Sociological Methods & Research
1-40

DOI: 10.1177/0049124119852386
journals.sagepub.com/home/smr
SAGE

The Social Integration of American Cities: Network Measures of Connectedness Based on Everyday Mobility Across Neighborhoods

Nolan E. Phillips[1], Brian L. Levy[1], Robert J. Sampson[1], Mario L. Small[1], and Ryan Q. Wang[2]

Abstract

The social integration of a city depends on the extent to which people from different neighborhoods have the opportunity to interact with one another, but most prior work has not developed formal ways of conceptualizing and measuring this kind of connectedness. In this article, we develop original, network-based measures of what we call "structural connectedness" based on the everyday travel of people across neighborhoods. Our principal index captures the extent to which residents in each neighborhood of a city travel to all other neighborhoods in equal proportion. Our secondary index captures the extent to which travels within a city are concentrated in a handful of receiving neighborhoods. We illustrate the value of our indices for the 50 largest American cities based on hundreds of millions of geotagged tweets over 18 months. We uncover important features of major American cities,

[1] Harvard University, Cambridge, MA, USA
[2] Northeastern University, Boston, MA, USA

Corresponding Author:
Robert J. Sampson, Harvard University, William James Hall, 33 Kirkland St., Cambridge, MA 02138, USA.
Email: rsampson@wjh.harvard.edu

including the extent to which their connectedness depends on a few neighborhood hubs, and the fact that in several cities, contact between some neighborhoods is all but nonexistent. We also show that cities with greater population densities, more cosmopolitanism, and less racial segregation have higher levels of structural connectedness. Our indices can be applied to data at any spatial scale, and our measures pave the way for more powerful and precise analyses of structural connectedness and its effects across a broad array of social phenomena.

Keywords

urban mobility, connectedness, networks, neighborhoods, integration

Social integration has always been important to sociological research. The extent to which societies are integrated has concerned thinkers as varied as Durkheim ([1893] 1984, [1897] 1966), Toqueville ([1835] 2000), and Tönnies ([1887] 1957) as well as contemporary researchers such as Blau (1977) and Fischer (1982, 2011). Socially integrated societies are expected to experience greater solidarity, trust, pro-social behavior, longevity, and well-being, as well as lower violence, conflict, and crime (Angell 1947; Putnam 2000). In recent decades, researchers have examined how integrated cities are based on how segregated their neighborhoods are by race or income (Lichter, Parisi, and Taquino 2015; Logan, Stults, and Farley 2004; Reardon and Bischoff 2011). This approach reflects the important point that social integration depends not only on how attached people feel to the collective but also, and more importantly, on how much interaction different groups have with one another (Blau 1977, 1994; Blau and Schwartz 1997).

Nevertheless, this research faces an important limitation. Social interaction requires groups to come into contact (Blau 1977), and to the extent that different groups lack opportunities to come into contact, they will have difficulty integrating socially (Moody 2001; Mouw and Entwisle 2006). Traditional measures of segregation assume, reasonably, that people are less likely to come into contact if they live in different neighborhoods (e.g., Logan et al. 2004; Massey and Denton 1993; Reardon and Bischoff 2011). For example, Reardon and O'Sullivan (2004:122) apply this assumption by conceptualizing segregation as "the extent to which individuals of different groups occupy or experience different social environments" (see also Wilson [1987] 2012).

Yet, living in different neighborhoods is not synonymous with experiencing different social environments. Over the course of their everyday lives, people can, and do, leave their residential neighborhoods; they travel throughout the city for work and leisure, creating multiple opportunities for contact with those in other neighborhoods and experiencing other social environments, local communities, and cultural practices (Browning, Calder, Krivo, et al. 2017; Krivo et al. 2013; Small 2004; Wikström et al. 2010). In addition, people have opportunities to interact with nonlocal residents who visit their neighborhood. As a result, residents of different neighborhoods may be more or less connected as a function of their own as well as others' everyday geographic movements. Although this movement between neighborhoods is essential to cities' social integration, the connectedness created by mobility has been largely ignored by scholars, due in part to data limitations.

Fortunately, rich, fine-grained data at large scales are increasingly available (Akhavan et al. 2019) to measure people's geographic movements as an important part of social integration. This article takes those movements seriously. Drawing on Blau (1977, 1994), Blau and Schwartz (1997), and others, we posit that (a) opportunities for contact are essential to a city's social integration and (b) they depend on people's quotidian mobility patterns across its neighborhoods. We propose that a city's *structural connectedness* is the extent to which its neighborhoods are tied to one another by the movement of their residents, and we conceive of the city as a network in which neighborhoods are vertices and residents' travels between neighborhoods are edges (Sampson 2012:309-28). We develop two measures of cities' structural connectedness: one based on the degree to which neighborhoods visit each of the others in equal proportion, what we call the "equitable mobility index" (EMI); the other based on the extent to which travels are concentrated in a handful of receiving neighborhoods, what we call the "concentrated mobility index" (CMI). As we discuss below, these measures capture unique and important aspects of the structural connectedness of the city, which, to our knowledge, have not been previously conceptualized at this level of refinement. Thus, this article contributes to the theoretical understanding of social integration based on a perspective wherein residents' everyday mobility patterns are central.

Our measures are also general and can be applied to any data sets that capture patterns of geographic mobility such as data based on social media posts, cell phone movement, or GPS tracking. We illustrate the value of our new measures by applying them to a data set of 650 million geocoded tweets sent by 1.3 million Twitter users over 18 months, building on a recent approach to estimating social isolation (Wang et al. 2018a). We uncover important new features of U.S. cities and provide the tools for researchers to apply our measures to other data sets, other cities, and other levels of aggregation.

Conceptualizing Structural Connectedness

Our article combines a geographic mobility-based approach to cities' integration with a growing literature that examines the connectedness between neighborhoods from a network perspective (Browning, Calder, Stoller, et al. 2017; Hipp and Boessen 2017; Neal 2012; Papachristos and Bastomski 2018; Sampson 2012). Although analyses of mobility patterns have increased rapidly with the emergence of large data sets that contain geographic information, these studies frequently focus on patterns among individuals rather than between neighborhoods (Barthelemy 2016; Gabrielli et al. 2014; Jiang et al. 2016; Jurdak et al. 2015; Wang et al. 2018a). Prior research at levels higher than the individual has typically examined travel between rather than within cities (Lenormand et al. 2015) or focused on travel within a city exclusively between home and work (Graif, Lungeanu, and Yettera 2017; Louail et al. 2015). Studies that have examined travel patterns between areal units within cities generally use larger boundaries and small numbers of cities (Bora, Chang, and Maheswaran 2014; Shelton, Poorthuis, and Zook 2015; Zhong et al. 2014). By contrast, we examine neighborhood (i.e., block group) connectedness within cities.

Understanding the structural connectedness of cities is important for several reasons. First, the greater the movement of people between two neighborhoods, the more one can expect social ties to form across them. The formation and existence of such ties are important for several outcomes. Sampson (2012) found that in Chicago, when households relocated to different neighborhoods, they moved not just to places that were nearby but also to places in which residents likely had prior social connections. Between-neighborhood contact plays a role in understanding how people make movement decisions even in the midst of core social processes, such as gentrification, or major policy interventions, such as the demolition of housing projects or the Moving to Opportunity experiments (Sampson 2008). Research on crime in Chicago has also found that people commit crimes regularly with co-offenders who live not necessarily in nearby neighborhoods but in those that are socially connected (Papachristos and Bastomski 2018; see also Graif et al. 2017). The formation of such ties allows for information to travel between neighborhoods, an important issue for theories positing that the lack of access to information about jobs is a consequence of neighborhood poverty (Wilson [1987] 2012; but see Wang et al. 2018a).

Second, the greater the movement of people between two neighborhoods, the more one can expect diffusion across them, separate from the formation of social ties. Important forms of diffusion do not require people in different neighborhoods to know one another. One is disease. Many diseases, such as influenza, depend on physical contact or merely proximity, and the extent of contact between neighborhoods is related to the ease with which diseases of this kind travel between them (Balcan et al. 2009; Colizza et al. 2007). Another is cultural diffusion, as manifested in tastes, fashion, values, attitudes, and cultural practices (Small, Harding, and Lamont 2010). Cultural practices, such as the acceptance of openly gay couples or of interracial marriages, are displayed publicly in neighborhood sidewalks, cafés, parks, and other public places (Brown-Saracino 2017; Ghaziani 2014), and the extent to which people travel between two neighborhoods will shape the extent to which the cultural practices travel as well. Similarly, contacts in one's "activity spaces," or the places visited on a regular basis, can predict social values and mores (Browning, Calder, Stoller, et al. 2017).

Third, such interneighborhood connections shape the movement of ideas, information, attitudes, beliefs, and practices about the city as a whole. Conceiving of the city as a network of neighborhoods with greater or lesser movement across them creates the possibility of understanding a vast array of larger scale questions: Whether ideas, information, cultural practices, and even diseases are likely to spread faster in some cities than others; how much cities depend on the presence of neighborhood hubs or "hot spots," such as Grand Central Terminal in New York City or Grant Park in Chicago, for their residents to come in contact with one another; how much the traditional obstacles to integration, such as racial or income residential segregation, are likely to limit the formation of social networks and the building of community in a city; how successful cities are likely to be in countering isolation with interventions that incentivize residents to leave their neighborhoods; and many more. In short, adequately measuring what we call cities' "structural connectedness" allows an understanding of integration that is at once far deeper, much subtler, and substantially more powerful.

Measurement Strategies

Connectedness has been conceived and measured in numerous ways. Wellman (1979) proposes that a community's connectedness can be understood from a network perspective as a function of the structure of ties in the community. Although his analysis focuses on ties between individuals within two cities, his approach proves instructive, since it points to the usefulness of network measures to capture integration (see also Fischer 1982). Sampson (2012:309-28) measures connectivity as the ties between neighborhoods, focusing on both a neighborhood's outdegree (moves out of the neighborhood) and indegree (moves into the neighborhood). Outdegree and indegree are classic measures of network structure that can be thought of as "expansiveness" and "popularity," respectively, though their interpretation depends on the particular network's context (Wasserman and Faust 1994).

Some of the most important nonnetwork measures of connectedness to date measure the opposite of integration: segregation, specifically residential segregation. Measures of residential segregation reflect a long body of theoretical and empirical research on what segregation means and how best to measure it (e.g., Duncan and Duncan 1955; James and Taeuber 1985; Massey and Denton 1988; Reardon and Bischoff 2011; Reardon and Firebaugh 2002). In their seminal article, Massey and Denton (1988) propose five dimensions of residential segregation: evenness, exposure, concentration, centralization, and clustering. They note that evenness and exposure account for a large share of the variance among a host of segregation measures.

We develop network indicators of structural connectedness that bear a resemblance to Massey and Denton's (1988) measures of evenness and concentration but that are nonetheless distinct. Specifically, we conceive of structural connectedness as the extent to which neighborhoods in a city are tied to one another by the travel of residents across neighborhoods, where travel creates the opportunity for social contact. We do not presume that opportunities for contact guarantee contact—only that the absence of these opportunities precludes it. In this sense, structural connectedness may be thought of as an essential precursor to macrosocial integration in a city (see also Blau 1977). Moreover, although (static) residential segregation likely influences (dynamic) macrosocial integration (or the lack thereof), our measures are based on mobility patterns *between* neighborhoods to account for the broader, lived experiences of residents beyond the boundaries of their residential neighborhood.

We note that the meaning of the term "structure" is subfield-specific. In the networks literature, the term is widely understood to refer to the number of vertices and nature of edges in a network; in the urban sociology literature, the term has diverse applications that range from urban policies, institutional quality, and physical boundaries to economic and educational opportunities. Here, we embrace the network conceptualization and refer to structural connectedness as the extent to which neighborhoods are tied by flows of people between them—independent of spatial proximity and institutional, social, or economic similarity. Our conceptualization and measurement of connectedness is distinct from Krackhardt's (1994), which quantifies connectedness as the reachability between entities in a network.

Defining Network Measures

Networks consist of vertices (entities of a similar type) and edges, which indicate a relation between two vertices (Wasserman and Faust 1994). A network's edges can be binary (representing the presence or absence of a relation) or valued (representing the frequency or proportion of a relation). Additionally, the edges can express undirected relations, such as marriage (which are symmetric between the vertices), or directed relations, such as international exports (which flow from one vertex to the other). The movement between neighborhoods is best represented via valued, directed networks because (a) the frequency or proportion of all travels from one neighborhood to another is more meaningful than whether any resident from one neighborhood visited another and (b) visits between two neighborhoods are not necessarily reciprocated (Balcan et al. 2009; Colizza et al. 2007). Whereas individuals' movement patterns constitute the connections between neighborhoods, the separate visits can be aggregated to construct edge weights between two neighborhoods, either as the frequency of visits or the proportion of visits sent from one neighborhood. As a proportion, the edge weights indicate the proclivity of a neighborhood's residents to visit a particular neighborhood relative to all other neighborhoods. This approach illuminates the overall structure, or the consistent patterns in the relationships, of a city's mobility network.

Our first measure of cities' structural connectedness quantifies the concentration of visits between neighborhoods by calculating the Gini coefficient of neighborhoods' indegree centrality. Degree centrality quantifies the prominence of vertices in a network based on their edges (connections) to all other vertices; additionally, the dispersion and variance of a network's degree distribution represents heterogeneity or inequality in vertices' degree. In a directed network, vertices' degree can be separated into their indegree and outdegree. The former is the sum of all edges sent to a vertex; the latter is the sum of all edges sent by a vertex. For a directed, weighted network, indegree and outdegree centralities are typically found by summing the weighted edges (Newman 2001).[1] For a directed network $N\!: = (V, E, W)$, such that V is the set of all vertices, E is the set of edges, and W is the weighted adjacency matrix, the indegree centrality of each vertex is calculated using equation (1):

$$C_D^W(i) = \sum_{i}^{V_n} w_{i,j}\ ,\ i \neq j. \tag{1}$$

Given our interest in interneighborhood ties, equation (1) stipulates that $i \neq j$ because the network does not contain loops (edges from a vertex to itself). Networks' degree distributions depend on the size of the network. Therefore, the distributions should be normalized. For example, the variance of the distribution can be divided by the maximum possible variance to generate a dimensionless index or the distribution can be assessed using Freeman's (1978) formula for graph centralization that produces values bound between 0 and 1 (Snijders 1981; Wasserman and Faust 1994). These procedures create values that are comparable across networks of different sizes, which is important, given that cities vary in their numbers of neighborhoods.

Gini coefficients, which are frequently used in social stratification research, measure dispersion in distributions and generate values bound between 0 (maximum equality) and 1 (maximum inequality). We derive our CMI by calculating the Gini coefficients of neighborhoods' indegree centralities for each city's network. CMI values approaching 1 denote high levels of concentration in visits between neighborhoods; in other words, a few neighborhoods receive a much larger proportion of all possible visits relative to the other neighborhoods. Inversely, CMI values closer to 0 denote less concentrated shares of visits across a city's neighborhoods. Figure 1 illustrates three example mobility networks along with their adjacency matrices to demonstrate the utility and shortcoming of the measure.

All of the mobility networks contain six neighborhoods (vertices) that have outdegree centralities of 1. Panel A displays an even connectedness network, panel B a maximum uneven connectedness, and panel C a hub connectedness network. In panels A and B, the neighborhoods have indegree centralities of 1, but the two differ greatly despite having identical indegree and outdegree distributions. In the even connectedness network (panel A), each neighborhood's visits are evenly distributed across all other neighborhoods, but in the maximum uneven connectedness network (panel B), each neighborhood visits only one other neighborhood. Despite the clear disparity between their structures, the CMI, centralization index, and normalized

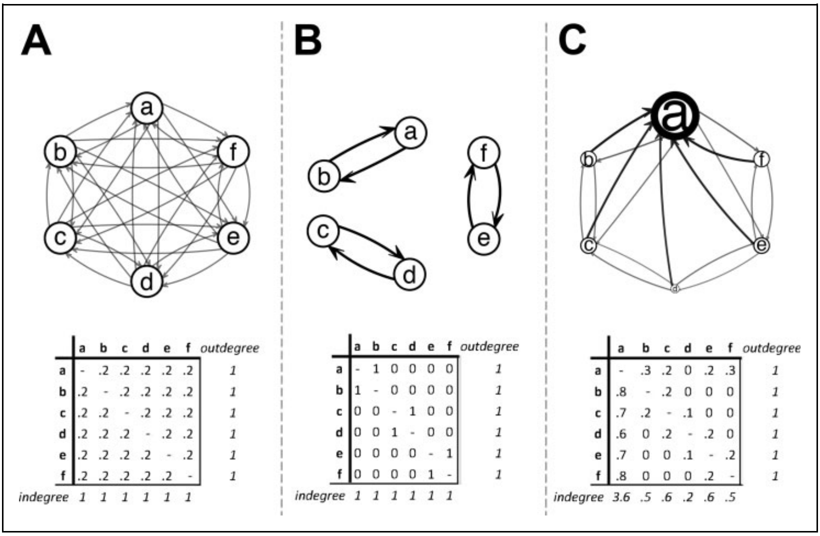

Figure 1. Example networks of connectedness. (A) Even connectedness. (B) Maximum uneven connectedness. (C) Hub connectedness.

variance all equal 0. The hub connectedness network (panel C) depicts a hypothetical mobility network, such that neighborhoods send more than half of their visits to neighborhood "a"; we refer to this as hub connectedness since one neighborhood has a much greater indegree centrality relative to the others.[2] Its CMI is 0.58. The CMI (as well as indegree centrality) can illuminate when one (or a few) neighborhood(s) is disproportionately visited or when it receives a higher concentration of visits relative to all other neighborhoods in a city. Yet, the measure poorly quantifies a city's equitableness or the extent to which neighborhoods are connected in equal proportion to each other. This motivates the development of our second measure.

The CMI, or nodes' indegree centralities, reduces the adjacency matrix to a vector, which can obfuscate important differences between networks' structures even if they have identical degree distributions (see panels A and B of Figure 1). Accounting for differences in the adjacency matrices requires comparing the matrices element-wise to preserve the second dimension. Figure 2 illustrates the utility of examining the elements of the adjacency matrices compared to the indegree centralities. Using the same networks from Figure 1, the top row shows the cumulative frequency distributions

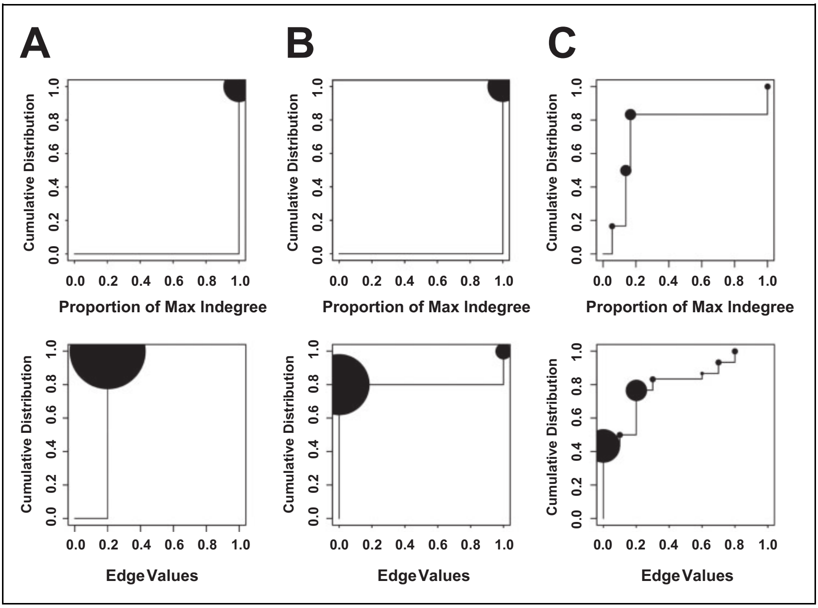

Figure 2. Cumulative frequency distributions of example networks' indegree and edge values. (A) Even connectedness. (B) Maximum uneven connectedness. (C) Hub connectedness.

of neighborhoods' indegree centralities normalized by the maximum indegree for each mobility network, and the bottom row shows the cumulative frequency distributions of the elements in each adjacency matrix (i.e., visits between neighborhoods or edge values). The black circles are scaled by the number of observations for each value.

In the top row, panels A and B show identical indegree distributions. All vertices have an indegree equal to 1, and their CMI (Gini coefficients) equal 0. The distribution for the hub connectedness network (panel C) shows a skewed indegree distribution, such that five of the neighborhoods have indegree centralities that are less than 20 percent of the highest indegree. The distributions of edge values, however, exhibit marked contrasts from the indegree distributions. Panel A on the bottom row displays an even distribution of visits between neighborhoods since all of the values are the same; all 30 edges equal 0.2. In contrast, the distribution of edge values in panel B is highly uneven with 24 edges equal to 0 and 6 equal to 1. In panel C, the distribution of edge values is less skewed than the indegree distribution; 13 of the edges equal 0, but 7 equal the evenness values (0.2), while the right tail is shorter than in panel B.

Taken together, the cumulative frequency distributions of indegree and edge values underscore the importance of accounting for the edge weights to assess structural connectedness. Despite having identical indegree distributions, the even connectedness and maximum uneven connectedness networks differ substantially, which is shown in the bottom row. Additionally, the distribution of edge values is less skewed than the distribution of indegree centralities for the hub connectedness networks, even though the max indegree is highest for this network. The CMI better identifies the degree of concentration in neighborhoods' total visits received, but it omits important information regarding the equity in the neighborhoods' visits to each other. Our second index illuminates the latter more effectively. Both are important for understanding cities' mobility-based structural connectedness, though we believe that examining visits between neighborhoods (the elements of an adjacency matrix) yields greater insights into mobility networks' structures overall.

We use Hamming distance to measure equality in visits between pairs of neighborhoods in each city's mobility network and construct our second measure of cities' structural connectedness: the Equitable Mobility Index (EMI). Hamming distance is the absolute value of the element-wise differences between two adjacency matrices of the same size; put another way, it is the sum of the edge changes (deletions and additions) necessary to make two networks of the same size identical (Butts and Carley 2005; Hamming 1950). The formula for calculating Hamming distance appears in equation (2):

$$\sum i \sum j |W_A^{i,j} - W_B^{i,j}| \text{ , } i \neq j. \tag{2}$$

Keeping with the previous notation, W is the weighted adjacency matrix for networks A and B, respectively; i and j are vertices in the adjacency matrices. As mobility networks, W contains the proportions of visits between neighborhoods (i and j) in hypothetical cities A and B.

Intuitively, if the adjacency matrices are identical, then the Hamming distance equals 0. Returning to Figure 1, the Hamming distance between the mobility networks in panels A and B is 9.6 since 0.8 needs to be removed from each element in adjacency matrix B that equals 1 and redistributed equally across the off-diagonal elements in the same row (i.e., each of the four elements add 0.2). This creates a Hamming distance of 1.6 for each row, and there are six rows. The Hamming distance between networks A and C is 5.6, and it is 8.4 between B and C.

The Hamming distance between the even connectedness network (panel A) and the maximum uneven connectedness network (panel B) quantifies the maximum observable Hamming distance between a mobility network and the even connectedness network. We leverage this quantity to create our EMI measure. Hamming distance requires two networks of equal size, so we compare each city's mobility-based network to the even connectedness network of the same size. The even connectedness network has a value of $\frac{1}{N-1}$ in all of the off-diagonal elements of its adjacency matrix, where N is the number of neighborhoods; put another way, each neighborhood visits all other neighborhoods in equal proportion.

The Hamming distance quantifies how much the observed mobility network would have to be altered to become the even connectedness network; we define this as $\mathrm{HD}_{\mathrm{Obs}}$. Similar to Freeman's (1978) seminal centralization index, we then divide $\mathrm{HD}_{\mathrm{Obs}}$ by the maximum possible Hamming distance ($\mathrm{HD}_{\mathrm{Max}}$) for a network of that size to facilitate comparisons across cities of different sizes. The formula for the latter value is given in equation (3):

$$\mathrm{HD}_{\mathrm{Max}} = \frac{2N*(N-2)}{N-1}. \tag{3}$$

If each neighborhood only visits one other neighborhood, then the city's mobility network is the maximum distance from the even connectedness network. To make the maximum uneven connectedness network identical to the even connectedness network, all but $\frac{1}{N-1}$ of each neighborhood's outgoing visits must be redistributed to $(N-2)$ neighborhoods in order to evenly distribute the visits across all possible neighborhoods. The numerator is multiplied by $2N$ because each neighborhood's visits needs to be subtracted from the neighborhood that previously received all of the visits from a neighborhood and added to all of the other neighborhoods. Using these values, we calculate the EMI using equation (4):

$$\mathrm{EMI} = 1 - \frac{\mathrm{HD}_{\mathrm{Obs}}}{\mathrm{HD}_{\mathrm{Max}}}. \tag{4}$$

Dividing the observed Hamming distance by the maximum Hamming distance bounds the values between 0 and 1. Subtracting this quotient from 1 generates an index, such that EMI values closer to 0 are less even and values closer to 1 are more even.[3] Similar to the denominator in Freeman's centralization index, the maximum Hamming distance is influenced by the size of the network (Anderson, Butts, and Carley 1999; Butts 2006).[4] This aspect is inconsequential, though, since we scale observed values by the theoretical maximum value for the network given its size. Thus, both EMI and CMI indicate the proportion of the observable value out of the maximum possible value.

In sum, the CMI illuminates when a neighborhood receives a disproportionate concentration of visits, whereas the EMI quantifies how much observed mobility networks would need to be altered to have the mobility patterns evenly distributed across neighborhoods. Together, the measures elucidate the structural connectedness of mobility networks and can be used to compare sets of networks. Importantly, the methods delineated above are applicable to any mobility data with fine-grained temporal and spatial resolution; they can also be applied at different geographic scales.

Measurement Assumptions

Any attempt at measuring connectedness on a large scale requires simplifying assumptions to make the analyses tractable. Our first assumption is that each resident of a city is equally important for a neighborhood's contact with other neighborhoods. Practically, to create our edge lists, we normalize each observed resident to have an outdegree of one by dividing their visits to all other neighborhoods by their total numbers of visits outside of their residential neighborhoods. This step makes each resident's visits to a neighborhood a proportion of their total visits.

Second, we assume that each neighborhood has equal importance in the structural connectedness of a city. Accordingly, we normalize each neighborhood to have an outdegree of 1 by dividing its residents' aggregated proportions of visits to other neighborhoods by the sending neighborhood's total number of residents. The normalization aligns with our focus on a city's structural connectedness at the neighborhood level.[5]

Third, we assume that travelers to a city play a different and less important role than residents in the structural connectedness of a city, as we define it. Accordingly, we remove travelers from the calculation of CMI and EMI. This decision aligns with prior work on residential segregation, which focuses on the population of individuals whose primary residence is within the location under analysis. Hence, this assumption allows us to make a stronger link between our connectedness measures and static segregation measures.

Our final assumption is that equitable connectedness within a city would result from all neighborhoods having equivalent values for their outdegree to, and indegree from, all other neighborhoods. Stated another way, where neighborhoods' residents equally visit and are visited by all other neighborhoods' residents. This assumption may be problematic if some visits are transitory and yield little interaction. Yet, exposure and the potential for interaction are important in their own right (Blau 1977, 1994; Moody 2001; Mouw and Entwisle 2006). At a minimum, visiting another neighborhood, no matter how fleeting, is a precondition for exposure. Our aim is to directly measure this type of exposure.

Estimating Everyday Geographic Mobility: An Application

Thus far, we have used fictitious, purposefully small mobility networks to demonstrate how we derive our measures. Constructing mobility-based measures of neighborhood connectedness for multiple cities requires large-scale geographical information on individuals' residential locations and movement patterns over an extended period of time. Although conceptually straightforward, few data sources meet these requirements. Twitter, however, provides high-resolution data on when and where micromessages, tweets, are sent by users that opt in to Twitter's geotagging service; this includes the latitude and longitude for a tweet to a fraction of a second (Luo et al. 2016; Sutton et al. 2015). We use the same corpus of tweets as Wang et al. (2018a), which comprises over 650 million geotagged tweets sent over 18 months by 1.3 million Twitter users (from October 1, 2013, to March 31, 2015) in the areas surrounding the 50 largest cities of the United States.[6] The spatial granularity and temporal scale of the data provide a high level of detail regarding where people move within the cities over 500 days. Below and in the Appendix (which can be found at http://smr.sagepub.com/supplemental/), we present further details on the reliability and validity of Twitter data for present purposes.

Procedures

To construct cities' mobility networks, we need urban residents' estimated home locations and the neighborhoods they visit. Wang et al. (2018a) used machine learning to estimate individuals' residential block groups from the latitude and longitude provided by Twitter for publicly available, geotagged tweets. We, too, use the density-based spatial clustering of applications with noise (DBSCAN*) algorithm because it deterministically identifies clusters, enables user-specified minimum cluster sizes as well as distances between points comprising a cluster, and efficiently handles large data sets (Birant and Kut 2007; Campello, Moulavi, and Sander 2013). Although geotagged tweets provide high geographic resolution, noise caused by device factors or weather can affect the precision of the location. DBSCAN* ameliorates this potential precision issue by comparing the distances between points and then classifying points as part of a cluster or as noise based on the user-specified parameters. We use the *DBSCAN* package in R with the specifications that clusters (a) contain at least three data points (tweets), such that (b) each point is no more than 0.0004 degrees (approximately 56 meters) apart from two other points within the cluster (Hahsler and Piekenbrock 2017). Points that do not meet both criteria are classified as noise (or outliers) and are not part of any cluster; this includes border points, which makes the algorithm fully deterministic (Luo et al. 2016; Soliman et al. 2015).[7]

For each city, we collect all tweets within boundaries larger than the city, so that we can properly distinguish between suburban residents that regularly commute to a city, tourists, and urban residents. The larger areas are bounding boxes that extend beyond cities' commuting zones according to the U.S. Department of Agriculture Economic Research Service (Parker 2012). To identify individuals' home locations, we use the subset of their tweets that are sent between 8 p.m. and 12 a.m. local time on Monday through Thursday, assuming that most individuals send most of their tweets from their residence during this time (see also Jiang et al. 2016). Then, we apply DBSCAN* to all locations within this subset to identify clusters.[8] Next, we find the centroid of the largest cluster (based on the number of points) and assign that centroid as the individual's approximate home location. If an individual has more than one cluster equal to the maximum cluster size, then we assign their home cluster based primarily on the length of time between the first and last tweet in the cluster and secondarily on the geographic compactness of the cluster. After each individual is assigned a single home cluster, the centroid of the cluster is spatially joined with its block group using PostgreSQL v9.3 and PostGIS 2.4.7.[9] This process provides precision in estimated locations at the block group level, hence restricting the ability to pinpoint an individual's home address (for further details, see Wang et al. [2018a:7736 and SI]).

The procedure estimates individuals' residential block groups in ways that improve upon previous research. Yet, these steps do not adequately address the presence of travelers to a city, who have divergent mobility patterns from residents (Gabrielli et al. 2014). We remove these travelers from a city's data for two reasons. First, the 50 cities experience different rates of tourism, so the effects would be heterogeneous across the cities. Second, this aligns with prior work on residential segregation, which focuses on the population of individuals whose primary residence is within the location of analysis. To identify the residents of a city, we remove individuals who have less than

30 days between their first and last tweet within a city. If an individual has home clusters in multiple cities, we then compare the number of tweets they sent during the home period (8 p.m. to 12 a.m. from Monday through Thursday) and the number of days between their first and last tweet in each of the cities. We then assign individuals to only one city based on these characteristics. After all individuals have only one home location, we retain individuals whose home block groups are inside the city boundaries of the 50 most populous cities. These additional data processing steps increase our confidence that we have more accurate estimates of individuals' home block groups than previous research using geotagged tweets (Jiang, Li, and Ye 2018; Malik et al. 2015).

Our procedures generate a data set for each city comprising individuals with estimated residences in cities' block groups and each time they uniquely visited (tweeted from) any block group in the city. The data set contains 133,766,610 geotagged tweets sent by 375,504 individuals. As described above, to control for differences between individuals' tweeting rates, we divide individuals' numbers of visits (tweets) to any block group in the city by the number of tweets each individual sent, excluding tweets sent in their home block group. For each block group, we find the mean proportions of visits to all other block groups by summing residents' proportions to visited block groups and dividing each sum by the number of residents (Twitter users in our sample) in the block group. This yields a data set of normalized visits from each block group to all other block groups within a city with values bound between 0 and 1. These data are valued edge lists, which we use to construct each city's directed, weighted network, whose size equals the number of block groups. Each edge indicates the proportion of visits from one block group to another.

Data Assumptions

Our use of Twitter-based mobility data requires additional assumptions beyond those general to our measures. First, we assume that Twitter data may be used to make valid inferences about interneighborhood mobility. This assumption requires addressing three potential sources of bias: Twitter users may not be representative of the population; individuals who geotag their tweets may not be representative of Twitter users or the population; and geotagged tweets may not be representative of travels between neighborhoods. As we discuss in the Appendix (which can be found at http://smr.sagepub.com/supplemental/), the third of these, which ultimately encompasses the first two, is the core threat to our inferences, since it pertains to general travel patterns in cities rather than individuals' travel patterns. In the Appendix (which can be found at http://smr.sagepub.com/supplemental/), we review existing studies that address the issue by comparing mobility patterns derived from Twitter data with those derived from other data sources such as travel diaries and commuting patterns. The comparison studies suggest that Twitter data are appropriate for making our inferences.

Second, we assume that individuals send geotagged tweets evenly during their daily activities. If individuals tweet more frequently from particular areas or when engaging in particular activities, then this could bias our estimates of their mobility patterns. We use the same corpus of tweets as Wang et al. (2018a), and their results for individuals' travel distances and interactions across demographic groups aligned with prior work using GPS, cell phone data, and travel diaries (Jones and Pebley 2014; Krivo et al. 2013; Palmer et al. 2013).

We also conducted two empirical tests of these assumptions. In the most extensive, we apply our method to an alternative data source based on cell phone GPS signals for millions of users in Houston, and we successfully replicate a key finding reported below. We discuss this replication study in greater depth in the Appendix (which can be found at http://smr.sagepub.com/supplemental/). These consistent findings across multiple types of data lend credence to our use of geotagged tweets to study geographic mobility.

To assess whether or not individuals would selectively enable and disable the geotagging service, thereby undermining representativeness, we randomly selected 5,000 individuals who sent at least one geotagged tweet from June 9 to June 15, 2018. Using the streaming application programming interface from Twitter, we followed these users for about a one-month period (June 17 to July 19, 2018) and collected all of their tweets. We found that all tweets from these users were geotagged during this period.[10] Thus, we are quite confident that the vast majority of individuals who geotag their tweets do not alternate between geotagging and not geotagging their tweets.

Finally, along with these Twitter-specific assumptions, our general assumptions for the measures have some implications in the context of geotagged Twitter data. Regarding our first methodological assumption (individual-level normalization), we assume that individuals' propensities to tweet outside of their neighborhoods are unrelated to their levels of mobility. In other words, we believe that individuals' mobility patterns are more equal than their tweeting patterns. Not making this assumption would imply that people who tweet more travel more and that the two have a linear relationship. The notion that the number of tweets an individual sends is a linear function of their mobility is clearly dubious. Failure to normalize and treat individuals equally would dramatically overweight individuals that more frequently tweet.[11] Regarding the second general assumption (neighborhood-level normalization), neighborhoods exhibit sizable variation in numbers of observed Twitter accounts. This difference might reflect variation in population across neighborhoods to some extent, but the correlation between neighborhoods' populations and numbers of Twitter users is modest (0.48). These two normalizations are necessary to generate comparable, unbiased mobility networks. Again, we stress that our main contribution is the development of connectedness measures; geotagged tweets are merely the application.

Results

Our first measure of structural connectedness, the CMI, quantifies the disproportionate concentration of visits to a few neighborhoods within a city. Cities' CMIs range from 0.415 (Miami) to 0.585 (San Diego), indicating that visits from residents of other neighborhoods are most concentrated in a small share of neighborhoods in San Diego. The mean and median values are 0.506 and 0.505, respectively. Additionally, the distribution is quite compact with 38 of the values lying within one standard deviation of the mean. Our second measure of structural connectedness, the EMI, quantifies how closely neighborhoods' mobility patterns in a city approximate a scenario in which each neighborhood visits all other neighborhoods in equal proportion. Cities' EMIs range from 0.048 (New York City) to 0.226 (Raleigh), with mobility patterns being more evenly distributed in the latter. The mean and median values are 0.137 and 0.142, respectively. The distribution of cities' EMIs is also quite compact with 35 of the cities' values lying within one standard deviation of the mean.

The correlation between EMI and CMI is only −0.033, revealing that the two measures, indeed, capture distinct aspects of the structural connectedness of the cities' mobility networks. Each city's EMI and CMI values appear in Figure 3. Larger cities tend to have smaller EMI values, which is expected because it is more difficult for residents of larger cities to visit all neighborhoods (and unlikely they would do so in equal proportions), especially those that are further away from their residential neighborhoods. In contrast, cities' CMI values do not evince as strong of a relationship with size. Although many of the cities are in the center of the plot, four cities stand out based on their EMI and CMI values: New York City, San Francisco, Detroit, and

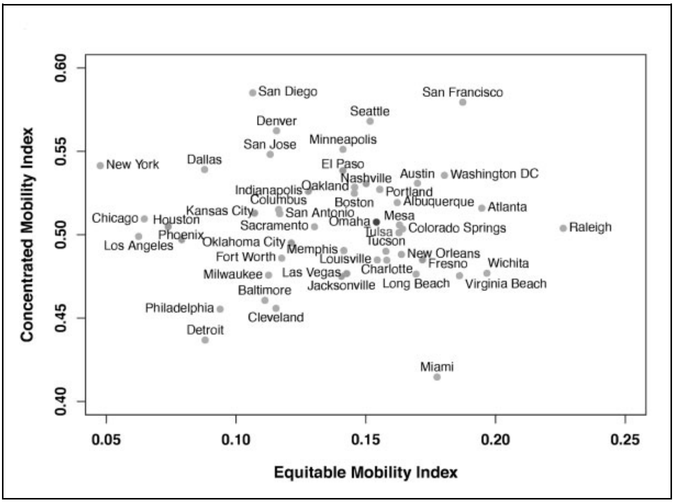

Figure 3. Relationship between cities' equitable and concentrated mobility indices.

Miami. New York City has the lowest EMI value and the seventh highest CMI value. San Francisco has the second highest CMI value but the fourth highest EMI value. Detroit has the seventh lowest EMI value but the second lowest CMI value. Miami has the lowest CMI value and the seventh highest EMI value. Their positions outside of the central cluster of cities bear further scrutiny, and we delve deeper into the structures of these cities' mobility networks to deduce what factors might drive these differences.

In Figure 4, we show neighborhoods' indegree colored by their values in these four cities. The neighborhoods are colored from low (dark blue) to high (dark red) indegree, and black indicates neighborhoods with indegree centralities greater than 5. The spatial clustering of low and high indegree neighborhoods is immediately apparent in all four panels. Yet, noticeable differences across the cities are clear. The heat maps for New York and San Francisco contain more compact clusters of similar degree values than Detroit and Miami. Furthermore, the numbers and spatial distributions of hubs (neighborhoods with indegree centralities greater than 5) are discernible. In New York City and San Francisco, 2.3 percent and 2.1 percent of all neighborhoods are hubs, whereas only 1.4 percent and 1.5 percent of neighborhoods are hubs in Detroit and Miami. Conversely, 16.3 percent and 20.9 percent of neighborhoods in New York City and San Francisco have indegree values below 0.25 compared to 9.9 percent and 4.9 percent of neighborhoods

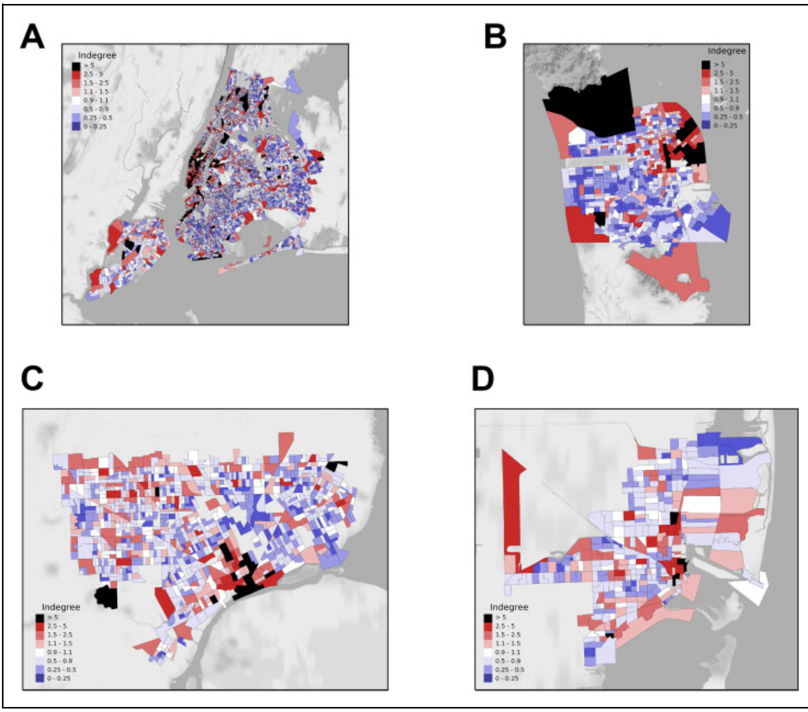

Figure 4. Heat maps of neighborhoods' indegree in four cities. (A) New York. (B) San Francisco. (C) Detroit. (D) Miami. Four block groups in San Francisco that border the water are excluded from map B because the city's shape file for them includes water extending across the bay. They are included in the analyses, however. We retain the Presidio and Golden Gate Bridge block group in the map, the large black hub in the northwest part of the city.

in Detroit and Miami. The index of concentrated mobility captures these differences.

For New York, there is a concentration of high indegree values and hubs in Manhattan, and there is greater heterogeneity in Brooklyn and Queens. The neighborhood with the highest indegree (49.5) contains Penn Station. Hubs are present across the city, but the bulk of them are in Manhattan. This pattern indicates a polycentric urban form where multiple hubs exist (Zhong et al. 2014). In San Francisco, Outer Sunset and Sunset District show a higher concentration of low indegree values. Conversely, the Financial District and Telegraph Hill show the greatest concentration of high indegree values, and the neighborhood with the highest indegree value (24.1) is the Mission Street area near the Financial District. The hubs are highly concentrated in the downtown area with other neighborhoods that also have indegree centralities that exceed the threshold of equality, suggesting a monocentric urban form. Two additional hubs are surrounded by neighborhoods with low indegree centralities. In these two cities, the figures clearly illustrate that certain areas receive a substantially higher concentration of visits than other neighborhoods, which is why their CMI values are among the highest.

Detroit evinces greater spatial heterogeneity in neighborhoods' indegree. The downtown area displays a greater concentration of neighborhoods with high indegree values, whereas the neighborhoods east of downtown illustrate a concentration of low indegree values. However, Brightmoor and Warrendale (to the west of downtown) contain neighborhoods with heterogeneous indegree values. In Detroit, the most visited neighborhoods are primarily located downtown with three additional areas of concentration on the outskirts of the city. Similar to Detroit, Miami's downtown area has a high concentration of neighborhoods with high indegree value, and the neighborhood with the highest indegree (7.1) is in the middle of downtown. Also similar to Detroit, Miami's neighborhoods display less similarity for indegree values to their proximate neighborhoods compared to New York City and San Francisco. The overall lack of indegree similarity for nearby neighborhoods reveals that several neighborhoods receive more visits than their surrounding neighborhoods, such that hubs are not spatially compact.[12]

The neighborhoods visited most frequently in the four cities also provide face validity to our data. As we would expect, the downtown areas illustrate higher indegree values overall, and the neighborhoods most visited are sensible and reflect our ground truth for these cities. Similarly, we found the neighborhood most visited in Los Angeles contains the Staples Center and Pershing Square, and a neighborhood in the Northwest inner loop is most visited in Chicago. These consistent patterns with high face validity increase our confidence in Twitter as a data source for measuring mobility. Moreover, nearly all of the 38,505 neighborhoods in the 50 cities are visited at least once by another neighborhood, a substantially better coverage rate than could be achieved using travel diaries or surveys.

We further assess the differences in neighborhoods' indegree in these four cities with cumulative frequency distributions. These are shown in Figure 5, and each neighborhood's indegree is divided by the maximum indegree in the city, so that the distributions are on the same scale. In each city's plot, the

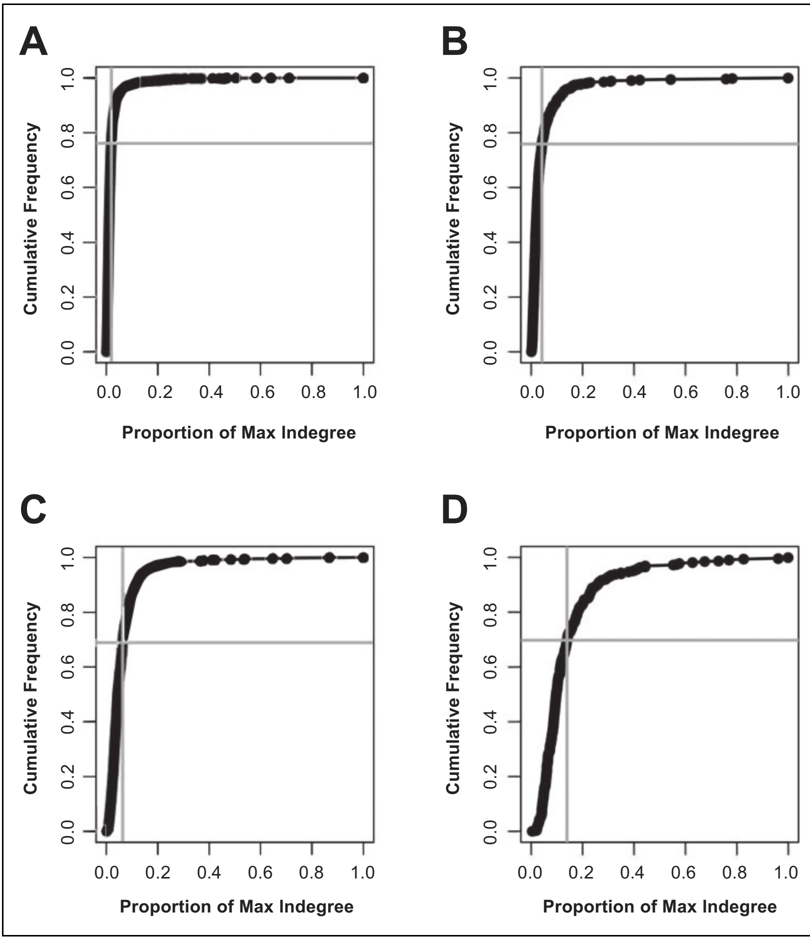

Figure 5. Cumulative distributions of proportional indegree in four cities. (A) New York City. (B) San Francisco. (C) Detroit. (D) Miami.

vertical, gray line indicates the indegree value for equal visits in a city; the horizontal line indicates the percentage of neighborhoods falling below this threshold. The distributions for New York City and San Francisco rise much more rapidly than the distributions for Detroit and, especially, Miami. Additionally, in New York City and San Francisco, 76.1 percent and 75.9 percent of neighborhoods, respectively, are below the equitable visits threshold compared to 68.9 percent in Detroit and 69.8 percent in Miami. The shape of the distributions for the first two elucidates why they have higher CMI values; a large majority of the neighborhoods in these cities receive far fewer visits than the most visited neighborhoods. In other words, New York City and San Francisco have more concentrated mobility networks. In contrast, the distributions for Detroit and Miami increase more gradually. Detroit and Miami have less concentrated distributions of visits across neighborhoods; hence, they have lower CMI values.

As shown in Figures 1 and 2, however, this lower concentration does not necessarily entail equitability in the overall exposure of residents from different neighborhoods to each other. Figure 6 shows the cumulative frequency distributions of the weighted edges (i.e., the elements of the adjacency matrices) for the four cities. The enlarged black circles reflect the (large) percentages of edges that equal 0. New York City still has a sharp increase and a long tail, but now Detroit's distribution takes this shape as well. The large number of zeros indicates a lack of connectedness in both cities where many neighborhoods are not visiting each other. Taken together with Figure 5, this means that a few neighborhoods receive a disproportionate share of visits in New York City, and many neighborhoods do not visit each other at all. For Detroit, the low CMI and EMI values indicate that the mobility network is cleaved, such that residents of the city neither travel to the same neighborhoods en masse nor do they travel to many of the neighborhoods in the city overall. Essentially, not only are Detroit neighborhoods disconnected, but they also lack an area where many of their residents come together. The combination of our measures therefore reveals distinct insights about the nature of a city's structural integration based on mobility.

The distributions for San Francisco and Miami have fewer zeros and increase more gradually, though the distribution for San Francisco does have more weight in its upper tail. For these two cities, residents visit a relatively larger share of the cities' neighborhoods. Taking into account Figure 5, a small number of neighborhoods in San Francisco receive a large share of residents' visits, but the remaining share of each neighborhood's residents' visits is distributed fairly equitably across neighborhoods. For Miami, neighborhoods' residents visit many of the neighborhoods in the city with few neighborhoods visited at exceptionally high rates. In other words, Miami's CMI and EMI indicate a high level of connectedness compared to other cities, as residents are highly exposed to one another.

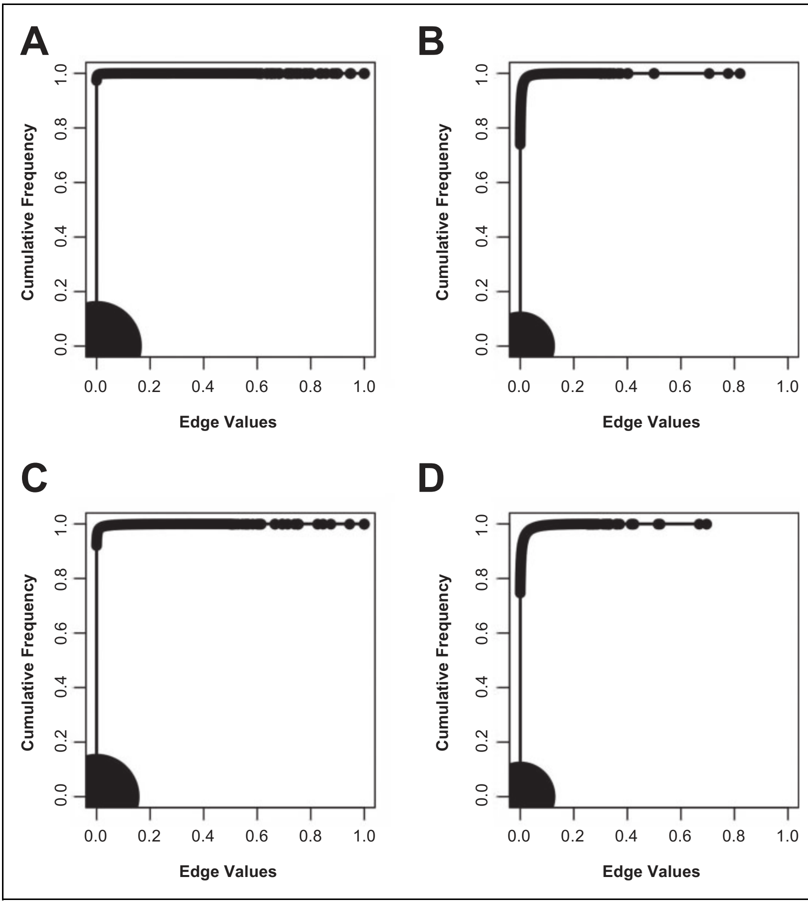

Figure 6. Cumulative distributions of edge values in four cities. (A) New York City. (B) San Francisco. (C) Detroit. (D) Miami.

Predictors of Connectedness

To better understand what features of cities correspond to their levels of connectedness, we estimate variation in the EMI across the 50 cities using ordinary least squares regression models. Space constraints preclude us from properly discussing the predictors involved in understanding our two separate outcomes. For this reason, we focus only on the EMI, which is a more comprehensive measure of a city's structural integration.[13] At the end of this section, we also discuss the general pattern of results we obtained for CMI.

Our EMI estimates include a parsimonious set of predictors reflecting core sociological theories of social integration. Social scientists have long argued that the demographic and spatial characteristics of cities affect the character of their interactions. Foremost, cities with large populations and land mass will likely evince less connectedness, as the likelihood for interaction between any two neighborhoods is structurally more difficult (Blau 1977, 1994) and a large metropolis can be overwhelming (Simmel 1950). Relatedly, cities that are larger in land area may experience lower connectedness, as distance increases the time costs to travel between neighborhoods. Public transit may obviate some of these costs and facilitate interneighborhood flows. We measure public transit as the percentage of adult workers who use public transit to commute to work using data from the 2011–2015 American Community Survey (ACS).

We also include several measures of the cities' social characteristics based on 2011–2015 ACS data. Merton (1968) argued that cosmopolitans are highly likely to exhibit an extralocal orientation and view themselves as key components in the outside world. In the context of large urban cities and their neighborhoods, we expect that more cosmopolitan cities will have neighborhoods with greater connectedness to each other. We measure cosmopolitanism as the percentage of adults with a bachelor's degree. Blau (1977) theorizes that diversity increases the likelihood of intergroup contact. Still, diversity may not lead to neighborhood connectedness, as individuals often exhibit preferences for racial and ethnic homophily (McPherson, Smith-Lovin, and Cook 2001; Wang et al. 2018a). Indeed, social trust and solidarity tend to be lower in ethnically diverse neighborhoods (Putnam 2007). We measure diversity using Blau's diversity index, which is calculated as one minus the Herfindahl Index (see equation [5]).[14]

$$B = 1 - \sum P_r^2. \tag{5}$$

Finally, racial and income segregation may spatially and socially cleave cities, thereby reducing connectedness. Although segregation may counteract the tendency to "hunker down" that Putnam (2007) observes in diverse neighborhoods, it may also yield the pernicious consequence that city residents are less likely to visit a wide range of neighborhoods. We calculate a Theil index of multigroup racial segregation as:

$$H_{\text{race}} = \frac{1}{E}\sum_{r=1}^{R} \pi_r \sum_{j-1}^{J} \frac{t_j}{T} p_{jr} \ln(p_{jr}), \tag{6}$$

where π_r is the proportion of individuals in racial/ethnic group r in the city, t_j is the total count of individuals in neighborhood j, T is the total count of individuals in the city, $p_{jr} = \pi_{jr}/\pi_r$, and $E = \sum_{r=1}^{R} \pi_r \ln\left(\frac{1}{\pi_r}\right)$.[15]

Finally, we calculate residential segregation by income using the block-group-specific household counts across the 16 income ranges included in the ACS.[16] Given the ordinal nature of this variable, we use Reardon and Bischoff's (2011) preferred rank ordered information theory (Theil) index to calculate income segregation.

$$E(i) = i \times \ln\left(\frac{1}{i}\right) + (1 - i) \times \ln\left(\frac{1}{1 - i}\right), \tag{7}$$

$$H(i) = 1 - \sum_{j=1}^{J} \frac{t_j E_j(i)}{TE(i)}, \tag{8}$$

$$H_{\text{income}} = 2 \times \ln(2) \int_0^1 E(i)H(i)di. \tag{9}$$

At any given value of income i, $E(i)$ is the entropy of the population divided into groups above and below the income threshold, $H(i)$ is the traditional information theory (Theil) index, and H_{income} is a weighted average of income segregation across the distribution. Reardon and Bischoff (2011) provide further information on calculating this index.

Our regression models are not causal estimates. Rather, our intent is to provide a description of predictive patterns based on externally derived measures emphasized in urban theory, even though our statistical power is limited with a sample size of 50 cases. Summary statistics of the variables appear in Table 1. Table 2 presents the regression results.

Model 1 estimates cities' EMI values based on these theoretically relevant covariates. As Blau and Simmel predict, more populous cities show significantly less connectedness. This relationship is not, however, a function of land area, which is unrelated to connectedness. Further, public transit seems to contribute little to evenness in neighborhood ties. Two social features of cities also emerge as salient predictors of connectedness. As Merton theorized, cosmopolitan cities show greater structural integration of neighborhoods. On the other hand, cities that are more residentially segregated by race demonstrate less connectedness. Diversity and income segregation are unrelated to EMI.

Table 1. Summary Statistics.

	Mean	Standard Deviation	Minimum	Maximum
Equitable mobility index	0.137	.038	0.048	0.226
Ln(population)	13.64	.634	12.902	15.955
Ln(land area)	5.785	.981	3.771	8.032
Percent public transit[a]	0.094	.122	0.006	0.596
Percent bachelor's[b]	0.341	.097	0.147	0.581
Theil race	0.288	.092	0.141	0.499
Theil income	0.16	.023	0.114	0.212
Blau heterogeneity	0.603	.091	0.348	0.766

[a]Calculated using workers not working from home.
[b]Calculated using individuals age 25 or older.

Table 2. Ordinary Least Squares Models of Equitable Mobility Index.

	Model 1	Model 2	Model 3
Ln(population)	−0.0447*** (0.0078)	−0.0465*** (0.0059)	−0.0435*** (0.0057)
Ln(land area)	0.0028 (0.0055)	0.0012 (0.0038)	0.0014 (0.0042)
Percent public transit	0.0407 (0.0464)		
Percent bachelor's	0.1047* (0.0393)		0.1082** (0.0365)
Theil race	−0.1013* (0.0476)		−0.0742† (0.0429)
Theil income	0.1723 (0.1543)		
Blau heterogeneity	−0.0283 (0.0395)		
Constant	0.7103*** (0.0844)	0.7649*** (0.0775)	0.7067*** (0.0678)
N	50	50	50
R^2	0.721	0.583	0.707
BIC	−218.1	−217.6	−227.4

Note: BIC = Bayesian information criterion.
***$p \leq .001$. **$p \leq .01$. *$p \leq .05$. †$p \leq .1$.

Model 2 investigates the amount of variance in cities' connectedness levels that is explained by basic demographics: population and land area. These variables explain 58 percent of the variance, and population accounts for nearly all of this. Adding measures of racial segregation and cosmopolitanism in model 3 explains an additional 12 percent of the variance. In sum, cities' populations and social characteristics explain a large share of the variation in their neighborhoods' connectedness levels. The pattern of these relationships aligns with long-standing sociological theories regarding interactions and integration in cities and countries.[17]

Extensions and Implications

In this article, we build on a sociological canon highlighting the importance of social integration (Blau 1977; Durkheim [1893] 1984, [1897] 1966), as well as a recent set of studies analyzing neighborhood connectedness from a networks perspective (Browning, Calder, Stoller, et al. 2017; Graif et al. 2017; Papachristos and Bastomski 2018; Sampson 2012), by developing measures of structural connectedness that identify the equitability and concentration of residents' visits between a city's neighborhoods. Our measures are applicable at any areal scale and with any data set containing movement data. Although we believe that the EMI offers a more comprehensive assessment of connectedness, the CMI can supplement the other measure by indicating that neighborhoods are disproportionately visited in a city.

Connectedness as a phenomenon is related, though conceptually distinct, from residential segregation by race or income. Although segregation can be thought of, broadly, as the extent to which individuals both reside in and experience different social spaces (Reardon and O'Sullivan 2004), research on residential segregation investigates the former almost exclusively and neglects exposure from or to outside neighborhoods. The focus on static measures of residential segregation stems in large part from a lack of available data on individuals' day-to-day mobility patterns for a large number of places and cities. Thus, while sociological research on residential segregation is impressively long running and empirically rich (Du Bois 1899; Duncan and Duncan 1955; James and Taeuber 1985; Lichter et al. 2015; Logan et al. 2004; Massey and Denton 1988; Massey and Denton 1993; Reardon and Bischoff 2011; Reardon and Firebaugh 2002), research on connectedness is much more limited. The general measures we develop, and subsequently apply to Twitter data, contribute to this fallow research area.[18]

We demonstrate the utility of our measures using a data set of approximately 650 million geotagged tweets that were sent in the nation's 50 largest cities. We uncover differences in the equitability and concentration of visits between neighborhoods in these cities, which has implications for social capital and cohesion, as well as the diffusion of culture, ideas, information, crime, diseases, and other social outcomes. Future research analyzing variation in neighborhood outcomes between or within cities should consider the role of residents' movements across neighborhoods. For example, recent research demonstrates that dynamic population changes of a neighborhood each day are much more important for its odds of experiencing a crime than static residential population counts (Boivin and Felson 2018). Moreover, the demographic and economic characteristics of individuals physically located in a neighborhood may vary quite dramatically throughout the day (Le Roux, Vallée, and Commenges 2017; Vallée 2018). These temporal fluctuations in cities' connectedness and exposure should be accounted for, when the data facilitate it (Wang et al. 2018b), and our measures could be further applied to mobility patterns during different times of the day.

Interestingly, we find that a city's demographic characteristics are important predictors of its levels of connectedness. In line with classic structuralist theory (e.g., Blau 1977; Simmel 1950), more populous cities are less structurally connected. Population size can account for over half of the variance in cities' levels of neighborhood connectedness. In addition, cities that are more cosmopolitan and less racially segregated tend to be more structurally connected. Whether racial segregation might lead to less neighborhood connectedness or neighborhood connectivity might reduce racial segregation is worthy of further exploration. Again, this analysis was descriptive and does not support causal conclusions. Nevertheless, the strong relationship between several theoretically relevant characteristics of a city and its level of social integration provides confidence in our measure of neighborhood connectedness.

Moving forward, we envision research that examines additional network properties, such as clique membership based on demographic attributes of block groups. Additionally, research could identify the presence and spatial distribution of hubs across cities or use community detection algorithms to assess another dimension of cities' structural connectedness (Zhong et al. 2014). The network data could be further analyzed using exponential family random graph models to illuminate the generative mechanisms that produce the observed mobility networks and then distinguish which characteristics of cities create similarities or differences between the networks (Handcock et al. 2008; Hunter et al. 2008). Conceptualizing urban mobility patterns as networks opens a host of research possibilities that we hope other scholars will pursue.

Conclusion

We have developed two indices of structural connectedness that quantify theoretically important, yet currently unmeasured, features of cities' social integration. We illustrated the value of these indices using Twitter data. Importantly, our measures do not depend on the use of these particular data; in the future, data from cell phones or other sources may generate better or larger samples for assessing neighborhood connectedness. Our indices can quantify structural connectedness with any mobility data that have sufficient coverage and resolution, such as GPS signals from cell phones, census commuting patterns, or public transit "tap-in" data. Applied to these data, the indices can further illuminate the structural connectedness of cities or other geographical areas, such as metropolitan statistical areas. Another aspect of our methods and measurements is that they can be applied to different geographic boundaries and at different geographic scales. In the age of big data, we are at an important moment to broaden our understanding of social integration to include a range of measures that reflect not just the places people live but also the spaces they experience together.

Acknowledgment

The authors thank the anonymous reviewers of *Sociological Methods & Research* for especially helpful comments.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by MacArthur Foundation Grant 14-107644-000-USP, National Science Foundation Grants SES-#1637136 and SES #1735505, and the Project on Race, Class and Cumulative Adversity at Harvard University funded by the Ford Foundation and the Hutchins Family Foundation.

ORCID iD

Robert J. Sampson https://orcid.org/0000-0003-4259-8146

Supplemental material

Supplemental material for this article is available online.

Notes

1. The number of vertices connected as well as edges' weights can be taken into account with a tuning parameter, but what value the parameter should be is not well defined (Opsahl, Agneessens, and Skvoretz 2010).
2. We are defining hubs as neighborhoods that receive a disproportionately large share of visits (i.e., significantly larger indegree centralities). Others have defined hubs in scale-free networks by assessing the degree centralities relative to a

power law distribution (Barabási and Bonabeau 2003) or based on vertices' degree and betweenness centralities (Butts, Petrescu-Prahova, and Cross 2007).

3. This transformation is similar to Blau's (1977) calculation of the ethnic heterogeneity index by subtracting the value of the Herfindahl index from 1.
4. For example, directed star networks with one additional edge have indegree centralization values of 0.9375, 0.9877, and 0.9999 for networks of sizes 5, 10, and 100, respectively. Analogously, the values of $\frac{HD_{Obs}}{HD_{Max}}$ for networks of sizes 5, 10, and 100 (where all but one neighborhoods send all of their visits to one neighborhood; the one neighborhood visits all others equally) are 0.8, 0.9, and 0.99, respectively. The smallest city in our analyses has 250 neighborhoods and has a value of 0.996, while the largest city has 6,220 neighborhoods and a value of 0.9998. It is unsurprising that the maximum distance is greater for larger networks because it is inherently more difficult for residents of each neighborhood to visit all other neighborhoods in larger cities.
5. This assumption may be undesirable for particular outcomes such as flu epidemics. Since neighborhoods have different population sizes, they could potentially have different saliences for a city's structural connectedness. Although differences in neighborhood size could be accounted for with census data—as differential weighting of residents could be accomplished based on time in or visits to other neighborhoods, as well as salient demographic characteristics available with the individual-level data—we do not recommend this strategy unless one is confident they have spatially stratified samples and strong theoretical motivations to weight differently. One could easily remove this normalization when constructing the mobility edge lists. A city's concentrated mobility index (CMI) would still be calculated in the same way, but the equitable mobility index (EMI) would need to account for differences in neighborhoods' outdegree. This could be done by row-wise adjusting the adjacency matrices for the even connectedness and maximum uneven connectedness networks before calculating the Hamming distances. For even connectedness, each neighborhood's outdegree would still be divided by the $N - 1$. For uneven, each vertex's outdegree would be multiplied by the numerator of equation (3), but the first term would only be 2 rather than $2N$, since the Hamming distance is being calculated row-wise. The sum of these products is the HD_{Max}. This flexibility furthers the measure's utility.
6. Wang et al. (2018a) investigate whether residents of minority or impoverished neighborhoods experience social isolation—an inherently individual-level question. To do this, they analyze how individuals' distances traveled, numbers of neighborhoods visited, and types of neighborhoods visited in a city vary based on the demographics of their home neighborhood. In contrast, we study social integration of cities based on the connectedness between their neighborhoods

and the equity of visits across their neighborhoods. Our substantive interest here is in the city neighborhood networks—not individual-level mobility patterns.

7. Border points are within the specified minimum distance of points from two different clusters but are not within that distance for the specified minimum number of points. As a result, the point could be assigned to either cluster based on the ordering of the data, which makes the algorithm nondeterministic. This would affect our estimates of individuals' home locations.
8. Each tweet represents a unique visit to a location by a user. Still, Twitter (2017) enables users to automatically post tweets. If a user posts more than one tweet at the same time and location, then we regard all but one of the tweets as duplicates. We must cull these tweets because, otherwise, each tweet would no longer indicate a unique visit, and our counts of individuals' mobility patterns that automatically post tweets would suffer reporting bias. Moreover, this would affect our ability to estimate these individuals' home block groups. Additionally, we remove accounts if more than 10 percent of all their tweets, or 25 percent of their home period tweets, are duplicates. This step also addresses the potential issue of bots that geotag their tweets.
9. The commuting zones boundaries are based on counties' boundaries. The boundaries of the cities are from the "places" shapefiles in the "Places" data. The "block groups" shapefiles are provided by 2010 Census Data. We overlaid the shapefiles for the block groups and city boundaries, and any block groups that were within or overlapping with a city's boundaries were treated as being in that city. We removed block groups that had populations under 300 based on data from the 2011–2015 American Community Survey.
10. Of the 5,000 sampled users, 4,842 sent tweets during the period of observation, leaving 158 users who did not send tweets. We thank the anonymous reviewer for suggesting this validation check.
11. In our data, the most prolific individuals tweet 1,000 times more than individuals at the low end of the distribution. The correlation between the number of neighborhoods visited and tweets sent by individuals in our analyzed data set is modest at .54. Analyses of mobility using other social media or cell phone data, particularly if aggregated to areal units, also make a similar assumption.
12. For Detroit and Miami, the urban form is less structured outside of downtown, but the heterogeneity in neighborhoods' indegree, coupled with the relative lack of hubs, does not indicate a polycentric form similar to New York City. Rather, they have disproportionately visited downtown areas with more diffuse visits dispersed throughout the cities, which is why they have lower CMI values.
13. CMI compares the distribution of visits to neighborhoods to determine whether a city's visits are spread evenly across neighborhoods or concentrated in a small share of neighborhoods. EMI, on the other hand, compares the sending and

receiving of visits for each neighborhood with all other neighborhoods in a city against an equitable network; in this sense, it captures the average extent to which any neighborhood in a city is connected to all other neighborhoods.

14. We measure diversity using eight racial and ethnic groups: whites, blacks, Hispanics, Asians, American Indians or Alaskan Natives, Native Hawaiians or Other Pacific Islanders. The correlation between the eight-category measure and a five-category (white, black, Hispanic, Asian, other) measure is 0.99.
15. We use an eight-category measure. Its correlation with a five-category measure is 0.99.
16. Income cut points for the categories are (in thousands of dollars) as follows: 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 75, 100, 125, 150, and 200 plus.
17. In a regression of CMI with the same predictors included in model 1, CMI is positively associated with the share of adults with bachelor's degrees and negatively associated with racial residential segregation. Both associations are statistically significant at the $p < .05$ level. CMI is not significantly related to our other predictors, although population has a positive coefficient and approaches statistical significance. The model explains 65.4 percent of the cross-city variance in CMI. A model including only population, residential segregation, and share of adults with bachelor's degrees as predictors explains 64.3 percent of the variance. All three variables are significant at the $p < .01$ level with coefficients substantively similar to those in the full model.
18. Our measures for the 50 cities are available in the online Appendix Table A1.

References

Akhavan, Armin, Nolan Edward Phillips, Jing Du, Jiayu Chen, Bita Sadeghinasr, and Qi Wang. 2019. "Accessibility Inequality in Houston." *IEEE Sensors Letters* 3(1): 1-4.

Anderson, Brigham S., Carter Butts, and Kathleen Carley. 1999. "The Interaction of Size and Density with Graph-level Indices." *Social Networks* 21:239-67.

Angell, Robert C. 1947. "The Social Integration of American Cities of More than 100,000 Population." *American Sociological Review* 12:335-42.

Balcan, Duygu, Vittoria Colizza, Bruno Gonçalves, Hao Hu, José J. Ramasco, and Alessandro Vespignani. 2009. "Multiscale Mobility Networks and the Spatial Spreading of Infectious Diseases." *Proceedings of the National Academy of Sciences* 106:21484-89.

Barabási, Albert-László and Eric Bonabeau. 2003. "Scale-free Networks." *Scientific American* 288:60-69.

Barthelemy, Marc. 2016. *The Structure and Dynamics of Cities*. Cambridge, MA: Cambridge University Press.

Birant, Derya and Alp Kut. 2007. "St-Dbscan: An Algorithm for Clustering Spatial–Temporal Data." *Data & Knowledge Engineering* 60:208-21.

Blau, Peter M. 1977. *Inequality and Heterogeneity: A Primitive Theory of Social Structure*. New York: Free Press.

Blau, Peter M. 1994. *Structural contexts of opportunities*. Chicago, IL: University of Chicago Press.

Blau, Peter M. and Joseph E. Schwartz. 1997. *Crosscutting Social Circles: Testing a Macrostructural Theory of Intergroup Relations*. New Brunswick, NJ: Transaction.

Boivin, Remi and Marcus Felson. 2018. "Crimes by Visitors versus Crimes by Residents: The Influence of Visitor Inflows" *Journal of Quantitative Criminology* 34: 465-80.

Bora, Nibir, Yu-Han Chang, and Rajiv Maheswaran. 2014. "Mobility Patterns and User Dynamics in Racially Segregated Geographies of US Cities." Pp. 11-18 in *International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction*, edited by William G. Kennedy, Nitin Agarwal, and Shanchieh Jay Yang. Cham, Switzerland: Springer.

Brown-Saracino, Japonica. 2017. *How Places Make Us: Novel LBQ Identities in Four Small Cities*. Chicago: University of Chicago Press.

Browning, Christopher R., Catherine A. Calder, Lauren J. Krivo, Anna L. Smith, and Bethany Boettner. 2017. "Socioeconomic Segregation of Activity Spaces in Urban Neighborhoods: Does Shared Residence Mean Shared Routines?" *The Russell Sage Foundation Journal of the Social Sciences* 3:210-31.

Browning, Christopher R., Catherine A. Calder, Brian Stoller, Aubrey L. Jackson, and Jonathan Dirlam. 2017. "Ecological Networks and Neighborhood Social Organization." *American Journal of Sociology* 122:1939-88.

Butts, Carter T. 2006. "Exact Bounds for Degree Centralization." *Social Networks* 28: 283-96.

Butts, Carter T. and Kathleen M. Carley. 2005. "Some Simple Algorithms for Structural Comparison." *Computational & Mathematical Organization Theory* 11: 291-305.

Butts, Carter T., Miruna Petrescu-Prahova, and B. Remy Cross. 2007. "Responder Communication Networks in the World Trade Center Disaster: Implications for Modeling of Communication within Emergency Settings." *Mathematical Sociology* 31:121-47.

Campello, Ricardo J. G. B., Davoud Moulavi, and Jörg Sander. 2013. "Density-based Clustering Based on Hierarchical Density Estimates." Pp. 160-72 in *Pacific-Asia Conference on Knowledge Discovery and Data Mining*. edited by Jian Pei, Vincent S. Tseng, Cao Longbing, Motoda Hiroshi, and Guandong Xu. Berlin, Heidelberg: Springer.

Colizza, Vittoria, Alain Barrat, Marc Barthelemy, Alain-Jacques Valleron, and Alessandro Vespignani. 2007. "Modeling the Worldwide Spread of Pandemic Influenza: Baseline Case and Containment Interventions." *PLoS Medicine* 4:e13.

Du Bois, W. E. B. 1899. *The Philadelphia Negro*. Philadelphia: University of Pennsylvania Press.

Duncan, Otis Dudley and Beverly Duncan. 1955. "A Methodological Analysis of Segregation Indexes." *American Sociological Review* 20:210-17.

Durkheim, Emile. [1893] 1984. *Division of Labour in Society*. New York: Free Press.

Durkheim, Emile. [1897] 1966. *Suicide*. New York: Free Press.

Fischer, Claude S. 1982. *To Dwell among Friends: Personal Networks in Town and City*. Chicago: University of Chicago Press.

Fischer, Claude S. 2011. *Still Connected: Family and Friends in America Since 1970*. New York: Russell Sage Foundation.

Freeman, Linton C. 1978. "Centrality in Social Networks Conceptual Clarification." *Social Networks* 1:215-39.

Gabrielli, L., S. Rinzivillo, F. Ronzano, and D. Villatoro. 2014. "From Tweets to Semantic Trajectories: Mining Anomalous Urban Mobility Patterns." Pp. 26-35 in *Citizen in Sensor Networks. Lecture Notes in Computer Science*, Vol. 8313, edited by J. Nin and D. Villatoro. Cham, Switzerland: Springer.

Ghaziani, Amin. 2014. *There Goes the Gayborhood?* Princeton, NJ: Princeton University Press.

Graif, Corina, Alina Lungeanu, and Alyssa M. Yettera. 2017. "Neighborhood Isolation in Chicago: Violent Crime Effects on Structural Isolation and Homophily in Inter-neighborhood Commuting Networks." *Social Networks* 51:40-59.

Hahsler, Michael and Matthew Piekenbrock. 2017. "Density Based Clustering of Applications with Noise (DBSCAN) and Related Algorithms." R Package Version 1.1-1.

Hamming, Richard W. 1950. "Error Detecting and Error Correcting Codes." *Bell System Technical Journal* 29:147-60.

Handcock, Mark S., David R. Hunter, Carter T. Butts, Steven M. Goodreau, and Martina Morris. 2008. "Statnet: Software Tools for the Representation, Visualization, Analysis and Simulation of Network Data." *Journal of Statistical Software* 24:1548.

Hipp, John R. and Adam Boessen. 2017. "Neighborhoods, Social Networks, and Crime." Pp. 275-298 in *Challenging Criminological Theory: The Legacy of Ruth Rosner Kornhauser*, edited by Francis T. Cullen, Pamela Wilcox, Robert J. Sampsom, and Brendan D. Dooley. New York: Routledge.

Hunter, David R., Mark S. Handcock, Carter T. Butts, Steven M. Goodreau, and Martina Morris. 2008. "ergm: A Package to Fit, Simulate and Diagnose

Exponential-family Models for Networks." *Journal of Statistical Software* 24: nihpa54860.
James, David R. and Karl E. Taeuber. 1985. "Measures of Segregation." *Sociological Methodology* 15:1-32.
Jiang, Yuqin, Zhenlong Li, and Xinyue Ye. 2018. "Understanding Demographic and Socioeconomic Biases of Geotagged Twitter Users at the County Level." *Cartography and Geographic Information Science* 46:1-15.
Jiang, Shan, Yingxiang Yang, Siddharth Gupta, Daniele Veneziano, Shounak Athavale, and Marta C. González. 2016. "The Timegeo Modeling Framework for Urban Mobility without Travel Surveys." *Proceedings of the National Academy of Sciences* 113:5370-78.
Jones, Malia and Anne R. Pebley. 2014. "Redefining Neighborhoods Using Common Destinations: Social Characteristics of Activity Spaces and Home Census Tracts Compared." *Demography* 51:727-52.
Jurdak, Raja, Kun Zhao, Jiajun Liu, Maurice AbouJaoude, Mark Cameron, and David Newth. 2015. "Understanding Human Mobility from Twitter." *PLoS One* 10: e0131469.
Krackhardt, David. 1994. "Graph Theoretical Dimensions of Informal Organizations." Pp. 89-111 in *Computational Organization Theory*, edited by Kathleen M. Carley and Michael J. Prieteula. Mahwah, NJ: Lawrence Erlbaum Associates.
Krivo, Lauren J., Heather Washington, Ruth D. Peterson, Christopher R. Browning, Catherine Calder, and Mei-Po Kwan. 2013. "Social Isolation of Disadvantage and Advantage: The Reproduction of Inequality in Urban Space." *Social Forces* 92: 141-64.
Lenormand, Maxime, Bruno Gonçalves, Antonia Tugores, and José J. Ramasco. 2015. "Human Diffusion and City Influence." *Journal of the Royal Society Interface* 12:20150473.
Le Roux, Guillaume, Julie Vallée, and Hadrien Commenges. 2017. "Social Segregation around the Clock in the Paris Region (France)." *Journal of Transport Geography* 59:134-45.
Lichter, Daniel T., Domenico Parisi, and Michael C. Taquino. 2015. "Toward a New Macro Segregation? Decomposing Segregation within and between Metropolitan Cities and Suburbs." *American Sociological Review* 80:843-73.
Logan, John R., Brian J. Stults, and Reynolds Farley. 2004. "Segregation of Minorities in the Metropolis: Two Decades of Change." *Demography* 41:1-22.
Louail, Thomas, Maxime Lenormand, Miguel Picornell, Oliva García Cantú, Ricardo Herranz, Enrique Frias-Martinez, José J. Ramasco, and Marc Barthelemy. 2015. "Uncovering the Spatial Structure of Mobility Networks." *Nature Communications* 6:6007.

Luo, Feixiong, Guofeng Cao, Kevin Mulligan, and Xiang Li. 2016. "Explore Spatio-temporal and Demographic Characteristics of Human Mobility via Twitter: A Case Study of Chicago." *Applied Geography* 70:11-25.

Malik, Momin M., Hemank Lamba, Constantine Nakos, and Jurgen Pfeffer. 2015. "Population Bias in Geotagged Tweets." Pp. 18-27 in *Standards and Practices in Large-scale Social Media Research: Papers from the 2015 ICWSM Workshop*, edited by Derek Ruths and Juergen Pfeffer. Palo Alto, California: The AAAI Press.

Massey, Douglas S. and Nancy A. Denton. 1988. "The Dimensions of Residential Segregation." *Social Forces* 67:281-315.

Massey, Douglas S. and Nancy A. Denton. 1993. *American Apartheid: Segregation and the Making of the Underclass*. Cambridge, MA: Harvard University Press.

McPherson, Miller, Lynn Smith-Lovin, and James M. Cook. 2001. "Birds of a Feather: Homophily in Social Networks." *Annual Review of Sociology* 27:415-44.

Merton, Robert K. 1968. *Social Theory and Social Structure*. New York: The Free Press.

Moody, James. 2001. "Race, School Integration, and Friendship Segregation in America." *American Journal of Sociology* 107:679-716.

Mouw, Ted and Barbara Entwisle. 2006. "Residential Segregation and Interracial Friendship in Schools." *American Journal of Sociology* 112:394-441.

Neal, Zachary P. 2012. *The Connected City: How Networks Are Shaping the Modern Metropolis*. New York: Routledge.

Newman, Mark E. J. 2001. "Scientific Collaboration Networks. II. Shortest Paths, Weighted Networks, and Centrality." *Physical Review E* 64:016132.

Opsahl, Tore, Filip Agneessens, and John Skvoretz. 2010. "Node Centrality in Weighted Networks: Generalizing Degree and Shortest Paths." *Social Networks* 32:245-51.

Palmer, John R. B., Thomas J. Espenshade, Frederic Bartumeus, Chang Y. Chung, Necati Ercan Ozgencil, and Kathleen Li. 2013. "New Approaches to Human Mobility: Using Mobile Phones for Demographic Research." *Demography* 50: 1105-28.

Papachristos, Andrew V. and Sara Bastomski. 2018. "Connected in Crime: The Enduring Effect of Neighborhood Networks on the Spatial Patterning of Violence." *American Journal of Sociology* 124:517-68.

Parker, Timothy. 2012. *USDA Economic Research Service-commuting Zones and Labor Market Areas*. Washington, DC: United States Department of Agriculture, Economic Research Service.

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

Putnam, Robert D. 2007. "E Pluribus Unum: Diversity and Community in the Twenty-first Century. The 2006 Johan Skytte Prize Lecture." *Scandinavian Political Studies* 30:137-74.

Reardon, Sean F. and Kendra Bischoff. 2011. "Income Inequality and Income Segregation." *American Journal of Sociology* 116:1092-153.

Reardon, Sean F. and Glenn Firebaugh. 2002. "Measures of Multigroup Segregation." *Sociological Methodology* 32:33-67.

Reardon, Sean F. and David O'Sullivan. 2004. "Measures of Spatial Segregation." *Sociological Methodology* 34:121-62.

Sampson, Robert J. 2008. "Moving to Inequality: Neighborhood Effects and Experiments Meet Social Structure." *American Journal of Sociology* 114:189-231.

Sampson, Robert J. 2012. *Great American City: Chicago and the Enduring Neighborhood Effect*. Chicago: University of Chicago Press.

Shelton, Taylor, Ate Poorthuis, and Matthew Zook. 2015. "Social Media and the City: Rethinking Urban Socio-spatial Inequality Using User-generated Geographic Information." *Landscape and Urban Planning* 142:198-211.

Simmel, Georg. 1950. *The Sociology of Georg Simmel: Translated, Edited, and with an Introduction by Kurt H. Wolff*. New York: The Free Press.

Small, Mario L. 2004. *Villa Victoria: The Transformation of Social Capital in a Boston Barrio*. Chicago: University of Chicago Press.

Small, Mario L., David Harding, and Michele Lamont. 2010. "Reconsidering Culture and Poverty." *Annals of the American Academy of Political and Social Science* 629:6-27.

Snijders, Tom A. B. 1981. "The Degree Variance: An Index of Graph Heterogeneity." *Social Networks* 3:163-74.

Soliman, Aiman, Junjun Yin, Kiumars Soltani, Anand Padmanabhan, and Shaowen Wang. 2015. "Where Chicagoans Tweet the Most: Semantic Analysis of Preferential Return Locations of Twitter Users." Pp. 55-58 in *Proceedings of the 1st International ACM SIGSPATIAL Workshop on Smart Cities and Urban Analytics*, edited by Mohamed Ali. ACM.

Sutton, Jeannette, C. Ben Gibson, Nolan Edward Phillips, Emma S. Spiro, Cedar League, Britta Johnson, Sean M. Fitzhugh, and Carter T. Butts. 2015. "A Cross-hazard Analysis of Terse Message Retransmission on Twitter." *Proceedings of the National Academy of Sciences* 112:14793-98.

Tönnies, Ferdinand. [1887] 1957. *Gemeinschaft und Gesellschaft* [Community and Society]. Translated and edited by Charles P. Loomis. East Lansing: Michigan State University Press.

Toqueville, de Alexis. [1835] 2000. *Democracy in America*. Translated and edited by Harvey Mansfield and Delba Winthrop. Chicago: University of Chicago Press.

Twitter. 2017. Retrieved January 5, 2018. (https://help.twitter.com/en/rules-and-policies/twitter-automation).

Vallée, Julie. 2018. "Urban Isolation and Daytime Neighborhood Social Composition from Twitter Data." *Proceedings of the National Academy of Sciences* 115: E11886-87.

Wang, Qi, Nolan Edward Phillips, Mario L. Small, and Robert J. Sampson. 2018a. "Urban Mobility and Neighborhood Isolation in America's 50 Largest Cities." *Proceedings of the National Academy of Sciences* 115:7735-40.

Wang, Qi, Nolan Edward Phillips, Mario L. Small, and Robert J. Sampson. 2018b. "Reply to Vallée: Different Questions for Different Data." *Proceedings of the National Academy of Sciences* 115:E11888-89.

Wasserman, Stanley and Katherine Faust. 1994. *Social Network Analysis: Methods and Applications.* Vol. 8. Cambridge, England: Cambridge University Press.

Wellman, Barry. 1979. "The Community Question: The Intimate Networks of East Yorkers." *American Journal of Sociology* 84:1201-31.

Wikström, Per-Olof H., Vania Ceccato, Beth Hardie, and Kyle Treiber. 2010. "Activity Fields and the Dynamics of Crime: Advancing Knowledge about the Role of the Environment in Crime Causation." *Journal of Quantitative Criminology* 26:55-87.

Wilson, William J. [1987] 2012. *The Truly Disadvantaged: The Inner City, the Underclass and Public Policy*. Chicago: The University of Chicago Press.

Zhong, Chen, Stefan Müller Arisona, Xianfeng Huang, Michael Batty, and Gerhard Schmitt. 2014. "Detecting the Dynamics of Urban Structure through Spatial Network Analysis." *International Journal of Geographical Information Science* 28: 2178-99.

Author Biographies

Nolan E. Phillips is a postdoctoral fellow in the Department of Sociology at Harvard University. His research develops and uses network analytic methods to understand social structure and its effects.

Brian L. Levy is a postdoctoral fellow in the Department of Sociology at Harvard University. He studies neighborhoods, segregation, and social stratification. His current research takes a life course perspective to study neighborhood effects on economic and educational outcomes; a second project analyzes heterogeneity in the neighborhood attainment process.

Robert J. Sampson is the Henry Ford II Professor of the Social Sciences at Harvard University. His books include *Crime in the Making: Pathways and Turning Points through Life, Shared Beginnings, Divergent Lives: Delinquent Boys to Age 70*, and *Great American City: Chicago and the Enduring Neighborhood Effect*. His research interests include crime, disorder, the life course, neighborhood inequality, civic engagement, “ecometrics,” and the social structure of the city.

Mario L. Small, Ph.D., is Grafstein Family Professor in the Department of Sociology at Harvard University. He works on urban poverty, inequality, personal networks, and the relationship between qualitative and quantitative methods. His books include *Villa Victoria: The Transformation of Social Capital in a Boston Barrio, Unanticipated Gains: Origins of Network Inequality in Everyday Life*, and, most recently, *Someone To Talk To*, a study of how actors mobilize their networks.

Ryan Qi Wang is an assistant professor in the Department of Civil and Environmental Engineering at Northeastern University and the Associate Director of Research on Social Media at the Boston Area Research Initiative (BARI). His research focuses on two interrelated areas: human movement perturbation under the influence of natural and manmade disasters, and mobility equality in big cities.