Causal Inference for Social Network Data

Elizabeth L. Ogburn

Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA

E-mail: eogburn@jhsph.edu

Oleg Sofrygin

Department of Biostatistics, University of California, Berkeley, CA, USA

Iván Díaz

Division of Biostatistics and Epidemiology, Weill Cornell Medicine, New York, NY, USA

Mark J. van der Laan

Department of Biostatistics, University of California, Berkeley, CA, USA

Summary. We describe semiparametric estimation and inference for causal effects using data from a single social network. Our asymptotic results allow for dependence of each observation on a growing number of other units as sample size increases. While previous methods have generally implicitly focused on one of two possible sources of dependence among social network observations, we allow for both dependence due to transmission of information across network ties, and for dependence due to latent similarities among nodes sharing ties. We describe estimation and inference for causal effects that are specifically of interest in social network settings.

Keywords: Statistical dependence, Causal inference, Social networks, Semiparametric inference

1. Introduction

Many aspects of social networks are of interest to researchers, from the clustering of individuals into communities to the probability distributions that describe the generation of new relationships between individuals in the network. There is increasing interest in identifying and estimating causal effects in the contexts of social networks, that is causal effects that one individual's behavior, treatment assignment, beliefs, or health outcome could have on his or her social contacts' behaviors, exposures, beliefs, or health statuses. There have been a number of high profile articles that use standard methods like generalized linear models (GLM) and generalized estimating equations (GEE) to attempt to infer causal peer effects from network data (e.g. Christakis and Fowler, 2007, 2008, 2010), and this work has inspired several research programs that study peer effects using the same statistical methods (Ali and Dwyer, 2010; Cacioppo et al., 2009; Madan et al., 2010; Rosenquist et al., 2010; Wasserman, 2013). However, these methods have come under considerable criticism from the statistical community (Cohen-Cole and Fletcher, 2008; Lyons,

2011; Shalizi and Thomas, 2011). These statistical models are not equipped to deal with dependence across individuals and are rarely appropriate for estimating effects using network data (Ogburn and VanderWeele, 2014). In some settings it may be possible to use them to test for the presence of network dependence, but some properties of such tests are unknown (VanderWeele et al., 2012; Shalizi, 2012; Ogburn and VanderWeele, 2014).

Methodology has not kept apace with interest in causal inference using data from individuals connected in a social network, and researchers have resorted to using standard statistical methods to analyze this new type of data, or to collecting multiple independent groups of individuals and using the groups, rather than the individuals, as the unit of statistical inference. The former strategy is not statistically appropriate and the latter strategy, while statistically valid, can be an inefficient use of data. Furthermore, it limits the settings and effects that one can study, because in some settings it may be too expensive or labor intensive to collect data from many independent groups, or independent groups may not exist (e.g. a global infectious disease epidemic).

Very recently, researchers interested in causal inference for interconnected subjects have begun to develop methods designed specifically for the network setting. Many methods for interference–the effect of one individual's treatment or exposure on others' outcomes–are relevant to the analysis of network data (Aronow and Samii, 2013; Athey et al., 2016; Bowers et al., 2013; Eckles et al., 2014; Forastiere et al., 2016; Graham et al., 2010; Halloran and Struchiner, 1995; Halloran and Hudgens, 2011; Hong and Raudenbush, 2006, 2008; Hudgens and Halloran, 2008; Jagadeesan et al., 2017; Liu and Hudgens, 2014; Liu et al., 2016; Rosenbaum, 2007; Rubin, 1990; Sobel, 2006; Tchetgen Tchetgen and VanderWeele, 2012; VanderWeele, 2010). However, the inferential methods developed in this context generally require observing multiple independent groups of units, which corresponds to observing multiple independent networks, or else they require treatment to be randomized. Ideally, we would like to be able to perform inference even when all observations are sampled from a single social network and in observational settings in addition to randomized experiments. Tchetgen Tchetgen et al. (2017), which was developed in parallel to this work, is the only other proposed solution to this problem of which we are aware. Their approach is quite different from ours, primarily because it assumes that the outcomes of interest comprise a single realization of a specific type of Markov random field over the network. This corresponds to certain types of equilibrium distributions and is incompatible with (though may in some settings approximate) the traditional causal data-generating mechanisms that we work with in this paper, namely causal structural equation models and directed acyclic graph (DAG) models (for a discussion of these compatibility issues see Lauritzen and Richardson, 2002).

In this paper we extend recent work by van der Laan (2014) on causal inference for causally connected units to more general social network settings. van der Laan (2014) introduced methods for causal inference from a single collection of interconnected units when each unit is known to be independent of all but a small number of other units. Asymptotic results rely on the number of dependent units being fixed as the total number of units goes to infinity. We propose similar causal and statistical inference without requiring the number of dependent units to be fixed as sample size increases. We are not aware of any

previous methods for inference about network members in observational settings that allow the number of ties per node to increase as the network grows. While previous methods have generally implicitly focused on one of two possible sources of dependence among social network observations, we allow for both dependence due to contagion or transmission of information across network ties, and dependence due to latent similarities among nodes sharing ties. We describe estimation and inference for causal effects that are specifically of interest in social network settings. A companion paper (Sofrygin and van der Laan, 2015), written in tandem with this one, provides more details about the implementation and computation of the estimation procedures but does not address the the new types of dependence or the new asymptotic regime considered here. Despite these important steps towards accounting for all of the complexities found in real social network data, our methods still rely on many strong and unrealistic assumptions, which we will point out as they come up. Until methods can be developed to account for all complexities, or data can be collected to match the assumptions of existing methods, we urge caution in drawing causal conclusions from observational network data.

In Section 2 we give some background on causal inference for social network data, discussing briefly the relationship between causal structural equation models and network edges, the types of statistical dependence likely to be found in social network data, and asymptotic growth. In Section 3 we define our target of inference, which is the expectation of the average counterfactual outcome across the individuals in a network, and present the identifying assumptions that we will use in the methods that follow. We give the efficient influence function for our target parameter under the conditional independence assumptions from van der Laan (2014). When these independence assumptions are relaxed, this will still be an influence function for our target parameter but it may not be efficient. Estimation is efficient under the stronger independence assumptions but still consistent and asymptotically normal under the weaker independence assumptions. In Section 3.5 we prove our main result, which is the asymptotic normality of our estimator under an asymptotic regime in which the number of ties per node grows with n. In Section 4 we discuss estimation of causal effects that are specifically of interest in social network settings. Section 5 demonstrates the performance of our methods in simulations.

2. Background and setting

2.1. Networks and structural equation models

A network is a collection of units, or nodes, and information about the presence or absence of pairwise ties between them. The presence of a tie between two units indicates that the units share some kind of a relationship; for example, in a social network we might define a tie to include familial relatedness, friendship, or shared place of work. Some types of relationships are mutual, for example familial relatedness; others, like friendship, can go in only one direction. For simplicity we will assume all networks are undirected in what follows, but our methods are equally applicable to directed networks. In an undirected network, the *degree* of a node is the number of ties it has. The *alters* of node i are the nodes with which i shares ties.

Underlying inquiries into causal effects across network nodes is a representation of the network as a structural equation model. Consider a network of n subjects, indexed by i, with binary undirected ties $A_{ij} \equiv I\{\text{subjects } i \text{ and } j \text{ share a tie}\}$. The matrix $\mathbf{A}$ with entries A_{ij} is the adjacency matrix for the network. Associated with each subject is a vector of random variables, O_i, including an outcome Y_i, covariates C_i, and an exposure or treatment variable X_i, all possibly indexed by time t. In numerous applications across the social, political, and health sciences, researchers are interested in ascertaining the presence of and estimating causal interactions across alter-ego pairs. Is there *interference*, i.e. does the treatment of subject i have a causal effect on the outcome of subject j when i and j share a network tie? Is there *peer influence*, i.e. does the outcome of subject i at time t have a causal effect on a future outcome of subject j when i and j are adjacent in the network? These inquiries can be formalized with the help of a causal structural equation model, informed by the network.

A structural equation model is a system of equations of the form $y_i = f_i[pa_i(Y), \varepsilon_i]$, where $pa_i(Y)$, the set of parents of Y, is a collection of variables that are causes of Y for subject i, and ε_i is an error term that may include omitted causes of Y. In general C_i and X_i will be included in $pa_i(Y)$ (Pearl, 2000). When causal inference is performed on network data, the network ties inform which variables are to be included in $pa_i(Y)$. For example, if interference might be present, then the collection of treatment variables for i's alters, $\{X_j : A_{ij} = 1\}$, must be included in the set $pa_i(Y)$ (Sussman and Airoldi, 2017). If contagion might be present then $\{Y_{j,t-k} : A_{ij} = 1\}$ must be included in the set $pa_i(Y_t)$, where t indexes time and k is an outcome-specific lag time such that no causal effect can be transmitted from one person to another in less than k time steps (**?**).

It is important that the network be completely and accurately specified; missing ties are akin to missing components of a multidimensional treatment vector because they result in important elements of exposure of interest being left out of the SEM. Whenever an inquiry into causal effects is informed by a social network, measurement error in the network is tantamount to measurement error in the exposure of interest, and missing edges or nodes may also result in unmeasured confounding. This is obviously a huge burden on data collection in many settings, but would be straightforward for online social networks. The network for which data is collected must be calibrated to the causal question of interest. If we are interested in peer effects on academic achievement among elementary school children and think that being in the same classroom is the relationship that determines whether or not two children affect one another's outcomes, then being in the same classroom is the relationship that determines whether or not a network tie exists, and a network that captures interaction during playground sports is not informative or useful. In other words, a tie between nodes i and j represents the possibility of a causal effect of an element of O_i on an element of O_j at a later time, and vice versa. These issues have not been made explicit in much of the existing literature on causal inference for network data; equating a network with the underlying SEM can help to make them precise.

2.2. Networks and dependence

Perhaps the greatest challenge and barrier to causal and statistical inference using observations from a single, interconnected social network is dependence among observations. The literature on statistics for dependent data is vast and multifaceted, but very little has been written about the dependence that arises when observations are sampled from a single network. Most of the literature on dependent random variables assumes that the domain from which observations are sampled (e.g. time or geographic space) has an underlying Euclidean geometry. The principles behind asymptotic results in the Euclidean dependence literature are simple and intuitive. They rely on a combination of stationarity assumptions, i.e. assumptions that certain features of the data generating process do not depend on an observation's location in the sample domain, and assumptions that bound the nature and the amount of dependence in the data. Most frequently these are mixing assumptions, which describe the decay of the correlation between observations as a function of the distance between them. Intuitively, in order to extract an increasing amount of information from a growing sample of dependent observations, old observations must be predictive of new observations, which is ensured by stationarity assumptions, and the amount of independence in the sample must grow faster than the amount of dependence, which is ensured by mixing conditions.

This literature is not immediately applicable to the network setting. Roughly, this is due to the difference between Euclidean and network topology. While it is possible to embed a network in $\mathbb{R}^d$ in such a way that preserves distances, to do so is to allow d to increase as n increases. Euclidean dependence results generally require d to be fixed, implying that, as new observations are sampled at the boundary of a Euclidean domain, the average and maximum pairwise distance between observations increases. Networks, on the other hand, often do not have a clear boundary to which we can add observations in such a way that ensures growth in the sample domain. In a large sample with Euclidean dependence, most observations will be distant from most other observations. This is not necessarily the case in networks. The maximum distance between two nodes can be small even in very large networks, and even if the maximum distance between two nodes is large, there may be many nodes that are close to one another. Therefore, mixing conditions do not necessarily result in more independence than dependence in a large sample from a network. Research indicates that social networks generally have the small-world property (sometimes referred to as the "six degrees of separation" property), meaning that the average distance between two nodes is small (Watts and Strogatz, 1998). Therefore distances in real-world networks may grow slowly with sample size. Of course some types of networks, e.g. lattices, embed in $\mathbb{R}^d$ as n grows, but these are generally trivial cases that are not useful for naturally occurring networks like social networks.

Dependence in networks is of two varieties–latent variable dependence and dependence due to direct transmission–each with its own implications for inference. In the literature on spatial and temporal dependence, dependence is often implicitly assumed to be the result of latent traits that are more similar for observations that are close in Euclidean distance than for distant observations. This type of dependence is likely to be present in many network contexts as well. In networks, edges present opportunities to transmit traits or

information, and this direct transmission is an important additional source of dependence that depends on the underlying network structure.

Latent variable dependence will be present in data sampled from a network whenever observations from nodes that are close to one another are more likely to share unmeasured traits than are observations from distant nodes. Homophily, or the tendency of people who share similar traits to form network ties, is a paradigmatic example of latent variable dependence. If the outcome under study in a social network has a genetic component, then we would expect latent variable dependence due to the fact that family members, who share latent genetic traits, are more likely to be close in social distance than people who are unrelated. If the outcome were affected by geography or physical environment, latent variable dependence could arise because people who live close to one another are more likely to be friends than those who are geographically distant. Of course, these traits can create dependence whether they are latent or observed. But if they are observed then conditioning on them renders observations independent; therefore the methodological challenges are greater when they are latent. Just like in the spatial and temporal dependence context, there is often little reason to think that we could identify, let alone measure, all of these sources of dependence. In order to make any progress towards valid inference in the presence of latent trait dependence, some structure must be assumed, namely that the range of influence of the latent traits is primarily local in the network and that any long-range effects are negligible. In a structural equation model, latent trait dependence would be captured by dependence among the error terms across subjects.

Dependence due to direct transmission will be present whenever one subject's treatments, outcomes, or covariates affect other subjects' treatments, outcomes, or covariates. This kind of dependence, which arises from causal effects between subjects, has structure lacking in latent trait dependence. Figure 1 depicts contagion in a network of three individuals. This diagram is the directed acyclic graph representation (Pearl, 1995; **?**) of the following structural equation model: At each time t, Y_i^t is affected by i's own past outcomes and those of i's social contacts. Individual 2 shares ties with 1 and 3 but individuals 1 and 3 are not connected. This structure implies conditional independences: $Y_1^{t-2} \perp Y_3^t \mid Y_2^{t-1}$ because any transmission from individual 1 to 3 must pass through 2; $Y_1^{t-2} \perp Y_2^{t-2}$ because information cannot be transmitted instantaneously. If observations are observed at closely spaced time intervals then these conditional independences can be harnessed for inference. There is no reason to think that any such conditional independences would hold with latent variable dependence. If some time points are not observed then the structure is lost and dependence due to direct transmission is indistinguishable from latent variable dependence.

In this paper, we accommodate both dependence due to direct transmission and dependence due to latent traits. We assume that both kinds of dependence are limited to *dependence neighborhoods* determined by the underlying social network: each subject, or node, i can directly transmit information, outcomes, or exposures to the nodes with which i shares a network tie, and each node i can share latent traits with the nodes with which i shares a network tie or a mutual connection. That a subject can only transmit to his or her immediate social contacts may be a reasonable assumption (indeed, our definition

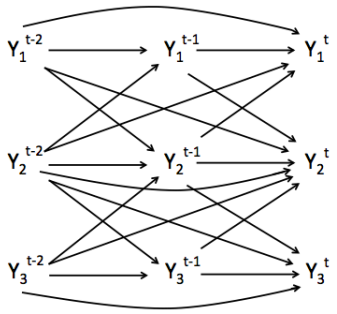

Fig. 1. A simple example of dependence due to direct transmission.

of network ties makes this true), but it is likely unrealistic to assume that latent variable dependence only affects nodes at a distance of one or two ties, as we assume throughout. Furthermore, harnessing the structure of direct transmission requires detailed data that may not be available in practice in many settings. This represents a first step towards valid statistical and causal inference under more realistic assumptions than have been required by previous work, but future work is needed to address more realistic–i.e. longer range–forms of latent variable dependence.

3. Methods

In this section we describe estimation of and inference about the causal effect of a treatment or exposure, X, including randomized and non-randomized exposures subject to interference. The approach we describe below is different from traditional approaches to interference in that it is justified when partial interference does not hold. As far as we are aware, this is the first approach to interference that references an asymptotic regime in which the number of ties for a given individual may grow with sample size. The estimating procedure that we describe in this section is based on van der Laan (2014), but we generalize the results to a broader class of causal effects and to more general and pervasive forms of dependence among observations. The conditions under which the resulting estimators are consistent and asymptotically normally distributed are different and weaker here than those in van der Laan (2014).

For the remainder of Section 3, we describe consistent and asymptotically normal (CAN) estimators of causal effects under two different sets of assumptions. One set of assumptions allows dependence due to direct transmission but not latent variable dependence, as in van der Laan (2014); under this set of assumptions our estimators inherit the efficiency properties from van der Laan (2014). The other set of assumptions allows dependence due to direct transmission and latent variable dependence; under this set of assumptions our estimators are CAN but may not be efficient. Our main result is asymptotic normality under an asymptotic regime in which the number of ties for a given individual may grow

with sample size in Section 3.5.

In Section 3.6 we describe statistical inference for the estimators introduced in Section 3.2. We consider two different classes of estimators: estimators that marginalize over baseline covariates and estimators that condition on baseline covariates. In some cases, variance estimation is facilitated by conditioning on covariates. Under the assumptions encoded in the structural equation model in Section 3.1, the conditional estimator is in fact consistent for the marginal estimand. However, conditional estimators have smaller variance and inference about the conditional estimand cannot be interpreted as inference about the marginal estimand. All of our estimands and estimators condition on the observed network as given by the adjacency matrix $\mathbf{A}$. A table summarizing the different assumptions and properties can be found in the Appendix.

We focus throughout on single time-point treatments. Longitudinal interventions are also possible under the theory introduced here but we leave the details for future work. We state our results under the assumption that all variables take values on discrete sets. Analogous results are valid for other types of random variables: it is straightforward to extend our notation and central limit theorem to continuous covariates and outcomes (though all efficiency results require discrete covariates), but continuous treatments are more complicated (see van der Laan, 2014).

3.1. Structural equation model

Let $K_i = \sum_{j=1}^{n} A_{ij}$, that is, K_i is the degree of node i, or the number of individuals sharing a tie with individual i. The degree of subject i and the degrees of i's alters may be included in the covariate vector C_i. We define $\mathbf{Y} = (Y_1, ..., Y_n)$ and $\mathbf{C}$ and $\mathbf{X}$ analogously. We use a structural equation model to define the causal effects of interest, as in Section 2, but note that analogous definitions may be achieved within the potential outcome framework (Pearl, 2012).

We assume that the data are generated by sequentially evaluating the following set of equations:

$$\begin{aligned} C_i &= f_C\left[\varepsilon_{C_i}\right] && i = 1, \ldots, n \\ X_i &= f_X\left[\left\{C_j : A_{ij} = 1\right\}, \varepsilon_{X_i}\right] && i = 1, \ldots, n \\ Y_i &= f_Y\left[\left\{X_j : A_{ij} = 1\right\}, \left\{C_j : A_{ij} = 1\right\}, \varepsilon_{Y_i}\right] && i = 1, \ldots, n, \end{aligned} \tag{1}$$

where f_C, f_X, and f_Y are unknown and unspecified functions and $\varepsilon_i = (\varepsilon_{C_i}, \varepsilon_{X_i}, \varepsilon_{Y_i})$ is a vector of exogenous, unobserved errors for individual i. This set of equations corresponds to observational settings when f_X depends on $\mathbf{C}$ and to randomized settings when it does not. Both X and Y may depend on $\mathbf{A}$ only through $\mathbf{C}$. Time ordering is a fundamental component of a structural causal model. For example, we assume that C is first drawn for all units, so that, in addition to C_i, the other components of the vector $\mathbf{C}$–corresponding to i's social contacts–may affect the value of X_i.

In addition, nonparametric identification of causal effects requires the following assump-

tions on the error terms from the SEM:

$$(\varepsilon_{X_1}, ..., \varepsilon_{X_n}) \perp (\varepsilon_{Y_1}, ..., \varepsilon_{Y_n}) \mid \mathbf{C}, \tag{A1}$$

$$\varepsilon_{X_1}, ..., \varepsilon_{X_n} \mid \mathbf{C} \text{ and } \varepsilon_{Y_1}, ..., \varepsilon_{Y_n} \mid \mathbf{C}, \mathbf{X} \text{ are identically distributed,} \tag{A2a}$$

$$\varepsilon_{X_i} \perp \varepsilon_{X_j} \mid \mathbf{C} \text{ and } \varepsilon_{Y_i} \perp \varepsilon_{Y_j} \mid \mathbf{C}, \mathbf{X} \text{ for } i, j \text{ s.t.}$$
$$A_{ij} = 0 \text{ and } \exists! k \text{ with } A_{ik} = A_{kj} = 1 \tag{A2b}$$

$$\varepsilon_{C_i}, i = 1, ..., n, \text{ are identically distributed, and} \tag{A3a}$$

$$\varepsilon_{C_i} \perp \varepsilon_{C_j} \text{ for } i, j \text{ s.t. } A_{ij} = 0 \text{ and } \exists! k \text{ with } A_{ik} = A_{kj} = 1. \tag{A3b}$$

Assumption (A1) ensures that $\mathbf{C}$ suffices to control for confounding of the effect of $\mathbf{X}$ on $\mathbf{Y}$. It implies that any latent variable dependence affects $\mathbf{X}$ and $\mathbf{Y}$ separately; in general a latent variable that affected $\mathbf{X}$ and $\mathbf{Y}$ jointly would constitute a violation of this assumption. Assumptions (A2b) and (A3b) ensure that any unmeasured sources of dependence–i.e. latent trait dependence–only affect pairs of observations up to a distance of two network ties–that is, friends or friends-of-friends. Assumption (A3) can be omitted if attention is restricted to causal effects conditional on $\mathbf{C}$.

Although our main result, given in Theorem 1 below, holds for inference in the SEM defined by assumptions (A1)–(A3b), some asymptotic properties are guaranteed only when stronger versions of assumptions (A2b) and (A3b) hold. We therefore introduce alternative assumptions

$$\varepsilon_{X_1}, ..., \varepsilon_{X_n} \mid \mathbf{C} \text{ are i.i.d. and } \varepsilon_{Y_1}, ..., \varepsilon_{Y_n} \mid \mathbf{C}, \mathbf{X} \text{ are i.i.d., and} \tag{A4}$$

$$\varepsilon_{C_i}, i = 1, ..., n, \text{ are i.i.d.} \tag{A5}$$

These assumptions are consistent with dependence due to direct transmission but not latent variable dependence.

Note that, although the variance-covariance structure of the SEM given in (1) is affected by the dependence allowed in (A2b) and (A3b), the mean structure is unaltered by the choice of assumptions (A2) and (A3) or (A4) and (A5). This rules out the possibility that any latent sources of dependence introduce confounding, and in particular while it allows limited forms of homophily to induce dependence it rules out confounding due homophily, which is a strong and often unrealistic assumption (Shalizi and Thomas, 2011). Therefore, any estimator that is unbiased under (A4) and (A5) will remain unbiased when these are relaxed to (A2) and (A3). In Section 3.2 we discuss nonparametric identification of causal parameters, which is agnostic to the choice of the weaker or stronger independence assumptions. In Section 3.3 we derive estimators under assumptions (A1), (A4), and (A5)–that is, in the presence of dependence due to direct transmission but not latent variable dependence. We use the stronger assumptions because the resulting model is amenable to familiar tools for deriving semiparametric estimators. In Section (3.5) we prove that the estimators derived under assumptions (A1), (A4), and (A5) are CAN under the weaker set of assumptions (A1)–(A3b). In Section 3.6 we discuss inference under each of the two sets of assumptions.

3.2. Definition and nonparametric identification of causal effects

In principle it is possible to perform statistical inference in the model defined by assumptions (A1)-(A3b) or by assumptions (A1), (A4), and (A5). However, in practice we may need to make dimension-reducing assumptions on the forms of f_X and f_Y. This is done by considering summary functions s_X and s_Y and random variables $W_i = s_{X,i}\left(\{C_j : A_{ij} = 1\}\right)$ and $V_i = s_{Y,i}\left(\{C_j : A_{ij} = 1\}, \{X_j : A_{ij} = 1\}\right)$ such that the model may be written as

$$\begin{aligned} C_i &= f_C\left[\varepsilon_{C_i}\right] && i = 1, \ldots, n \\ X_i &= f_X\left[W_i, \varepsilon_{X_i}\right] && i = 1, \ldots, n \\ Y_i &= f_Y\left[V_i, \varepsilon_{Y_i}\right] && i = 1, \ldots, n. \end{aligned}$$

For example, $s_{X,i}\left(\{C_j : A_{ij} = 1\}\right) = \left(C_i, \sum_{j:A_{ij}=1} C_j\right)$ implies that the treatment status of unit i only depends on i's own covariate value and on the sum of the covariate values of the units sharing a tie with i. Analogously, $s_{Y,i}\left(\{C_j : A_{ij} = 1\}, \{X_j : A_{ij} = 1\}\right) = \left(C_i, \sum_{j:A_{ij}=1} C_j, X_i, \sum_{j:A_{ij}=1} X_j\right)$ is an example of a summary function for f_Y. For convenience we use the notation $s_{X,i}(\mathbf{C})$ and $s_{Y,i}(\mathbf{C}, \mathbf{X})$ below; however, this notation should not undermine the important fact that W_i can only depend on the subset $\{C_j : A_{ij} = 1\}$ and V_i can only depend on the subsets $\{C_j : A_{ij} = 1\}$ and $\{X_j : A_{ij} = 1\}$ of $\mathbf{C}$ and $\mathbf{X}$, as these are the only components of $\mathbf{C}$ and $\mathbf{X}$ that are parents of $\mathbf{X}$ and $\mathbf{Y}$, respectively, in the network-as-structural-causal-model. For notational convenience, in what follows we augment the observed data random vector with V_i and W_i, recognizing that these are deterministic functionals of C_i and X_i, defined by $s_{Y,i}$ and $s_{X,i}$, and are therefore technically redundant.

A hypothetical intervention that deterministically sets X_i to a user-given value x_i^* for $i = 1, ..., n$ is given by

$$\begin{aligned} C_i &= f_C\left[\varepsilon_{C_i}\right] && i = 1, \ldots, n \\ X_i &= x_i^* && i = 1, \ldots, n \\ Y_i(\mathbf{x}^*) &= f_Y\left[V_i(\mathbf{x}^*), \varepsilon_{Y_i}\right] && i = 1, \ldots, n, \end{aligned}$$

where $\mathbf{x}^* = (x_1^*, \ldots, x_n^*)$. Here $Y_i(\mathbf{x}^*)$ denotes the potential or counterfactual outcome of individual i in a hypothetical world in which $P(\mathbf{X} = \mathbf{x}^*) = 1$. Analogously, $V_i(\mathbf{x}^*) = s_{Y,i}(\mathbf{C}, \mathbf{x}^*)$ is a counterfactual random variable in a hypothetical world in which $P(\mathbf{X} = \mathbf{x}^*) = 1$**.** Note that, although $V_i(\mathbf{x}^*)$ is counterfactual, its value is determined by the observed realization of $\mathbf{C}$ and by the user-specified value $\mathbf{x}^*$, and it is therefore known. In order to streamline notation as we describe increasingly complex interventions, we denote the counterfactual variables $V_i(\mathbf{x}^*)$ and $Y_i(\mathbf{x}^*)$ by V_i^* and Y_i^*, respectively.

The causal parameter of interest throughout is the expected average potential outcome in this same hypothetical world, i.e. $E\left[\bar{Y}_n^*\right]$, where $\bar{Y}_n^* = \frac{1}{n}\sum_{i=1}^n Y_i^*$. This parameter is conditional on the observed adjacency matrix and, unlike typical causal parameters in i.i.d. settings, is allowed to depend on n. Causal effects are contrasts for two different hypothetical intervention vectors $\mathbf{x}^*$. The *overall* effect of treatment (Hudgens and Halloran,

2008; Tchetgen Tchetgen and VanderWeele, 2012), for example, contrasts the intervention in which everyone is treated to the intervention in which nobody is treated. In Section 4 we discuss other types of causal effects of interest in social network settings.

We are now ready define notation that we will use throughout the remainder of the paper for functionals of the distribution of $\mathbf{O}$. Let $p_C(\mathbf{c}) = P(\mathbf{C} = \mathbf{c})$, $g(\mathbf{x}|\mathbf{w}) = P(\mathbf{X} = \mathbf{x} \mid \mathbf{W} = \mathbf{w})$, $g_i(x|w) = P(X_i = x|W_i = w)$, $p_Y(\mathbf{y}|\mathbf{v}) = P(\mathbf{Y} = \mathbf{y}|\mathbf{V} = \mathbf{v})$, and $p_{Y,i}(y|v) = P(Y_i = y|V_i = v)$. Define the two marginal distributions $h_i(v) = P(V_i = v)$ and $h_{i,x^*}(v) = P(V_i^* = v)$, noting that both h_i and h_{i,x^*} are determined by g and p_C and are therefore observed data quantities. Finally, $m(v) = \sum_y y\, p_Y(y|v)$ is the conditional expectation of Y given $V = v$.

In addition to assumptions (A1)-(A3b) or (A1), (A4), and (A5), identification of $E\left[\bar{Y}_n^*\right]$ requires the positivity assumption that, for all $\mathbf{c}$ in the support of $\mathbf{C}$,

$$P(V = v|\mathbf{C} = \mathbf{c}) > 0 \quad \text{for all } v \text{ in the range of } V_i^* \ . \tag{A6}$$

This assumption states that, within levels of $\mathbf{C}$, the values of V determined by the hypothetical intervention $\mathbf{x}^*$ have positive probability under the observed-data-generating distribution. Now the causal parameter $E\left[\bar{Y}_n^*\right]$ is identified by

$$\psi = \frac{1}{n}\sum_{i=1}^{n} E\left[m(V_i^*)\right] = \frac{1}{n}\sum_{i=1}^{n}\sum_{v} m(v)h_{i,x^*}(v). \tag{2}$$

This identification result is equivalent to

$$\psi = \frac{1}{n}\sum_{i=1}^{n}\sum_{\mathbf{c}} m(s_{Y,i}(\mathbf{c}, \mathbf{x}^*))p_C(\mathbf{c}). \tag{3}$$

From (3), it is clear that the conditional causal parameter $E\left[\bar{Y}_n^* \mid \mathbf{C} = \mathbf{c}\right]$ is identified by $\frac{1}{n}\sum_{i=1}^{n} m(s_{Y,i}(\mathbf{c}, \mathbf{x}^*))$.

3.3. Estimation

Estimation of and inference about $E\left[\bar{Y}_n^*\right]$ requires a statistical model $\mathcal{M}$ for the distribution of the observed data $P(\mathbf{O})$. That is, $\mathcal{M}$ is a collection of distributions over $\mathbf{O}$ of which one element is the true data-generating distribution. Our target of inference is a pathwise differentiable mapping $\Psi : \mathcal{M} \to \mathbb{R}$ such that ψ is $\Psi(P)$, the mapping evaluated at the true data-generating distribution. Under assumptions (A1), (A4), and (A5) the probability distribution of the observed data may be factorized as

$$P(\mathbf{O} = \mathbf{o}) = P(\mathbf{C} = \mathbf{c})\, g(\mathbf{x}|\mathbf{w})p_Y(\mathbf{y}|\mathbf{v}), \tag{4}$$

suggesting that $\mathcal{M}$ requires three components: a model for p_C, a model for g, and a model for $P(Y|V)$. Furthermore, the identification results in (2) and (3) indicate that ψ depends on $P(Y|V)$ only through m. The empirical distribution $\hat{p}_C$ can be used throughout to nonparametrically estimate p_C, but, when $\mathbf{C}$ is high-dimensional, g and m cannot be

non-parametrically estimated at rates of convergence that are fast enough to satisfy the regularity conditions of Theorem 1 (see Appendix). Therefore, in order to define the parameter mapping we require a statistical model $\mathcal{M} = \mathcal{M}_g \times \mathcal{M}_m$, where $\mathcal{M}_g$ is a collection of conditional distributions for X given W such that the true conditional distribution is a member, and $\mathcal{M}_m$ is a collection of expectations of Y relative to conditional distributions of Y given V such that the true conditional expectation of Y given V is a member. Estimation of ψ is based on the efficient influence function for the parameter mapping $\Psi : \mathcal{M} \to \mathbb{R}$. Under assumptions (A1), (A4), and (A5), the efficient influence function, D, evaluated at a fixed value $\mathbf{o}$ of $\mathbf{O}$ was derived by van der Laan (2014) and is given by

$$D(\mathbf{o}) = \sum_{j=1}^{n} \frac{1}{n} \sum_{i=1}^{n} E\left[m\left(V_i^*\right) \mid C_j = c_j\right] - \psi + \frac{1}{n} \sum_{i=1}^{n} \frac{\bar{h}_{x^*}(v_i)}{\bar{h}(v_i)} \left\{y_i - m\left(v_i\right)\right\}, \tag{5}$$

where $\bar{h}(v_i) = \frac{1}{n}\sum_{j=1}^{n} h_j(v_i)$, $\bar{h}_{x^*}(v_i) = \frac{1}{n}\sum_{j=1}^{n} h_{j,x^*}(v_i)$, $v_i = s_{Y,i}(\mathbf{c}, \mathbf{x})$, and $V_i^* = s_{Y,i}(\mathbf{C}, \mathbf{x}^*)$. The influence function has expected value equal to 0 at the true ψ; this fact can be used to generate unbiased estimating equations for ψ. van der Laan (2014) showed that estimating equations based on this efficient influence function are doubly robust: the right hand side of Equation (5) has expected value equal to 0 if $m(\cdot)$ is replaced with an arbitrary functional of V or if $g(\cdot)$ is replaced with an arbitrary functional of W, as long as one of the two remains correctly specified. (Recall that $g(\cdot)$, along with p_C, determines $\bar{h}_{x^*}(v_i)$ and $\bar{h}(v_i)$.) This implies that an estimating equation based on Equation (5) will be unbiased for ψ if either model $\mathcal{M}_m$ for $m(\cdot)$ or model $\mathcal{M}_g$ for $g(\cdot)$ is correctly specified, i.e. contains the truth, even if one is not. This influence function is efficient in that, when $m(\cdot)$ is correctly specified, it has the smallest variance among all influence functions in model $\mathcal{M}_g$. This sense of efficiency derives from the Convolution Theorem (Bickel et al., 1998), which holds under local asymptotic normality (Van der Vaart, 1998; van der Laan, 2014) and therefore in our setting.

The efficient influence function in a model that does not make any distributional assumptions about $\mathbf{C}$, that is under assumptions (A1) and (A4) only, is given in equation (6) below.

$$D'(\mathbf{o}) = \frac{1}{n} \sum_{i=1}^{n} \left(E\left[m\left(V_i^*\right) \mid \mathbf{C} = \mathbf{c}\right] - \psi + \frac{\bar{h}_{x^*}(v_i)}{\bar{h}(v_i)} \left\{y_i - m\left(v_i\right)\right\} \right). \tag{6}$$

We use this influence function in what follows. This is also the influence function used to derive estimators conditional on $\mathbf{C}$, in which case the first two terms cancel out; we will denote the conditional influence function with $D_c(\mathbf{o})$.

In the Appendix we describe a targeted maximum loss-based estimator (TMLE) of ψ, however all of the results that follow are equally applicable to a standard estimating equation approach. The estimator inherits the double robustness property we described above: it will be consistent for ψ if either the working model $\hat{g}$ for g or the working model $\hat{m}$ for m is correctly specified. This resulting estimator remains CAN for ψ under assumptions

(A2) and (A3) instead of (A4) and (A5), and the same procedure can be used to estimate the parameter conditional on $\mathbf{C}$.

3.4. A note on asymptotic growth

There are many complex issues surrounding asymptotic growth of networks (e.g. Diaconis and Janson, 2007; Shalizi and Rinaldo, 2013), and a large literature on graph limits (Lovász, 2012). These issues are largely beyond the scope of this paper, but we believe that our methods are consistent with realistic social networks. In particular, observed social networks and models proposed for generating social networks tend to have heavy-tailed degree distributions, with most nodes having low degree but a non-trivial proportion of nodes having high degree, with the maximum degree dependent on the size of the network, resulting in asymptotically sparse networks. Some researchers speculate that the heavy right tails of social network degree distributions tend to approximately follow power laws: $Pr(degree = k) \sim k^{-\alpha}$ for $2 < \alpha < 3$ (Barabási and Albert, 1999; Lovász, 2012; Newman and Park, 2003), in which case $Pr(degree > k) = O(k^{1-\alpha})$ for any fixed k. Even if degree distributions depart from power law distributions (Clauset et al., 2009) they are frequently incompatible with the assumption of bounded degrees, which has been used in previous methods for inference about observations sampled from a single social network. Our new methods are not able to accommodate the most highly connected nodes from a power law degree sequence, but they can nevertheless be used to perform inference about the other nodes in a network that has a power law degree distribution (see Section 4.4).

Our theoretical results require an asymptotic regime in which the number of nodes in the network, n, goes to infinity. Formalizing asymptotic growth of network-generating models, in particular for models with sparse limits, is an active area of research (Caron and Fox, 2017; Kolaczyk and Krivitsky, 2015) and is beyond the scope of this paper. We take for granted a sequence of networks with increasing n such that the structural equation model that specifies the distributions of covariates, treatment, and outcome is preserved, along with key features of the network topology.

The role of the central limit theorem below is to license the use of approximate 95% confidence intervals and normal approximations in finite samples, and as with any data-adaptive parameter we use asymptotic arguments to show that as $n \to \infty$, 95% confidence intervals approach nominal coverage rates. Because our parameter of interest is conditional on $\mathbf{A}$ and may depend on n, it is most natural to think of inference about the true causal parameter for the given, observed network. However, researchers may have reason to believe that the causal parameter does not depend on n or on $\mathbf{A}$ except through the distribution of $\mathbf{C}$ and $\mathbf{X}$, in which case inference about other similar networks may be warranted.

3.5. Asymptotic normality

In order to accommodate more realistic models of asymptotic growth in the network context, we consider an asymptotic regime in which K_i may grow as $n \to \infty$.

Theorem 1: *Let $K_{max,n} = max_i\{K_i\}$ for a fixed network with n nodes. Suppose that*

$K^2_{max,n}/n \to 0$ as $n \to \infty$. Under independence assumptions (A1) through (A3b), positivity assumption (A6), and regularity conditions (see Appendix),

$$\sqrt{C_n}\left(\hat{\psi} - \psi\right) \xrightarrow{d} N(0, \sigma^2),$$

$n/K^2_{max,n} \leq C_n \leq n$. The asymptotic variance of $\hat{\psi}$, σ^2, is given by the variance of the influence curve of the estimator.

In Section 4.4, below, we discuss settings in which the conditions for this theorem fail to hold, and ways to recover valid inference for conditional estimands in some of these settings. The proof of Theorem 1 is in the Appendix. Broadly, the proof has two parts: first, to show that the second order terms in the expansion of $\hat{\psi} - \psi$ are stochastically less than $1/\sqrt{C_n}$, and second, to show that the first order terms converge to a normal distribution when scaled by a factor of order $\sqrt{C_n}$. The proof that the second order terms are stochastically less than $1/\sqrt{C_n}$ is an extension of the empirical process theory of Van Der Vaart and Wellner (1996) and follows the same format as the proof in van der Laan (2014). For the proof that the first order terms converge to a normal distribution, we rely on Stein's method of central limit theorem proof (Stein, 1972). Stein's method allows us to derive a bound on the distance between our first order term (properly scaled) and a standard normal distribution; this bound depends on the degree distribution $K_1, ..., K_n$. We show that this bound converges to 0 as $n \to \infty$ under regularity conditions and our running assumption that $K^2_{max,n} = o(n)$.

When all nodes have the same number of ties, i.e. $K_i = K_{max,n}$ for all i, then the rate of convergence will be given $\sqrt{C_n} = \sqrt{n/K^2_{max,n}}$. When $K_{max,n}$ is bounded above as $n \to \infty$, as in van der Laan (2014), the rate of convergence will be $\sqrt{n}$. When $K_{max,n} \to \infty$ but some nodes have fewer than $K_{max,n}$ ties, the exact rate of convergence is between $\sqrt{n/K^2_{max,n}}$ and $\sqrt{n}$ but is difficult or impossible to determine analytically, as it may depend intricately on the structure of the network. The inferential procedures that we describe below do not require knowledge of the rate of convergence.

3.6. Inference

A 95% confidence interval for ψ is given by $\hat{\psi}_n \pm 1.96\sigma/\sqrt{C_n}$. In practice neither σ nor C_n are likely to be known, but available variance estimation methods estimate the variance of $\hat{\psi}_n$ directly, incorporating the rate of convergence without requiring it to be known a priori.

In principle, the variance of $\hat{\psi}$ can be estimated using the empirical average of the square of the influence function, substituting $\hat{\psi}$ for ψ and the fitted values from the working models $\hat{g}$ and $\hat{m}$ for g and m. Although this variance may be anticonservative if one, but not both, of the working models $\hat{g}$ and $\hat{m}$ is correctly specified, using flexible or non-parametric specifications for these models increases opportunities to estimate both consistently. However, unlike in i.i.d. settings, the expectation of the square of the empirical version of the influence function given in Equation (5) does not reduce to the sum of squared influence

terms for each observation. Instead, it includes double sums for all pairs of observations that are not marginally independent of one another. These terms capture covariances between dependent observations; these extra covariance terms reflect a larger variance and a slower rate of convergence due to dependence across observations.

When dependence is due to direct transmission, that is, under assumptions (A1), (A4), and (A5), two alternative variance estimation procedures are available. One option is to estimate the variance of the influence function $D'(\mathbf{o})$ given by Equation (6). Our TMLE is based on $D'(\mathbf{o})$, but because this is the efficient influence function in a model that makes fewer assumptions than (A1), (A4), and (A5), it has larger variance than $D(\mathbf{o})$ and provides a valid (asymptotically conservative) variance estimate even when estimation is based on $D(\mathbf{o})$. For consistent and computationally feasible estimators for the variance of $D'(\mathbf{o})$ see Sofrygin and van der Laan (2015).

An alternative approach to estimate the variance of $\hat{\psi}$ under assumptions (A1), (A4), and (A5) is to employ the following version of a parametric bootstrap, which might offer improvements in finite-sample performance over the previously described approach. For each of B bootstrap iterations, indexed by $b = 1, \ldots, B$, first n covariates $\mathbf{C}^b = (C_1^b, \ldots, C_n^b)$ are sampled with replacement, then the existing model fit $\hat{g}$ is applied to sampling of n exposures $\mathbf{X}^b = (X_1^b, \ldots, X_n^b)$, followed by a sample of n outcomes $\mathbf{Y}^b = (Y_1^b, \ldots, Y_n^b)$ based on the existing outcome model fit $\hat{m}$. The corresponding bootstrap random summaries W_i^b and V_i^b, for $i = 1, \ldots, n$, are constructed by applying the summary functions s_X and s_Y to $\mathbf{C}^b$ and $(\mathbf{C}^b, \mathbf{X}^b)$, respectively. This bootstrap sample is then used to obtain the predicted values from the existing auxiliary covariate fit $(\hat{\bar{h}}_{x^*}/\hat{\bar{h}})(V_i^b)$, for $i = 1, \ldots, n$, followed by a bootstrap-based fitting of ϵ, and finally, evaluation of bootstrap TMLE. Note that the TMLE model update is the only model fitting step needed at each iteration of the bootstrap, which significantly lowers the computational burden of this procedure. The variance estimate is then obtained by taking the empirical variance of bootstrap TMLE samples $\hat{\psi}^b$. Because the parametric bootstrap relies on known or assumed independences, and because only the TMLE model (i.e. not the full likelihood) is fit at each iteration, this procedure consistently estimates the variance of the first order terms in the expansion of $\hat{\psi} - \psi$, and we prove in the Appendix that the higher order terms are asymptotically neglible. However, due to dependence across observations, one must be judicious with applications of the bootstrap. For example, the parametric bootstrap procedure described above requires conditional independence of X_i given W_i and Y_i given V_i, along with the consistent modeling of the corresponding factors of the likelihood. It may seem natural to sample V_i directly from its corresponding auxiliary model fit, but this is likely to result in an anti-conservative variance estimates, since the conditional independence of V_i is unlikely to hold by virtue of its construction as a summary measure of the network.

When latent variable dependence is present, that is under assumptions (A1) through (A3), consistent and computationally feasible variance estimation procedures are not currently available for either $D'(\mathbf{o})$ or $D(\mathbf{o})$, because existing methods require bootstrapping some of the observed data. Without latent variable dependence we can take advantage of marginal and conditional independences to employ i.i.d. or parametric bootstrap methods,

but latent variable dependence requires new methods for dependent data bootstrap. For this reason, we instead estimate the conditional parameter with influence function $D_C(\mathbf{o})$. A simple plug-in estimator is available for the variance of this influence function (see the Appendix and van der Laan, 2014).

4. Extensions

In this section we extend the estimation procedure to two causal effects of great interest in the context of social networks: social contagion, or peer effects, and interventions on the network structure itself, i.e. interventions on $\mathbf{A} = [A_{ij} : i, j \in \{1, \ldots, n\}]$ where, as above, $A_{ij} \equiv I\{\text{subjects } i \text{ and } j \text{ share a tie}\}$. First we introduce dynamic and stochastic interventions.

4.1. Dynamic and stochastic interventions

A dynamic intervention assigns treatment as a user-specified function $d_X(\cdot)$ of $\mathbf{C}$; this corresponds to substituting $d_{X,i}(\mathbf{C})$ for x_i^* in the intervention model, definitions, and estimating procedure above. Treatment is deterministically specified conditional on covariates but is but allowed to depend ("dynamically") on covariates. A stochastic intervention (Muñoz and van der Laan, 2012; Haneuse and Rotnitzky, 2013; Young et al., 2014) that replaces f_X with a new, user-specified function r_X is represented by an intervention SEM that replaces the equation for X_i with $X_i^* = r_X[W_i^*, \varepsilon_{X_i}]$. The intervention changes the distribution of X but does not eliminate the stochasticity introduced by ε_X. In the social network setting, stochastic interventions that change the dependence of X_i on $\mathbf{C}$ and of and Y_i on $\mathbf{C}$ and $\mathbf{X}$ are of particular interest. For example, consider data generated by an SEM in which f_X depends on $\mathbf{C}$ only through $W_i = \frac{1}{|A_i|}\sum_{j:A_{ij}=1} C_j$, i.e. the mean of C among the set of alters of i. We might be interested in the mean counterfactual outcome under a stochastic intervention that forces f_X to depend instead on $W_i^* = max_{j:A_{ij}=1}\{C_j\}$, i.e. the maximum value C among the alters of i. This particular stochastic intervention modifies f_X only through W; it is represented by an intervention SEM that replaces the equation for X_i with $X_i^* = f_X[W_i^*, \varepsilon_{X_i}]$. For each x in the support of X, X_i is set by the intervention to x with probability $P\left[X = x | W = max_{j:A_{ij}=1}\{C_j\}\right]$.

We formally define the class of stochastic interventions that alter the dependence of X_i on $\mathbf{C}$ and of and Y_i on $(\mathbf{C}, \mathbf{X})$, discuss identifying assumptions and estimation procedures, and then describe some such interventions of particular interest. Let $s_{X,i}^*(\cdot)$ and $s_{Y,i}^*(\cdot,\cdot)$ be user-specified functionals. They are denoted by an asterisk because they index hypothetical interventions rather than realized data-generating mechanisms. Let $W_i^* = s_{X,i}^*(\mathbf{C})$ and $V_i^* = s_{Y,i}^*(\mathbf{C}, \mathbf{X}^*)$. We are concerned with the class of stochastic interventions given by

$$\begin{aligned} C_i &= f_C[\varepsilon_{C_i}] && i = 1, \ldots, n \\ X_i^* &= f_X[W_i^*, \varepsilon_{X_i}] && i = 1, \ldots, n \\ Y_i^* &= f_Y[V_i^*, \varepsilon_{Y_i}] && i = 1, \ldots, n. \end{aligned} \tag{7}$$

This can be interpreted as an intervention where, for each x^* in the support of X and for $i = 1, ..., n$, X_i is set to x^* with probability $P\left[X = x^* | W = s^*_{X,i}(\mathbf{C})\right]$ and V_i is set to $s^*_{Y,i}(\mathbf{C}, \mathbf{x}^*)$ deterministically for each possible realization $\mathbf{x}^*$. Because Y depends on $\mathbf{X}$ only through V, this is equivalent to an intervention that sets V_i to v with probability $P\left[\mathbf{X} \in \left\{\mathbf{x}^* : s^*_{Y,i}(\mathbf{C}, \mathbf{x}^*) = v\right\} \mid \mathbf{W} = \mathbf{s}^*_X(\mathbf{C})\right]$, where $\mathbf{s}^*_X(\mathbf{C}) = \left(s^*_{X,1}(\mathbf{C}), ..., s^*_{X,n}(\mathbf{C})\right)$.

This intervention is identified under the same assumptions as the deterministic interventions described above, with the exception of a positivity assumption that is a slight modification of (A6). Define $\mathcal{X}^* = \{\mathbf{x}^* : P[\mathbf{X} = \mathbf{x}^* | \mathbf{W} = \mathbf{s}^*_X(\mathbf{C})] > 0\}$ to be the set of treatment vectors $\mathbf{x}^*$ that have positive probability under the stochastic intervention defined by (7). We assume that, for all $\mathbf{c}$ in the support of $\mathbf{C}$,

$$min_{v \in \mathcal{V}^*} P(V = v | \mathbf{C} = \mathbf{c}) > 0 \quad \text{for } \mathcal{V}^* = \left\{ s^*_{Y,i}(\mathbf{C}, \mathbf{x}^*) : \mathbf{x}^* \in \mathcal{X}^* \right\} \tag{8}$$

That is, the conditional support of V^* must be included in the conditional support of V in order for the intervention to be supported by the data. Note that, in order for this positivity assumption to hold, the supports of $s^*_X(\cdot)$ and $s^*_Y(\cdot, \cdot)$ must be of the same dimensions as the supports of $s_X(\cdot)$ and $s_Y(\cdot, \cdot)$, respectively.

The causal parameter of interest is the expected average potential outcome under this hypothetical intervention, $E\left[\bar{Y}^*_n\right]$. Define $h^*_i(v) = P[V^*_i = v] = P\left[s^*_{Y,i}(\mathbf{C}, \mathbf{X}^*) = v\right]$. Then $E\left[\bar{Y}^*_n\right]$ is identified by

$$\begin{aligned}
\psi &= \frac{1}{n} \sum_{i=1}^{n} \sum_{\mathbf{c}, \mathbf{x}} E\left[Y_i | s^*_{Y,i}(\mathbf{c}, \mathbf{x})\right] P[\mathbf{X} = \mathbf{x} | \mathbf{W} = \mathbf{s}^*_X(\mathbf{c})] \, p_C(\mathbf{c}) \\
&= \frac{1}{n} \sum_{i=1}^{n} E[m(V^*_i)] = \frac{1}{n} \sum_{i=1}^{n} \sum_{v} m(v) h^*_i(v).
\end{aligned}$$

An influence function for ψ, evaluated at a fixed value of the observed data, $\mathbf{o}$, is given by

$$D^\dagger(\mathbf{o}) = \sum_{j=1}^{n} \frac{1}{n} \sum_{i=1}^{n} E\left[m\left(V^*_i\right) \mid C_j = c_j\right] - \psi + \frac{1}{n} \sum_{i=1}^{n} \frac{\bar{h}^*(v_i)}{\bar{h}(v_i)} \left\{y_i - m\left(v_i\right)\right\},$$

where $\bar{h}^*(v_i) = \frac{1}{n} \sum_{j=1}^{n} h^*_j(v_i)$. (When h^* is known, this is the efficient influence function under assumptions (A4) and (A5).) Estimation of $\bar{h}^*$ is carried out by substituting $\hat{g}$ and $\hat{p}_C$ for g and p_C in the expression

$$\bar{h}^*(v) = \frac{1}{n} \sum_{j} \sum_{\mathbf{c}, \mathbf{x}} I\left(s^*_{Y,i}(\mathbf{c}, \mathbf{x}) = v\right) g\left(\mathbf{x} | \mathbf{s}^*_X(\mathbf{c})\right) p_C(\mathbf{c}).$$

Since $\hat{p}_C$ is an empirical distribution that puts mass one on the observed value $\mathbf{c}$, the

estimator $\hat{\bar{h}}^*$ reduces to

$$\hat{\bar{h}}^*(v) = \frac{1}{n}\sum_{j=1}^{n}\sum_{\mathbf{x}} I\left(s^*_{Y,i}(\mathbf{x},\mathbf{C}) = v\right)\hat{g}(\mathbf{x}|\mathbf{s}^*_X(\mathbf{C})).$$

We denote by $\hat{\bar{h}}$ and $\hat{\bar{h}}^*$ the corresponding estimates of $\bar{h}$ and $\bar{h}^*$. Now the TMLE of ψ is computed according to the steps outlined in Section 3, but with V^* and Y^* defined as immediately above.

A special case of this class of stochastic interventions intervenes only on s_X, like the example discussed above in which the intervention forces f_X to depend on $W_i^* = max_{j:A_{ij}=1}\{C_j\}$ but does not alter the functional form of s_Y. $E[\bar{Y}_n^*]$ under this type of intervention is identified by

$$\begin{aligned}\psi &= \frac{1}{n}\sum_{i=1}^{n}\sum_{\mathbf{c},\mathbf{x}} E\left[Y_i|\mathbf{C}=\mathbf{c},\mathbf{X}=\mathbf{x}\right]P\left[\mathbf{X}=\mathbf{x}|\mathbf{W}=s^*_X(\mathbf{c})\right]p_C(\mathbf{c})\\ &= \frac{1}{n}\sum_{i=1}^{n} E\left[m(V_i^*)\right] = \frac{1}{n}\sum_{i=1}^{n}\sum_{v} m(v)h_i^*(v).\end{aligned}$$

With V_i^* defined as $s_{Y,i}(\mathbf{C},\mathbf{X}^*)$, estimation of this class of intervention proceeds as immediately above. The fact that $\mathbf{X}^*$ is random does not affect the estimation algorithm.

4.2. Peer effects

Define Y_i^0 to be the outcome variable measured at a time previous to the primary outcome measurement Y_i. Peer effects are the class of causal effects of Y_j^0 on Y_i for $A_{ij} = 1$: the effects of individuals' outcomes on the subsequent outcome of their alters. We can operationalize peer effects as the effects of dynamic interventions where the counterfactual exposure for subject i is given by a user-specified function $d_X(\cdot)$ of $\{Y_j^0 : A_{ij} = 1\}$. In order to maintain the identifying assumptions A2b and A3b, the time elapsed between Y^0 and Y must permit transmission only between nodes and their immediate alters. Otherwise, if the outcome could have spread contagiously more broadly, there will be more dependence present than our methods can account for, and also possible confounding of the effect of Y_i^0 on Y_j for $A_{ij} = 1$ due to mutual friends.

4.3. Interventions on network structure

An intervention on the network, i.e. an intervention that adds, removes, or relocates ties in the network, is a special case of a joint intervention on $s_X(\cdot)$ and $s_Y(\cdot)$. To see this, note that the network structure, codified by the adjacency matrix $\mathbf{A}$, enters the data-generating SEM (1) only through $s_X(\cdot)$ and $s_Y(\cdot)$; therefore we can represent any modification to $\mathbf{A}$ via the corresponding modification to $s_X(\cdot)$ and $s_Y(\cdot)$. This represents a strong assumption; if network structure can affect $\mathbf{Y}$ not through $s_X(\cdot)$ and $s_Y(\cdot)$ then estimating these effects

is more challenging (Ogburn et al., 2014; Toulis et al., 2018). Consider an intervention that replaces the observed adjacency matrix $\mathbf{A}$ with a user-specified adjacency matrix $\mathbf{A}^*$. This is a stochastic intervention, with $s^*_{X,i}(\mathbf{C})$ replaced by $s^{\mathbf{A}^*}_{X,i}(\mathbf{C}) \equiv s_{X,i}\left(\left\{C_j : A^*_{ij} = 1\right\}\right)$ and $s^*_{Y,i}(\mathbf{C}, \mathbf{X}^*)$ by $s^{\mathbf{A}^*}_{Y,i}(\mathbf{C}, \mathbf{X}^*) \equiv s_{Y,i}\left(\left\{X^*_j : A^*_{ij} = 1\right\}, \left\{C_j : A^*_{ij} = 1\right\}\right)$. The intervention SEM differs from the data-generating SEM only in that X_i depends on the covariate values for the individuals with whom i shares ties in the intervention adjacency matrix $\mathbf{A}^*$ and Y_i depends on the counterfactual treatments and observed covariate values for those same individuals.

Interventions on summary features of the adjacency matrix can also be viewed as stochastic interventions. Instead of replacing $\mathbf{A}$ with $\mathbf{A}^*$, an intervention on features of the network structure replaces $\mathbf{A}$ with the members of a class $\mathcal{A}^*$ of $n \times n$ adjacency matrices that share the intervention features, stochastically according to some probability distribution $g_{\mathbf{A}^*}$ over $\mathcal{A}^*$. For example, we might be interested in interventions that constrain the degree distribution of the network, e.g. fixing the maximum degree to be smaller than some D. We might specify $g_{\mathbf{A}^*}(A) = \frac{1}{|\mathcal{A}^*|} I\{A \in \mathcal{A}^*\}$, giving equal weight to each realization in the class $\mathcal{A}^*$. Effectively, this kind of intervention sets V_i to v with probability

$$P\left[\mathbf{X} \in \left\{\mathbf{x}^* : s^{\mathbf{T}^*}_{Y,i}(\mathbf{C}, \mathbf{x}^*) = v\right\} \mid \mathbf{W} = \mathbf{s}^{\mathbf{A}^*}_X(\mathbf{C}) \text{ for some } \mathbf{A}^* \in \mathcal{A}^*\right],$$

where $\mathbf{s}^{\mathbf{A}^*}_X(\mathbf{C}) = \left(s^{\mathbf{A}^*}_{X,1}(\mathbf{C}), ..., s^{\mathbf{A}^*}_{X,n}(\mathbf{C})\right)$.

As with the stochastic interventions discussed in the previous section, positivity is a crucial assumption for identifying interventions on $\mathbf{A}$: the support of V^* must be the same as the support of V. If replacing $\mathbf{A}$ with $\mathbf{A}^*$ (either deterministically or as a random selection from the class $\mathcal{A}^*$) assigns to unit i a value of V that not observed in the real data for a unit in the same $\mathbf{C}$ stratum as i, then the effect of the intervention that that replaces $\mathbf{A}$ with $\mathbf{A}^*$ is not identified for unit i. In general it will be possible to identify interventions on local but not global features of network structure. Examples of local features of network structure include the degree of subject i and local clustering around subject i: they depend on $\mathbf{A}$ only through subject i and subject i's immediate contacts. A local clustering coefficient for node i can be defined as the proportion of potential triangles that include i as one vertex and that are completed, or the number of pairs of neighbors of i who are connected divided by the total number of pairs of neighbors of i (Newman, 2009). This measure of triangle completion captures the extent to which "the friend of my friend is also my friend": triangle completion is high whenever two subjects who share a mutual contact are more likely to themselves share a tie than are two subjects chosen at random from the network. Positivity could hold if, within each level of $\mathbf{C}$, subjects were observed to have a wide range of degrees and of triangle completion among their contacts. In contrast with degree and local clustering, network centrality is a node-specific attribute that nevertheless depends on the entire network structure. It captures the intuitive notion that some nodes are central and some nodes are fringe in any given network. It can be measured in many different ways, based, for example, on the number of network paths that intersect node i, on the probability that a random walk on the network will intersect node

i, or on the mean distance between node i and the other nodes in network (see Chapter 7 of Newman, 2009 for a comprehensive discussion of these and other centrality measures). Centrality is given by a univariate measure for each node in a network, but each node's measure depends crucially on the entire graph. In reality it is not generally possible to intervene on centrality without altering the entire adjacency matrix $\mathbf{A}$, and the positivity assumption is unlikely to hold.

4.4. Too many friends, too much influence

The conditions of Theorem 1 will be violated for any asymptotic regime in which the degree of one or more nodes grows at a rate equal to or faster than $\sqrt{n}$. This is problematic because social networks frequently have a small number of "hubs"–that is, nodes with very high degree (Newman, 2009), and the occurrence of hubs is a feature of many of the network-generating models that have been proposed for social networks. When a small number of individuals wield influence over a significant portion of the rest of the population, two problems arise for statistical inference. First, the number of hubs may stay small as n increases. If the hubs are systematically different from the rest of the population, then a fixed or slowly growing number of hubs would not allow for consistent inference about this distinct subpopulation. Second, and more importantly, the sweeping influence of hubs creates dependence among all of the influenced nodes that undermines inference. Our methods rely on the independence of Y_i and Y_j whenever nodes i and j do not share a tie or a mutual alter. When hubs are present, a significant proportion of nodes will share a connection to one of these hubs, undermining our methods.

We can recover valid inference using our methods if we condition on the hubs, treating them as features of the background network environment rather than as observations. This results in different causal effects or statistical estimands, as all of our inference is conditional on the identity and characteristics of the hubs. Imagine a social network comprised of the residents of a city in which a cultural or political leader is connected to almost all of the other nodes. It may be impossible to disentangle the influence of this leader, which affects every other node, from other processes simultaneously occurring among the other residents of the city. It will certainly be impossible to statistically learn about the hub, as the sample size for the hub subgroup is 1. But it may make sense to consider the hub as a feature of the city rather than a member of the network. We could then learn about other processes occurring among the other residents of the city, conditional on the behavior and characteristics of the leader. For example, we could evaluate the effect of a public health initiative encouraging residents to talk to their friends about the importance of exercise, but we could not evaluate a similar program targeting the leader's communication about exercise.

Practically speaking, for real and finite datasets, this implies that the methods we have proposed are inappropriate for networks in which the degree is large, compared to n, for one or more nodes. If many nodes are connected to a significant fraction of other nodes, this problem is intractable. However, if only a small number of nodes are highly connected we can condition on them to recover approximately valid inference using our methods for

conditional estimands. There is a theoretical tradeoff between the rate of convergence of our estimators and the order of K relative to n that, in finite samples, becomes a practical tradeoff between generality and variance. Increasing the number of nodes classified as hubs will increase the rate of convergence by decreasing the size of K for the remaining, non-hub nodes (assuming that the number of hubs remains small compared to n so that the sample size does not decrease significantly when we exclude hubs from the analysis). On the other hand, classifying more nodes as hubs results in analyses that are increasingly specific: conditioning on a single hub may preserve generalizability to other networks (similar cities with similar leaders), but conditioning on many hubs is likely to limit the generalizability of the resulting inference.

5. Simulations

We conducted a simulation study that evaluated the finite sample and asymptotic behavior of the TMLE procedure described in Section 3.3. We generated social networks of size $n = 500$, $n = 1,000$, and $n = 10,000$ according to the preferential attachment model (Barabási and Albert, 1999), where the node degree (number of friends) distribution followed a power law with $\alpha = 0.5$. We generated data with two different types of dependence: first with dependence due to direct transmission only, and second with both latent variable dependence and dependence due to direct transmission. Details of the simulations, along with results for networks generated under the small world model (Watts and Strogatz, 1998), are in the Appendix.

Our simulations mimicked a hypothetical study designed to increase the level of physical activity in a population comprised of members of a social network. For each community member indexed by $i = 1, \ldots, n$, the study collected data on i's baseline covariates, denoted C_i, which included the indicator of being physically active, denoted PA_i and the network of friends on each subject, F_i. The exposure or treatment, X_i, was assigned randomly to 25% of the community. For example, one can imagine a study where treated individuals received various economic incentives to attend a local gym. The outcome Y_i was a binary indicator of maintaining gym membership for a pre-determined follow-up period. We estimated the average of the mean counterfactual outcomes $E\left[\bar{Y}_n^*\right]$ under various hypothetical interventions g^* on such a community. First, we considered a stochastic intervention g_1^* which assigned each individual to treatment with a constant probability of 0.35; this differs from the observed allocation of treatment to 25% of the community members. We also considered a scenario in which the economic incentive was resource constrained and could only be allocated to up to 10% of community members. We estimated the effects of various targeted approaches to allocating the exposure. For example, we considered an intervention g_2^* that targeted only the top 10% most connected members of the community, as such a targeted intervention would be expected to have a higher impact on the overall average probability of maintaining gym membership among the community, when compared to purely random assignment of exposure to 10% of the community. Another hypothetical intervention g_3^* assigned an additional physically active friend to individuals with fewer than 10 friends. This is an intervention on the structure of the social network

itself. Finally, we estimated the combined effect of simultaneously implementing intervention g_2^* and the network-based intervention g_3^* on the same community . For simplicity, this simulation study only reports the expected outcome under each of these interventions; causal effects defined as contrasts of these interventions can be easily estimated based on the same methods.

We estimated the expected counterfactual outcomes under the four interventions and evaluated their finite sample biases. For the simulations under dependence due to direct transmission, we estimated the marginal parameter $E\left[\bar{Y}_n^*\right]$ and compared three different estimators of the asymptotic variance and the coverage of the corresponding confidence intervals. First, we looked at the naive plug-in i.i.d. estimator ("*IID Var*") for the variance of the influence curve which treated observations as if they were i.i.d. Second, we used the plug-in variance estimator based on the efficient influence curve which adjusted for the correlated observations ("*dependent IC Var*") (Sofrygin and van der Laan, 2015). Finally, we used the parametric bootstrap variance estimator ("*bootstrap Var*") described in Section 3.6. The simulation results showing the mean length and coverage of these three CI types are shown in Figure 2. The results from the simulations with latent variable dependence are in Figure 3. We estimated the conditional parameter $E\left[\bar{Y}_n^*\right]$ and we compared two plug in variance estimators based on the conditional influence function D_C: one that assumes conditionally i.i.d outcomes (conditional on $\mathbf{X}$ and $\mathbf{C}$), which would be true if all dependence were due to direct transmission but is violated in the presence of latent variable dependence ("*IID Var*"), and one that does not make this assumption ("*dependent IC Var*"). In the Appendix we compare histograms of the estimates to the predicted normal limiting distribution.

One of the lessons of our simulation study is that by leveraging the structure of the network it might be possible to achieve a larger overall intervention effect on a population level (Harling et al., 2016). For example, the results in the left panel of Figure 2 show that by targeting the exposure assignment to highly connected and physically active individuals, intervention g_2^* increases the mean probability of sustaining gym membership compared to the similar level of un-targeted coverage of the exposure. We also demonstrated the feasibility of estimating effects of interventions on the observed network structure itself, such as intervention g_3^*, which can be also combined with economic incentives, as it was mimicked by our hypothetical intervention $g_2^* + g_3^*$. These combined interventions could be particularly useful in resource constrained environments, since they may result in larger community level effects at the lower coverage of the exposure assignment.

Results from simulations with dependence due to direct transmission show that conducting inference while ignoring the nature of the dependence in such datasets generally results in anticonservative variance estimates and under-coverage of CIs, which can be as low as 50% even for very large sample sizes ("*IID Var*" in the right panel of Figure 2). The CIs based on the dependent variance estimates ("*dependent IC Var*") obtain nearly nominal coverage of 95% for large enough sample sizes, but can suffer in smaller sample sizes due to lack of asymptotic normality and near-positivity violations. Notably, the CIs based on the parametric bootstrap variance estimates provide the most robust coverage for smaller sample sizes, while attaining the nominal 95% coverage in large sample sizes for

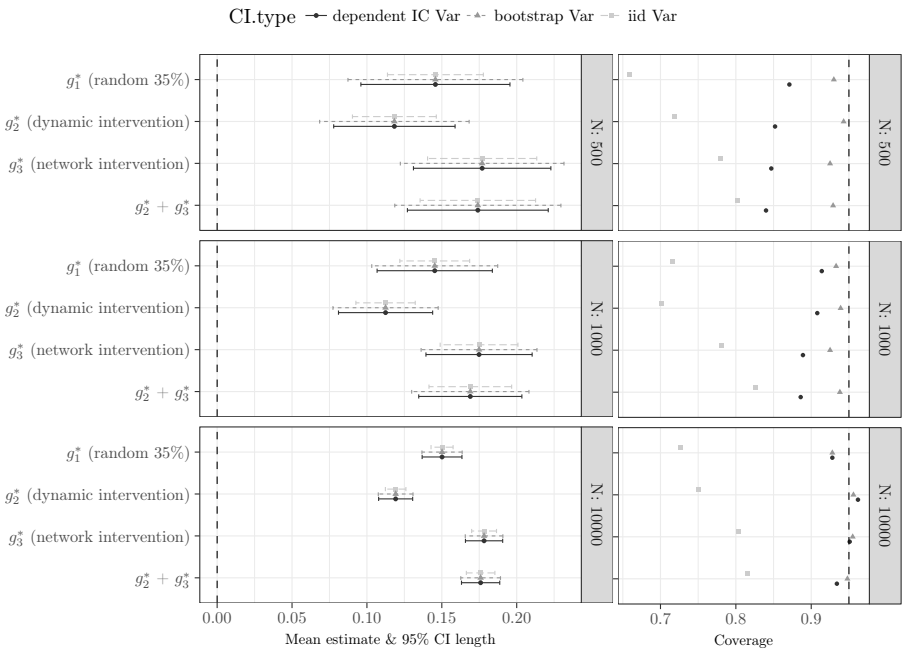

Fig. 2. Mean 95% CI length (left panel) and coverage (right panel) for the TMLE in preferential attachment network with dependence due to direct transmission, by sample size, intervention and CI type.

nearly all of the simulation scenarios ("*bootstrap Var*"). The apparent robustness of the parametric bootstrap method for inference in small sample sizes, even as low as $n = 500$, was one of the surprising finding of this simulation study. Future work will explore the assumptions under which this parametric bootstrap works and its sensitivity towards violations of those assumptions. Similarly, in the simulations with latent variable dependence the variance estimates that assume conditionally i.i.d. outcomes, i.e. that dependence may be due to direct transmission but not to latent variables, are anti-conservative.

6. Conclusion

We proposed new methods that allow for causal and statistical inference using observations sampled from members of a single interconnected social network when the observations evince dependence due to network ties. In contrast to existing methods, ours do not require randomization of an exogenous treatment and have proven performance under asymptotic regimes in which the number of network ties grows (slowly) with sample size. In future work we plan to address a key limitation of the present proposal, namely the assumption that the network is observed fully and without error. We also plan to develop data-adaptive methods for estimating the summary measures s_X and s_Y, as it may be unreasonable to expect these to be known *a priori*. Finally, we plan to develop estimating algorithms for

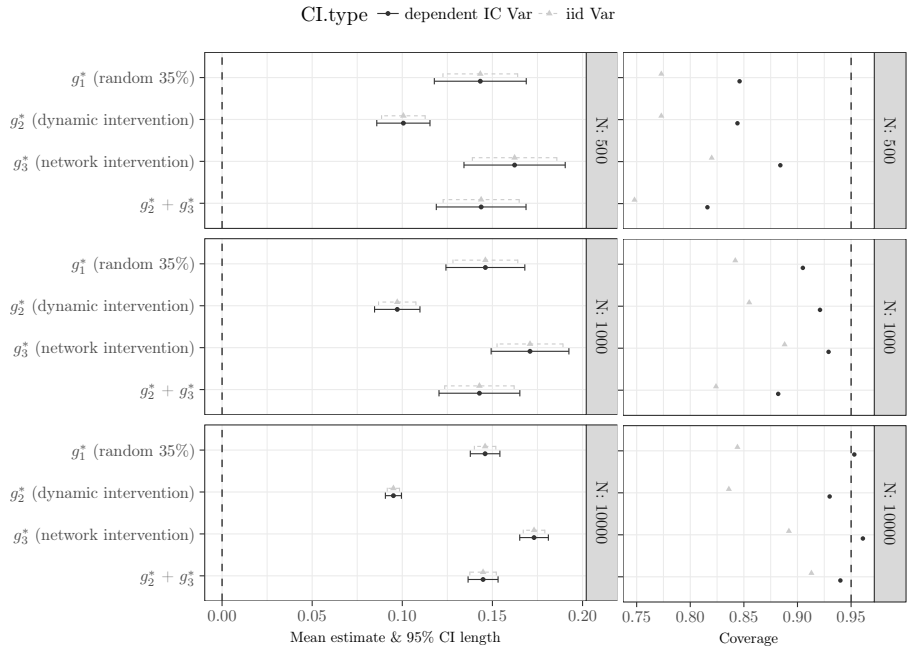

Fig. 3. Mean 95% CI length (left panel) and coverage (right panel) for the TMLE in preferential attachment network with latent variable dependence, by sample size, intervention and CI type.

longitudinal settings; the influence function and asymptotic results for these settings are straightforward extensions of the results presented here, but estimation can be challenging.

Acknowledgements

The authors are grateful to Caleb Miles, Eric Tchetgen Tchetgen and Victor De Gruttola for helpful comments. Elizabeth Ogburn was supported by ONR grant N000141512343. Oleg Sofrygin and Mark van der Laan were supported by NIH grant R01 AI074345-07.

References

Ali, M. M. and D. S. Dwyer (2010). Social network effects in alcohol consumption among adolescents. *Addictive behaviors 35*(4), 337–342.

Aronow, P. M. and C. Samii (2013). Estimating average causal effects under general interference. Technical report, Yale University.

Athey, S., D. Eckles, and G. W. Imbens (2016). Exact p-values for network interference*. *Journal of the American Statistical Association* (just-accepted).

Barabási, A.-L. and R. Albert (1999). Emergence of scaling in random networks. *science 286*(5439), 509–512.

Bickel, P. J., C. A. Klaassen, P. J. Bickel, Y. Ritov, J. Klaassen, J. A. Wellner, and Y. Ritov (1998). *Efficient and adaptive estimation for semiparametric models*, Volume 2. Springer New York.

Bowers, J., F. M. M, and P. C (2013). Reasoning about interference between units: A general framework. *Political Analysis 21*, 97–124.

Cacioppo, J. T., J. H. Fowler, and N. A. Christakis (2009). Alone in the crowd: the structure and spread of loneliness in a large social network. *Journal of personality and social psychology 97*(6), 977.

Caron, F. and E. B. Fox (2017). Sparse graphs using exchangeable random measures. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 79*(5), 1295–1366.

Christakis, N. and J. Fowler (2007). The spread of obesity in a large social network over 32 years. *New England Journal of Medicine 357*(4), 370–379.

Christakis, N. and J. Fowler (2008). The collective dynamics of smoking in a large social network. *New England journal of medicine 358*(21), 2249–2258.

Christakis, N. and J. Fowler (2010). Social network sensors for early detection of contagious outbreaks. *PloS one 5*(9), e12948.

Clauset, A., C. R. Shalizi, and M. E. Newman (2009). Power-law distributions in empirical data. *SIAM review 51*(4), 661–703.

Cohen-Cole, E. and J. Fletcher (2008). Is obesity contagious? social networks vs. environmental factors in the obesity epidemic. *Journal of Health Economics 27*(5), 1382–1387.

Diaconis, P. and S. Janson (2007). Graph limits and exchangeable random graphs. *arXiv preprint arXiv:0712.2749*.

Eckles, D., B. Karrer, and J. Ugander (2014). Design and analysis of experiments in networks: Reducing bias from interference. *arXiv preprint arXiv:1404.7530*.

Forastiere, L., E. M. Airoldi, and F. Mealli (2016). Identification and estimation of treatment and interference effects in observational studies on networks. *arXiv preprint arXiv:1609.06245*.

Graham, B., G. Imbens, and G. Ridder (2010). Measuring the effects of segregation in the presence of social spillovers: A nonparametric approach. Technical report, National Bureau of Economic Research.

Halloran, M. and M. Hudgens (2011). Causal inference for vaccine effects on infectiousness. *The University of North Carolina at Chapel Hill Department of Biostatistics Technical Report Series*, 20.

Halloran, M. and C. Struchiner (1995). Causal inference in infectious diseases. *Epidemiology*, 142–151.

Haneuse, S. and A. Rotnitzky (2013). Estimation of the effect of interventions that modify the received treatment. *Statistics in medicine 32*(30), 5260–5277.

Harling, G., R. Wang, J.-P. Onnela, and V. DeGruttola (2016). Leveraging contact network structure in the design of cluster randomized trials. *Harvard University Biostatistics Working Paper Series* (Working Paper 199).

Hong, G. and S. Raudenbush (2006). Evaluating kindergarten retention policy. *Journal of the American Statistical Association 101*(475), 901–910.

Hong, G. and S. Raudenbush (2008). Causal inference for time-varying instructional treatments. *Journal of Educational and Behavioral Statistics 33*(3), 333–362.

Hudgens, M. and M. Halloran (2008). Toward causal inference with interference. *Journal of the American Statistical Association 103*(482), 832–842.

Jagadeesan, R., N. Pillai, and A. Volfovsky (2017). Designs for estimating the treatment effect in networks with interference. *arXiv preprint arXiv:1705.08524*.

Kolaczyk, E. D. and P. N. Krivitsky (2015). On the question of effective sample size in network modeling: An asymptotic inquiry. *Statistical science: a review journal of the Institute of Mathematical Statistics 30*(2), 184.

Lauritzen, S. L. and T. S. Richardson (2002). Chain graph models and their causal interpretations. *Journal of the Royal Statistical Society: Series B 64*(3), 321–348.

Liu, L. and M. G. Hudgens (2014). Large sample randomization inference of causal effects in the presence of interference. *Journal of the american statistical association 109*(505), 288–301.

Liu, L., M. G. Hudgens, and S. Becker-Dreps (2016). On inverse probability-weighted estimators in the presence of interference. *Biometrika 103*(4), 829–842.

Lovász, L. (2012). *Large networks and graph limits*, Volume 60. American Mathematical Soc.

Lyons, R. (2011). The spread of evidence-poor medicine via flawed social-network analysis. *Statistics, Politics, and Policy 2*(1).

Madan, A., S. T. Moturu, D. Lazer, and A. S. Pentland (2010). Social sensing: obesity, unhealthy eating and exercise in face-to-face networks. In *Wireless Health 2010*, pp. 104–110. ACM.

Muñoz, I. D. and M. van der Laan (2012). Population intervention causal effects based on stochastic interventions. *Biometrics 68*(2), 541–549.

Newman, M. (2009). *Networks: an introduction*. Oxford: Oxford University Press.

Newman, M. E. and J. Park (2003). Why social networks are different from other types of networks. *Physical Review E 68*(3), 036122.

Ogburn, E. L. and T. J. VanderWeele (2014). Vaccines, contagion, and social networks. *arXiv preprint arXiv:1403.1241*.

Ogburn, E. L., T. J. VanderWeele, et al. (2014). Causal diagrams for interference. *Statistical science 29*(4), 559–578.

Pearl, J. (1995). Causal diagrams for empirical research. *Biometrika 82*(4), 669–688.

Pearl, J. (2000). *Causality: models, reasoning and inference*. Cambridge Univ Press.

Pearl, J. (2012). The causal foundations of structural equation modeling. Technical report, CALIFORNIA UNIV LOS ANGELES DEPT OF COMPUTER SCIENCE.

Rosenbaum, P. (2007). Interference between units in randomized experiments. *Journal of the American Statistical Association 102*(477), 191–200.

Rosenquist, J. N., J. Murabito, J. H. Fowler, and N. A. Christakis (2010). The spread of alcohol consumption behavior in a large social network. *Annals of Internal Medicine 152*(7), 426–433.

Rubin, D. (1990). Comment: Neyman (1923) and causal inference in experiments and observational studies. *Statistical Science 5*(4), 472–480.

Shalizi, C. and A. Thomas (2011). Homophily and contagion are generically confounded in observational social network studies. *Sociological Methods & Research 40*(2), 211–239.

Shalizi, C. R. (2012). Comment on "why and when 'flawed' social network analyses still yield valid tests of no contagion". *Statistics, Politics, and Policy 5*(1).

Shalizi, C. R. and A. Rinaldo (2013). Consistency under sampling of exponential random graph models. *Annals of Statistics 41*(2), 508–535.

Sobel, M. (2006). What do randomized studies of housing mobility demonstrate? *Journal of the American Statistical Association 101*(476), 1398–1407.

Sofrygin, O. and M. J. van der Laan (2015). Semi-Parametric Estimation and Inference for the Mean Outcome of the Single Time-Point Intervention in a Causally Connected Population. *U.C. Berkeley Division of Biostatistics Working Paper Series* (Working Paper 344).

Stein, C. (1972). A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. In *Proc. Sixth Berkeley Symp. Math. Stat. Prob.*, pp. 583–602.

Sussman, D. L. and E. M. Airoldi (2017). Elements of estimation theory for causal effects in the presence of network interference. *arXiv preprint arXiv:1702.03578*.

Tchetgen Tchetgen, E. J., I. Fulcher, and I. Shpitser (2017, 09). Auto-g-computation of causal effects on a network. Technical report.

Tchetgen Tchetgen, E. J. and T. VanderWeele (2012). On causal inference in the presence of interference. *Statistical Methods in Medical Research 21*(1), 55–75.

Toulis, P., A. Volfovsky, and E. M. Airoldi (2018). Propensity score methodology in the presence of network entanglement between treatments. *arXiv preprint arXiv:1801.07310*.

van der Laan, M. J. (2014). Causal inference for a population of causally connected units. *Journal of Causal Inference 0*(0), 2193–3677.

Van der Vaart, A. W. (1998). *Asymptotic statistics*, Volume 3. Cambridge university press.

Van Der Vaart, A. W. and J. A. Wellner (1996). Weak convergence. In *Weak Convergence and Empirical Processes*, pp. 16–28. Springer.

VanderWeele, T. (2010). Direct and indirect effects for neighborhood-based clustered and longitudinal data. *Sociological Methods & Research 38*(4), 515–544.

VanderWeele, T. J., E. L. Ogburn, and E. J. T. Tchetgen (2012). Why and when 'flawed' social network analyses still yield valid tests of no contagion. *Statistics, Politics, and Policy 3*(1).

Wasserman, S. (2013). Comment on "social contagion theory: Examining dynamic social networks and human behavior" by nicholas christakis and james fowler. *Statistics in Medicine 32*(4), 578–580.

Watts, D. J. and S. H. Strogatz (1998). Collective dynamics of 'small-world' networks. *Nature 393*(6684), 440–442.

Young, J. G., M. A. Hernán, and J. M. Robins (2014). Identification, estimation and approximation of risk under interventions that depend on the natural value of treatment using observational data. *Epidemiologic Methods 3*(1), 1–19.