Environment and Planning B: Urban Analytics and City Science
():1–13

Reprints and permission:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/ToBeAssigned
www.sagepub.com/
SAGE

The Local Structures of Human Mobility in Chicago

James Saxon[1]

Abstract

A large literature establishes the role of mobility in the maintenance of neighborhood social structures. Jane Jacobs famously argued that social capital is maintained through "cross-use of space," and James Coleman formalized its dependence on the "closure" of human interactions. Since many of these interactions entail human movement, neighborhoods with higher social capital should be distinguishable by more cohesive mobility networks. I observe the mobility of Chicago residents through a large dataset of smartphone users. I construct a neighborhood-level mobility network for the city and characterize neighborhoods according to their local graph structure. Neighborhoods that are well integrated with their surroundings have higher income and educational attainment. Consistent with social capital theory and routine activity theory in criminology, higher local network integration independently predicts lower levels of violent and property crime. The methodologies presented provide a meaningful, replicable, and inexpensive approach to the structural measurement of neighborhood networks and social structure.

Keywords

neighborhoods, closure, clustering, social capital, networks, urban, crime, cell phones, big data.

In *The Death and Life of Great American Cities*, Jane Jacobs articulated the related social and spatial behaviors necessary for the effective self-government of urban neighborhoods. For Jacobs, a neighborhood's social trust and control are developed and maintained through its social and spatial networks: the "intricate ballet" of city streets and the continuous cross-use between adjacent spaces. These "neighborhood networks," she wrote, "are a city's irreplaceable social capital." A neighborhood's social and economic relations are at once imprinted on residents' physical trajectories in space and, themselves, affected by the neighborhood's physical configuration. In short, neighborhood social vitality and safety are inseparable from the structures of residents' routines in physical space. Those structures – and thus that capital – are observable through movement on the street. (Jacobs, 1961)

Sociologists, planners, and urban economists have embraced and formalized Jacobs' ideas. Coleman (1988) established the theoretical foundations of social capital and its dependence on the "closure" – the graph clustering coefficient – of social relations. The projection of these social relations into physical space sits at the heart of urban sociology. Over the past decade, Sampson has called for quantitative analysis of neighborhood networks, constructed from human spatial flows or social ties (e.g., 2011; 2012). Responding to this call, Browning has constructed "econetworks" of neighborhoods and adopted Coleman's project of measuring closure. (2014; 2017a; 2017b; 2017c) But these studies simply have not been constructed from actual measurements of quotidian, human trajectories in space. For half a century, Jacobs' intricate ballet has evaded direct, large-scale, quantitative measurement.

The fine-grained spatial data required for this work are increasingly available to researchers. In this project, I leverage a large dataset of GPS locations of smartphone users to measure the quotidian destinations of

[1]Harris School of Public Policy and Center for Spatial Data Science, University of Chicago

Corresponding author:
James Saxon Harris School of Public Policy 1307 E 60th Street, Chicago, IL 60637, USA
Email: jsaxon@uchicago.edu

residents of Chicago neighborhoods. From these data, I construct a mobility network for the city. Researchers in transportation and network science have used similar data, *inter alia*, to demonstrate the consistency between individuals' mobility and that of their social relations (Toole et al., 2015) and to show that certain characteristics – notably network diversity – correlate with socioeconomic class. (Eagle et al., 2010; Florez et al., 2017; Pappalardo et al., 2016) This paper differs from past work in its focus on neighborhood processes and physical space. The *identities* of neighborhoods matter. The character of neighborhoods is determined by interactions with specific, physically adjacent neighbors and by the relationships among the destinations frequented by residents. Poor integration among physical neighbors or among visited locations augurs ill for neighborhood health.

Drawing from Jacobs' and Coleman's social theory as well as from geographic analysis, I construct simple network characterizations of neighborhoods' local mobility behaviors: a (node- and edge-weighted) clustering coefficient and a spatially-local out degree. These constructions of long-theorized concepts in fact yield "comprehensible" results in data. They exhibit strong, expected correlations with established neighborhood observables. I then reach beyond the correlations shown in past work, between socioeconomic class and spatial network observables (diversity, in particular). The constructed variables are not just elaborate reconstructions of traditional observables, already better-measured through official surveys. They are distinct in both concept and information content. I illustrate this through a simple model of neighborhood crime levels. Consistent with social disorganization and routine activity theories of criminality, the constructed variables offer explanatory power beyond traditional controls. Higher levels of interactions with neighbors are associated with significantly lower crime. Notwithstanding, crime per se is not the intended focus of this paper; the interest of the model is to demonstrate the power of the new observables.

My aim is to establish that these new constructions of long-recognized concepts are strongly related to but usefully distinct from other familiar neighborhood observables. The local structure of neighborhood mobility is a distinct dimension in the massively correlated matrix of experiences of urban poverty and privilege. This work is performed with inexpensive, national data sources, so that the new variables are reproducible across cities and across time.

Literature Review: Why measure neighborhood networks?

Social ties in neighborhoods require and enforce routines and behaviors in physical space. Modern telecommunications notwithstanding, maintenance of formal and informal ties requires individuals to traverse space. The closure of social ties, which Coleman argued is fundamental to the social capital of a neighborhood, is necessarily mirrored in the clustering coefficient of its mobility network. (1988) This topology of actions has been used to characterize behaviors in neighborhoods since the early years of urban sociology. In *The City*, Burgess diagrammed a "delinquency triangle," to show how the physical location of a young couple's date affected social closure and behaviors. (Park and Burgess, 1925)

More recently, in a perspective closely mirroring Jacobs, Forrest and Kearns (2001) have elaborated the role of quotidian routines in maintaining neighborhood health. Neighborhoods are "series of overlapping social networks," and these residence-based networks govern everyday routines that are "arguably the basic building-blocks of social cohesion." This argument jibes with Small's (2009) work on the organizational embeddedness of social capital. Individuals' social ties (capital) develop within institutional and organizational contexts that affect the "returns" that those ties yield. I extend Small's position very slightly: social ties may be embedded in organizations, but those organizations are themselves often embedded in space. Ties require movement.

On the other hand, this perspective of dyadic, identifiable interpersonal ties is neither a necessary nor a complete view of neighborhood social capital. Forrest and Kearns also describe the quieter processes of "repair work" and "normalisation" that follow implicitly as fellow urbanites attend to adjacent routines. Anderson (2011) describes these processes in his ethnography of city-wide "cosmopolitan canopies": urban spaces where people from diverse backgrounds share positive experiences and observe others at ease in a pleasant habitat. In a sense, successfully-shared spaces provide a "ballast" against prejudices and inevitable social slights that would otherwise shear the social fabric. Do denizens' neighborhoods serve this role at a local level? In order to do so, the neighborhood spaces must be shared. Are they? In short, insofar as movement reflects social structures, it offers a behavioral vantage point for observing those structures.

Of course, mobility does not directly map to social interaction. Social ties are not identical with spatial flows. Participation in economic, social, and civic organizations may all require movement while entailing different social behaviors. Putnam's (1995) fabled bowling leagues illustrate that a single physical location can have ambiguous social content. One can bowl alone or with friends, and one can visit a park to walk a dog or to partake in a game of chess. On their own, physical locations betray only the aggregate forces that motivate individuals across space. Nevertheless, even devoid of explicitly social interactions, physical trajectories reflect the environments that humans experience. A solitary, aimless stroll expresses confidence in the security of an environment or appreciation of its aesthetic flavor.

Moreover, existing empirical work attests to the strong social content of spatial behaviors. For example, cell phone call data records have been used repeatedly to show that users who interact socially (exchange calls) are more likely to frequent similar locations (use the same cell towers). (Bagrow and Lin, 2012; Toole et al., 2015) These analyses have focused on users instead of neighborhoods, and have not assessed the social meaning of variation across neighborhoods in spatial integration. To that end, Browning and colleagues have, in a series of papers, found that Los Angeles neighborhoods where residents reported more shared locations had higher reported trust and collective efficacy. (Browning et al., 2017b,c) Using *simulated* mobility data for Columbus, Ohio, they characterize the closure of the "econetwork" of activity spaces: the degree to which households share multiple spaces, conditional on sharing one. They show that Los Angeles neighborhoods with better closure have lower crime rates. (Browning and Soller, 2014; Browning et al., 2017a) The analysis that follows mirrors and reproduces this result using measured behaviors in Chicago.

The paper builds on the neighborhood network approach advanced by Sampson and Graif in the context of the Project on Human Development in Chicago Neighborhoods (PHDCN) and the associated Key Informant study (Graif et al., 2014; Sampson, 2011, 2012), and Raudenbush and Sampson's "ecometric" work, more broadly. (1999) The ecometric approach entails a focus on the neighborhood as a unit of observation in its own right, analytically distinct from simple aggregates of residents. Neighborhood character arises from collective processes and the neighborhood's spatial situation: it is defined, among other things, by the social relations among residents (social capital, etc.) and their (physical) routines. In addition to the empirical work already mentioned, Browning et al. have reinforced these theoretical foundations, in an "econetwork" approach to neighborhood mobility that integrates ecometrics with Coleman's work on social capital and closure (1988), and a broad geographic literature on "activity spaces." (2017c)

In 2011, Sampson lamented that "the ways in which cross-neighborhood networks are tied into the social structural web of the city are not well understood and virtually never studied empirically." This is no longer true. Location data like those used here are increasingly available to researchers and support a burgeoning literature. Perhaps most pertinent is work showing that increasing geographic network diversity of social ties (Eagle et al., 2010) as well as mobility behaviors (Florez et al., 2017; Pappalardo et al., 2016) are correlated with higher socioeconomic status. Over the last few years, Twitter and GPS location data have been used to measure segregation in daily routines (Athey et al., 2019; Wang et al., 2018), and to quantify the concentration and diversity in destinations, across cities. (Phillips et al., 2019) Graif and collaborators have studied the interplay of criminality with commuting networks and found that network ties in Chicago are less-likely to exist or persist in neighborhoods with high violent crime. (2017; 2019) This paper continues this empirical work, with large-scale, fine-grained data on physically-local mobility behaviors, analyzed at the neighborhood-scale.

The preceding pages suggest that measured mobility will capture information about social structure. To spoil the surprise: it does. The network observables I construct correlate strongly with socioeconomic status; the measurements show that residents of high and low-status neighborhoods navigate their local spaces differently. But this leaves a legitimate question: to the extent that my characterizations of the network structure *differ* from already-available observables, are those differences meaningful or consequential? Are the network variables simply rococo representations of wealth and education, identifying walkable neighborhoods affordable only to elites? Or do they offer any new information? In short: why go to the trouble of constructing the networks?

There are a many social processes where network structure of mobility may be expected to affect outcomes. This paper demonstrates the added predictive value of network information in just one context: criminality. Two of the dominant theories of criminality emphasize how communities' social and physical structures affect the prevalence of crime. Social disorganization theory posits that crime rates depend not only on the personal characteristics of neighborhood residents, but on the influences of the neighborhood itself and the capacity for social control of the structures connecting residents. Alternately, in the routine activities framework, crime occurs when a motivated offender coincides in space and time with a target without an effective guardian present. (Cohen and Felson, 1979) This perspective underscores the need for an accurate accounting of the population "at risk" for criminal behavior or victimization in a location, as first suggested by Boggs in 1965. The two theories thus privilege different controls, but they make the consistent prediction that guardianship or social control and cohesion will depress crime.

The beginning of this Section described the role of mobility in the maintenance of social cohesion and as a necessary, physical imprint of social closure. Residents' local mobility behaviors are characterized formally in the *Methods* Section. According to the social disorganization view, these behaviors should correlate with lower crime through mechanisms of social cohesion and control. In the routine activities perspective, one would argue analogously that locals might be better attuned to their surroundings and more motivated to protect them than visitors; they might be better guardians. Reversing the direction of causality – positing that crime would scare people off of the streets – would lead to the same behavior. (In this case, the theoretical arguments reviewed above would imply that crime would depress social capital by decreasing social contact and shared routines, above and beyond any direct effect.)

The practical hypothesis is thus that, with full controls, neighborhoods with greater local activity or cross-use or that share more location activity with their neighbors (closure) should see lower crime. We will again find that the data are consistent with the hypothesis.

Data

This project aims first to meaningfully characterize the mobility behaviors of urban neighborhoods and then to present the relationship between these behaviors and other social observables. This entails two classes of data: (1) a record of locations from individuals' cell phones and (2) a raft of controls for the socioeconomic and physical environment.

Cell Phone Locations

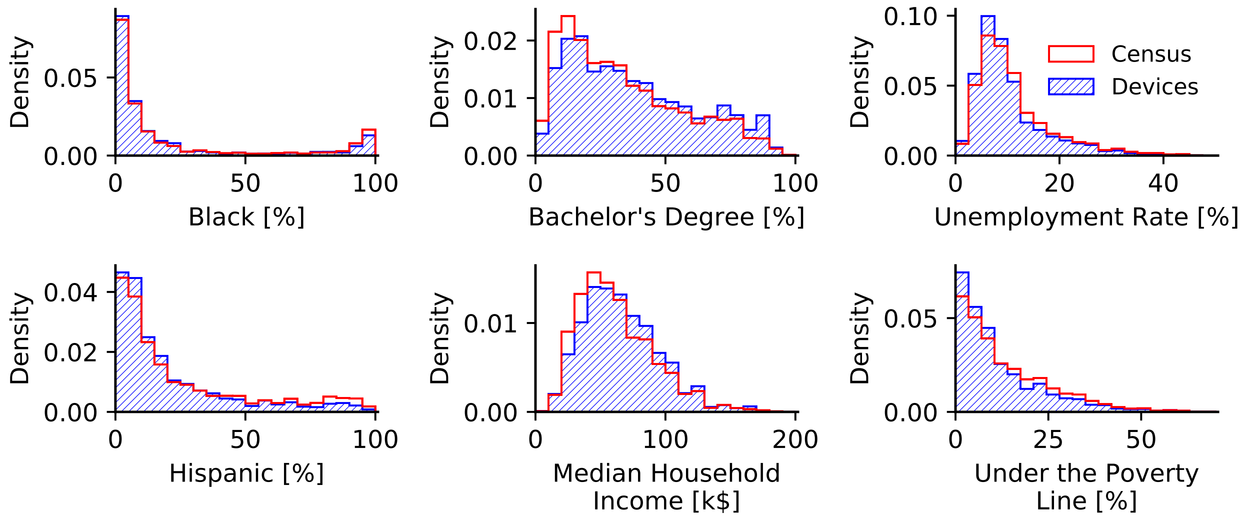
The primary dataset consists of over 600 million GPS locations from smart phone users in Chicago and its western suburbs. The data were recorded in May 2017 by active applications on users' phones, aggregated provided by Carto. Several dozen applications generate locations. The specific applications are not available, but the product category is given in some cases. These apps cover a range of categories including navigation, travel, dating, music, and weather. The bulk of the data is generated through software development kits for advertising in unknown apps. Both iOS and Android devices are included. These and other similar sources have recorded consistent data across the entire United States, for years. This means that the strategies developed are in practice repeatable across time and space. The project thus responds to prominent calls for "big data" methods for in the social sciences (e.g., Lazer et al., 2009).

Substantial processing is required to construct this network. Some applications have access only to the user's approximate location; I discard these imprecise data. Using OpenStreetMap data (OpenStreetMap contributors, 2018), I also suppress data where individuals are in transit: points within 10 meters of motorways, trunk, primary or secondary roads, as well as railways or subways. After data cleaning, more than 268 million user locations remain.

I associate these points to 2017 Census tracts with a simple point-in-polygon merge. For each user (device) in the dataset, I define the "home" Census tract as the modal location between midnight and 6am. This location can be defined for 309 thousand devices, out of 845 thousand unique identifiers in the data. This amounts to a roughly 3% sample of the 10 million-person data area. These users collectively account for 238 million points; the median user records 200 locations over the month. I drop "residents" of airports as well as the two devices assigned to the now-vacant site of Robert Taylor Homes. One weakness of the nocturnal assignment method is that places with large tourist populations or 24-hour operations are over-represented, presumably with visitors and workers mis-assigned as residents.

Unlike most surveys, the location data is obviously a convenience sample. Nevertheless, the data appear to be fairly representative. The Pew Research Center (2018) found that, as of the study period in 2017, 83% of urban residents owned smartphones. This ownership is skewed towards the wealthy and educated, and away from the elderly. Biases by race, ethnicity, and sex are small. Other studies have found that Census tabulations correlate closely with locations from Twitter and cell phone call data records. (Lenormand et al., 2014) One

Figure 1. Demographic observables of Census tracts from the American Community Survey are weighted by corresponding population estimates or by device populations. Biases towards wealthier, whiter and more-educated tracts are apparent but less severe than one might expect.



reaches a similar impression by comparing the devices' home locations with population estimates from the American Community Survey (ACS). Figure 1 displays normalized histograms of race, ethnicity, income, education, poverty, and unemployment, for Census tracts in the study area. The tracts in each histogram are weighted first by the ACS tract population estimates and then by the number of GPS devices that "live" there. This comparison reproduces the expected shift towards whiter, wealthier, and more educated tracts, but the bias is not nearly so severe as one might expect. This bias is presumably reproduced *within* (as well as *among*) tracts, but this is less readily measurable.

Results are robust however, to dramatic redefinitions of the sample of devices or locations. Correlations and regressions are substantively consistent if the constructed variables are re-evaluated after removing simultaneous pings on a device from multiple apps, or requiring more pings per user, more days observed in sample, and more nights at home. (See supplementary materials.)

Socioeconomic Covariates

Correlations between constructed, mobility observables and "traditional" neighborhood characteristics are made based on the 5-year estimates of the 2017 American Community Survey. For the models of crime, I use three additional data sources from the US Census and the City of Chicago. Crime counts and zoning are drawn from the city. Workplace populations come from the Census's LEHD Origin Destination Employment Statistics (LODES). The LODES data are derived from administrative records of unemployment insurance for wage and salary jobs. They cover approximately 95 percent of these jobs. (Graham et al., 2014)

Methods

Characterizing Neighborhoods

Using the cell phone location data, I characterize the local structure of mobility between Chicago neighborhoods. To do so, I first construct a network of neighborhoods: an interaction matrix representing the (normalized) rates at which residents of each neighborhood (Census tract) visit every other neighborhood in the city. For a user device u, I call the fraction of locations generated in Census tract ℓ, A_u^ℓ. For home neighborhood h, the set of devices based there is $\mathcal{R}_h$, and the number of such devices is n_h. The neighborhood-averaged mobility of users based at location h is thus $\hat{A}_{h\ell} = \sum_{u \in \mathcal{R}_h} A_u^\ell / n_h$. The matrix $\hat{A}_{h\ell}$ represents a directed network of neighborhoods in the city. Cells in the matrix (edges in the network) represent the average rates at which residents of h visit location ℓ.

I next construct properties of urban neighborhoods (nodes) and evaluate them on this network. Most Census tracts have between 2,500 and 5,000 residents, but the range in Chicago extends from just hundreds to nearly 18,000. Urban tracts are on average less populous than those in the suburbs. The intensive character of neighborhoods does not depend on the Census Bureau's decisions to split or merge Census tracts. The measures proposed below are therefore designed to be independent of tract population; they are "node-split invariant." (Heitzig et al., 2012)

Clustering Coefficient. The clustering coefficient has long been recognized as a critical property of social networks. It quantifies the degree of closure in destinations. Informally, it measures whether a node's neighbors are also neighbors with each other. More formally, it quantifies the number of closed triangles out of all triplets on the graph (three nodes connected by at least two edges). This definition suffices for simple graphs but is inadequate for graphs with edge and node weights (interactions and populations). The definition has been extended to graphs with weighted (Saramäki et al., 2007) and directed edges (Fagiolo, 2007), but not to graphs with weighted nodes as in the present case.

For any three nodes, there are eight directed triangles. I focus on the two triangles with two edges outgoing from the reference node, so the question is whether the residents of a node's neighbors themselves interact. Do residents at my destinations visit the (other) places that I visit? If I visit locations A and B, do residents of A visit B? The interactions $\hat{A}_{h\ell}$ are intensive characteristics of h; they are averaged over the population of the node. But person-to-person interactions are diluted by the population at the destination. Dividing cell i's interaction with neighbor j by the neighbor's population, I define weights as $w_{ij} \equiv \hat{A}_{ij}/n_j$. This is effectively "doubly-normalized" over the number of potential pairs of individuals over whom it is spread. It may be thought of as the likelihood of a person in i interacting with a *specific* person in j. A larger tract is likely to attract more visitors, but is less affected by each individual visit. A nearly node-split invariant measure may then be constructed as

$$c_h \equiv \sum_{i \neq h} \sum_{j \neq h} \hat{A}_{hi} \hat{A}_{hj} w_{ij} \Bigg/ \sum_{i \neq h} \sum_{j \neq h} \hat{A}_{hi} \hat{A}_{hj},$$

though it remains in principle sensitive to the population of the home tract. (See supplementary materials.) The clustering coefficient mirrors the "econetwork intensity" of Browning et al. (2014; 2017a; 2017b; 2017c) in the sense that it quantifies the closure of movement behaviors.

Local Out-Degree. The "local out-degree" is intended to capture Jacobs' cross-use of neighborhoods by residents. It also measures the "consistency" of Tobler's "First Law of Geography," that "everything is related to everything else, but near things are more related than distant things." (1970) This "law" is at once a credo for the field and its methodological foundation. Is there variation in the "respect" for Tobler's Law, or social consequences for contraventions?

I define the value for neighborhood h as the fraction of residents' behaviors that take place within the neighborhood's immediate vicinity. This "vicinity" is defined by analogy with k nearest neighbors. However, since people interact with people and not Census tracts, node-split invariance requires that the vicinity be defined with constant population. I set this population to $N = 40{,}000$ people. I call the cumulative population of the nearest k tracts N_k and the population of the next nearest tract n_{k+1}. I define k as the smallest value for which $N_k + n_{k+1} > N$, and call the corresponding set of k tracts the vicinity of h, $\mathcal{V}_h$. I take the notational liberty of indexing $\hat{A}_{h\ell}$ by distances from h, so that $\hat{A}_{h,k+1}$ is the interaction of $k+1$'th neighbor. The local out-degree can then be written as

$$\text{local out-degree}_h \equiv \left[\left(\sum_{\ell \in \mathcal{V}_h} \hat{A}_{h\ell} \right) + \hat{A}_{h,k+1} \frac{N - N_k}{n_{k+1}} \right] \Bigg/ \left(1 - \hat{A}_{hh} \right).$$

This represents the sum of the k interaction fractions in $\mathcal{V}_h$, corrected with the interaction with tract $k+1$ scaled by the remaining population difference $(N - N_k)/n_{k+1}$, and normalized to the total out-of-home activity, $1 - \hat{A}_{hh}$.

Because individuals do tend to spend time in the vicinity of their own residence, the local out-degree is correlated with the clustering coefficient. In this sense, the local out-degree codifies a similar characteristic of neighborhoods in a simpler way and offers a check of the robustness of results. But the local degree also differs, since it explicitly incorporates the spatial distribution of interactions while the clustering coefficient does not.

Ambient Population. Neighborhoods' populations vary by hour and day; at a given moment, the population is neither the count of residents nor workers. Averaging over the observation window, I estimate the actual population present – the ambient population – by summing over all users the fraction of time spent (by user) in each tract. I normalize this population by the number of users U and tracts T:

$$\text{ambient}\,(\ell) = \frac{T}{U}\sum_u A_u^\ell.$$

This is obviously *not* node-split invariant and it is not meant to be: it is an extensive measure of the size and usage of the tract. Its logarithm is therefore included as a covariate in the logged crime model of the following section, to establish the "base rate."

Modeling Criminality.

I model criminality for census tracts in Chicago using a spatial autoregressive (SAR) error model (Anselin, 1988) of the form

$$\boldsymbol{y} = \boldsymbol{\beta X} + \lambda \boldsymbol{W u} + \boldsymbol{\varepsilon}.$$

The outcome $\boldsymbol{y}$ is the logarithm of the violent or property crime counts, for each populated tract in the city. The first term on the right-hand side represents standard covariates, and the second two the autoregressive and idiosyncratic parts of the error. The covariates $\boldsymbol{X}$ include the clustering and local out-degree, along with controls from both social disorganization and routine activities theories. For the routine activities controls, tract area is used along with three population normalizations: residents, workers, and "ambient" cell phone devices. To facilitate comparisons and freeze parameters in the model, social disorganization controls are constructed to align with Browning et al. (2017a) and Morenoff et al. (2001) to the extent practicable. A complete description of these controls is provided in the supplementary materials.

The weights matrix $\boldsymbol{W}$ is constructed as the (normalized) "Queen" contiguity matrix, such that neighbors are defined by shared perimeters or corners. The SAR error model is chosen over the spatial lag model based on the results of the LM test. Despite somewhat large heteroskedasticity, I evaluate models through maximum likelihood estimation in order to show the informational interpretation of the results. I show in the supplementary materials that parameter estimates and pseudo-R^2 are consistent, with results derived using the generalized method of moments, or alternative weight specifications. All models are implemented using pySAL. (Anselin and Rey, 2014)

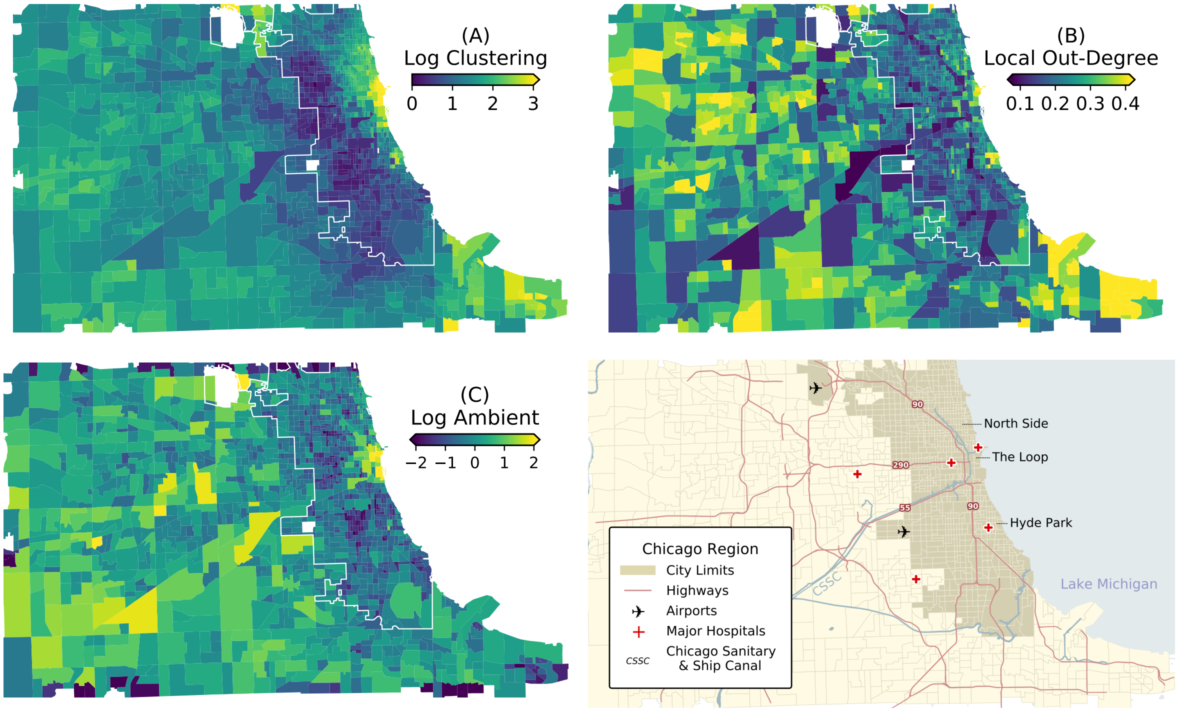
The hypothesis tested is that the two constructed variables – clustering and local out-degree – will each be associated with significantly lower crime, and that including them in the model will reduce the model's information loss (or improve its explained variance). This would occur under both the social disorganization or routine activities models, or a "reversed" view in which residents simply avoid streets they know to be dangerous. Either way, the parameters should have the "right" sign and contribute information not previously available.

Results: Network Structure and Social Status

Figure 2 displays choropleths of the constructed variables: (a) the logarithm of the clustering (shifted to a minimum of 0, for legibility), (b) the local out-degree, and finally (c) the logarithm of the normalized ambient population. The Figure also includes a map of the study area. Each of the variables shows significant variation across the city. Higher values in (a) and (b) denote places with higher integration of movement behaviors. The clustering (a) quantifies the strength of ties between a neighborhood's destinations while the local out-degree (b) can be interpreted as the strength of the relation to physical neighbors. The population-weighted correlation between the two measures is 0.52; it it somewhat higher for a more-restricted sample of devices (see supplementary materials). The ambient population (c) highlights the active regions of the city.

Observers familiar with the geography of Chicago will find in Figures 2(a) and (b) a confirmation of their priors for the city. The wealthier North Side and the suburbs have "better" outcomes while the West and South Sides are more depressed. Major highways (I-90 and I-55), the Chicago Shipping and Sanitation Canal,

Figure 2. Network characteristics of Chicago neighborhoods; see Section for details on their construction. The city of Chicago is outlined in white. The bottom right panel shows a schematic map, including landmarks discussed in the text.



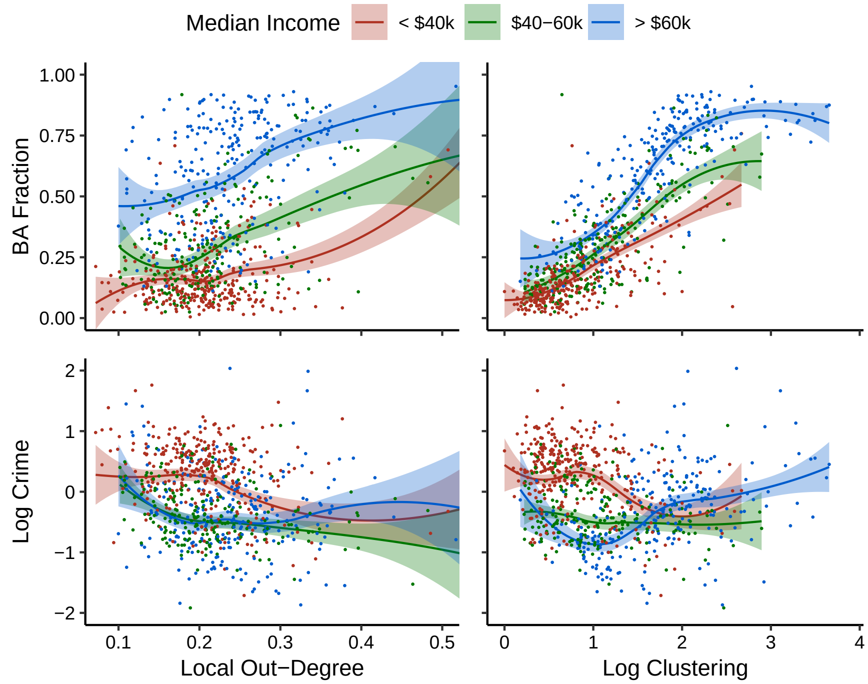
and the North Branch of the Chicago River all show up as cleavages in the local out-degree: physical borders unsurprisingly affect mobility. Proceeding south along Lake Michigan, the notable outlier is the neighborhood of Hyde Park. This is no surprise: anchored by the University of Chicago, Hyde Park is economically and institutionally isolated from its neighbors. (Boyer, 2015; Jacobs, 1961; Levi, 1961; Sampson, 2012; Winling, 2017) For the ambient populations, the Loop – the center city – is the clear and unsurprising outlier. But a few multimodal shipping hubs in the suburbs also stand out: a revolving group of "residents" (likely, truckers) spend *all* of their (limited) time there.

Figure 3 presents the relationship of the constructed mobility variables with logged crime rates and adult bachelor's degree attainment, for Census tracts in the city. The crime rates in that plot combine violent and property crime and, for illustration, are normalized by residential population. The data are divided into three bins of roughly equal populations, by median annual household income: tracts with income less than \$40k, between \$40k and \$60k, and above \$60k. There are strong (non-linear) relationships between the variables. The population-weighted correlation between the bachelor's attainment of the population aged 25 years and older and the log clustering is 85%; the corresponding correlation for the local out-degree is 43%. In the lower-income bin, crime rates are significantly higher at moderate clustering and out-degree, but converge towards the trends for higher-income tracts at high clustering or out-degree.

This point – that the socioeconomic pattern of the city is reflected in the daily trajectories of individuals in their communities – is worth emphasizing. The strong relationship with established outcomes lends credence to the clustering coefficient and local out-degree as measures of local-scale neighborhood integration: they are not just noise. It may be expected that privileged households will choose safe, walkable neighborhoods; the data show that residents of privileged and poorer neighborhoods in fact engage their local environments in measurably different ways. Denizens of wealthy neighborhoods exhibit more activity in the vicinity of their place of residence than their counterparts from poorer or less-educated neighborhoods. Variation in the "relatedness" of adjacent environments tracks with socioeconomic status.

It is worth lingering to unpack this finding in light of existing literatures. It squares neatly with Jacobs' view of the necessary process of an "intricate ballet" on city streets, with which I began this paper. Neighborhoods chosen by those with greater means tend to have more vibrant cross-use and more "cohesive" mobility networks: residents exhibit more local behaviors, and the destinations they frequent are more consistent with

Figure 3. Local out-degree and clustering coefficient are plotted against the bachelor's degree attainment of the adult population and combined violent and property crime rate (normalized by residential population), for Census tracts in Chicago.



"peers" in the neighborhoods that they visit. This represents an interesting counterpoint with the findings of greater network diversity in high-status neighborhoods. (Florez et al., 2017; Pappalardo et al., 2016) Higher network diversity, measured as greater entropy, means lower predictability in the distribution of visited locations. By contrast, I find that interactions between neighborhoods (visits by residents) are *better* predicted by physical proximity and peers' behaviors in high- than in low-status locations. This tension is resolved by the fact that my prediction relies either on an implicit model (the expectation of higher interactions with spatial neighbors) or new information (peers' destinations). Per Jacobs, perhaps both diversity and consistency are necessary for healthy spaces: the diversity of uses stimulates activity, making a safe environment where weak ties can form. Said differently, Granovetter's weak ties (1973) and Jacobs' ballet are reconcilable and even reinforcing.

The variation in the local out-degree can also be understood as demonstrating heterogeneity in the degree of conformity to Tobler's First Law of Geography (that nearer places are more related than distant places). In the City of Chicago, the intensity of relations with neighbors depends on where you are. The purely spatial implication of the Law – high relatedness with physical neighbors – is more true for the rich and educated than for the poor. This perspective is possible only with modern datasets.

The relationships presented and described suggest that the constructed variables are not just noise. The question then becomes whether the mobility data provide any independent explanatory power for predicting other social outcomes. Are expensive, walkable, and cohesive neighborhoods simply an amenity affordable only to educated elites? Or are these neighborhoods safer and more vibrant, beyond what population composition would suggest?

Results: Cross-Use and Crime

Table 1 shows the results of the regressions for crime in Chicago neighborhoods. The first three columns present results for violent crime, first with only the population and social disorganization covariates, and then adding either the clustering coefficient or the local out-degree. Columns three through six repeat this pattern for property crime.

Table 1. Log violent and property crime counts, regressed on the clustering coefficient and local out-degree, with "standard" controls for social disorganization and routine activities. See supplemental materials for all covariates.

Dependent	Violent	Violent	Violent	Property	Property	Property
Log Clustering		-0.24***			-0.20***	
		(0.05)			(0.04)	
Local Out-Degree			-0.86**			-0.51*
			(0.27)			(0.22)
N	780	780	780	780	780	780
Pseudo-R^2	0.64	0.68	0.65	0.66	0.68	0.66
AIC	664	643	656	351	326	348

Standard errors in parentheses; $*p < 0.05$, $**p < 0.01$, $***p < 0.001$.

Consistent with their theoretical motivations, the constructed variables significantly and independently predict crime rates in the city. This conclusion holds with a rich array of controls for social organization, with three alternative definitions of the population or risk set, and for both violent and property crime. The network effects are significant, above and beyond the base activity level of the space (its ambient population). The clustering and out-degree provide distinct alternatives on the measurement but yield consistent qualitative results. Both reduce the Akaike information criterion with respect to the base model, although the gain is larger for clustering than local out-degree. It is natural to ask whether simpler network constructions (like fractional time out-of-home), available Census measures of physical mobility (commuting mode or duration), or measures of app use (mean or median "ping" rate) offer the same explanatory power. I show in the supplementary materials that they do not.

These findings are consistent with work on "econetworks" by Browning et al. (2017a) already discussed. They support both the social disorganization and routine activity theories of criminality. In the latter perspective, higher local out-degree and clustering would be suggestive of better guardianship. Locals may be better attuned to their surroundings and more motivated to protect them, which would depress crime. Although the home Census tract is itself excluded from the local out-degree (to avoid in-residence activity), higher local out-degree does represent more locals circulating in the immediate vicinity of their residence. By contrast, the clustering coefficient measures whether individuals share activity locations with the residents of the neighborhoods that they visit. This consistency of use might also be considered a "shared community" to protect. Though the clustering coefficient is not explicitly spatial, it does capture consistent communities of users who may be effective guardians.

This work was motivated by the idea that local mobility behaviors provide a window into levels of neighborhood social capital. This capital is expected to suppress crime, but the results presented show only that these physical behaviors correlate negatively with crime levels. Causality need not flow from social capital to crime rates, nor need social capital be even involved. A more prosaic, alternative explanation is that residents are aware of the relative safety of their environment, and respond by using or avoiding it. The mobility data would then capture this local knowledge. In that case, to the extent that physical mobility is required for the maintenance of social bonds, criminality would reduce social capital instead of the other way around.

This section has elaborated the relationship between crime and the constructed mobility variables: with a slew of controls, communities with more-cohesive mobility have lower crime. But the broader conclusion is different: the clustering coefficient, out-degree, and ambient population as measured in cell phone data have strong relationships with other sociological observables, but they are distinct concepts.

Conclusions

This project has used new, large data sources to characterize the local structure of human mobility in Chicago neighborhoods. The variables defined – the clustering and local out-degree – are motivated from close antecedents in geographic analysis and sociological theory intended to encode spatial proximity and the network topology of social capital. As measured in the data, these variables are strongly correlated with adult educational attainment, but also provide independent explanatory power for the prediction of crime levels. Outliers are comprehensible in the social and economic context of the city. In short: the variables as constructed have good face validity and represent new and distinct concepts. They quantify consequential variation in the local structure of Chicago neighborhoods that has not before been observable.

This window on neighborhoods is possible thanks to new, large data sources. These data sources and the methods used to analyze them allow replication over time and space. Detailed work is still needed to ground-truth the relationship between mobility and social interactions across socioeconomic classes and physical spaces. But these data hold immense promise for continuous and consistent measurement of empirical mobility behaviors and local networks at national scope.

References

Anderson, E., *The Cosmopolitan Canopy: Race and Civility in Everyday Life*, Kindle (2011). W. W. Norton & Company, New York.

Anselin, L., *Spatial Econometrics: Methods and Models* (1988). Kluwer Academic Publishers, Dordrecht.

Anselin, L. and Rey, S. J., *Modern Spatial Econometrics in Practice: A Guide to GeoDa, GeoDaSpace and PySAL* (2014). GeoDa Press LLC, Chicago.

Athey, S., Ferguson, B., Gentzkow, M., and Schmidt, T., *Experienced Segregation.* (February 2019). Stanford University Working Paper. pp. 1-72

Bagrow, J. P. and Lin, Y.-R., "Mesoscopic Structure and Social Aspects of Human Mobility," *PLOS ONE,* **7**:5, 1-6 (05, 2012). DOI: 10.1371/journal.pone.0037676.

Boggs, S. L., "Urban Crime Patterns," *American Sociological Review,* **30**:6, 899-908 (1965). DOI: 10.2307/2090968.

Boyer, J. W., *The University of Chicago: A History* (2015). University of Chicago Press, Chicago.

Browning, C. R. and Soller, B., "Moving Beyond Neighborhood: Activity Spaces and Ecological Networks as Contexts for Youth Development.," *Cityscape,* **9**:1, 165-196 (2014).

Browning, C. R., Calder, C. A., Boettner, B., and Smith, A., "Ecological Networks and Urban Crime: The Structure of Shared Routing Activity Locations and Neighborhood-Level Informal Control Capacity," *Criminology,* **55**:4, 754-778 (2017a). DOI: 10.1111/1745-9125.12152.

Browning, C. R., Calder, C. A., Krivo, L. J., Smith, A. L., and Boettner, B., "Socioeconomic Segregation of Activity Spaces in Urban Neighborhoods: Does Shared Residence Mean Shared Routines?" *The Russell Sage Foundation Journal of the Social Sciences,* **3**:2, 210-231 (2017b). DOI: 10.7758/rsf.2017.3.2.09.

Browning, C. R., Calder, C. A., Soller, B., Jackson, A. L., and Dirlam, J., "Ecological Networks and Neighborhood Social Organization," *American Journal of Sociology,* **122**:6, 1939-1988 (2017c). DOI: 10.1086/691261.

Cohen, L. E. and Felson, M., "Social Change and Crime Rate Trends: A Routine Activity Approach," *American Sociological Review,* **44**:4, 588-608 (1979). DOI: 10.2307/2094589.

Coleman, J. S., "Social Capital in the Creation of Human Capital," *American Journal of Sociology,* **94**, S95-S120 (1988). DOI: 10.2307/2780243.

Eagle, N., Macy, M., and Claxton, R., "Network Diversity and Economic Development," *Science,* **328**:5981, 1029-1031 (2010). DOI: 10.1126/science.1186605.

Fagiolo, G., "Clustering in complex directed networks," *Physical Review E,* **76**, 026107 (Aug, 2007). DOI: 10.1103/PhysRevE.76.026107.

Florez, M. A., Jiang, S., Li, R., Mojica, C. H., Rios, R. A., and González, M. C.. "Measuring the impacts of economic well being in commuting networks – A case study of Bogota, Colombia," at *TRB annual meeting*, (January, 2017). 1–18

Forrest, R. and Kearns, A., "Social Cohesion, Social Capital and the Neighbourhood," *Urban Studies,* **38**:12, 2125-2143 (2001). DOI: 10.1080/00420980120087081.

Graham, M. R., Kutzbach, M. J., and McKenzie, B., *Design Comparison of LODES and ACS Commuting Data Products.* Technical Report 14-38 (October 2014). Center for Economic Studies, U.S. Census Bureau.

Graif, C., Gladfelter, A. S., and Matthews, S. A., "Urban Poverty and Neighborhood Effects on Crime: Incorporating Spatial and Network Perspectives," *Sociology Compass,* **8**:9, 1140-1155 (2014). DOI: 10.1111/soc4.12199.

Graif, C., Lungeanu, A., and Yetter, A. M., "Neighborhood isolation in Chicago: Violent crime effects on structural isolation and homophily in inter-neighborhood commuting networks," *Social Networks,* **51**, 40-59 (2017). DOI: 10.1016/j.socnet.2017.01.007. Crime and Networks

Graif, C., Freelin, B. N., Kuo, Y.-H., Wang, H., Li, Z., and Kifer, D., "Network Spillovers and Neighborhood Crime: A Computational Statistics Analysis of Employment-Based Networks of Neighborhoods," *Justice Quarterly,* **0**:0, 1-31 (2019). DOI: 10.1080/07418825.2019.1602160.

Granovetter, M. S., "The Strength of Weak Ties," *American Journal of Sociology,* **78**:6, 1360-1380 (1973). DOI: 10.2307/2776392.

Heitzig, J., Donges, J. F., Zou, Y., Marwan, N., and Kurths, J., "Node-weighted measures for complex networks with spatially embedded, sampled, or differently sized nodes," *The European Physical Journal B,* **85**:1, 38 (Jan 23, 2012). DOI: 10.1140/epjb/e2011-20678-7.

Jacobs, J., *The Death and Life of Great American Cities*, Kindle (1961). Knopf Doubleday Publishing Group, New York.

Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabási, A.-L., Brewer, D., *et al.*, "Computational Social Science," *Science,* **323**:5915, 721-723 (2009). DOI: 10.1126/science.1167742.

Lenormand, M., Picornell, M., Cantú-Ros, O. G., Tugores, A., Louail, T., Herranz, R., *et al.*, "Cross-Checking Different Sources of Mobility Information," *PLOS ONE,* **9**:8, 1-10 (08, 2014). DOI: 10.1371/journal.pone.0105184.

Levi, J., *The neighborhood program of the University of Chicago* (1961). University of Chicago Press, Chicago.

Morenoff, J. D., Sampson, R. J., and Raudenbush, S. W., "Neighborhood Inequality, Collective Efficacy, and the Spatial Dynamics of Urban Violence," *Criminology,* **39**:3, 517-558 (2001). DOI: 10.1111/j.1745-9125.2001.tb00932.x.

Pappalardo, L., Vanhoof, M., Gabrielli, L., Smoreda, Z., Pedreschi, D., and Giannotti, F., "An analytical framework to nowcast well-being using mobile phone data," *International Journal of Data Science and Analytics,* **2**:1, 75-92 (Dec 01, 2016). DOI: 10.1007/s41060-016-0013-2.

Park, R. E. and Burgess, E. W., *The City (Heritage of Sociology Series)*, Kindle Edition (1925). University of Chicago Press, Chicago, IL.

Phillips, N. E., Levy, B. L., Sampson, R. J., Small, M. L., and Wang, R. Q., "The Social Integration of American Cities: Network Measures of Connectedness Based on Everyday Mobility Across Neighborhoods," *Sociological Methods & Research,* **0**:0, 1-40 (2019). DOI: 10.1177/0049124119852386.

Putnam, R. D., "Bowling Alone: America's Declining Social Capital," *Journal of Democracy,* **6**:1, (January, 1995). DOI: 10.1353/jod.1995.0002.

Raudenbush, S. W. and Sampson, R. J., "Ecometrics: Toward a Science of Assessing Ecological Settings, With Application to the Systematic Social Observation of Neighborhoods," *Sociological Methodology,* **29**:1, 1-41 (1999). DOI: 10.1111/0081-1750.00059.

Sampson, R. J., "Neighborhood effects, causal mechanisms and the social structure of the city." In *Analytical Sociology and Social Mechanisms*. Demeulenaere, P. ed. (Cambridge University Press, 2011), pp. 227–249.

Sampson, R. J., *Great American City: Chicago and the Enduring Neighborhood Effect*, Kindle (2012). University of Chicago Press, Chicago, IL.

Saramäki, J., Kivelä, M., Onnela, J.-P., Kaski, K., and Kertész, J., "Generalizations of the clustering coefficient to weighted complex networks," *Phys. Rev. E,* **75**, 027105 (Feb, 2007). DOI: 10.1103/PhysRevE.75.027105.

Small, M. L., *Unanticipated Gains: Origins of Network Inequality in Everyday Life*, Kindle (2009). Oxford University Press, Oxford, UK.

Tobler, W., "A Computer Movie Simulating Urban Growth in the Detroit Region," *Economic Geography,* **46**, 234-240 (1970). DOI: 10.2307/143141.

Toole, J. L., Herrera-Yaqüe, C., Schneider, C. M., and González, M. C., "Coupling human mobility and social ties," *Journal of The Royal Society Interface,* **12**:105, (February 25, 2015). DOI: 10.1098/rsif.2014.1128.

Wang, Q., Phillips, N. E., Small, M. L., and Sampson, R. J., "Urban mobility and neighborhood isolation in America's 50 largest cities," *Proceedings of the National Academy of Sciences,* **115**:30, 7735-7740 (2018). DOI: 10.1073/pnas.1802537115.

Winling, L. C., *Building the Ivory Tower: Universities and Metropolitan Development in the Twentieth Century* (2017). University of Pennsylvania Press, Philadelphia.

OpenStreetMap contributors, "Planet data retrieved from http://overpass-api.de." Accessed 2018. Accessed March 7, 2018.

Pew Research Center, "Mobile Fact Sheet." Accessed 2018. Published Feburary 5, 2018; Accessed October 15, 2018.

The Local Structures of Human Mobility in Chicago: Supplementary Materials

James Saxon

Illustration of the Constructed Variables

Figure 1 illustrates the clustering and local out-degree. The average share of residents' out-of-home behaviors are shown in blue. Expressed in the notation used in the main text, this is $\hat{A}_{h\ell}/(1-\hat{A}_{hh})$, where $\hat{A}_{h\ell}$ is the interaction level in tract ℓ by residents of h, and $\hat{A}_{hh}$ is the self-interaction rate. The "local out-degree" is the sum of these fractions that fall within the "vicinity" $\mathcal{V}_h$ of the origin h that contains the nearest $N = 40{,}000$ residents. In the Figure, that region is bounded in a thick black line. It does not include the origin tract itself. The most-distant tract in that region (tract "$k+1$," bottom left of the vicinity) is shown with a hatched pattern. The contribution to the local out-degree from that tract is downweighted by the fraction of residents needed to reach the 40,000-person population threshold. The full expression is then:

$$\text{local out-degree}_h \equiv \left[\left(\sum_{\ell\,\in\,\mathcal{V}_h} \hat{A}_{h\ell}\right) + \hat{A}_{h,k+1}\frac{N-N_k}{n_{k+1}}\right] \Big/ \left(1-\hat{A}_{hh}\right).$$

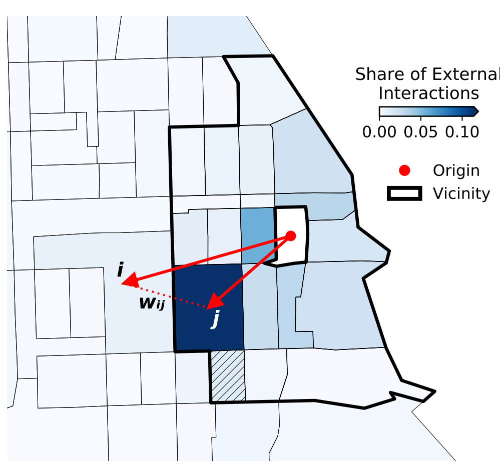
In the visual language of the Figure, this is "the share of the blue interactions in the bounded vicinity" (discounted for tract $k+1$). Intuitively, this is the rate at which residents visit the spaces that are immediately proximate to their homes.

By contrast, the clustering coefficient c_h evaluates the extent to which individuals' mobility choices coincide with the choices of residents of the neighborhoods that they visit. This evaluates the degree of closure in destinations: do residents at my destinations visit the (other) places that I visit. This is represented in Figure 1 by the dotted line, $w_{ij} = \hat{A}_{ij}/n_j$. These destination-to-destination interactions are weighted or directed towards the spaces that residents actually use, rather than the ones that happen to be geographically proximate:

$$c_h \equiv \sum_{i\neq h}\sum_{j\neq h} \hat{A}_{hi}\hat{A}_{hj}w_{ij} \Big/ \sum_{i\neq h}\sum_{j\neq h} \hat{A}_{hi}\hat{A}_{hj},$$

This concept differs from the local out-degree in that it is not geographically bounded. However, because people *do* tend to use the spaces around their homes, the two measures are strongly correlated, as shown in Table 3.

Figure 1. The two constructed variables are illustrated for a Census tract in the neighborhood of Hyde Park.



Node Split Invariance of the Clustering Coefficient

We require a node split invariant clustering coefficient, for graphs with weighted edges and nodes. It is claimed that

$$c_h = \sum_{i \neq h} \sum_{j \neq h} \hat{A}_{hi} \hat{A}_{hj} w_{ij} \Bigg/ \sum_{i \neq h} \sum_{j \neq h} \hat{A}_{hi} \hat{A}_{hj}$$

satisfies this.

Splitting the destinations would divide the interactions $\hat{A}_{ij}$, but this is compensated by the additional terms. Consider a node k of population n_k split into two nodes of populations $n_{k'}$ and $n_{k''}$. A node i that interacted with k at a level $\hat{A}_{ik}$, would interact with the two new nodes at rates proportional to their shares of the original population: $\hat{A}_{ik'} = \hat{A}_{ik} n_{k'}/n_k$ and $\hat{A}_{ik''} = \hat{A}_{ik} n_{k''}/n_k$. Note that $\hat{A}_{ik} = \hat{A}_{ik'} + \hat{A}_{ik''}$. Before splitting k, the clustering contains a term

$$\hat{A}_{ij} \hat{A}_{ik} w_{jk} = \hat{A}_{ij} \hat{A}_{ik} \hat{A}_{jk} / n_k.$$

Afterwards, this is

$$\hat{A}_{ij} \left(\hat{A}_{ik'} \hat{A}_{jk'} / n_{k'} + \hat{A}_{ik''} \hat{A}_{jk''} / n_{k''} \right).$$

Substituting $\hat{A}_{jk'} = \hat{A}_{jk} n_{k'}/n_k$ and $\hat{A}_{jk''} = \hat{A}_{jk} n_{k''}/n_k$ shows the two expressions to be equivalent:

$$\hat{A}_{ij} \left(\hat{A}_{ik'} \hat{A}_{jk} n_{k'} / n_k / n_{k'} + \hat{A}_{ik''} \hat{A}_{jk} n_{k''} / n_k / n_{k''} \right)$$

$$= \hat{A}_{ij} \left(\hat{A}_{ik'} + \hat{A}_{ik''} \right) \hat{A}_{jk} / n_k = \hat{A}_{ij} \hat{A}_{ik} \hat{A}_{jk} / n_k.$$

Splitting an origin changes nothing for interactions outside the home: the "intensive" preferences are the same on either side of the split.

But the "home location" itself is special – it is excluded from the denominator and the numerator. The clustering and local out-degree are both constructed in this way, because the social implications of staying "closed-in" at home are very different from being outside. In fact, the fraction of time at-home is higher in Chicago's more depressed neighborhoods. However, though a large piece of a user's home interactions are actually at the place of residence, some of them are in the immediate outside environment. Because the immediate outside location is the most-likely out-of-home space, the clustering coefficient – and the local out-degree – will be biased low for large home Census tracts where a higher fraction of immediately-local, non-residence locations are suppressed. To this extent, the node-split invariance is imperfect for the home location.

This weakness could be addressed by identifying home locations at finer granularity, but doing this begins to come up against the resolution of the data.

Variable Definitions for Crime Model

This Section describes the control and exogenous variables used in the two models of crime levels from the main text. There are two sets of controls, corresponding to the two major theories of criminality: routine activities and social disorganization.

Routine Activities. According to the *routine activities* perspective, crime happens when a motivated offender interacts with a target without a capable guardian present. (Boggs, 1965; Cohen and Felson, 1979) This emphasizes the need for accurate counts of all populations present at a location. Criminologists and sociologists have long converted crime counts to rates by dividing by residential populations, but there are clearly circumstances where this normalization is inadequate or inappropriate. For example, the "denominator" for a victimization rate in a city's core business district is not simply the number of residents; it includes as well the workers and visitors. In addition to changing the numerical effective population, residents, workers, and visitors may differ in their propensity to be victimized, offend, or act as guardians.

As viable alternative normalizations have become available, Andresen (2006) and others have shown that this choice of denominator is consequential in practice: it can affect parameter estimates. I reproduce this effect in the following Section, and therefore take a somewhat "aggressive" group of routine activities controls. I include the logarithms of the workplace (daytime), residential (nighttime), and ambient (cell-phone) populations, as well as the tract area. The workplace population is from the Census's LEHD Origin Destination Employment Statistics (LODES), and the residential populations are from the American Community Survey (ACS). The LODES data are derived from administrative records of unemployment insurance for wage and salary jobs. They cover approximately 95 percent of these jobs. (Graham et al., 2014) Ambient populations are defined as in the main text as the normalized share of smartphones locations that are recorded in each tract. It represents the total population present in a space. It includes people in all of their functions: work and home, but also leisure and transit, for instance.

In the routine activities perspective, the network variables may be considered as indicators of the availability of capable guardians.

Social Disorganization. All "social disorganization" covariates are derived from the American Community Survey 2017 5-year estimates, except for the fraction of tracts' areas zoned commercial. To facilitate comparisons and freeze parameters in the model, the variables are constructed to align with Browning et al. (2017a) and Morenoff et al. (2001) to the extent practicable.

Disadvantage and *residential instability* are both the first principal component of a set normalized inputs. Each of these PCAs is computed for all tracts in Cook County, Illinois.

- *Disadvantage*: combines fraction not working (unemployed or out of the labor force), log median household income, fraction of the adult population with a bachelor's degree, fraction of workers in managerial positions, and fraction of households with a female head. The sign is set to be anti-correlated with bachelor's degree attainment.
- *Residential instability*: combines fraction of residents who were living in the same house one year ago, fraction of units occupied by renters, and fraction of units vacant. The sign is set to ensure a positive correlation with fraction of units occupied by renters.

The final constructed variable is:

- *Racial heterogeneity* is the sum of the squares of all fractional races: White, Black, Native American, Asian, Hawaiian/Pacific Islander (all non-Hispanic), or Hispanic. Similarly, the variable *Fraction Black* does not include Hispanics.

Two variables require minimal manipulation or explanation:

- *Fraction young*: share of the population between 15 and 24 years old, inclusive. This is slightly more expansive than Browning et al. It is what was available in the Census profile API.
- *Fraction black*: non-Hispanic Blacks.
- *Married households*: fraction of households a married-couple family.
- *Recent immigrants*: foreign-born, entered the US 2010 or later.

Finally, a single variable is based on data from the City of Chicago:

- *Fraction commercial* is the share of a tract's area that is zoned commercial. I define commercial zoning loosely, as anything that is neither a park nor residential (single, twin, multiple, or downtown). This includes most downtown spaces, planned developments like airports, ballparks, and universities, as well as traditional business, commerce, manufacturing, and transportation designations.

Crime Data. Geocoded crime counts and zoning data are from the City of Chicago, for the 5-year period from 2013 to 2017. This time window significantly exceeds the single-month observation from the cell phone data, but aligns with the ACS data, and it allows for logged violent and property crimes at the Census tract level. The aim of the criminality analysis is to demonstrate the information content of the constructed values; it is not a causal argument. For this reason, the statistical precision of the longer sample period is preferred over one temporally aligned with or offset from the cell phone data. The longer window also allows for logged crime levels by Census tract. Violent and property offenses are categorized by FBI Code following the definitions of the Chicago Police Department.

Crime data are from the City of Chicago. Offenses are categorized by FBI Code according to the classification of the Chicago Police Department.

- Violent crime includes: homicide of both first and second degree (code 1A), criminal sexual assault (2), robbery (3), aggravated assault (4A), and aggravated battery (4B).
- Property crime includes: burglary (5), larceny (6), motor vehicle theft (7), and arson (9).

Estimation Robustness

Complete Specification

The regressions from the main text are shown with full controls in Table 1.

Estimation and Weights

Table 2 shows that the main results on mobility variables are robust to alternative weights definitions or GMM estimation.

Alternative Sample Definitions for Constructed Variables

A persistent concern with the present data is that the results are a relic of still-novel data collection methods, instead of what they are substantively "intended" to represent. In particular, one would worry that devices are not representative of the human populations of the tracts from which they originate. One way to probe this, is to change how devices are selected.

I do this in two ways. First, multiple applications may record a location simultaneously on a single device. I remove all such "duplicate" pings, as well as pings with estimated precision worse than 500 meters. Second, I additionally require that devices show up more-consistently in the sample: that they appear at home on at least three nights, that they register locations in the beginning, middle, and final thirds of study period (month), and that they record at least a hundred unique locations. I refer to these samples as "deduplicated" and "restricted." Deduplication and the additional precision requirement reduces the number of pings from 238 million to 132 million. The restricted samples contains 94 thousand devices – less than a third of the "nominal" sample. These restricted devices record a total of 93 million unique locations, and the median device in this sample records 556 locations (as compared with 200 locations in the nominal sample).

Based on the deduplicated sample, the correlations between the clustering and local out-degree rises from the nominal level of 0.52 to 0.61. Using the restricted sample, it rises yet further to 0.73. All population-weighted correlations among the constructed variables, and with the adult bachelor's attainment rate, are shown in Table 3. Tables 4 and 5 present the criminality regressions of the paper, for clustering and local out-degree derived from the alternative samples.

Alternative Variables

It is also natural to ask whether simpler network variables or already-available Census variables capture the information content of the clustering coefficient and local out-degree, as constructed.

From the network side, a natural candidate is base level of activity: the number of trips or total out-degree. Because the cell phone applications do not record locations continuously, I do not construct "trips" per se. But it is trivial to construct the total fractional time out of the home; using the notation of the main text, it is $1 - \hat{A}_{hh}$. Since my definition of the local out-degree is constructed as a fraction of the *out-of-home* exposures, it is constructed to be independent of the total out-of-home time. Table 6 shows that, empirically, the full out-fraction is not nearly so performant: the parameter estimates are not significant, and including the variable degrades the Akaike Information Criterion.

For the Census variables, two natural candidates describe neighborhood commuting behaviors: the likelihood of walking to work, and the length of a commute. Are physical behaviors near home simply driven by work locations, themselves already measured by the Census. These variables are missing for one Census tract in Chicago, so the Akaike Information Criterion (AIC) is not directly comparable. But any gain is small, the pseudo-R^2 is unaffected, the parameter estimates are insignificant, and the parameters for the "benchmark" network observables are unaffected (and remain very significant) – again, see Table 6

Finally, one could include the rate of app use in the regressions. There is very large variability in the number of "pings" (location records) per device (Figure2). The constructed variables are based on tract-wide average behaviors, with each device receiving one vote. But one might imaging that the constructed variables

Table 1. Log violent and property crime counts, regressed on the clustering coefficient and local out-degree, with "standard" controls for social disorganization and routine activities.

Dependent	Violent	Violent	Violent	Property	Property	Property
Constant	1.34***	1.52***	1.55***	3.21***	3.33***	3.31***
	(0.31)	(0.30)	(0.31)	(0.25)	(0.25)	(0.25)
Log Clustering		-0.24***			-0.20***	
		(0.05)			(0.04)	
Local Out-Degree			-0.86**			-0.51*
			(0.27)			(0.22)
Log Res.	0.39***	0.42***	0.40***	0.34***	0.37***	0.35***
	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)
Log Ambient	0.28***	0.26***	0.26***	0.31***	0.31***	0.30***
	(0.03)	(0.03)	(0.04)	(0.03)	(0.03)	(0.03)
Log Work	0.03*	0.02	0.03	0.06***	0.05***	0.05***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Log Area	0.16***	0.14***	0.14***	0.20***	0.17***	0.18***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Disadvantage	0.07***	0.04*	0.07***	-0.03*	-0.07***	-0.04*
	(0.02)	(0.02)	(0.02)	(0.01)	(0.02)	(0.01)
Married Households	-0.34	-0.43*	-0.33	-0.26	-0.37*	-0.26
	(0.19)	(0.19)	(0.19)	(0.16)	(0.16)	(0.16)
Residential Instability	0.10***	0.10***	0.10***	0.06***	0.07***	0.06***
	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)
Fraction Young	-0.17	-0.19	-0.18	0.08	0.05	0.08
	(0.21)	(0.21)	(0.21)	(0.18)	(0.17)	(0.18)
Recent Immigrants	-1.42**	-1.36**	-1.34**	-1.29***	-1.23***	-1.23***
	(0.44)	(0.44)	(0.44)	(0.37)	(0.37)	(0.37)
Fraction Black	0.85***	0.82***	0.81***	0.52***	0.51***	0.51***
	(0.11)	(0.11)	(0.11)	(0.08)	(0.08)	(0.08)
Racial Heterogeneity	0.02	0.02	0.00	-0.03	-0.03	-0.04
	(0.11)	(0.11)	(0.11)	(0.09)	(0.09)	(0.09)
Log Fraction Commercial	-0.27***	-0.27***	-0.28***	-0.04	-0.03	-0.05
	(0.06)	(0.06)	(0.06)	(0.05)	(0.05)	(0.05)
Log Fraction Commercial Sq.	-0.08***	-0.08***	-0.08***	-0.03	-0.02	-0.03
	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)
λ (Error)	0.86***	0.84***	0.85***	0.69***	0.69***	0.69***
	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)
N	780	780	780	780	780	780
Pseudo-R^2	0.64	0.68	0.65	0.66	0.68	0.66
AIC	664	643	656	351	326	348
Weights	Queen	Queen	Queen	Queen	Queen	Queen
Routine	ML Error	ML Error	ML Error	ML Error	ML Error	ML Error

Standard errors in parentheses; $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

would be sensitive to neighborhoods' overall data rates. This does not seem to be the case. Tract ping rates are minimally correlated with the local out-degree (-0.005 for median and -0.02 for mean pings). The correlation is low, but somewhat larger for clustering (-0.23 and -0.05 for median and mean pings) but these correlations evaporate (-0.08, and -0.006) in the restricted sample described above. Moreover, Table 7 shows that controlling for tracts' mean or median deduplicated user ping rate, does not affect the core parameters, Coefficients for the ping rates are themselves insignificant. In short: simpler constructions, existing variables, and app use do not perform nearly so well as the theoretically-motivated variables.

Table 2. Estimates of are consistent or slightly more-pronounced, with alternative definitions of the spatial weights matrix and with heteroskedastic-robust GMM estimation.

Dependent	Log Violent Crime				
Constant	1.52***	1.43**	1.23**	1.20**	1.11*
	(0.30)	(0.43)	(0.43)	(0.43)	(0.44)
Log Clustering	-0.24***	-0.28***	-0.28***	-0.28***	-0.27***
	(0.05)	(0.06)	(0.06)	(0.06)	(0.06)
Log Res.	0.42***	0.44***	0.45***	0.46***	0.46***
	(0.04)	(0.06)	(0.06)	(0.06)	(0.06)
Log Ambient	0.26***	0.27***	0.30***	0.29***	0.29***
	(0.03)	(0.05)	(0.05)	(0.05)	(0.05)
Log Work	0.02	0.02	0.02	0.02	0.02
	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)
Log Area	0.14***	0.08	0.06	0.07	0.06
	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)
Disadvantage	0.04*	0.07***	0.07**	0.07**	0.07**
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Married	-0.43*	-0.49*	-0.49*	-0.42	-0.37
Households	(0.19)	(0.24)	(0.24)	(0.24)	(0.25)
Residential	0.10***	0.13***	0.13***	0.13***	0.14***
Instability	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Fraction Young	-0.19	-0.08	-0.17	-0.17	-0.23
	(0.21)	(0.32)	(0.32)	(0.30)	(0.31)
Recent	-1.36**	-1.55*	-1.56*	-1.52*	-1.61*
Immigrants	(0.44)	(0.69)	(0.67)	(0.64)	(0.65)
Fraction Black	0.82***	0.94***	0.98***	0.98***	1.00***
	(0.11)	(0.14)	(0.15)	(0.16)	(0.16)
Racial	0.02	0.01	-0.01	0.00	-0.04
Heterogeneity	(0.11)	(0.13)	(0.15)	(0.14)	(0.14)
Log Fraction	-0.27***	-0.21**	-0.19**	-0.20**	-0.19*
Commercial	(0.06)	(0.07)	(0.07)	(0.08)	(0.08)
Log Fraction	-0.08***	-0.08***	-0.08***	-0.07***	-0.07***
Commercial Sq.	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
λ (Error)	0.84***	0.75	0.76	0.81	0.84
	(0.02)				
N	780	780	780	780	780
Pseudo-R^2	0.68	0.73	0.74	0.73	0.73
AIC	643				
Weights	Queen	Queen	KNN6	KNN8	KNN10
Routine	ML Error	GMM Error with Heteroskedasticity			

Standard errors in parentheses; $*p < 0.05$, $**p < 0.01$, $***p < 0.001$.

The Risk Sets for Urban Crime

The routine activities perspective emphasizes the need for an appropriate and complete measure of the risk set. (Boggs, 1965; Cohen and Felson, 1979) Though working populations have been available from the Census' Longitudinal Employer-Household Dynamics "LODES" program since 2002, non-residential populations are often ignored by researchers. Entering into the modern era, (Andresen, 2006) showed that altering the available normalization "risk sets" can change parameter estimates. I reproduce this effect here. Table 8 present spatial error model regressions for the logarithm of property crime, first controlling for the residential, ambient, and working populations alone, and then with all three together. The regression also includes the "social disorganization" controls used for the main results.

Table 3. Population-weighted correlations among constructed variables from the nominal (nom.), deduplicated (dedup.), and restricted (restr.) samples, and data from the American Community Survey (ACS).

		Log Clustering			Local Out-Degree			Bachelor's
	Sample	Nom.	Dedup.	Restr.	Nom.	Dedup.	Restr.	(ACS)
Log Clustering	Nominal	1.00	0.99	0.89	0.52	0.62	0.63	0.85
	Deduplicated	0.99	1.00	0.88	0.51	0.61	0.62	0.87
	Restricted	0.89	0.88	1.00	0.60	0.68	0.73	0.80
Local Out-Degree	Nominal	0.52	0.51	0.60	1.00	0.95	0.80	0.43
	Deduplicated	0.62	0.61	0.68	0.95	1.00	0.86	0.51
	Restricted	0.63	0.62	0.73	0.80	0.86	1.00	0.55
Bachelor's Attainment	(ACS)	0.85	0.87	0.80	0.43	0.51	0.55	1.00

Table 4. Log violent crime counts regressed on the clustering coefficient and local out-degree, with full controls (parameters not shown). Three different samples are used to construct the clustering and out-degree. The "nominal" version, used in the main text, simply assigns devices' home locations according to their modal night-time location. The second "deduplicated" version removes pings recorded for a single users by multiple applications at the same time. The final, "restricted" version requires that users record at least three nights at home and one hundred total pings, and that they show up in the data set in beginning, middle, and end of the study period (all three thirds). Parameters remain quite significant, and are consistent across models. All models are evaluated using the ML Error routine with Queen weights.

Dependent	Violent	Violent	Violent	Violent	Violent	Violent	Violent
Log Clustering		-0.24***	-0.22***	-0.24***			
		(0.05)	(0.05)	(0.05)			
Local Out-Degree					-0.86**	-1.04***	-1.12***
					(0.27)	(0.28)	(0.25)
Variable Version		Nom.	Dedup.	Restr.	Nom.	Dedup.	Restr.
N	780	780	780	780	780	780	780
Pseudo-R^2	0.64	0.68	0.67	0.66	0.65	0.65	0.65
AIC	664	643	644	646	656	652	646

Standard errors in parentheses; $^*p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

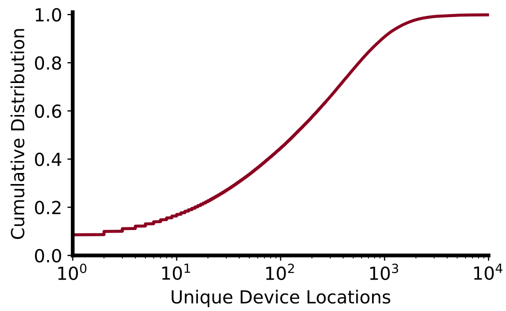
Table 5. Log property crime counts regressed on the clustering coefficient and local out-degree, with full controls (parameters not shown). Three different samples are used to construct the clustering and out-degree. The "nominal" version, used in the main text, simply assigns devices' home locations according to their modal night-time location. The second "deduplicated" version removes pings recorded for a single users by multiple applications at the same time. The final, "restricted" version requires that users record at least three nights at home and one hundred total pings, and that they show up in the data set in beginning, middle, and end of the study period (all three thirds). Parameters remain quite significant, and are consistent across models. All models are evaluated using the ML Error routine with Queen weights.

Dependent	Property	Property	Property	Property	Property	Property	Property
Log Clustering		-0.20***	-0.20***	-0.18***			
		(0.04)	(0.04)	(0.04)			
Local Out-Degree					-0.51*	-0.68**	-0.71***
					(0.22)	(0.23)	(0.21)
Variable Version		Nom.	Dedup.	Restr.	Nom.	Dedup.	Restr.
N	780	780	780	780	780	780	780
Pseudo-R^2	0.66	0.68	0.67	0.67	0.66	0.66	0.67
AIC	351	326	324	336	348	344	341

Standard errors in parentheses; $^*p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

There are two conclusions to draw from this exercise. The first is that the ambient population as constructed is serviceable as a population. On its own it in fact offers the lowest AIC of the three populations; adding it

Figure 2. Cumulative distribution of unique, precise locations per device.



and/or the working population substantially improves the pseudo-R^2 of the model. In short, the ambient (or work-time) population is a straightforward and necessary control in establishing relevant risk set for crime rates. The second point is that the coefficients of the other parameters, and qualitative conclusions, are not robust to these changes in the controls available for the risk set. In particular, the table shows significant changes in estimates of the coefficients for commercial zoning and social disadvantage.

The changes in the commercial zoning variables are to be expected since they provide a crude proxy for the "missing" populations, which change from column to column. That is the intent, in the model by Browning et al. (2017a). The variables respond predictably to the presence or absence of components of the effective population. Lacking a direct measure of local employment, commercial spaces stand in as a proxy for the working population. Conversely, absent a control for the residential population, the fraction commercial indicates the space remaining for residences. In the model, the coefficients for fraction of tracts' area zoned commercial and its square are positive and significant with residential population alone, negative and significant for work population alone, and insignificant with the ambient population alone.

But the zoning variables make an unsatisfactory proxy: they do not make the parameter estimates robust to the inadequate risk set. The changes in the social disadvantage estimate and residential instability estimates are, while understandable, undesirable and unexpected. The incomplete population controls lead to unstable and misleading estimates. This motivates the full set of population controls (residential, work, and ambient) in the main results of this paper. It is worth acknowledging that the model's sensitivity to the risk normalization is stronger for property than violent crime; in the latter case, residential population is a better measure of the risk.

Table 6. Log violent crime counts, regressed on the clustering coefficient and local out-degree, with "standard" controls for social disorganization and routine activities. In addition, this table simpler commuting and network variables. These variables are not themselves significant and do not affect the clustering and local out-degree variables. All models are evaluated using the ML Error routine with Queen weights.

Dependent	Violent	Violent	Violent	Violent	Violent	Violent
Constant	1.52***	1.56***	1.50***	1.55***	1.55***	1.69***
	(0.30)	(0.32)	(0.35)	(0.31)	(0.33)	(0.37)
Log Clustering	-0.24***	-0.24***	-0.24***			
	(0.05)	(0.05)	(0.05)			
Local Out-Degree				-0.86**	-0.88***	-0.91***
				(0.27)	(0.27)	(0.27)
Log Res.	0.42***	0.42***	0.42***	0.40***	0.40***	0.39***
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Log Ambient	0.26***	0.26***	0.26***	0.26***	0.26***	0.25***
	(0.03)	(0.03)	(0.03)	(0.04)	(0.04)	(0.04)
Log Work	0.02	0.03	0.02	0.03	0.03*	0.03
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Log Area	0.14***	0.14***	0.14***	0.14***	0.14***	0.14***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Disadvantage	0.04*	0.05*	0.04*	0.07***	0.08***	0.07***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Married	-0.43*	-0.38*	-0.43*	-0.33	-0.28	-0.33
Households	(0.19)	(0.19)	(0.19)	(0.19)	(0.19)	(0.19)
Residential	0.10***	0.11***	0.10***	0.10***	0.10***	0.10***
Instability	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Fraction Young	-0.19	-0.25	-0.19	-0.18	-0.23	-0.18
	(0.21)	(0.22)	(0.21)	(0.21)	(0.22)	(0.21)
Recent	-1.36**	-1.41**	-1.36**	-1.34**	-1.40**	-1.34**
Immigrants	(0.44)	(0.44)	(0.44)	(0.44)	(0.44)	(0.44)
Fraction Black	0.82***	0.84***	0.82***	0.81***	0.82***	0.81***
	(0.11)	(0.11)	(0.11)	(0.11)	(0.11)	(0.11)
Racial	0.02	0.02	0.02	0.00	-0.00	0.01
Heterogeneity	(0.11)	(0.11)	(0.11)	(0.11)	(0.11)	(0.11)
Log Fraction	-0.27***	-0.28***	-0.27***	-0.28***	-0.29***	-0.28***
Commercial	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
Log Fraction	-0.08***	-0.08***	-0.08***	-0.08***	-0.08***	-0.08***
Commercial Sq.	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Walkers		0.02			0.06	
		(0.29)			(0.29)	
Commute Time		-0.00			-0.00	
		(0.00)			(0.00)	
Out Fraction			0.03			-0.17
			(0.24)			(0.25)
λ (Error)	0.84***	0.84***	0.84***	0.85***	0.85***	0.85***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
N	780	779	780	780	779	780
Pseudo-R^2	0.68	0.68	0.68	0.65	0.66	0.66
AIC	643	641	645	656	654	657

Standard errors in parentheses; $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

Table 7. Log violent crime counts, regressed on the clustering coefficient and local out-degree, with "standard" controls for social disorganization and routine activities. In addition, this table includes mean and median ping rates at the tract level as explanatory variables. These variables are not themselves significant and do not affect the clustering and local out-degree variables. All models are evaluated using the ML Error routine with Queen weights.

Dependent	Violent	Violent	Violent	Violent	Violent	Violent
Constant	1.52***	1.45***	1.51***	1.55***	1.48***	1.54***
	(0.30)	(0.31)	(0.30)	(0.31)	(0.31)	(0.31)
Log Clustering	-0.24***	-0.23***	-0.23***			
	(0.05)	(0.05)	(0.05)			
Local Out-Degree				-0.86**	-0.87**	-0.86**
				(0.27)	(0.27)	(0.27)
Log Res.	0.42***	0.41***	0.42***	0.40***	0.38***	0.40***
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Log Ambient	0.26***	0.28***	0.26***	0.26***	0.28***	0.26***
	(0.03)	(0.04)	(0.03)	(0.04)	(0.04)	(0.04)
Log Work	0.02	0.02	0.02	0.03	0.03	0.02
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Log Area	0.14***	0.14***	0.14***	0.14***	0.13***	0.14***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Disadvantage	0.04*	0.05*	0.04*	0.07***	0.07***	0.07***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Married	-0.43*	-0.43*	-0.42*	-0.33	-0.34	-0.32
Households	(0.19)	(0.19)	(0.19)	(0.19)	(0.19)	(0.19)
Residential	0.10***	0.10***	0.10***	0.10***	0.10***	0.10***
Instability	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Fraction Young	-0.19	-0.19	-0.19	-0.18	-0.18	-0.17
	(0.21)	(0.21)	(0.21)	(0.21)	(0.21)	(0.21)
Recent	-1.36**	-1.31**	-1.34**	-1.34**	-1.29**	-1.32**
Immigrants	(0.44)	(0.44)	(0.44)	(0.44)	(0.44)	(0.44)
Fraction Black	0.82***	0.82***	0.83***	0.81***	0.80***	0.82***
	(0.11)	(0.11)	(0.11)	(0.11)	(0.11)	(0.11)
Racial	0.02	0.03	0.02	0.00	0.02	0.01
Heterogeneity	(0.11)	(0.11)	(0.11)	(0.11)	(0.11)	(0.11)
Log Fraction	-0.27***	-0.28***	-0.27***	-0.28***	-0.29***	-0.28***
Commercial	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
Log Fraction	-0.08***	-0.08***	-0.08***	-0.08***	-0.08***	-0.08***
Commercial Sq.	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Median Pings		0.00			0.00	
		(0.00)			(0.00)	
Mean Pings			0.00			0.00
			(0.00)			(0.00)
λ (Error)	0.84***	0.85***	0.84***	0.85***	0.86***	0.86***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
N	780	780	780	780	780	780
Pseudo-R^2	0.68	0.67	0.67	0.65	0.65	0.65
AIC	643	642	643	656	655	656

Standard errors in parentheses; $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

Table 8. Regressions of logged crime counts yield qualitatively different estimates for parameters of social disorganization, depending on the available "risk sets." This conclusion holds for crime rates with varying normalizations, but the normalizations affect pre-fit variance and lead to variation in the R^2 and AIC.

Dependent	Log Property Crime			
Constant	1.37***	4.23***	5.66***	3.21***
	(0.25)	(0.15)	(0.17)	(0.25)
Log Res.	0.82***			0.34***
	(0.03)			(0.03)
Log Ambient		0.69***		0.31***
		(0.02)		(0.03)
Log Work			0.27***	0.06***
			(0.01)	(0.01)
Disadvantage	-0.07***	-0.00	0.00	-0.03*
	(0.01)	(0.02)	(0.02)	(0.01)
Married	-0.26	0.11	-0.34	-0.26
Households	(0.19)	(0.18)	(0.23)	(0.16)
Residential	0.06***	0.04*	0.04	0.06***
Instability	(0.02)	(0.02)	(0.02)	(0.01)
Fraction Young	0.09	0.12	0.10	0.08
	(0.22)	(0.21)	(0.27)	(0.18)
Recent	-2.24***	-1.59***	-0.99	-1.29***
Immigrants	(0.46)	(0.43)	(0.55)	(0.37)
Fraction Black	0.79***	0.53***	0.59***	0.52***
	(0.09)	(0.09)	(0.11)	(0.08)
Racial	-0.08	-0.06	-0.00	-0.03
Heterogeneity	(0.11)	(0.10)	(0.13)	(0.09)
Log Fraction	0.36***	-0.09	-0.33***	-0.04
Commercial	(0.06)	(0.06)	(0.08)	(0.05)
Log Fraction	0.05**	-0.03*	-0.07***	-0.03
Commercial Sq.	(0.02)	(0.02)	(0.02)	(0.01)
λ (Error)	0.49***	0.62***	0.54***	0.69***
	(0.05)	(0.04)	(0.04)	(0.03)
N	780	780	780	780
Pseudo-R^2	0.59	0.59	0.39	0.66
AIC	708	594	978	351
Weights	Queen	Queen	Queen	Queen
Routine	ML Error	ML Error	ML Error	ML Error

Standard errors in parentheses; $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

References for Supplementary Materials

Andresen, M. A., "Crime Measures and the Spatial Analysis of Criminal Activity," *The British Journal of Criminology,* **46**:2, 258-285 (2006). DOI: 10.1093/bjc/azi054.

Boggs, S. L., "Urban Crime Patterns," *American Sociological Review,* **30**:6, 899-908 (1965). DOI: 10.2307/2090968.

Browning, C. R., Calder, C. A., Boettner, B., and Smith, A., "Ecological Networks and Urban Crime: The Structure of Shared Routing Activity Locations and Neighborhood-Level Informal Control Capacity," *Criminology,* **55**:4, 754-778 (2017a). DOI: 10.1111/1745-9125.12152.

Cohen, L. E. and Felson, M., "Social Change and Crime Rate Trends: A Routine Activity Approach," *American Sociological Review,* **44**:4, 588-608 (1979). DOI: 10.2307/2094589.

Graham, M. R., Kutzbach, M. J., and McKenzie, B., *Design Comparison of LODES and ACS Commuting Data Products.* Technical Report 14-38 (October 2014). Center for Economic Studies, U.S. Census Bureau.

Morenoff, J. D., Sampson, R. J., and Raudenbush, S. W., "Neighborhood Inequality, Collective Efficacy, and the Spatial Dynamics of Urban Violence," *Criminology,* **39**:3, 517-558 (2001). DOI: 10.1111/j.1745-9125.2001.tb00932.x.

Chicago Police Department, "Crime Type Categories: Definition & Description." Accessed April 8, 2019.