Theme issue article

Urban Analytics and City Science

Environment and Planning B: Urban Analytics and City Science
0(0) 1–19

DOI: 10.1177/2399808318785375
journals.sagepub.com/home/epb
SAGE

Large-scale spatial network models: An application to modeling information diffusion through the homeless population of San Francisco

Zack W Almquist
University of Minnesota, USA

Abstract

To address the effects of increasing homeless populations, planners must understand the size and distribution of their homeless populations, as well as how information and resources are diffused throughout homeless communities. Currently, there is limited publicly available information on the homeless population, e.g. the estimates of the homeless, gathered annually by the US Housing & Urban Development point in time survey. While it is theorized in the literature that the networks of homeless individuals provide access to important information for planners in areas such as health (e.g. needle exchanges) or access (e.g. information diffusion about the location of new shelters), it is almost never measured, and if measured, only at a very small scale. This research addresses the question of how planners can leverage publicly available data on the homeless to better understand their own homeless networks (e.g. relations among the homeless themselves) in a cost-effective and reliable way. To this end, we provide a method for simulating realistic networks of a social relation among the homeless population and perform a diffusion analysis over the resultant homeless-to-homeless networks, and also over a simulated homeless youth Facebook network. We validate the former through novel use of historical data, while the latter is based on recent work that demonstrated that the homeless youth have similar size Facebook networks and usage. We see much stronger spatial hopping and quicker diffusion over the youth network, i.e. we expect information to pass among the youth network much faster than the homeless-to-homeless network. This finding implies that non-government organizations and public health efforts that seek to provide information, goods or services to the homeless should start with the homeless youth, given the potential for faster diffusion when homeless youth are the initial transmitters. Overall, these methods and analysis provide a unique opportunity for visualizing, characterizing and inferring information for large-scale and hard to measure social networks.

Corresponding author:
Zack W Almquist, University of Minnesota, 267, 19th Avenue South, Minneapolis, MN 55455, USA.
Email: almquist@umn.edu

Keywords
Spatial networks, networks, social networks, Bernoulli graphs, spatial modeling, homeless, social cohesion, facebook, youth networks

Introduction

In recent years, the prevalence of homelessness in the United States (US), and its well-established contributions to adverse outcomes – including social and economic inequality – has led to what many refer to as a nationwide emergency. Perhaps most notably, the state of Hawaii and the cities of Los Angeles, California, Portland, Oregon and Seattle, Washington have all made emergency declarations typically reserved for natural disasters such as earthquakes or fires. In combating homelessness in the US, organizations – city, county, state, federal and nongovernmental organizations (NGOs) – attempt to engage homeless individuals at the micro-level; most large-scale (public) data on homelessness in the US are based primarily on the US Housing & Urban Development point in time (PiT) survey conducted every January, which provides only a count of the homeless population in spatial aggregates once a year. Further, this is limited to simple count data of the sheltered and unsheltered homeless populations in the US, and sometimes a limited survey run at the same time; however, the more detailed survey is not typically made public or coordinated with other geographic areas. One data source that would be of particular interest to planners would be that of the social support networks of the homeless population. In particular, we know from Wolch and Rowe (1992) and others that the social support networks (e.g. homeless-to-homeless) are particularly important for planning issues related to public health (e.g. sanitation), information (e.g. provider locations) or safety (e.g. information on safe places to sleep at night). Networks can be particularly important in a health context. For example, a recent outbreak of Hepatitis A in San Diego County over the course of January 2018 resulted in 577 infections with 20 deaths (3.5%) and 395 hospitalizations (68.5%). Hepatitis A is spread via person-to-person contact and through contact with a fecally contaminated environment, where the typical public health response is the distribution of hand-wash stations and sanitizers. To contain this outbreak, public health workers need to engage the network of homeless individuals to encourage them and to let them know where these stations will be located.[1] Further, we know that social support networks (e.g. Coleman, 1988; Putnam, 1995) are important mechanisms for how information and goods diffuse through the population, e.g. finding out about new shelters, access to resources, public health information and so forth. These networks also underlie modern methods for enumeration of hard-to-reach populations like the homeless.

To date, there have been numerous small-scale qualitative studies assessing the level of access that homeless populations have to social capital and organizational resources (e.g. Elvy et al., 1986). Scholars such as Mitchell (1987) have investigated the *strength of weak ties* (Granovetter, 1974) in the context of homeless women – describing the multiplex social support network (e.g. emotional support or practical aid) of these women based on their weak ties (e.g. relationships with care providers or other homeless) and strong ties (e.g. kin or close friends). There also exists an extensive literature of small-scale health-based research of the homeless and the importance of social support for mental illness (e.g. Hwang et al., 2009; Irwin et al., 2008; Rice et al., 2011, 2012; Rowe and Wolch, 1990; Shinn et al., 1991), but there continues to be a dearth of population-level measures of the social relations of the homeless.

Thus, the goal of this work is to show the ability to leverage knowledge of spatial location (or approximate location) to understand and predict social interaction. Here, we focus on a spatial network framework, referred to as *spatial Bernoulli networks*, which have been used for obtaining intuition and inference on problems with minimal data and no network information, e.g. homeless-to-homeless social networks in the US. This work takes advantage of new developments in spatial networks (Onnela et al., 2011) and resulting integration with GIS Systems (Andris, 2016) to provide a basic understanding and initial hypotheses of what homeless-to-homeless social networks may look like and their resulting characteristics.

This article is laid out as follows. First, we introduce the statistical and computational methods needed for simulating and inferring the characteristics of a spatially distributed homeless-to-homeless network and the online social network of homeless youth. Next, we review the simulation procedures and simulate the homeless population locations from PiT count data and then simulate spatial networks. Finally, we perform some validation of the models and a diffusion simulation to understand how goods or information passing through these two homeless-to-homeless networks might occur. We conclude with a discussion of how this could be used to inform public planners.

Spatial network models for extrapolative simulation of homeless-to-homeless networks

Wolch and Rowe (1992) and others have found that the coping abilities of the homeless are impacted by their social networks (homeless-to-homeless networks) and their daily paths. Specifically, homeless-to-homeless networks allow for access to food, clothes, shelter, emotional support, security, etc. While this work and others have detailed the importance of homeless-to-homeless networks, it is unfortunately expensive and complex data to collect. So the question becomes: can we gain traction on this problem using existing publicly available data? This is largely limited to count data in spatial aggregate units (e.g. cities or counties; Burt et al., 2007). Hence to answer this question, we need a way to take aggregate spatial data and simulate a social network with the desired properties. Almquist and Butts (2012) have shown this can be done for some populations and both Almquist and Butts (2015) and Hipp et al. (2013) have further demonstrated that *spatial Bernoulli networks* with an appropriate data set can produce spatially embedded networks that correlate with social outcomes of interest such as regional identification (e.g. I am a Parisian or I am from New York) or crime. To apply this method, we do not need to know an individual's exact location, but we do need to assume they have some sort of stable "home" location. If, again, we take an insight from Wolch and Rowe (1992) that homeless maintain a set of stable relationships and tend to have stable locations to which they return (e.g. shelters, tent camps, etc.), and we combine this with Tobler's First Law, "everything is related to everything else, but near things are more related than distant things" (Tobler, 1970), then we might imagine that given a homeless individual's base location, we might be able to get leverage on their social network.

In the next section, we will review *spatial Bernoulli networks* and their properties before applying them to gain traction on two homeless-to-homeless networks.

Methods

Spatially embedded networks

Human social systems often reside in a geographic space, which can constrain social structure (e.g. friendships or even needle exchanges) in complex ways. In this section, we will focus on the formal definitions of a spatial network as well as key components in the larger geographic and planning literature that have looked into how both space and place may structure human interaction such as road systems or neighborhood associations.

To define a spatially embedded network, we need to first define a *network* (Wasserman and Faust, 1994). A network in this case refers to a collection of entities (e.g. individuals or organizations) and a set of relations between these entities (e.g. friendship or collaboration). We typically represent networks as mathematical objects known as *graphs*. A graph (G) is a set composed of two subsets: (1) a vertex (V) or node set (e.g. a homeless individual) and (2) an edge (E) set (e.g. friendship or communication). Together these sets are referred to as a graph, $G = (V, E)$. The graph can also be represented as a binary adjacency matrix $Y_{ij} = 1$ or 0 with size n vertices. Self-ties (i.e. Y_{ii}) are typically assumed to be NA or 0, but can be considered in the analysis if desired. This resulting network is said to be directed when Y_{ij} does not have to equal Y_{ji}, and is said to be undirected if $Y_{ij} = Y_{ji}$ (i.e. the adjacency matrix is symmetric).

Examples of directed networks include relations such as communication or self-reported friendship while undirected networks include relations such as Facebook friendship or co-occurrence in a geographic location. A *spatially embedded* network is one where the vertices (and/or edges) have a geographically associated property (e.g. the longitude and latitude location of a home).[2] It is worth pointing out that, generally speaking, spatial networks can refer to brain networks, cosmology networks and other non-geographic networks; however, in this work, we will not distinguish between spatial and non-geographic networks.

From the point of view of researchers, geolocated information is either a powerful covariate/control for modeling/predicting social action or is, itself, the endpoint of the analysis. Recently, there has been work on studying the latter case. For example, large-scale complex systems arise naturally in many social science and public planning applications, especially when studying large-scale systems such as cities or states, and large-scale social structure is inevitably shaped by geographic factors through its ability to inhibit or encourage interaction (e.g. Bossard, 1932; Festinger et al., 1950; Freeman et al., 1988; Hägerstrand, 1966; Latané et al., 1994). Further, there is a growing literature focusing on various methods, models and concepts of spatial networks, which includes numerous studies on migration and location-based social networks (e.g. Cho et al., 2011; Csáji et al., 2013), issues of social contagion and propagation (e.g. Christakis and Fowler, 2013; Pentland, 2014) and social movements (e.g. Almquist and Bagozzi, 2016; Conover et al., 2013). Much of this is made available by extensive developments in software and computing, as well as developments in the theory of spatial networks (e.g. Acedo et al., 2017b; Acedo et al, 2017b; Li et al., 2017; Pelcer-Vujačić and Kovačević, 2016; Psyllidis, 2016; Xu et al., 2017). Other key GIS network activities include investigating geographic communities (Comber et al., 2012), mapping kinship networks (Koylu et al., 2014) and exploring network centrality/influence measures (Barthélemy, 2011; Crucitti et al., 2006; Gil, 2016; Wang et al., 2011). Most of this work is empirically driven through modern data collection (e.g. Carrasco et al., 2008) and computational storage considerations (e.g. Câmara et al., 1996) as well as new methods for plotting and displaying spatial network data (e.g. Heer and Perer, 2014).

Spatial Bernoulli networks

There is a history of *network-related* spatial models in the social science literature, such as gravity models (e.g. Haynes and Fortheringham, 1984; Wilson, 2011). Recently, there has been a great deal of development in the spatial network subfield (e.g. Butts et al., 2012), including the development of a holistic framework for spatial network models based on a mapping $\mathcal{F}_d(d, \psi)$ that (for some real parameter ψ) indicates the distance between two vertices (d) into the probability of an i, j edge. This function is generic and can take on a number of reasonable functional forms, which are discussed in detail in the next section. These models are collectively known as *spatial Bernoulli graphs* (Butts, 2003) and rely on a kernel function ($\mathcal{F}_d(d, \psi)$) referred to as the *spatial interaction function* (SIF), which models the form of the relationship between the spatial distribution of nodes and the network structure (details discussed in the following subsection). These models have been shown to correlate with social processes of interest and to study spatial networks in a variety of settings such as organizational collaboration during disasters (Butts et al., 2012) or interpersonal relations (e.g. Butts, 2002; Butts and Acton, 2011; Hipp and Perrin, 2009). We begin first by defining the canonical parameters $\eta(\psi, d) = \text{logit}\mathcal{F}_d(d, \psi)$, which allow us to write the probability mass function (pmf) for random graph G with support $\mathcal{G}$ (represented via its adjacency matrix $Y \in \mathcal{Y}$) as a curved exponential family (Hunter and Handcock, 2006) of the form

$$\Pr(Y = y \mid D, \psi) = \prod_{\{i,j\}} B\big(y_{ij} \mid \mathcal{F}_d(d_{ij}, \psi)\big) \tag{1}$$

where B is the Bernoulli pmf and D is a distance matrix on the elements of V. Notice that this model can be extended to a fully spatial exponential random graph model with the addition of network sufficient statistics (i.e. statistics on the network such as clustering terms like triangles or cliques) (Robins et al., 2007). Further, these network models have been shown to provide good approximations of the behavior of more complex spatial network models in many empirical and simulated settings (Butts, 2011). Spatial Bernoulli graphs include gravity models as a special case.

The spatial interaction function

The key to being able to either estimate or simulate *spatial Bernoulli graphs* depends on the SIF, $\mathcal{F}_d$, that controls how distance relates to marginal tie probability. In other words, the SIF contains the information necessary to predict the likelihood of sharing a social relation (e.g. a friendship) between two individuals given their distance from each other. For example, if person A and person B live within a quarter mile of each other, then there may be a high probability that they are friends, but if person C lives 5000 miles away, then they are unlikely to be friends with either person A or B.

This work exploits the observation that marginal tie probability, e.g. ties of groups such as the homeless, in social networks is often structurally related to the distance between two individuals. Often as distance increases the likelihood of two individuals having a social relation decreases, i.e. there is an inverse relationship with distance. Many empirical research articles have shown the likelihood of a given social relation to decay with distance (e.g. Bossard, 1932; Festinger et al., 1950; Freeman et al., 1988; Hägerstrand, 1966; Latané et al., 1994; McPherson et al., 2001). The model in equation (1) treats edges as conditionally independent given distance, which is a simplification that approximates many aspects of complex human behavior such as communication or friendship (Butts, 2011). The choice of a given parametric form for the SIF can have large ramifications and is preferably found by measuring the network of interest. If this is not possible, selecting an SIF that has the right characteristics should be done with care. Butts et al. (2012) – implemented in software (Butts and Almquist, 2015) – have proposed six different forms (shown in Table 1). The choice of a given parametric form can have a very distinct effect on the resulting social network. To illustrate the importance of the underlying SIF, we plot three different social

Table 1. Six parametric models for the SIF kernel function from Butts (2002), where PL is power law, APL is attenuated power law, AT is arctangent law, ED is exponential decay, LP is logistic probability and CP is constant probability law.

Model	$\mathcal{F}_d(d)$
PL	$\frac{p_b}{(1+\alpha d)^{\gamma}}$
AP	$\frac{p_b}{1+(\alpha d)^{\gamma}}$
AT	$p_b\left(1-\frac{2}{\pi}\tan^{-1}(\alpha d)\right)$
ED	$\frac{p_b}{e^{\alpha d}}$
LP	$\frac{2p_b}{1+e^{\alpha d}}$
CP	p_b

SIF: spatial interaction function.

networks – see Figure 1 – resulting from a power law (PL), an attenuated power law (APL), and a uniform case – where the network is constructed to yield expected density equal to the mean of the other two network distributions. The center and right columns depict visualizations of the same simulated network, respectively, in non-spatial (Fruchterman–Reingold) and spatial form so as to allow the reader to see both the effect of the SIF (e.g. PL and APL versus the Uniform case with no spatial component). The network clustering (groups of highly connected nodes) can be seen to vary greatly depending on the choice of the functional form of the SIF. To perform this simulation, we use tools in the `networkSpatial` package (Butts and Almquist, 2015), which allows us to simulate a two-dimensional Halton sequence (an algorithm often used to generate points in space for numerical methods, which is deterministic, but appears random; see Braaten and Weller, 1979, for more details) for the vertex set. In the next subsection, we will discuss statistical estimation of an SIF using data from San Francisco.

Case study: Modeling the social support networks of the homeless in San Francisco

The simulation of spatial networks can be accomplished by either first attempting to estimate the SIF of the population of interest (e.g. Spiro et al., 2016) or to employ an SIF already computed from empirical data (see for example, Butts, 2003, 2002). One immediate issue is that these models require that all vertices have a spatial location. If this is known, we can use it directly. For example, we could use the address database maintained for tax records in a given city or county in the US for estimating or simulating a spatial network of households within a given city. If the specific location is not known, we can use a Poisson point process model as demonstrated by Almquist and Butts (2012). Here, we will simulate the location of homeless individuals and homeless youth population based on the Continuum of Care (CoC)[3] count for San Francisco. Using these two resultant data sets, we will simulate a social network for homeless individuals and a Facebook friendship network for homeless young adults. To fully appreciate the scale of these simulation models, two different simulations are performed: (1) a homeless-to-homeless social network, which we validate with the characteristics of a Respondent Driven Sample (RDS) of homeless in Atlanta; and (2) a youth network of a homeless population on Facebook based on Guadagno et al.'s

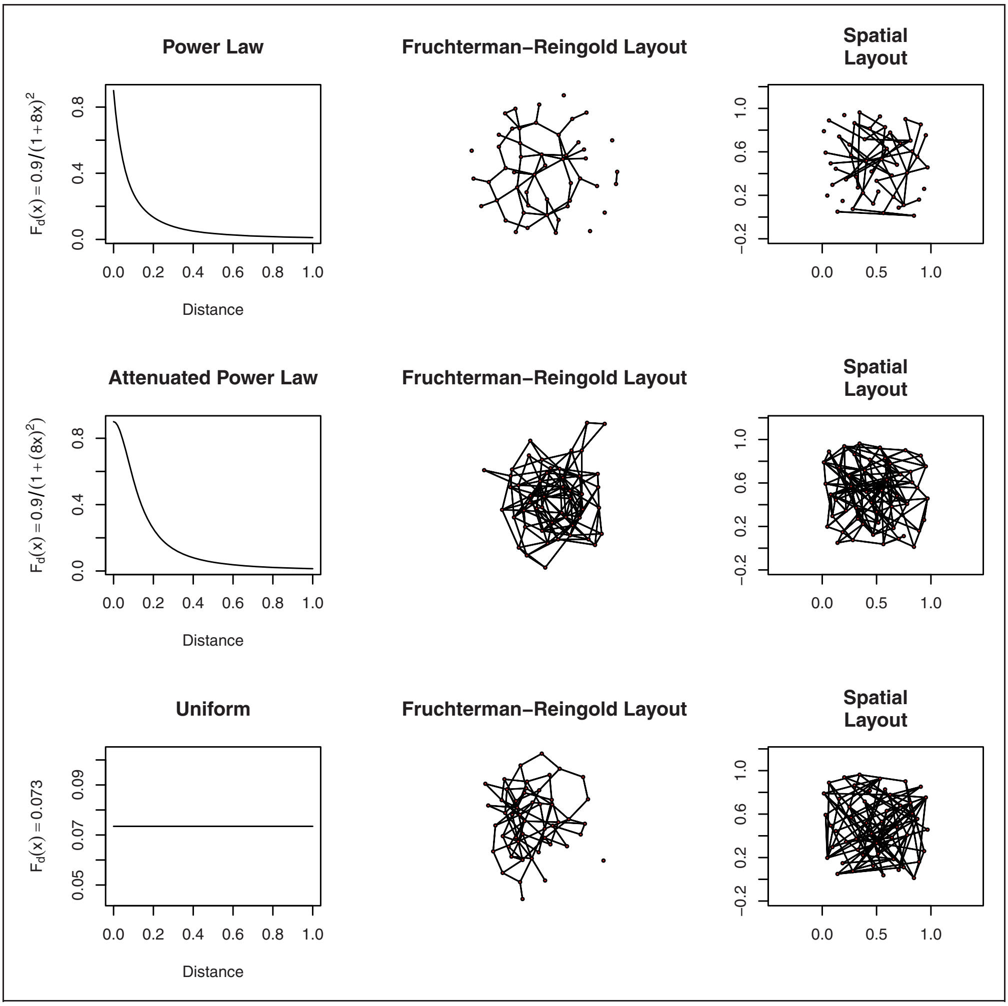

Figure 1. Three by three plot of SIF, standard network plot (e.g. Fruchterman–Reingold) and spatial plot (points placed on a plain via a Halton sequence). Plots use the `networkSpatial` package in R.
SIF: spatial interaction function.

(2013) finding that homeless youth's Facebook usage was similar to college students. Implications are discussed in the final section of the paper.

Simulating homeless locations

To simulate the location of homeless individuals for the year 2017, we use the PiT count data made available at the SF supervisor district level as reported in the *San Francisco Homeless Count & Survey Comprehensive report* (see Table 2 for details). While the ideal case would be neighborhood or US Census block aggregates, larger units can still provide reasonable boundary data for spatial point simulation for spatial Bernoulli networks as described in Almquist and Butts (2012). Further, the supervisor district level is the best approximation to neighborhood level data available and appears to capture the key clustering of homeless individuals in San Francisco. Here, we simulate the homeless individual locations with a

Table 2. PiT homeless counts by SF supervisor district in 2017 for sheltered, unsheltered and youth populations from the *San Francisco Homeless Count & Survey Comprehensive report*.

District	Sheltered	Unsheltered	Youth
1	79	57	37
2	7	53	5
3	96	293	30
4	0	31	7
5	316	143	85
6	1957	1723	477
7	17	74	7
8	65	236	116
9	271	281	96
10	174	1101	359
11	0	48	6

PiT: point in time.

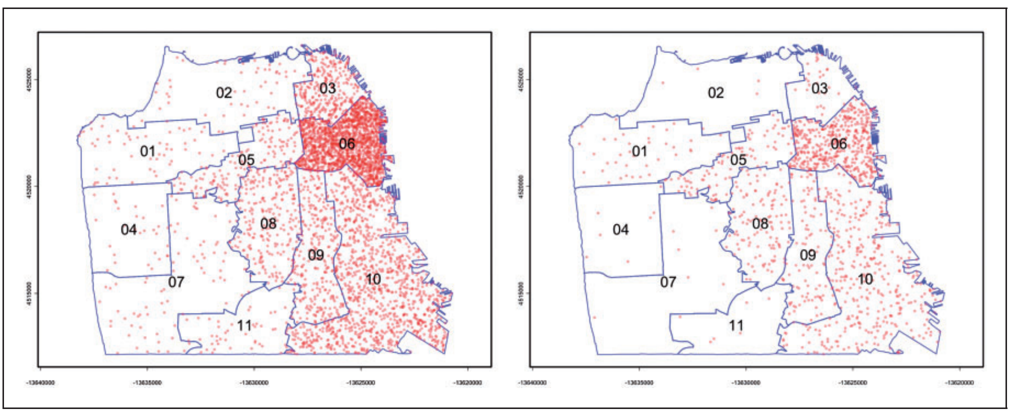

Figure 2. Poisson's simulated homeless locations based on PiT homeless counts by SF supervisor district in 2017 for unsheltered and youth populations for 2017. Supervisor districts are id'd 1 to 11 by the city.
PiT: point in time.

Poisson point process model, which is a maximum entropy solution in which the homeless are placed uniformly at random, subject to geographic constraints. This is a good solution when there is not much information available as in this case. We can visualize these distributions in Figure 2.

Simulation of a social network of homeless in San Francisco

Given the simulated homeless locations, we can now simulate a spatial network model given a plausible SIF. Here, we propose to use a face-to-face SIF estimated from a month of observations of individuals congregating in a large open area to best approximate the social space of the homeless network (PL $\psi = (0.859, 0.035, 6.437)$; Butts, 2002;

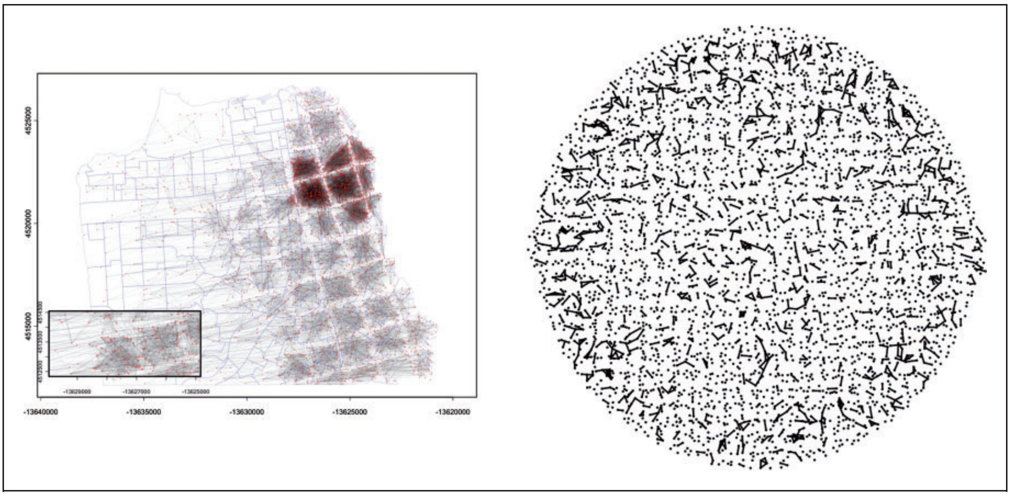

Figure 3. Large-scale network plot of face-to-face SIF with individual place location based on spatial point process model for the SF city supervisor districts and connected PiT Count data of the unsheltered homeless. Neighborhood shape-files provided by Open SF project. Network plot of the face-to-face SIF using standard Fruchterman and Reingold's force-directed placement algorithm.

SIF: spatial interaction function; PiT: point in time.

Freeman et al., 1989). We can visualize this network in Figure 3, where we can observe the spatial clustering around the center of the city and greater disbursement in outer neighborhoods. Figure 3 also provides a visualization of the standard network plot based on Fruchterman and Reingold's force-directed placement algorithm, showing a large variety of differing social clusters. This visualization provides a good sense of how we might expect connections between homeless individuals to be dispersed. Further, we will be able to use this simulation to make some predictions, and to develop hypotheses for future research and public health applications.

Comparison to a known homeless social network. Unfortunately for validation purposes, a complete network of homeless-to-homeless social interactions does not exist. However, a collection of historical RDS data sets is available, focused on HIV and drug use (Morris and Rothenberg, 2011). One of these data sets was from an urban environment, *the Atlanta Antiretroviral Adherence Study* (AAS). Included in this data set is a coding for a social relationship, and if the respondent (in network language, *ego*) was living on the ''streets'' and if the respondent's alter (social friend) was living on the ''streets.'' From this data set, we can build an egocentric sample of homeless-to-homeless social relations in an urban setting (visualized in Figure 4). This resulting egocentric network will be a degree-biased sample (Gile and Handcock, 2010b) of a true homeless-to-homeless social network. Because of the nature of this sampling scheme, there is somewhat limited network information provided; however, it does potentially provide a reasonable estimate of the observed degree distribution of a homeless-to-homeless network, albeit weighted to the higher degree cases (i.e. individuals with higher numbers of social ties). The degree distribution is an important set of statistics on a network (Wasserman and Faust, 1994), and is the table of counts of individuals (egos) with a particular degree score (see Figure 5 for an example of a boxplot of the degree distribution of a network).

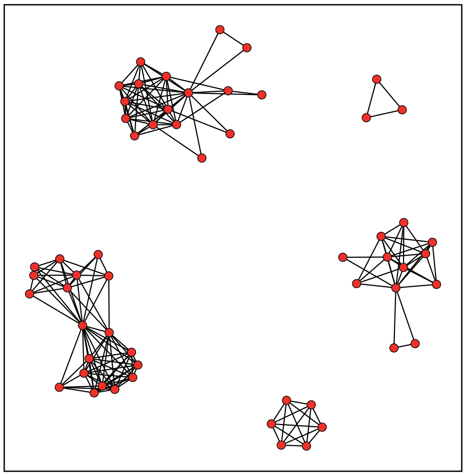

Figure 4. Egocentric representation of homeless-to-homeless social interaction derived from *the Atlanta Antiretroviral Adherence Study* and archival data, Morris and Rothenberg (2011). Resulting network is plotted in standard Fruchterman and Reingold's force-directed placement algorithm.

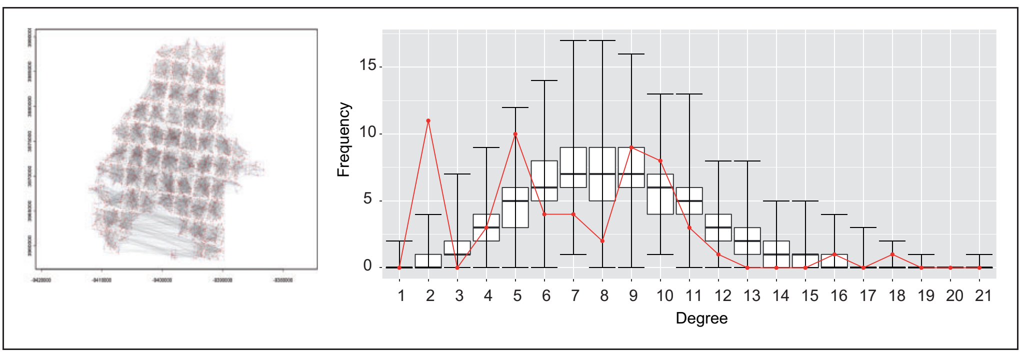

Figure 5. (On the left) Atlanta CoC simulated homeless-to-homeless network of 4063 homeless individuals. (On the right) Boxplots represent the predicted degree distribution estimates from a weighted sample of 1000 iterations of size 57 from the simulated face-to-face social network. Observed degree distribution – from the egocentric representation of homeless-to-homeless social interaction derived from *the Atlanta Antiretroviral Adherence Study* and archival data (Morris and Rothenberg, 2011) – is plotted in red. CoC: Continuum of Care.

We will do the validation test on the city of Atlanta (Figure 5). The Atlanta CoC in 2016 had 4063 homeless individuals. We simulate the homeless-to-homeless network in the same way as in the SF case using the same SIF. First, we compare the estimated population level mean degree of the AAS data ($\bar{d}_{AAS} = 6.9$) to the mean degree from our simulated network ($\bar{d}_{sim} = 7.1$), which has a difference p-value of approximately 0.99. Hence we do very well on an important aggregate measure. The degree distribution is particularly important for understanding network sampling schemes for reaching and measuring hard-to-reach populations, and for acquiring good diffusion simulations (e.g. to understand information flow over homeless-to-homeless populations; Watts, 2002), and thus the choice of this

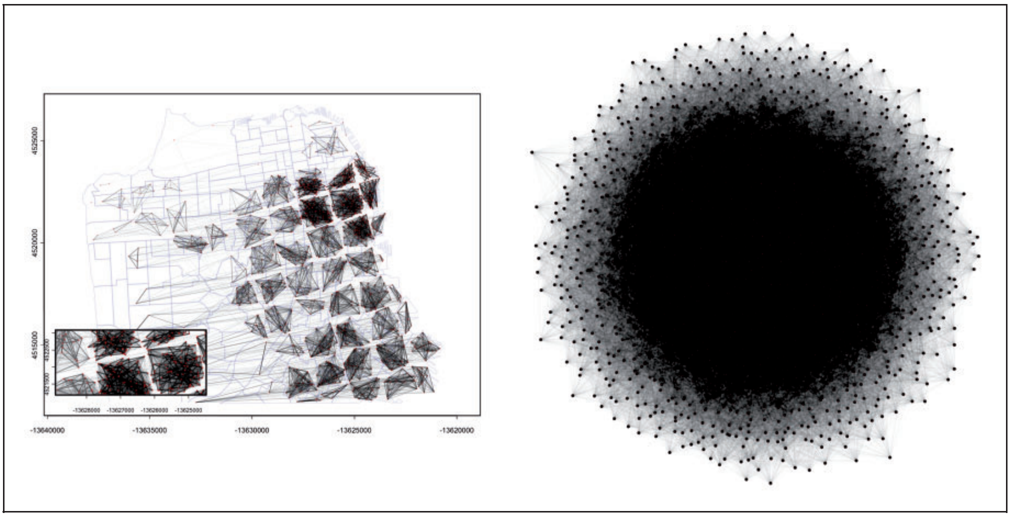

Figure 6. Large-scale network plot of Facebook friendships SIF with individual place location based on spatial point process model for the SF city supervisor districts and connected PiT Count data of the homeless youth population. Neighborhood shape-files provided by Open SF project. Network plot of the face-to-face SIF using standard Fruchterman and Reingold's force-directed placement algorithm.
SIF: spatial interaction function; PiT: point in time.

metric is defensible. To mimic this degree bias and limit the effects of having a sample and not a census of the network, we simulated 1000 draws of a degree-biased egocentric sample of size 57 (the size of the resulting homeless-to-homeless social network sample). We then computed the prediction interval of the degree distribution of this resulting process and compared it against the Atlanta data. This can be visualized in Figure 5. This procedure follows the same logic as Hunter et al. (2008) who introduced methods evaluating network models against observed data. In this case, we see that our degree-biased samples capture the observed data for all but degree two cases, but also include occasional higher degree cases (19, 20 and 21). Note that this is a very hard test as we are attempting to capture 21 different statistics with a parameter free model (i.e. a model not fit to the data). Overall the trend is generally captured and makes a reasonable attempt at capturing the weighted-degree distribution of the observed homeless-to-homeless social network.

Simulation of a Facebook network of homeless youth

In this example, we simulate the spatial network of a homeless Facebook network using an SIF estimated from college students in 2009 (APL with $\psi = (0.6266, 0.0489, 6.527)$; Spiro et al., 2016) and simulate the spatial distribution of homeless in San Francisco in 2017 using PiT Count data of the homeless youth population. This simulation is based on the work of Guadagno et al. (2013), who found that homeless young adults have similar Facebook networks and usage patterns as college students. This network can be visualized in Figure 6.

Analysis of information-diffusion over the simulated networks

To further investigate the properties of these two networks, we simulate a process of diffusion of information (e.g. location of a new shelter) or goods (e.g. condoms or needle exchange) throughout the two simulated networks of relations among the homeless and homeless youth of San Francisco.

We use the estimated interaction rates from the face-to-face SIF and Facebook SIF to inform our diffusion process models.[4] The diffusion process we consider follows a standard network model (Frambach, 1993), where an initial signal (or seed) will be provided to either the highest degree node(s) or randomly to several nodes. This type of diffusion is also known as "Poisson diffusion" and operates in the following manner. At a given time t, every individual is either "infected" or "uninfected". An individual (v), once infected, initiates an infection event for each of their out-neighbors (v_n), which occurs at time $t+X$ (where $X \sim e^{\lambda_{vv_n}}$ a Poisson distribution). An uninfected individual becomes infected at the first interaction event; subsequent infection events have no effect. The simulation terminates when all reachable neighbors have been infected. As suggested above, the speed of the diffusion process is governed by the edge-specific rates, λ, in this case assigned by the probability of interaction estimated from the spatial Bernoulli model based on the face-to-face SIF (on the unsheltered population) and Facebook SIF (on the homeless youth population).

Social network. We can visualize this process in Figure 7 over the simulated social network of homeless-to-homeless relations. Here, we observe a strong spatial component and an important social component as well, ranging down the center of the city with specialized hot spots, where the social interaction allows goods or information to skip around geographically. Starting from a single point (in this case, the highest degree individual as is often practiced in health RDS designs), we see that the network is almost completely diffused by 25 iterations, but takes about 30 for the network to be completely saturated (Figure 7). We also note that a random seed does not do much worse than picking the highest degree individual until about five iterations out, suggesting that low iteration RDS might be quite successful in practice. We can further analyze the relationship between this

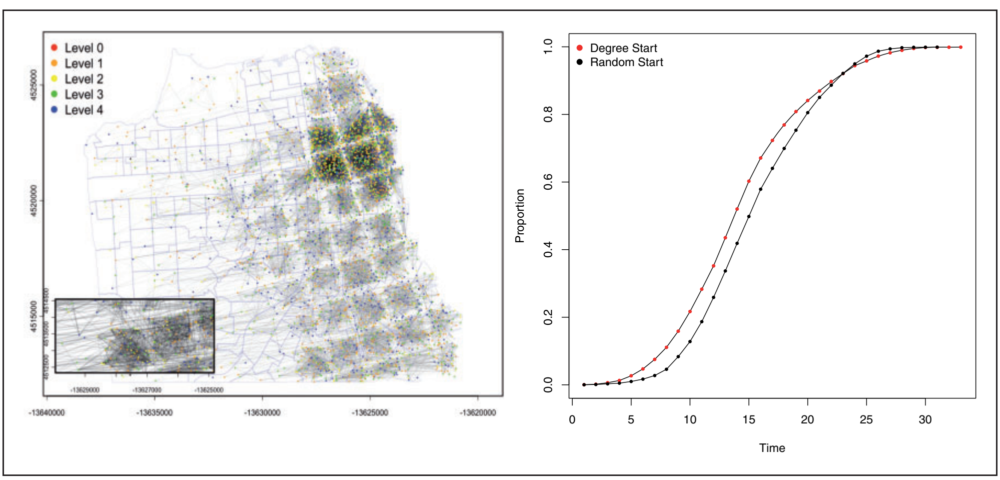

Figure 7. Large-scale network plot of the face-to-face SIF with individual place location based on their simulated location along with neighborhood polygons from the SF Open Data project. Quartile diffusion pattern plotted in five colors from red to blue.

SIF: spatial interaction function.

diffusion pattern and key landmarks of interest such as neighborhoods within the City of San Francisco. Table 3 in the online supplementary material shows the counts of homeless cross-tabulated with the quartile in which they received information or goods and the 41 SF neighborhoods reported by the city. A few key areas emerge, such as the critical neighborhoods of the Financial District, Bayview, Mission, Mission Bay, Potrero Hill, South of Market and the Tenderloin, which see strong spatial and social diffusion patterns (i.e. each continue to see large numbers of new diffusions in each quarter of the simulation procedure). We see that some neighborhoods (i.e. the Financial District) see high spatial cohesion – most of the individuals who receive goods or information do so within a short time interval. This further suggests that there are core neighborhoods in which one might pilot a public health survey as well as neighborhoods that may require a considerable amount of work to make sure information or goods are provided to the intended recipients (e.g. Outer Richmond).

Facebook network. We can visualize this process in Figure 8 over the simulated Facebook network of homeless youth. Here, we observe that the degree start is not nearly as effective. This is likely because of the very large average degree of a Facebook network. In addition, we see much stronger spatial hopping and a quicker diffusion process (i.e. we expect information to pass among the youth network much faster than the homeless-to-homeless network). This finding implies that non-government organizations (NGOs) and public health efforts that seek to provide information, goods or services to homeless should start with the homeless youth, given the potential for these information/goods to diffuse more quickly when homeless youth are the initial transmitters. Again, we can look at the diffusion pattern over the neighborhoods of San Francisco (Table 4 in the online supplementary material), which largely mimics the same pattern as the homeless-to-homeless network, but with a few notable areas such as Visitacion Valley and Golden Gate park standing out.

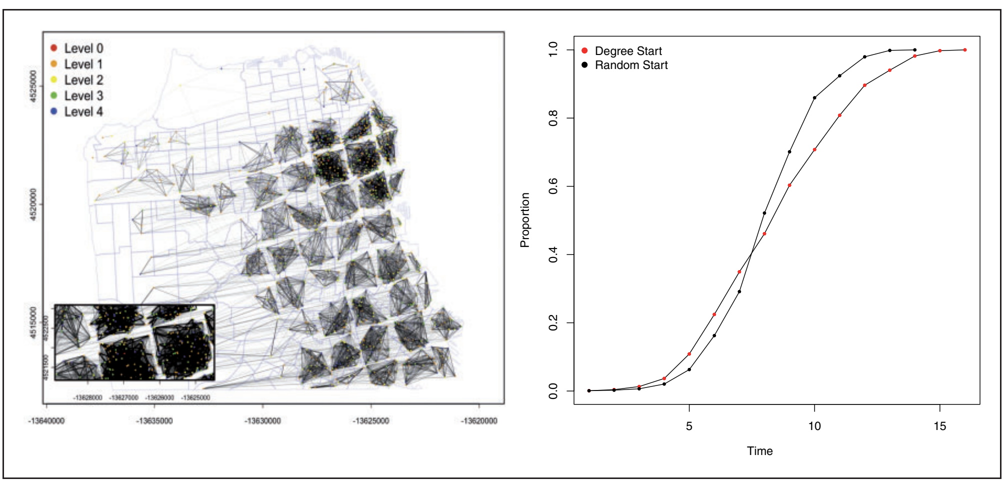

Figure 8. Large-scale network plot of the youth Facebook SIF with individual place location based on their simulated location along with neighborhood polygons from the SF Open Data project. Quartile diffusion pattern plotted in five colors from red to blue.

SIF: spatial interaction function.

Discussion

The use of large-scale spatial networks for urban planning and understanding the social sphere of entire cities is potentially quite extensive. However, one of the key areas where the inference and simulation of such networks may contribute is prior to analysis in hypotheses-testing of network-based sampling strategies. For example, RDS methods (e.g. Gile and Handcock, 2010a; Heckathorn, 1997; Salganik and Heckathorn, 2004), and more general random-walk methods (e.g. Gjoka et al., 2010), which are often used to measure hard-to-reach populations, are time-intensive and very expensive to run. Simulation strategies such as those presented here could be used by either government officials, NGOs or academic researchers as a method for developing key hypotheses before engaging in large (or small) RDS procedures. Similar methods have been used extensively for public health issues such as enumerating the population of individuals with HIV/AIDS (e.g. Magnani et al., 2005) or drug users (e.g. Hathaway et al., 2010), and for measuring properties of homeless populations (e.g. Salazar et al., 2007; Williams, 2010; Wincup et al., 2003). RDS and its comparable methods (e.g. capture-recapture) would certainly be the core method used to estimate the social relations of the homeless, as well as the social support of the homeless broadly. Employing these simulation studies and methods may allow researchers to acquire a sense of the number of individuals needed in order to both reach optimal penetration (e.g. to allow for large coverage of the social network) and for unbiased estimation of the parameters of interest (e.g. the average number of family support ties or friendship ties). The work presented here offers recommendations for such an analysis. In addition, this type of information is needed by large federal organizations (e.g. the National Institute of Health) or private NGOs to conduct a wider, more detailed examination of the homeless support system in the US.

A key finding includes the spread of information over the homeless networks within San Francisco. This analysis strongly suggests that one might expect information to flow faster over youth networks than with older residents. The interaction of these two networks could be engaged for both measurement of the population (virtual RDS) or to pass information using Facebook advertising targeted at homeless youth. One potential policy implication of this finding is that NGOs/government/health providers could most effectively target homeless-directed interventions and information passing to homeless youths, as this information will spread more quickly via such groups. However, this relies on the assumption that there will be (guaranteed) spillover between homeless youths and homeless communities more generally.

There are also limitations to this analysis. All statistical and simulation models must work with the data available and the mathematical assumptions made – this framework is no different. The key limitations in this framework are that the model requires location and network information for both inference and simulation. To this end, we must simulate both the homeless locations and the social networks although we have used reasonable empirically driven SIFs and have attempted to validate them with publicly available data and research.

Overall, these methods provide a unique opportunity for visualizing, characterizing and inferring information from large-scale social networks. This allows for a deeper understanding of the social fabric of different groups within urban centers, extrapolative simulation for use in hypothetical analysis, and preliminary work prior to implementing large expensive studies of hard-to-reach populations, or other traditionally marginalized and/or underrepresented groups.

Acknowledgements

The author would like to thank the editor and reviewers for their helpful reviews and critiques. Steven Manson for reading and commenting on the article; James Holland Jones and his lab for listening to my presentation of this work and helping with the public health angle; and lastly Carter Butts and his lab for letting me present this material to them and their resulting suggestions and critiques.

Declaration of conflicting interests

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

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Notes

1. www.sandiegocounty.gov/content/sdc/hhsa/programs/phs/community_epidemiology/dc/Hepatitis_A/outbreak.html
2. Note that there are more complex socio-demographic or ''latent'' embeddings that can also be of interest, such as the latent spaces in the statistical literature (e.g. Krivitsky and Handcock, 2008), anthro-spaces (Andris, 2016) or the ''Blau'' spaces (Blau, 1977; McPherson and Ranger-Moore, 1991) of the sociology literature.
3. CoCs are administrative units for homeless care in the US, more information can be found at US Department of Housing and Urband Development Office of Community Planning and Development (2009).
4. The simulation software used is from the `diffusion` package in `R` (Butts, 2008).

References

Acedo A, Mendoza G, Painho M and Casteleyn S (2017a) One tool to spatialize all: sense of place, social capital and civic engagement. In: Bregt A, Sarjakoski T, Lammeren R. V, & Rip F. (Eds.), *Societal Geo-innovation: short papers, posters and poster abstracts of the 20th AGILE Conference on Geographic Information Science* (pp. 1–5). (AGILE: The Annual International Conference on Geographic Information Science; No. 2017). Wageningen, the Netherlands: Wageningen University & Research.

Acedo A, Painho M and Casteleyn S (2017b) Place and city: Operationalizing sense of place and social capital in the urban context. *Transactions in GIS* 21(3): 503–520.

Almquist ZW and Bagozzi BE (2016) The spatial properties of radical environmental organizations in the UK: Do or die! *PLoS One* 11(11): e0166609.

Almquist ZW and Butts CT (2012) Point process models for household distributions within small areal units. *Demographic Research* 26(12): 593–632.

Almquist ZW and Butts CT (2015) Predicting regional self-identification from spatial network models. *Geographical Analysis* 47(1): 50–72.

Andris C (2016) Integrating social network data into GISystems. *International Journal of Geographical Information Science* 30(10): 2009–2031.

Barthélemy M (2011) Spatial networks. *Physics Reports* 499(1): 1–101.

Blau PM (1977) A macrosociological theory of social structure. *American Journal of Sociology* 83(1): 26–54.

Bossard JHS (1932) Residential propinquity as a factor in marriage selection. *American Journal of Sociology* 38(2): 219–224.

Braaten E and Weller G (1979) An improved low-discrepancy sequence for multidimensional quasi-Monte Carlo integration. *Journal of Computational Physics* 33(2): 249–258.

Burt MR, Pearson C and Montgomery AE (2007) Community-wide strategies for preventing homelessness: Recent evidence. *The Journal of Primary Prevention* 28(3–4): 213–228.

Butts CT (2002) *Spatial models of large-scale interpersonal networks*. Doctoral Dissertation, Carnegie Mellon University.
Butts CT (2003) Predictability of large-scale spatially embedded networks. In: Breiger RL, Carley KM and Pattison P (eds) *Dynamic Social Network Modeling and Analysis: Workshop Summary and Papers*. Washington, DC: National Academies Press, pp. 313–323.
Butts CT (2008) *Diffusion: Tools for Simulating Network Diffusion Processes. R Package Version 0.5*. Available at: http://lakshmi.calit2.uci.edu/ncasd/?page_id=19 (accessed 10 January 2017).
Butts CT (2011) Bernoulli graph bounds for general random graphs. *Sociological Methodology* 41: 299–345.
Butts CT and Acton RM (2011) Spatial modeling of social networks. In: Nyerges T, Couclelis H and McMaster R (eds) *The Sage Handbook of GIS and Society Research*. Thousand Oaks, CA: SAGE Publications, pp. 222–250.
Butts CT, Acton RM, Hipp JR, et al. (2012) Geographical variability and network structure. *Social Networks* 34: 82–100.
Butts CT and Almquist ZW (2015) *networkSpatial: Tools for the Generation and Analysis of Spatially-embedded Networks. R Package Version 1*. Available at: http://lakshmi.calit2.uci.edu/census2000/R/src/contrib/networkSpatial_1.00.tar.gz (accessed 10 January 2017).
Câmara G, Souza RCM, Freitas UM, et al. (1996) Spring: Integrating remote sensing and GIS by object-oriented data modelling. *Computers & Graphics* 20(3): 395–403.
Carrasco JA, Hogan B, Wellman B, et al. (2008) Collecting social network data to study social activity-travel behavior: An egocentric approach. *Environment and Planning B: Planning and Design* 35(6): 961–980.
Cho E, Myers SA and Leskovec J (2011, August) Friendship and mobility: user movement in location-based social networks. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 1082–1090). ACM. San Diego, California, USA. August 21 – 24.
Christakis NA and Fowler JH (2013) Social contagion theory: Examining dynamic social networks and human behavior. *Statistics in Medicine* 32(4): 556–577.
Coleman JS (1988) Social capital in the creation of human capital. *American Journal of Sociology* 94: S95–S120.
Comber A, Batty M, Brunsdon C, Hudson-Smith A, Neuhaus F and Gray S (2012) Exploring the geography of communities in social networks. In *Proceedings of the GIS Research UK 20th Annual Conference* (pp. 33–37). Lancaster, UK. April 11–13, 2012.
Conover MD, Davis C, Ferrara E, et al. (2013) The geospatial characteristics of a social movement communication network. *PLoS One* 8(3): e55957.
Crucitti P, Latora V and Porta S (2006) Centrality measures in spatial networks of urban streets. *Physical Review E* 73(3): 036125.
Csáji BC, Browet A, Traag VA, et al. (2013) Exploring the mobility of mobile phone users. *Physica A: Statistical Mechanics and its Applications* 392(6): 1459–1473.
Elvy M, MCAdam J, Scharer LK, et al. (1986) Homeless persons and health care. *Annals of Internal Medicine* 104: 405–409.
Festinger L, Schachter S and Back K (1950) *Social Pressures in Informal Groups: A Study of Human Factors in Housing*. Palo Alto, CA: Stanford University Press.
Frambach RT (1993) An integrated model of organizational adoption and diffusion of innovations. *European Journal of Marketing* 27(5): 22–41.
Freeman LC, Freeman SC and Michaelson AG (1988) On human social intelligence. *Journal of Social Biological Structure* 11: 415–425.
Freeman LC, Freeman SC and Michaelson AG (1989) How humans see social groups: A test of the Sailer-Gaulin models. *Journal of Quantitative Anthropology* 1: 229–238.
Gil J (2017) Street network analysis "edge effects": Examining the sensitivity of centrality measures to boundary conditions. *Environment and Planning B: Urban Analytics and City Science* 44(5): 819–836.
Gile KJ and Handcock MS (2010a) Modeling networks from sampled data. *Annals of Applied Statistics* 4(1): 5–25.

Gile KJ and Handcock MS (2010b) Respondent-driven sampling: An assessment of current methodology. *Sociological Methodology* 40(1): 285–327.

Gjoka M, Kurant M, Butts CT and Markopoulou A (2010, March) Walking in facebook: A case study of unbiased sampling of osns. In *Infocom, 2010 Proceedings IEEE* (pp. 1–9). Ieee. 14-19 March 2010. Location: San Diego, CA.

Granovetter MS (1974) *Getting a Job: A Study of Contacts and Careers*. Cambridge, MA: Harvard University Press.

Guadagno RE, Muscanell NL and Pollio DE (2013) The homeless use facebook?! Similarities of social network use between college students and homeless young adults. *Computers in Human Behavior* 29(1): 86–89.

Hägerstrand T (1966) Aspects of the spatial structure of social communication and the diffusion of information. *Papers in Regional Science* 16(1): 27–42.

Hathaway AD, Hyshka E, Erickson PG, et al. (2010) Whither RDS? An investigation of respondent driven sampling as a method of recruiting mainstream marijuana users. *Harm Reduction Journal* 7(1): 15.

Haynes KE and Fortheringham AS (1984) *Gravity and Spatial Interaction Models, Scientific Geography*, vol. 2. Beverly Hills: Sage Publications.

Heckathorn DD (1997) Respondent-driven sampling: A new approach to the study of hidden populations. *Social Problems* 44(2): 174–199.

Heer J and Perer A (2014) Orion: A system for modeling, transformation and visualization of multidimensional heterogeneous networks. *Information Visualization* 13(2): 111–133.

Hipp JR, Butts CT, Acton R, et al. (2013) Extrapolative simulation of neighborhood networks based on population spatial distribution: Do they predict crime? *Social Networks* 35(4): 614–625.

Hipp JR and Perrin AJ (2009) The simultaneous effect of social distance and physical distance on the formation of neighborhood ties. *City & Community* 8(1): 5–25.

Hunter DR, Goodreau SM and Handcock MS (2008) Goodness of fit of social network models. *Journal of the American Statistical Association* 103(481): 248–258.

Hunter DR and Handcock MS (2006) Inference in curved exponential family models for networks. *Journal of Computational and Graphical Statistics* 15: 565–583.

Hwang SW, Kirst MJ, Chiu S, et al. (2009) Multidimensional social support and the health of homeless individuals. *Journal of Urban Health* 86(5): 791–803.

Irwin J, LaGory M, Ritchey F, et al. (2008) Social assets and mental distress among the homeless: Exploring the roles of social support and other forms of social capital on depression. *Social Science & Medicine* 67(12): 1935–1943.

Koylu C, Guo D, Kasakoff A, et al. (2014) Mapping family connectedness across space and time. *Cartography and Geographic Information Science* 41(1): 14–26.

Krivitsky PN and Handcock MS (2008) Fitting position latent cluster models for social networks with latentnet. *Journal of Statistical Software* 24(2): 1–23.

Latané B, Nowak A and Liu JH (1994) Measuring emergent social phenomena: Dynamism, polarization, and clustering as order parameters of social systems. *Behavioral Science* 39: 1–24.

Li M, Ye X, Zhang S, Tang X and Shen Z (2018) A framework of comparative urban trajectory analysis. *Environment and Planning B: Urban Analytics and City Science* 45(3): 489–507.

Magnani R, Sabin K, Saidel T, et al. (2005) Review of sampling hard-to-reach and hidden populations for HIV surveillance. *AIDS* 19: S67–S72.

McPherson M, Smith-Lovin L and Cook JM (2001) Birds of a feather: Homophily in social networks. *Annual Review Sociology* 27: 415–444.

McPherson MJ and Ranger-Moore JR (1991) Evolution on a dancing landscape: Organizations and networks in dynamic Blau space. *Social Forces* 70(1): 19–42.

Mitchell JC (1987) The components of strong ties among homeless women. *Social Networks* 9(1): 37–47.

Morris M and Rothenberg R (2011) *HIV Transmission Network Metastudy Project: An Archive of Data from Eight Network Studies, 1988–2001*. Inter-university Consortium for Political and Social Research (ICPSR). Online. University of Michigan, Ann Arbor. Access date: 10/01/2017.

Okabe A (1992) *Spatial Tessellations*. [redacted]: Wiley Online Library.
Onnela JP, Arbesman S, González MC, et al. (2011) Geographic constraints on social network groups. *PLoS One* 6(4): e16939.
Pelcer-Vujačić O and Kovačević S (2016) A GIS Database of Montenegrin Katuns (Kuči Mountain and Durmitor). In: Ioannides M. et al. (eds) *Digital Heritage. Progress in Cultural Heritage: Documentation, Preservation, and Protection*. EuroMed 2016. Lecture Notes in Computer Science, vol 10059. Springer, Cham.
Pentland A (2015) *Social Physics: How Social Networks can Make us Smarter*. New York, New York: Penguin Publishing.
Psyllidis A (2016) Revisiting urban dynamics through social urban data. *Architecture and the Built Environment* (18): 1–334.
Putnam RD (1995) Bowling alone: America's declining social capital. *Journal of Democracy* 6(1): 65–78.
Rice E, Kurzban S and Ray D (2012) Homeless but connected: The role of heterogeneous social network ties and social networking technology in the mental health outcomes of street-living adolescents. *Community Mental Health Journal* 48(6): 692–698.
Rice E, Milburn NG and Monro W (2011) Social networking technology, social network composition, and reductions in substance use among homeless adolescents. *Prevention Science* 12(1): 80–88.
Robins G, Pattison P, Kalish Y, et al. (2007) An introduction to exponential random graph (p*) models for social networks. *Social Networks* 29(2): 173–191.
Rowe S and Wolch J (1990) Social networks in time and space: Homeless women in Skid Row, Los Angeles. *Annals of the Association of American Geographers* 80(2): 184–204.
Salazar LF, Crosby RA, Holtgrave DR, et al. (2007) Homelessness and HIV-associated risk behavior among African American men who inject drugs and reside in the urban south of the United States. *AIDS and Behavior* 11(2): 70–77.
Salganik MJ and Heckathorn DD (2004) Sampling and estimation in hidden populations using respondent-driven sampling. *Sociological Methodology* 34: 193–239.
Shinn M, Knickman JR and Weitzman BC (1991) Social relationships and vulnerability to becoming homeless among poor families. *American Psychologist* 46(11): 1180.
Spiro ES, Almquist ZW and Butts CT (2016) The persistence of division geography, institutions, and online friendship ties. *Socius: Sociological Research for a Dynamic World* 2: 1–15.
Tobler WR (1970) A computer movie simulating urban growth in the Detroit region. *Economic Geography* 46(Supp. 1): 234–240.
US Department of Housing and Urband Development Office of Community Planning and Development (2009) Hud's homeless assistance programs: Continuum of care 101. Technical report, US Department of Housing and Urban Development, DC.
Wang J, Mo H, Wang F, et al. (2011) Exploring the network structure and nodal centrality of china's air transport network: A complex network approach. *Journal of Transport Geography* 19(4): 712–721.
Wasserman S and Faust K (1994) *Social Network Analysis: Methods and Applications*. New York, NY: Cambridge University Press.
Watts DJ (2002) A simple model of global cascades on random networks. *Proceedings of the National Academy of Sciences* 99(9): 5766–5771.
Williams M (2010) Can we measure homelessness? A critical evaluation of 'capture-recapture'. *Methodological Innovations Online* 5(2): 49–59.
Wilson AG (2011) *Entropy in Urban and Regional Modelling*, vol.1. London, UK: Routledge.
Wincup E, Buckland G and Bayliss R (2003) *Youth homelessness and substance use: Report to the drugs and alcohol research unit*. London: Home Office London.
Wolch JR and Rowe S (1992) On the streets: Mobility paths of the urban homeless. *City & Society* 6(2): 115–140.
Xu Y, Belyi A, Bojic I, et al. (2017) How friends share urban space: An exploratory spatiotemporal analysis using mobile phone data. *Transactions in GIS* 21(3): 468–487.

Zack W Almquist is an assistant professor of Sociology and Statistics at the University of Minnesota. He holds a PhD in Sociology and a MA in Demography from the University of California, Irvine, a MS in Statistics from Northwestern University, and a BS in Mathematics from the University of Oregon. Almquist currently serves on editorial boards of the journals Social Networks and Sociological Methodology. His research focuses on understanding, modeling, and predicting the effects that space (geography) and time have on human interaction (e.g. sexual contact) and social processes (e.g. information passing or disease transmission).