). We rely on SafeGraph's Social Distancing Metrics Data, which contains digital trace everyday mobility information across different communities in 2019-2020. We aggregate the daily mobility data to the annual level for each census block group. SafeGraph's social distancing data shows that population mobility before and in the pandemic is highly racially segregated. A white dominated-community is densely connected with other local white communities and is highly disconnected with local black communities. A large proportion of New York black residents lives with low levels of residential exposures to white neighbours.

Given that we have daily visits from the $i-th$ and $j-th$ census block group, we can compute social connectedness index (SCI) for each pair of CBGs in a given period P as follows.

$$SCI_{ij} = \frac{\sum_{d=1}^{P} (Visits_{ijd} + Visits_{jid})}{\frac{1}{P}\sum_{d=1}^{P} (Devices_{id} + Devices_{jd})} \quad [1]$$

Where i and j denote the $i-th$ and $j-th$ census block group, d denotes the $d-th$ day, $Visits_{ij}$ indicates the daily visits from i to j, and devices denote the number of unique devices (i.e., individuals) in a given census block group. We then normalize social connectedness index by divided the maximum value of raw SCI score, so it ranges from 0 to 1. A large value indicates that two communities are densely connected. The preceding formula treats SCI as undirected, and we can also compute the SCI based on the origin-destination pattern simply as follows.

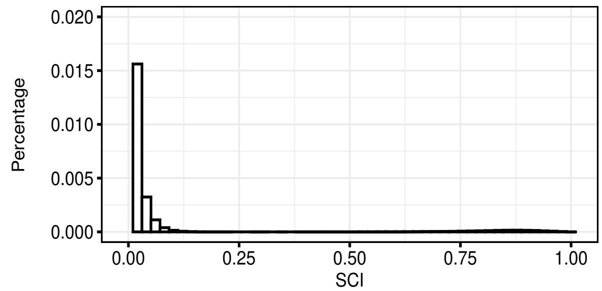

Fig. 1. Distributions for Social Connectedness Index in New York, 2020

$$SCI_{ij} = \frac{\sum_{d=1}^{P} Visits_{ijd}}{\frac{1}{P}\sum_{d=1}^{P} Devices_{id}} \quad [2]$$

Fig. 1 shows the distribution of the SCI based on Eq.1 for CBGs in New York after excluding disconnected communities. The average SCI between two CBGs is 0.0015 in 2020.

Measuring Spatial Segregation Socially

Spatial isolation index has been often used to measure the extent to which one social group is exposed to another group in its local environment. Spatial isolation index of blacks captures the extent to which an individual encounters neighbours from another racial group. A high value indicates that blacks are more isolated with whites living in neighboring communities. Let us first assume that in a spatial region R, it has N smaller units (i.e., in our case, census block groups) and M mutually exclusive racial groups. Thus, for any blacks in a CBG i, we define a spatially weighted isolation as follows.

$$Spatial\ Isolation_{im} = \frac{\sum_{k=1}^{N} \tau_{km} * \frac{1}{(D_{ik}+d)^a}}{\sum_{k=1}^{N} \frac{1}{(D_{ik}+c)^a}} \quad [3]$$

If we replace the distance weights with social connection index, a metric ranging from 0 to 1 capturing the connection between communities, then similarly we can compute the social isolation index as follows.

$$Social\ Isolation_{im} = \frac{\sum_{k=1}^{N} \tau_{km} * (S_{ik}+s)^b}{\sum_{k=1}^{N} (S_{ik}+s)^b} \quad [4]$$

We use the modified spatially and socially weighted isolation index to capture the isolation between blacks and whites to describe residential segregation in New York (24). Because SafaGraph does not provide individual-level mobility data, we focus on census block groups, the smallest geographic unit in the dataset. Note that our modified measure can be easily disaggregated to the individual level. We define the spatial and social isolation index (SSI) as follows.

$$SSI_{im} = \frac{\sum_{k=1}^{N} \tau_{km} * \frac{1}{(D_{ik}+d)^a} * (S_{ik}+s)^b}{\sum_{k=1}^{N} \frac{1}{(D_{ik}+c)^a} * (S_{ik}+s)^b} \qquad [5]$$

Where m and n denote the specific racial group m and n (in this case, they are blacks and whites) in M,i and k are indices for a given census block group $(i, k <= N)$, τ_{km} is the population density for a certain racial group m in a given census block group k, calculated by dividing its population by total population in the census block group k, B_k is the number of blacks in census block group k, W_k is the number of whites in census block group k, D_{ik} is the distance between census block group i and k, and S_{ik} is the social connection index between census block group i and k. d and s denote the constant adjustments which make sure the expression is defined when $D_{ik} = 0$ and $S_{ik} = 0$. a and b are two exponents used to control how much weights we tend to give to proximity or social connection in the measure. In this analysis, we set $a = b = d = s = 1$. Setting d higher would reduce the weights given to the smallest distances and setting s higher would increase the weights given to the smallest social connection.

Given that we can further restrict the limit of the physical distances between census block group i and k, spatial and social isolation (SSI) then can represent an individual's residential and social interaction experience in the local environment. We can further aggregate CBG level to county or city level. In this article, we use 10 km as the radius to define the local environment, although Wang et al.'s work reports the average travel radius within commuting zones for urban residents in the U.S. is 5.29 km with a SD of 1 km. We use a larger value as our neighborhood radius because our research includes census block groups in rural areas. We also computed 1km and 5km and results are consistent.

Fig.2 compares different indices. Panel A, B, C, and D show the distributions for proportion of blacks, spatial isolation index, social isolation index, and SSI index, respectively, in a given CBG. Note that Panel B and C shows that spatial isolation and social isolation indices show great disparities in terms of the distribution pattern. New York black residents show great social isolation with whites than spatial isolation in the pandemic.

This spatially and socially weighted segregation measure, SSI, has several merits. First, we can use spatial distance to define the geographic region R, and thus we treat scale as a variable instead of constant. This is particularly useful when scholars conceptualize the scale of neighborhood differently (25, 26). Second, SSI ranges from 0 to 1, and a larger value indicates that member in the focal group is completely spatially and socially isolated with members in other groups. Third, our measure is dynamic and can capture the temporal fluctuations in segregated mobility patterns. We use annual mobility data to capture the level of connectedness between communities, but scholars can also use monthly and quarterly mobility data to capture seasonal nature of racial residential segregation. Finally, scholars can also flexibly define the social connectedness index using different available social relation data between individuals or communities. For instance, scholars can use virtual relations to measure the strength of connectedness between two communities.

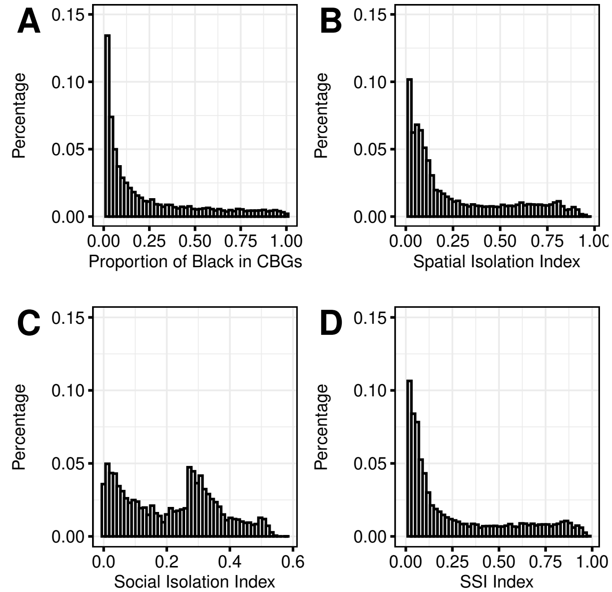

Fig. 2. Distributions for Proportion of Blacks, Spatial Isolation Index, Social Isolation Index, and SSI Index.

Results

Population mobility is highly racially segregated before and during COVID-19. We first analyzed population mobility in the pandemic from SafeGraph social distancing metrics data in New York.We define a community with over 60% blacks as black community and over 60% whites as white community. We compute the social connection index (SCI) from original CBG to neighboring destination CBG using average annual visits based on Eq.2. We use 10 km radius as the threshold to define a neighboring community. Fig.3 reports the observed SCI in 10 thousands for different origin-destination patterns. It shows that residents in black communities travel more to other black communities, and residents in white communities travel more to white communities. This demonstrates that social interactions between communities are highly racially segregated. Black and white communities are highly disconnected.

I then use OLS models to examine how racial composition of the origin CBG and the destination CBG affects the population mobility in 2019 and 2020. The key independent variables are segregated patterns: From Black to White, From Black to Black, From White to Black. The reference group is From White to White. We also control for distance, average number of devices, and total populations. Table 1 shows that after holding other factors constant, the connection from white to black and black to white communities is statistically significant weaker than the connection from white to white communities. Results in Table 1 also reveals that black communities is highly disconnected with neighboring black communities. These results hold for before and in the pandemic.

Being socially segregated is distinct from being spatially segregated. We compared the socially weighed version of residential segregation (i.e., social isolation index) with conventional spatial isolation index for blacks. The average spatial isolation index in New York is 0.209, while the average social isolation index is .214. This means that blacks are more socially iso-

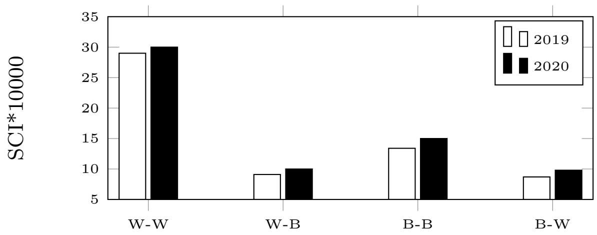

Fig. 3. Racial Patterns of Population Mobility across Communities within 10 km in New York, 2019-2020. We plot the observed SCI for different origin-destination patterns based on Safegraph social distancing data. W-B denotes the connection from white community to black community. We define a black community if the black population is over 40%.

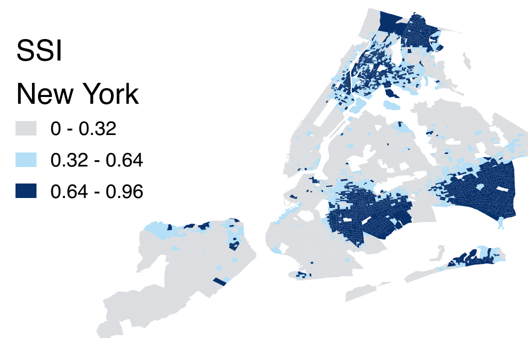

Fig. 4. Spatial and Social Isolation Index for Blacks in 2020

Table 1. OLS models estimating mobility in New York State

Variables	2019	2020
Black-White	-0.001***	-0.001***
	(0.000)	(0.000)
Black-Black	-0.002***	-0.002***
	(0.000)	(0.000)
White-Black	-0.001***	-0.001***
	(0.000)	(0.000)
Distance ln	-0.003***	-0.003***
	(0.000)	(0.000)
Controls	Yes	Yes
Constant	0.030***	0.030***
	(0.000)	(0.000)
Num.Obs.	8675436	6887835
R2 Adj.	0.074	0.072

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

lated with whites relative to spatial isolation in the pandemic. Although these two measures are strongly correlated (Pearson's r = 0.591), a large proportion of the variance in social interaction cannot be explained by psychical distance.

Black residents in Bronx, Kings, and Queens County are most spatially and socially isolated with white New Yorkers. We calculate the SSI index for all counties in New York State. Note that we set the 10 km radium as the local environment of interest. Consistent with prior studies, Black residents in urban areas are more isolated than those in rural areas. The average SSI indices for Bronx, Kings, and Queens county are 0.57, 0.35, and 0.29, respectively, then followed by Monroe County (0.22), New York County (0.21), Erie County (0.20), Westcherster County (0.19), Albany County (0.17), Rockland County (0.16), and Onondaga County (0.15).

Fig. 4 shows the spatial and social isolation measure for blacks in New York City. A darker blue shows greater spatial and social isolation. The Upper East Side of Manhattan shows the clustering of black New Yorkers with a lower exposure to white residents, while the southern extreme shows a lower isolation of blacks with whites due to the clustering whites in this area. The most segregation census block groups are located in Bronx, Brooklyn, and Queens areas.

Discussion

The ethnic and racial residential segregation has been an enduring feature of American society. It is the linchpin of racial stratification and a legacy of the *Plessy v. Ferguson*. For the past six decades, scholars have developed a variety of segregation indices to capture residential segregation, but most of them are aspatial and static. In this article, we introduced a novel spatial and dynamic measure of residential segregation, accounting for spatial proximity and social interactions. We fine-tuned the traditional measure of spatial exposure–spatial isolation index by adding a social interaction layer using residents' everyday travel data among communities.

Following recent work, we argue that social isolation is distinct from spatial isolation, even though social isolation is a function of spatial proximity (9, 17). We first analyzed the population mobility data in the pandemic and find that New York residents' travel patterns are racially segregated. White communities are highly connected with other local white communities, but are highly disconnected with other black communities. This indicates that the conventional isolation index only including spatial proximity cannot account for social connectedness between communities. We also compared spatial isolation index with social isolation index and find that these two measures are modestly but not completely correlated. Thus, we need a geographically and socially weighed measure of residential segregation to address the spatial and dynamic nature of racial segregation.

Using the novel spatial and social interaction index, we then quantify each CBG's isolation level in New York State in the pandemic. We find that black residents in New York City, especially in Bronx, Kings and Queens are most isolated with their white New Yorkers in neighboring communities in the pandemic. The spatial and social isolation might be beneficial regarding preventing the spread of COVID19 across communities, but it might have other issues, for instance, mental health problems and lack of medical resources. In this article, we only use New York State as an illustrative case and future studies should extend our approach to other states.

Our spatial and social isolation index focuses on census block group level due to data availability issue. A more nuanced approach should be defining spatially and socially weighted segregation measures at the individual level. A recent example is Brown and Enos' seminal work on partisan sorting for 180 million voters (24). Their partisan segregation accounted for spatial proximity but again not social interactions between Republicans and Democrats in neighboring communities. We are in an era of data explosion, but accessing large-scale individual-level digital trace data is still difficult, and it contains ethical and privacy issues. We still need to further strengthen the collaboration between the industry and the academia in order to harness the benefits of big data for the public good.

Materials and Methods

Here we briefly summarize how we calculate social connection index.

Social connection data. To combat COVID19, SafeGraph inc. has released places. geometry, and patterns data. We use its social distancing metrics at the census block group level (https://docs.safegraph.com/docs/social-distancing-metrics). It documents the origin CBG, the destination CBG, and the number of devices traveling from the origin to the destination.

SafeGraph has excluded census block groups with fewer than five devices visiting a place in a month and its data products and maps are aggregated and no human subjects would be re-identified.

Racial composition data. Open census data. we use the matched census data from SafeGraph, which is based on American Community Survey 2019 5 year estimates.

Geographic distance data. Distance was calculated using QGIS' distance matrix module.

Data and code availability. All aggregated data used in the analysis are available via https://osf.io/pvbxw/.

ACKNOWLEDGMENTS. Y.Z. wishes to thank the Institute of Advanced Computational Science at Stony Brook University for the access to the High-Performance Computing.

1. L Quillian, Segregation and poverty concentration: The role of three segregations. *Am. Sociol. Rev.* **77**, 354–379 (2012).
2. A Owens, Inequality in children's contexts: Income segregation of households with and without children. *Am. Sociol. Rev.* **81**, 549–574 (2016).
3. JE Fiel, Decomposing school resegregation: Social closure, racial imbalance, and racial isolation. *Am. Sociol. Rev.* **78**, 828–848 (2013).
4. JE Fiel, Y Zhang, With all deliberate speed: The reversal of court-ordered school desegregation, 1970–2013. *Am. J. Sociol.* **124**, 1685–1719 (2019).
5. ES Shihadeh, N Flynn, Segregation and crime: The effect of black social isolation on the rates of black urban violence. *Soc. forces* **74**, 1325–1352 (1996).
6. LJ Krivo, RD Peterson, DC Kuhl, Segregation, racial structure, and neighborhood violent crime. *Am. journal Sociol.* **114**, 1765–1802 (2009).
7. DS Massey, Reflections on the dimensions of segregation. *Soc. Forces* **91**, 39–43 (2012).
8. CZ Charles, The dynamics of racial residential segregation. *Annu. review sociology* **29**, 167–207 (2003).
9. Q Wang, NE Phillips, ML Small, RJ Sampson, Urban mobility and neighborhood isolation in america's 50 largest cities. *Proc. Natl. Acad. Sci.* **115**, 7735–7740 (2018).
10. OD Duncan, B Duncan, A methodological analysis of segregation indexes. *Am. sociological review* **20**, 210–217 (1955).
11. MJ White, The measurement of spatial segregation. *Am. journal sociology* **88**, 1008–1018 (1983).
12. DS Massey, NA Denton, The dimensions of residential segregation. *Soc. forces* **67**, 281–315 (1988).
13. J Iceland, DH Weinberg, E Steinmetz, *Racial and ethnic residential segregation in the United States 1980-2000.* (Bureau of Census) Vol. 8, (2002).
14. SF Reardon, G Firebaugh, Measures of multigroup segregation. *Sociol. methodology* **32**, 33–67 (2002).
15. SF Reardon, D O'Sullivan, Measures of spatial segregation. *Sociol. methodology* **34**, 121–162 (2004).
16. F Echenique, RG Fryer Jr, A measure of segregation based on social interactions. *The Q. J. Econ.* **122**, 441–485 (2007).
17. J Candipan, NE Phillips, RJ Sampson, M Small, From residence to movement: The nature of racial segregation in everyday urban mobility. *Urban Stud.*, 0042098020978965 (2021).
18. H Theil, AJ Finizza, A note on the measurement of racial integration of schools by means of informational concepts. (1971).
19. P Rich, J Candipan, A Owens, Segregated neighborhoods, segregated schools: Do charters break a stubborn link? *Demography* **58**, 471–498 (2021).
20. DW Wong, Comparing traditional and spatial segregation measures: A spatial scale perspective1. *Urban Geogr.* **25**, 66–82 (2004).
21. DS Massey, NA Denton, Hypersegregation in us metropolitan areas: Black and hispanic segregation along five dimensions. *Demography* **26**, 373–391 (1989).
22. M Bailey, R Cao, T Kuchler, J Stroebel, A Wong, Social connectedness: Measurement, determinants, and effects. *J. Econ. Perspectives* **32**, 259–80 (2018).
23. MJ Salganik, *Bit by bit: Social research in the digital age.* (Princeton University Press), (2019).
24. JR Brown, RD Enos, The measurement of partisan sorting for 180 million voters. *Nat. Hum. Behav.*, 1–11 (2021).
25. SF Reardon, et al., Race and space in the 1990s: Changes in the geographic scale of racial residential segregation, 1990–2000. *Soc. Sci. Res.* **38**, 55–70 (2009).
26. BA Lee, et al., Beyond the census tract: Patterns and determinants of racial segregation at multiple geographic scales. *Am. Sociol. Rev.* **73**, 766–791 (2008).