ANNUAL REVIEWS

Annual Review of Sociology

Causal Network Analysis

Weihua An,[1,2,*] Roberson Beauvile,[1,*] and Benjamin Rosche[3,*]

[1]Department of Sociology, Emory University, Atlanta, Georgia, USA; email: weihua.an@emory.edu

[2]Department of Quantitative Theory and Methods, Emory University, Atlanta, Georgia, USA

[3]Department of Sociology, Cornell University, Ithaca, New York, USA

Annu. Rev. Sociol. 2022. 48:23–41

First published as a Review in Advance on April 6, 2022

The *Annual Review of Sociology* is online at soc.annualreviews.org

https://doi.org/10.1146/annurev-soc-030320-102100

*The authors are listed alphabetically and contributed equally to the article

Keywords

causal inference, network formation, network effects, social capital, social contagion, peer effect

Abstract

Fueled by recent advances in statistical modeling and the rapid growth of network data, social network analysis has become increasingly popular in sociology and related disciplines. However, a significant amount of work in the field has been descriptive and correlational, which prevents the findings from being more rigorously translated into practices and policies. This article provides a review of the popular models and methods for causal network analysis, with a focus on causal inference threats (such as measurement error, missing data, network endogeneity, contextual confounding, simultaneity, and collinearity) and potential solutions (such as instrumental variables, specialized experiments, and leveraging longitudinal data). It covers major models and methods for both network formation and network effects and for both sociocentric networks and egocentric networks. Lastly, this review also discusses future directions for causal network analysis.

INTRODUCTION

Social network analysis (SNA) is a major breakthrough in social science research. It shifts the focus from individual units (e.g., persons, organizations, or words) to their connections (e.g., friendships, collaborations, or co-occurrences). It brings about both theoretical and methodological innovations. Interest in network analysis has exploded, especially in the past decade or so, due to new advancements in statistical modeling and the rapid growth of network data.

Broad overviews of SNA are provided by Wasserman & Faust (1994), Borgatti et al. (2009), Lusher et al. (2013), and Yang et al. (2016), among others. The literature predominantly focuses on descriptive analysis and correlational analysis. However, in our view, more attention needs to be devoted to causal inference in network analysis, not only as a response to the causal inference revolution in social sciences (Angrist & Pischke 2009, Morgan & Winship 2014, Imbens & Rubin 2015, Hernán & Robins 2020), but also for the purpose of improving network analysis itself. This is necessary because, first, the disciplines from which researchers move into the field differ significantly in the amount of knowledge accumulation on causal inference, which sometimes leads to unsubstantiated causal claims. Second, network analysis often entails important practical and policy implications, which naturally requires stronger causal scrutiny.

This review differs from prior reviews (An 2011, VanderWeele & An 2013, Bramoullé et al. 2020) on the subject in a few aspects. First, unlike prior reviews that only focus on peer effects, this one covers both network formation and network effects. Second, unlike prior reviews that only focus on sociocentric network analysis, this one also covers egocentric network analysis. Third, this review presents more updated literature on the subject by incorporating a considerable amount of new work that has been done in the past decade. Lastly, this review provides more detailed accounts of the various threats to causal network analysis, including measurement error, missing data, network endogeneity, contextual confounding, simultaneity, and collinearity, among others. We note that this review focuses on cross-sectional network data analysis but also touches on longitudinal network data analysis where it is useful to identify causal effects.

The article proceeds as follows. In the next section, we present a brief overview of SNA to provide some background. In the subsequent section, we review models for studying network formation, in particular, the exponential random graph model (ERGM). Then, we review models for identifying causal network effects, in particular, the social capital model and the social contagion model. Lastly, we provide conclusions and discussions.

A BRIEF HISTORY OF SOCIAL NETWORK ANALYSIS

Based on prior literature (Scott 1991, Freeman 2004), we can divide the history of SNA into five periods. First was the emergence period (1930s–1950s), in which Moreno (1934) introduced sociograms to represent relations between individual actors in a small group. Second was the quiet period (1950s–1970s). The most notable achievement of this period was that the Columbia School (e.g., Paul Lazarsfeld, James Coleman, Elihu Katz, Herbert Menzel, and Everett Rogers) studied social diffusion from a network perspective. Third was the take-off period (1970s–1990s), in which Harrison White and his students started the Harvard revolution in SNA that popularized the use of matrix algebra for network analysis. The fourth period was the breakthrough period (1990s–2000s), which saw the development of advanced statistical network models such as the ERGM and the stochastic actor-oriented model (SAOM). The last period is the deepening period (2010s–present), in which SNA has become more interdisciplinary and more engaged with causality, network-based interventions and predictions, big network analysis, text network analysis, and so on.

Network analysis typically employs egocentric or sociocentric research designs (Marsden 2005). In egocentric network designs, researchers often sample subjects (egos) from a large

population and elicit nodes directly connected to the egos (alters). Sociocentric network designs first identify a group of interest and then gather all the ties formed in that group. Researchers typically use observations (e.g., Whyte 1943), surveys (e.g., An 2022a), archives (Padgett & Ansell 1993), or digital or social media (Wimmer & Lewis 2010) to collect network data.

NETWORK FORMATION

In this section, we discuss popular models for modeling sociocentric networks and ego networks and point out potential threats to causal inference and possible solutions.

Modeling Sociocentric Networks

Suppose we have a sociocentric network or, equivalently, an adjacency matrix that describes the ties among actors. For example, if there are five actors, the adjacency matrix, with 1 indicating a tie from the row actor to the column actor and 0 the absence of the tie, is given by $\boldsymbol{w}$:

$$\boldsymbol{w} = \begin{bmatrix} 0\,1\,1\,0\,0 \\ 0\,0\,0\,1\,0 \\ 1\,1\,0\,1\,0 \\ 0\,0\,0\,0\,1 \\ 0\,1\,0\,0\,0 \end{bmatrix}.$$

Given the adjacency matrix, the task is to understand the determinants of tie formation—namely, why some cells in the adjacency are ones while others are zeros.

Theoretical perspectives. Studies in the literature tend to model network formation from three perspectives. First, past research has identified various individual covariates that could affect tie formation, such as age (Marsden 1988, McPherson et al. 2001), sex and gender (Gest et al. 2007, An 2022a), race and ethnicity (Wimmer & Lewis 2010), cultural taste (Lizardo 2006), value and beliefs (Lazarsfeld & Merton 1954), family background (Malacarne 2017, An 2022a), and genetic factors (Fowler et al. 2009). Past studies typically model three types of covariate effects: (*a*) receiver effects (i.e., effects of covariates on incoming ties), (*b*) sender effects (i.e., effects of covariates on outgoing ties), and (*c*) homophily effects (i.e., the tendency to affiliate with similar others) (Laumann 1965; McPherson et al. 2001; Wimmer & Lewis 2010; Leszczensky & Pink 2015, 2017). An (2022a) further shows how to transform these effects into tie-mixing effects to model status differential and differential homophily.

The second perspective focuses on structural opportunities that dictate the likelihood and frequency of two actors' meetings and interactions. For example, Feld (1981, 1982) argues that actors with shared activities/affiliations are more likely to form ties with each other. Small & Adler (2019) argue that spatial context (e.g., spatial propinquity, spatial composition, and spatial configuration) plays an important role in tie formation. Wimmer & Lewis (2010) show that college students living in the same dormitories are more likely to be friends with each other than with those living in different residences. Zeng & Xie (2008) show that group composition can also affect tie formation. In terms of operationalization, most variables describing structural opportunities, such as an individual's affiliation or residence, can be treated as individual covariates.

The third perspective is that ties can be formed due to other existing ties, leading to endogenous tie formation (Goodreau et al. 2009, Kossinets & Watts 2009, Wimmer & Lewis 2010, Block 2015). Prior studies have identified reciprocity (the tendency to reciprocate ties), transitivity (the tendency for an ego to link to alters' alters), preferential attachment (the tendency to link to a popular actor), and differential sociability (the tendency for some actors to send out large numbers of ties), among others, as important endogenous tie formation processes.

Prior research (Wimmer & Lewis 2010, An 2022a) shows that all three perspectives are important elements of a full account of network formation, and ignoring any one could lead to inaccurate understanding of the network formation processes. Hence, depending on the type of ties being studied, researchers should include appropriate individual covariates, structural opportunities, and endogenous tie formation processes to prevent omitted variable bias.

Models and estimation. A simple approach to modeling the (off-diagonal) cells in an adjacency matrix (assuming no self-loops) is to use logistic regression. An example is

$$\begin{aligned}\text{logit}\left[P(w_{ij}=1 \mid \boldsymbol{w}_{ij}^{\boldsymbol{L}}, \boldsymbol{x})\right] = \theta_1 + \ \theta_2 \times x_i + \theta_3 \times x_j + \theta_4 \times I_{x_i=x_j} \\ + \theta_5 \times \text{reciprocity} + \theta_6 \times \text{transitivity}.\end{aligned} \quad 1.$$

This model posits that the chance (i.e., log-odds) of having a tie from i to j, conditioned on $\boldsymbol{w}_{ij}^{\boldsymbol{L}}$ (the local network of the tie w_{ij}), depends on a sender effect from a binary covariate x_i; a receiver effect x_j; a homophily effect $I_{x_i=x_j}$, which indicates whether i and j have the same covariate value; and two endogenous tie formation processes: reciprocity (as measured by the tie status w_{ji}) and transitivity (as measured by the number of common contacts between i and j).

The dyadic logistic regression model can account for Markov dependence in tie formation—namely, the dependence between ties that share a common node. However, this model cannot account for higher-order dependence in tie formation, assuming ties are independent from one another conditional on their local dependence. Because of this nature of the model, the estimation of this model is termed the maximum pseudo-likelihood estimation (MPLE).

ERGMs model the entire adjacency matrix at once without any independence assumption, which makes ERGMs generally preferable to the dyadic logistic regression model. Specifically, an ERGM assumes that the probability of observing a network is as follows (Wasserman & Faust 1994, Handcock et al. 2008, Snijders et al. 2006, Goodreau 2007, Robins et al. 2007):

$$\Pr(\boldsymbol{W}=\boldsymbol{w}|\boldsymbol{X}) = \exp\left\{\boldsymbol{\theta}^{\mathrm{T}} g(\boldsymbol{w},\boldsymbol{X})\right\}/K, \quad 2.$$

where $\boldsymbol{W}$ represents a random network, $\boldsymbol{w}$ the observed network, $\boldsymbol{X}$ the covariates, $g(\boldsymbol{w},\boldsymbol{X})$ a list of modeled effects (also called network statistics) such as covariate effects and endogenous tie formation processes, $\boldsymbol{\theta}$ the model coefficients, and K a normalizing factor to ensure the probabilities sum to one. Hunter et al. (2008) show that the ERGM is equivalent to a conditional logit model in which the log-odds of forming a tie from i to j, conditioning on the rest of the network ($\boldsymbol{w}_{ij}^{c}$), is determined by $\delta_{ij}(\boldsymbol{w},\boldsymbol{X}) = \Delta g(\boldsymbol{w},\boldsymbol{X})$, namely, changes in the network statistics when the i-j tie in the network flips from absence to presence:

$$\text{logit}\left[P(w_{ij}=1|\boldsymbol{w}_{ij}^{c},\boldsymbol{X})\right] = \boldsymbol{\theta}^{\mathrm{T}}\delta_{ij}(\boldsymbol{w},\boldsymbol{X}). \quad 3.$$

ERGMs are typically estimated by the Monte Carlo maximum likelihood estimation (MCMLE) method (Hunter & Handcock 2006).

Scalability of exponential random graph models. Because the MCMLE method relies on Monte Carlo Markov chains for estimation when an ERGM includes endogenous tie formation processes, it is usually slower than the MPLE. This limitation is especially notable when fitting ERGMs on big networks. Moreover, the MCMLE method implicitly assumes that actors form ties with full knowledge of all other ties, which seems less plausible for big networks. Hence, for both computational and theoretical reasons, alternative strategies are needed for fitting ERGMs on big networks. An (2016) presents several strategies for this purpose. They range from speeding up the MCMLE (Calderhead 2014), using the MPLE as an approximation (Wasserman & Pattison

1996, van Duijn et al. 2009), taking the graph limit to approximate the normalizing factor (He & Zheng 2015), and fitting ERGMs on the subnetworks of a big network and then combining the estimates, to other data reduction and model reparameterization approaches.

Measurement error and missing data. Network data may contain various degrees of measurement error and missing data (Marsden 2005, Handcock & Gile 2007, An 2011). For example, measurement error in network ties can arise if respondents report their ties inaccurately (An & Schramski 2015) or forget about some of their ties (Brewer & Webster 2000), or network data are recorded inaccurately. Generally speaking, measurement error in network ties leads to larger than necessary standard errors, similar to measurement error in the dependent variable. One method to reduce measurement error in network ties is to employ multiple reports on the same network ties to triangulate and impute the true value (Krackhardt 1987, An & Schramski 2015). Similarly, An & Doan (2015) and An (2022b) suggested using multiple peer reports to triangulate and correct self-reports of sensitive behaviors to reduce measurement error in self-reported covariates.

Missing ties can arise when, for example, respondents do not share their tie information, respondents in the network are absent, or there is a cap on the maximum number of ties respondents are allowed to report. The first step in dealing with missing ties is to assess the degree of missingness. For example, if a cap is placed on the number of ties respondents can nominate, then researchers can check the proportion of respondents who have nominated the maximum number of ties to gauge the degree of missingness. If that proportion is small, then the degree of missingness is likely small. Second, researchers should distinguish missing ties from non-ties. Non-ties are coded as zeros in an adjacency matrix, but missing ties should be treated as structural zeros in an ERGM to leave them unmodeled. Third, sensitivity analysis can be employed to assess the robustness of the results. An & Doan (2015), for instance, fit ERGMs both on the full network and on the reduced networks, where only the first few nominated ties per respondent were used, to assess whether and how much missing ties impact the results. Lastly, researchers can impute missing ties. For example, one may fit ERGMs on the subnetwork with no missing ties, use the estimates to impute missing ties, and then fit ERGMs on the full network with both observed and imputed ties. This approach (possibly augmented with multiple imputations) may not improve the estimates themselves but can help increase the precision of the estimates.

Omitted variable bias, collinearity, and endogeneity. As with conventional regression models, researchers must include all relevant variables in an ERGM to avoid omitted variable bias in the estimates (Duxbury 2021b). Contextual confounding is a particular concern because context (e.g., neighborhood characteristics) can affect both tie formation and explanatory variables. Meanwhile, including too many variables, especially if they are highly correlated, can lead to multicollinearity problems (Duxbury 2021a).

ERGMs also face an endogeneity issue if tie formation has a reverse effect on the covariates. Suppose happiness is one of the covariates in an ERGM. While happiness may affect tie formation, the reverse could also be true. To identify the causal effect of happiness on tie formation, researchers may need to leverage exogenous variation in happiness through instrumental variables (Clarke & Windmeijer 2012) or experiments (Jiang et al. 2014).

ERGMs are parametric models that require that the function linking explanatory variables to network formation be correctly specified. The estimates could be severely biased if the functional form is misspecified. Hence, it would be valuable to explore nonparametric methods, such as matching to estimate counterfactual causal effects on network formation (e.g., by constructing counterfactual groups with the same covariate values except for the covariate of interest).

Modeling Egocentric Networks

Perry et al. (2018) offer a comprehensive guide to egocentric network analysis, and thus we keep this section brief. Egocentric networks (or ego networks for short) are typically collected via surveys and consist of a focal node (ego) and the nodes to whom ego is directly connected (alters). Therefore, when modeling ego networks using conventional regression analysis, the clustering of alter ties by egos must be taken into account. This can be done using clustered standard errors by egos (Comfort et al. 2021) or using multilevel regression models with egos at the higher level and alters at the lower level (Perry et al. 2018). In some cases, egos also report ties between alters, and thus there are both ego-alter ties and alter-alter ties. Researchers then should cluster standard errors by both egos and alters or use cross-classified models with random effects for both egos and alters (Perry et al. 2018, Snijders & Bosker 1999).

Recently, Krivitsky & Morris (2017) have extended ERGMs to model ego networks. Their approach, however, is currently only applicable to undirected ego networks with no alter-alter ties and does not account for any endogenous tie formation processes due to the data structure.

Ego network data are used to model the size, composition, or other ego network features more often than to model the ego networks themselves. In the former case, the statistical problem reduces to conventional regression analysis, and accordingly, the standard challenges of causal inference in regression analysis naturally apply (e.g., confounding and endogeneity).

Compared with sociocentric network data, ego network data have two major advantages. First, ego network data are more efficient to collect, because researchers do not need to survey all members in a particular group. Second, random samples of egos are easier to draw, and so representative ego network data are easier to collect, as opposed to representative sociocentric network data, where researchers have to randomly sample networks instead of individuals.

However, ego network data also face several limitations. First, many ties may be missing by design. For instance, in many ego network datasets, alter-alter ties, alter-ego ties, and indirect ties (to alters with path length larger than 1) are usually missing by design. This limits the potential of using ego networks to study network statistics that rely on alter-initiated ties (e.g., reciprocity) or full network information (e.g., betweenness and closeness centrality measures, network density). To alleviate this problem, researchers may use research designs that allow egos to report alter-initiated ties. Researchers may also use ego network statistics to approximate or estimate sociocentric network statistics. For example, Marsden (2002) and Everett & Borgatti (2005) show that under certain conditions, egocentric centrality measures (e.g., betweenness) approximate sociocentric measures well. Smith & Gauthier (2020) use sampled ego network data to estimate global network features such as network density. Second, ego network ties are often recorded as undirected, which precludes calculating network statistics that are based on tie direction or analyzing the directional effects of covariates. Third, there is probably more measurement error in ego network ties because the ties reported by egos are typically not validated by alters, and thus egos may be more liberal in reporting their ties.

In **Table 1** (for ERGMs) and **Table 2** (for ego networks), we provide readers with concise summaries of the problems, solutions, and examples mentioned in this section.

NETWORK EFFECTS

We now turn to models that treat networks as predictors to examine how social networks affect individual (or group) outcomes. Three types of network effects can be differentiated: relational effects from possessing certain network ties, positional effects from occupying certain network positions, and structural effects from the overall network structure (VanderWeele & An 2013).

Table 1 Causal inference threats to exponential random graph models and selected solutions

Problem	Selected solution(s)	Examples
Measurement error	1. Improve survey instruments 2. Combine multiple sources of data	An & Schramski (2015) and Krackhardt (1987) use multiple reports to impute the true network ties. An & Doan (2015) and An (2022b) use peer reports to refine the self-reported smoking measure.
Missing ties	1. Distinguish missing ties from non-ties 2. Perform sensitivity analysis 3. Impute missing ties	Missing ties should be treated as structural zeros in ERGMs. Researchers can check if results from ERGMs with randomly removed ties (An 2015a) or based on the first few nominated ties (An & Doan 2015) are similar to the original ERGMs. Researchers can fit an ERGM on the subnetwork with no missing ties and then impute the missing ties based on the estimated ERGM.
Reverse causality	4. Utilize some exogenous variation in the predictor 5. Use lagged predictor	Researchers may use experiments to introduce exogenous variation in the predictor of interest (Jiang et al. 2014).
Multicollinearity	Build the model from simple to complex; include fewer endogenous tie formation processes	Duxbury (2021a) provides examples for diagnosing multicollinearity in ERGMs.
Omitted variables bias	Use instrumental variables or experiments to introduce exogenous variation in the predictor	Clarke & Windmeijer (2012) describe a series of instrumental variables methods for binary outcomes. The methods are generally applicable to ERGMs.

Abbreviation: ERGMs, exponential random graph models.

Table 2 Causal inference threats to ego network models and selected solutions

Problem	Selected solutions	Examples
Clustering by egos when there are only ego-alter ties	1. Cluster standard errors by egos 2. Use multilevel regression models	An example of clustered standard errors by egos is provided by Comfort et al. (2021). Perry et al. (2018) use multilevel models with egos at the higher level and alters at the lower level.
Clustering by egos and alters when there are both ego-alter ties and alter-alter ties	1. Cluster standard errors by both egos and alters 2. Use cross-classified models with random effects for both egos and alters	See Comfort et al. (2021) and An (2022b) for clustering standard errors by both egos and alters. Perry et al. (2018) and Snijders & Bosker (1999) provide examples of cross-classified models.
Missing ties	1. Use research designs that allow egos to report alter-initiated ties 2. Use ego network statistics to approximate or estimate sociocentric network statistics 3. Perform sensitivity analysis	Marsden (2002) and Everett & Borgatti (2005) show that under certain conditions, egocentric centrality measures approximate the sociocentric measures well. Smith & Gauthier (2020) use sampled ego network data to estimate global network features such as network density. Researchers can check the robustness of their results by randomly removing ties (An 2015a) or using only the first few nominated ties (An & Doan 2015).
Other problems, such as measurement error, omitted variable bias, and endogeneity	Similar solutions are shown in **Table 1**.	Examples are shown in **Table 1**.

Relational Effects

The literature presents two popular models for studying relational effects: the social capital model and the social contagion model.

The social capital model. The social capital model examines how social ties provide access to social resources and social support (Cook 2014). Economic sociology has shown that social ties (e.g., knowing someone in a hiring organization) matter for getting a job or higher pay (Granovetter 1973, Bian 1997, Fernandez & Weinberg 1997, Lin 2001, Pinheiro & Dowd 2009, Dowd & Pinheiro 2013), while medical sociology has shown that social ties can provide various forms of social support (or stress) (Song 2011, Thoits 2011, Small 2013, Perry & Pescosolido 2015, Small & Sukhu 2016).

The social capital model typically takes the following form, where y_i represents an outcome of individual i, e.g., wage, and $\boldsymbol{X}_i$ represents of a series of control variables:

$$y_i = \beta_0 + \beta_1 \text{ social capital}_i + \boldsymbol{\beta}_2 \boldsymbol{X}_i + \varepsilon_i. \quad 4.$$

There are three broad critiques of the social capital model. One is how to measure social capital. There have been various measures of social capital in social networks, ranging from the existence of a tie and the number and strength of ties (Bian 1997) to the diversity of ties (Uzzi 1999). Different survey instruments are in use to measure the specific resources available through an ego's network (e.g., name, position, and resource generators) (Molina et al. 2020). Selecting measures that have substantive meaning in a specific context is critical, as different measures may lead to different empirical results and levels of statistical significance.

The second critique is about confounding. People with more social capital may differ from those with less social capital in unobserved ways, which will bias the estimated effect of social capital (Mouw 2006). Exogenous variation in social capital, such as by randomization or through instrumental variables, is necessary to address this issue.

The final critique is about reverse causality. Particularly with cross-sectional data, it is difficult to discern whether social capital causes the outcome or whether the outcome played some role in forming social capital. If longitudinal data are available, a lagged rather than contemporaneous effect of social capital can be estimated to rule out reverse causality (Shalizi & Thomas 2011, VanderWeele et al. 2012). We summarize the critiques of the social capital model and possible solutions in **Table 3**.

The social contagion model. The social contagion (or social influence or peer effect) model is used to examine whether and how social behaviors or other outcomes can transmit through social

Table 3 Causal inference threats to the social capital model and selected solutions

Problem	Selected solution	Examples
Measurement of social capital: different measures can lead to different results	Select a measure that is substantively meaningful and accurately measurable in a specific context	Bian (1997), Uzzi (1999), Lin (2001), Song (2011), and Molina et al. (2020) discuss the measurement of social capital.
Confounding: social capital may be correlated with unobserved covariates	Utilize exogenous variation in social capital created by experiments or instrumental variables	Examples include random assignment of social ties (Carrell et al. 2009, Rohrer et al. 2021, Sacerdote 2001) and quasi-random assignment of social ties (Carvalho et al. 2021, Mohnen 2021, Waldinger 2012).
Reverse causality: the outcome affects social capital	Use lagged social capital as a predictor to rule out reverse causality	Shalizi & Thomas (2011) and VanderWeele et al. (2012) discuss the issue of reverse causality.

ties (Marsden & Friedkin 1993, Friedkin & Johnsen 2011). A popular version of this model is the so-called best friend model, which is used to study whether an ego's best friend's outcome can socially transmit to the ego (Christakis & Fowler 2007, Cohen-Cole & Fletcher 2008, An 2015a). For example, An (2015a) specifies the following model to examine a best friend effect on smoking, where i has nominated j as a best friend:

$$y_i = \beta_0 + \beta_1 y_j + \boldsymbol{\beta}_2 \boldsymbol{X}_i + \boldsymbol{\beta}_3 \boldsymbol{X}_j + \varepsilon_i. \quad 5.$$

In this model, y_i is ego's outcome (smoking status), β_1 is called the endogenous peer effect, which measures direct social transmission of the outcome, $\boldsymbol{\beta}_2$ is the effects of ego's own covariates, and $\boldsymbol{\beta}_3$ is called the exogenous peer effects, which reflect how a best friend's covariates can affect an ego's outcome. The conceptual difference between endogenous and exogenous peer effects is important insofar as the endogenous peer effect can give rise to a social multiplier or ripple effect (Boucher & Fortin 2016).

The social contagion model can be extended in several ways. One way is to use the average outcome of all peers (e.g., all students in a classroom) instead of just the best friend's outcome. This will lead to the so-called linear-in-means model (Manski 1993). Another extension is to use the average outcome of peers to which ego is connected in a network. This will help model the effect of multiple directly connected peers and lead to the spatial autoregressive model or the spatial lag model (Friedkin & Cook 1990, O'Malley & Marsden 2008). The model can also be extended to model the effect of indirectly connected peers in a network (Christakis & Fowler 2007) or the effects of structurally equivalent peers (e.g., peers with similar network connections) (Burt 1987, Moody & Mucha 2013).

The literature on social contagion has focused on estimating the causal endogenous peer effect and points out four broad challenges for causal inference.

1. Confounding due to peer selection: Rather than y_j causing y_i, their correlation is due to i selecting j as a friend based on their similarity in the outcome. The linear-in-means model may additionally suffer from a confounding negative correlation between the outcomes of egos and peers because, by design in this model, egos are not part of their own peer group (Caeyers & Fafchamps 2020).
2. Confounding due to omitted variables: This can include omitting the effects of shared contexts (e.g., neighborhood or school features), omitting the effects of common friends of i and j, and omitting relevant exogenous peer effects.
3. Simultaneity: For mutual friends, y_i can cause y_j while the latter also affects the former. This will lead y_j to be correlated with ε_i and cause simultaneity bias in the estimate of the endogenous peer effect.
4. Measurement error: This can occur both in the outcome and in the network ties.

There are two broad solutions to the issue of confounding in the social contagion model (challenges 1 and 2): (*a*) leveraging exogenous variation or (*b*) correcting for confounding.

In solution *a*, exogenous variation is leveraged through instrumental variables or specially designed experiments. The instrumental variable strategy is to identify an exogenous variable that directly affects y_j and indirectly affects y_i through its effect on y_j. For example, An (2015a) uses the smoking environment at the best friend's home as instrumental variables for the best friend's smoking status. Bramoullé et al. (2009) use the status of the indirect friend in a triad as an instrument for the directly connected friend's status. O'Malley et al. (2014) use the genetic factors of the best friend as instruments for the best friend's outcome. Estrada et al. (2021) and Reza et al. (2021) use exogenous networks to instrument for the endogenous network.

Experiments with special designs can also be used to identify peer effects (An 2011, VanderWeele & An 2013). The first type of design utilizes randomized peer treatment, where researchers randomly assign treatment to alters but not to egos (An 2015b). If the treatment effect is then observed to diffuse from alters to egos, there is proof of peer effects. An issue with this design is that egos might be affected by multiple alters simultaneously, and the treatment of alters might interfere with each other, which makes it difficult to cleanly estimate peer effects. For a strategy to address this issue, readers are directed to An (2018). Another issue with this design is that treatment may, in some cases, alter the underlying social network, possibly leading to a recursive effect on the outcome (An 2015b, Comola & Prina 2021).

The other type of experimental design to identify causal peer effects is the random assignment of peers. Students, for instance, have been randomly assigned to desks (Rohrer et al. 2021), classrooms (Carrell et al. 2009), or dorm rooms (Sacerdote 2001). Furthermore, quasi-randomizations of social ties based on deaths, expulsions, relocations, or natural disasters have been leveraged in prior work (e.g., Waldinger 2012, Carvalho et al. 2021, Mohnen 2021). A limitation of this design is that an ego's outcome can be affected by multiple behaviors of the randomly assigned alters. Hence, to identify casual peer effects of one behavior, researchers need to control for other competing behaviors of the alters. Moreover, shared environmental factors may affect both egos' and alters' behaviors, which are usually not controlled in this design.

The two types of experimental designs estimate different kinds of peer effects. The first one estimates peer effects conditioned on the existent social ties, while treatment in the second one includes not only peers' outcomes but also new social ties.

The second broad solution (solution *b*) to the issue of confounding, which is employed in the absence of exogenous variation, is to use one of several correction strategies. First, to address peer selection, researchers have proposed modeling tie formation in a first stage, e.g., by using ERGMs (An 2011, Hsieh et al. 2020), and then correcting for peer selection in the network effect model (the second stage) using a type of Heckman selection model or treatment selection model (Goldsmith-Pinkham & Imbens 2013, Arduini et al. 2015, Hsieh & Lee 2016) or a matching approach (Aral et al. 2009). Researchers have also used lagged outcomes for both egos and alters to account for peer selection (Christakis & Fowler 2007). Shalizi & Thomas (2011) argue that controlling for lagged outcomes may not be sufficient to eliminate peer selection due to homophily antecedent to the lagged period. Boucher & Fortin (2016) show that with a rich set of controls, the impact of latent homophily may be small.

When longitudinal network data are available, another approach to account for peer selection is to jointly model the evolution of networks and behavior over time with the SAOM (Snijders 2001, 2011; Steglich et al. 2010). SAOM models network and behavior dynamics with two stochastic processes. The network process is akin to a dynamic ERGM which can account for peer selection (e.g., based on homophily in a behavioral outcome). The behavioral process is akin to the social contagion model in which exogenous and endogenous peer effects can be specified. In this way, SAOM structurally separates peer selection from peer influence by jointly estimating their effects. Prior work has examined the statistical properties of SAOM (Block et al. 2018, Leifeld & Cranmer 2019) and applied it in various contexts (Schaefer et al. 2012, An 2015b). Despite its capabilities, SAOM has several limitations: It is computationally expensive as it relies on Monte Carlo Markov chains for estimation, it makes strong behavioral assumptions (e.g., that actors have full information of the network when making decisions), and omitted variable bias (e.g., due to contextual confounding) can still bias the estimates.

To address contextual confounding in observational data, researchers have proposed controlling for the outcomes of neighbors (Christakis & Fowler 2007) or neighborhood features (Block et al. 2011) or using neighborhood/school fixed effects (Cohen-Cole & Fletcher 2008, Fortin &

Yazbeck 2015). Some researchers criticize the fixed effects model because the within estimator can generate correlations between the lagged outcomes and the transformed disturbance (Nickell 1981), which may bias the endogenous peer effect estimate.

To circumvent the simultaneity problem (challenge 3), researchers can estimate a lagged rather than a contemporaneous peer effect or utilize exogenous variation in peers' behaviors (as provided by instrumental variables or experiments). In the linear-in-means model, the problem of simultaneity additionally causes the reflection problem (Manski 1993), namely, the impossibility of separating endogenous peer effects from exogenous peer effects because the mean of peer outcomes and the mean of peer covariates are linearly dependent. Several solutions to the reflection problem have been proposed, including utilizing variation in group size (Lee 2007, Davezies et al. 2009), utilizing variation in individual-specific network structure (Bramoullé et al. 2009), or imposing parameter constraints (Graham 2008, Arcidiacono et al. 2012, Rose 2017).

Measurement error (challenge 4) in the outcome can be doubly detrimental as the outcome appears on both sides of the model (An 2015a). It may not only attenuate the estimated peer effect but also reduce its precision, making it more difficult to reject the null of no peer effect. Measurement error in network ties can also bias the peer effect estimate (Advani & Malde 2018a, Micklewright et al. 2012). To reduce measurement error, the general advice is to improve the measurement itself (e.g., by refining the survey instruments) or to combine measures from separate reports (An & Doan 2015, An & Schramski 2015, An 2022b).

There are still several open issues in this area of study. In terms of substantive progress, more work is needed to uncover the heterogeneity of peer effects, such as by ego and alter characteristics (e.g., race and gender) and by ego-alter dyadic characteristics (e.g., friendship duration and strength) (Cools et al. 2019). More work is also needed to identify the mechanisms underlying peer effects (Lin 2001, DiMaggio & Garip 2012, An 2015a). In terms of methodological progress, more work is needed to provide standard errors that can account for outcome dependence across units beyond social contagion (Advani & Malde 2018b, An 2018, 2021, Lee & Ogburn 2020). Finally, more research is also needed to explore nonparametric identification of network effects (Egami 2020; Ogburn et al. 2020a,b).

The threats to causal inference and possible solutions with the social contagion model are summarized in **Table 4**.

Positional Effects

Positional effects refer to whether and to what degree the position that individuals occupy in a network affects their outcomes. Prior work has identified popularity and isolation effects (Schaefer et al. 2010, Copeland et al. 2017) and effects of brokerage positions that bridge units in social networks (Gould & Fernandez 1989, Burt 1995, Cornwell 2009, Everett & Valente 2016, Greenberg 2021), to give a few examples.

In terms of modeling, studying the causal effects of network positions amounts to replacing the social capital indicators in the social capital model with positional measures. Accordingly, network endogeneity (e.g., individuals occupying special positions in a network select into these positions) and contextual confounding (e.g., variables that determine both network positions and the outcome are omitted) are the two major obstacles of causally identifying positional effects.

Structural Effects

Individual outcomes can also be influenced by the macro features of a network, such as cohesion, hierarchy, clustering, and composition. Cohesion describes how densely connected a network is

Table 4 Causal inference threats to the social contagion model and selected solutions

Problem	Selected solutions	Examples
Confounding due to peer selection	1. Utilize exogenous variation in peers' outcomes provided by instrumental variables or experiments 2. Model peer selection and control for it in the social contagion model	The instrumental variables approach is discussed by Bramoullé et al. (2009), O'Malley et al. (2014), and An (2015a). Randomized peer treatment is discussed by An (2011, 2015b). For random assignment of peers, see Sacerdote (2001), Carrell et al. (2009), and Rohrer et al. (2021). Examples of quasi-randomizations of peers are provided by Waldinger (2012), Carvalho et al. (2021), and Mohnen (2021). Correction approaches are proposed by Christakis & Fowler (2007), Aral et al. (2009), Steglich et al. (2010), Snijders (2011), and Hsieh & Lee (2016).
Confounding due to omitted variables (e.g., contextual factors)	1. Utilize exogenous variation in peers' behaviors provided by instrumental variables or experiments 2. Control for contextual confounding directly or statistically	Instrumental variable and experimental approaches are shown in the top row of this table. To address contextual confounding, researchers have proposed controlling for neighbors' outcomes (Christakis & Fowler 2007) or neighborhood features (Block et al. 2011) or using neighborhood/school fixed effects (Cohen-Cole & Fletcher 2008, Fortin & Yazbeck 2015).
Simultaneity bias	1. Utilize exogenous variation in peers' outcomes provided by instrumental variables or experiments 2. Use lagged peers' behaviors	Instrumental variable and experimental approaches are shown in the top row of this table. The lagged peer effect strategy is discussed by Yeung & Nguyen-Hoang (2016) and VanderWeele et al. (2012).
Measurement error	1. Improve survey instruments 2. Combine multiple sources of data	An & Schramski (2015) and Krackhardt (1987) use multiple reports to impute the true network ties. An & Doan (2015) and An (2022b) use peer reports to refine the self-reported smoking status.

and is usually measured by the density coefficient or mean geodistance (i.e., the shortest path between two nodes) of a network (Moody & White 2003). Hierarchy describes the extent to which network ties are unevenly distributed across actors and is typically measured by the centralization coefficient and reciprocity coefficient of a network. For example, prior work has shown that more centralized networks facilitate simple diffusion that requires only one-time exposure for adoption (Barrat et al. 2008). Clustering describes the degree to which network ties are concentrated in local actors and is measured by the global or local clustering coefficients of a network. It has been shown that local clustering facilitates complex diffusion that requires reinforcement from multiple contacts for adoption (Centola 2010). Finally, network compositions (e.g., in terms of race, sex, age, and other factors) can also affect individual outcomes (DiMaggio & Garip 2012, McFarland et al. 2014).

Randomization tests have been used to assess whether a network exhibits certain structural features. This can be done by randomly permuting the rows and/or columns of the adjacency matrix and comparing selected features of the observed network against the randomized networks (Bearman et al. 2004, Baldassarri & Diani 2007).

To study causal structural effects, one may use the structural features to replace the social capital indicators in the social capital model. Hence, similar critiques apply. In particular, contextual confounding and reverse causality (i.e., individuals affect the network structures) are of concern. To improve statistical inferences, one may use a multivariate regression model to account for outcome correlations across units (An 2018, 2021).

CONCLUSION AND DISCUSSION

Empirical models and methods for studying how social networks are formed and how social networks influence individual outcomes have made tremendous progress in the past decade. However, because of their technical nature, some of the models and methods are currently only accessible to methodologists, hindering their acceptance and application in the board research community. One goal of this paper is to help disseminate the knowledge at an accessible level.

We identify four areas that are critical for future development of causal network analysis. First, more accurate measurement of network ties and individual outcomes is crucial to research on both network formation and network effects. More precision in determining the meaning of ties (Kitts & Leal 2021) and the use of multiple sources of data, including reports from multiple respondents, timed interaction data, social media data, and multiplex tie information, can help improve the measurement (An & Doan 2015, An & Schramski 2015, Kitts & Quintane 2020, An 2022b). Moreover, future research should consider employing sensitivity analysis (VanderWeele 2011, An 2015a, An & Doan 2015) to address uncertain or missing ties (Fisher 2019).

Second, more work is needed to understand the mechanisms underlying network formation and network effects. Regarding network formation mechanisms, Wimmer & Lewis (2010) argue that endogenous tie formation processes can be either competing or mediating mechanisms for covariate effects. Future work is needed to distinguish between the two types of mechanisms. Moreover, the mechanisms underlying tie dissolution should be examined more as they can differ from tie formation mechanisms (McDermott et al. 2013). Regarding mechanisms for network effects, An (2015a), for instance, shows that sharing cigarettes can be one potential mechanism for peer influence on smoking. More of such work will aid our understanding of how behavioral imitation, normative pressure, and social exchange can explain social contagion. In addition, future research may study network effects by including relational, positional, and structural effects in one integrated model to examine how these effects simultaneously operate.

Third, more experimental work (conducted in a lab, in the field, or online) is needed to better identify causality in network analysis and triangulate findings across different settings. Network experiments may also be specifically designed to improve social interventions (An 2015b).

Lastly, more work is necessary to model and explain the tremendous heterogeneity in network formation and network effects. This includes allowing modeled effects to change across time, space, covariate values, different types of ties, and different regions in a network. Theories on network ecology (Entwisle et al. 2007, Small 2007, McFarland et al. 2014) may help researchers to explore certain causes of such heterogeneity.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

AUTHOR CONTRIBUTIONS

W.A. planned the article and revised the first draft, R.B. drafted the section titled Network Formation, and B.R. drafted the section titled Network Effects. All authors participated in later revisions.

ACKNOWLEDGMENTS

We thank Mario Small for the invitation to contribute and *Annual Review of Sociology* Co-Editor Karen Cook for helpful guidance on the manuscript. We also thank Kenneth Bollen, Bet Caeyers,

Ben Cornwell, Felix Elwert, Marcel Fafchamps, James Moody, Alison Schultz, and the anonymous reviewer for helpful comments.

LITERATURE CITED

Advani A, Malde B. 2018a. Credibly identifying social effects: accounting for network formation and measurement error. *J. Econ. Surv.* 32(4):1016–44

Advani A, Malde B. 2018b. Methods to identify linear network models: a review. *Swiss J. Econ. Surv. Stat.* 154:12

An W. 2011. Models and methods to identify peer effects. In *The SAGE Handbook of Social Network Analysis*, ed. J Scott, PJ Carrington, pp. 514–32. London: SAGE

An W. 2015a. Instrumental variables estimates of peer effects in social networks. *Soc. Sci. Res.* 50:382–94

An W. 2015b. Multilevel meta network analysis with application to studying network dynamics of network interventions. *Soc. Netw.* 43:48–56

An W. 2016. Fitting ERGMS on big networks. *Soc. Sci. Res.* 59:107–19

An W. 2018. Causal inference with networked treatment diffusion. *Sociol. Methodol.* 48(1):152–81

An W. 2021. A tale of twin dependence: a new multivariate regression model and an FGLS estimator for analyzing outcomes with network dependence. *Sociol. Methods Res.* **https://doi.org/10.1177/00491241211031263**

An W. 2022a. Friendship network formation in Chinese middle schools: patterns of inequality and homophily. *Soc. Netw.* 68:218–28

An W. 2022b. You said, they said: a framework on informant accuracy with application to studying self-reports and peer-reports. *Soc. Netw.* 70:187–97

An W, Doan L. 2015. Health surveillance through social networks. *Soc. Netw.* 42:8–17

An W, Schramski S. 2015. Analysis of contested reports in exchange networks based on actors' credibility. *Soc. Netw.* 40(1):25–33

Angrist JD, Pischke J. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton, NJ: Princeton Univ. Press

Aral S, Muchnik L, Sundararajan A. 2009. Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *PNAS* 106(51):21544–49

Arcidiacono P, Foster G, Goodpaster N, Kinsler J. 2012. Estimating spillovers using panel data, with an application to the classroom. *Quant. Econ.* 3(3):421–70

Arduini T, Patacchini E, Rainone E. 2015. *Parametric and semiparametric IV estimation of network models with selectivity*. EIEF Work. Pap. 1509, Einaudi Inst. Econ. Finance, Rome

Baldassarri D, Diani M. 2007. The integrative power of civic networks. *Am. J. Sociol.* 113(3):735–80

Barrat A, Barthelemy M, Vespignani A. 2008. *Dynamical Processes on Complex Networks*. Cambridge, UK: Cambridge Univ. Press

Bearman PS, Moody J, Stovel K. 2004. Chains of affection: the structure of adolescent romantic and sexual networks. *Am. J. Sociol.* 110(1):44–91

Bian Y. 1997. Bringing strong ties back in: indirect ties, network bridges, and job searches in China. *Am. Sociol. Rev.* 62(3):366–85

Block JP, Christakis NA, O'Malley AJ, Subramanian SV. 2011. Proximity to food establishments and body mass index in the Framingham Heart Study offspring cohort over 30 years. *Am. J. Epidemiol.* 174(10):1108–14

Block P. 2015. Reciprocity, transitivity, and the mysterious three-cycle. *Soc. Netw.* 40:163–73

Block P, Koskinen J, Hollway J, Steglich C, Stadtfeld C. 2018. Change we can believe in: comparing longitudinal network models on consistency, interpretability, and predictive power. *Soc. Netw.* 52:180–91

Borgatti SP, Mehra A, Brass DJ, Labianca G. 2009. Network analysis in the social sciences. *Science* 323(5916):892–95

Boucher V, Fortin B. 2016. Some challenges in the empirics of the effects of networks. In *The Oxford Handbook of the Economics of Networks*, ed. Y Bramoullé, A Galeotti, BW Rogers, pp. 276–302. Oxford, UK: Oxford Univ. Press

Bramoullé Y, Djebbari H, Fortin B. 2009. Identification of peer effects through social networks. *J. Econom.* 150(1):41–55

Bramoullé Y, Djebbari H, Fortin B. 2020. Peer effects in networks: a survey. *Annu. Rev. Econ.* 12:603–29
Brewer DD, Webster CM. 2000. Forgetting of friends and its effects on measuring friendship networks. *Soc. Netw.* 21(4):361–73
Burt RS. 1987. Social contagion and innovation: cohesion versus structural equivalence. *Am. J. Sociol.* 92(6):1287–335
Burt RS. 1995. *Structural Holes: The Social Structure of Competition*. Cambridge, MA: Harvard Univ. Press
Caeyers B, Fafchamps M. 2020. *Exclusion bias in the estimation of peer effects*. CEPR Discuss. Pap. 14386, Cent. Econ. Policy Res., London
Calderhead B. 2014. A general construction for parallelizing Metropolis–Hastings algorithms. *PNAS* 111(49):17408–13
Carrell SE, Fullerton RL, West JE. 2009. Does your cohort matter? Measuring peer effects in college achievement. *J. Labor Econ.* 27(3):439–64
Carvalho VM, Nirei M, Saito YU, Tahbaz-Salehi A. 2021. Supply chain disruptions: evidence from the great East Japan earthquake. *Q. J. Econ.* 136(2):1255–321
Centola D. 2010. The spread of behavior in an online social network experiment. *Science* 329(5996):1194–97
Christakis NA, Fowler JH. 2007. The spread of obesity in a large social network over 32 years. *N. Engl. J. Med.* 357(4):370–79
Clarke PA, Windmeijer F. 2012. Instrumental variable estimators for binary outcomes. *J. Am. Stat. Assoc.* 107:1638–52
Cohen-Cole E, Fletcher JM. 2008. Is obesity contagious? Social networks versus environmental factors in the obesity epidemic. *J. Health Econ.* 27(5):1382–87
Comfort AB, Harper CC, Tsai AC, Perkins JM, Moody J, et al. 2021. The association between men's family planning networks and contraceptive use among their female partners: an egocentric network study in Madagascar. *BMC Public Health* 21(1):209
Comola M, Prina S. 2021. Treatment effect accounting for network changes. *Rev. Econ. Stat.* 103(3):597–604
Cook KS. 2014. Social capital and inequality: the significance of social connections. In *Handbook of the Social Psychology of Inequality*, ed. J McLeod, E Lawler, M Schwalbe, pp. 207–27. Dordrecht, Neth.: Springer
Cools A, Fernández R, Patacchini E. 2019. *Girls, boys, and high achievers.* NBER Work. Pap. 25763
Copeland M, Bartlett B, Fisher JC. 2017. Dynamic associations of network isolation and smoking behavior. *Netw. Sci.* 5(3):257–77
Cornwell B. 2009. Good health and the bridging of structural holes. *Soc. Netw.* 31(1):92–103
Davezies L, D'Haultfoeuille X, Fougère D. 2009. Identification of peer effects using group size variation. *Econom. J.* 12(3):397–413
DiMaggio P, Garip F. 2012. Network effects and social inequality. *Annu. Rev. Sociol.* 38:93–118
Dowd TJ, Pinheiro DL. 2013. The ties among the notes: the social capital of jazz musicians in three metro areas. *Work Occup.* 40(4):431–64
Duxbury SW. 2021a. Diagnosing multicollinearity in exponential random graph models. *Sociol. Methods Res.* 50(2):491–530
Duxbury SW. 2021b. The problem of scaling in exponential random graph models. *Sociol. Methods Res.* **https://doi.org/10.1177/0049124120986178**
Egami N. 2020. Identification of causal diffusion effects under structural stationarity. arXiv:1810.07858 [stat.ME]
Entwisle B, Faust K, Rindfuss RR, Kaneda T. 2007. Networks and contexts: variation in the structure of social ties. *Am. J. Sociol.* 112 (5):1495–533
Estrada J, Huynh KP, Jacho-Chavez DT, Sanchez-Aragon L. 2021. *On the identification and estimation of endogenous peer effects in multiplex networks*. Work. Pap., Dep. Econ., Emory Univ., Atlanta, GA. **https://kphuynh.pages.iu.edu/rsch/multiplex_ecomt.pdf**
Everett M, Borgatti SP. 2005. Ego network betweenness. *Soc. Netw.* 27(1):31–38
Everett MG, Valente TW. 2016. Bridging, brokerage and betweenness. *Soc. Netw.* 44:202–8
Feld SL. 1981. The focused organization of social ties. *Am. J. Sociol.* 86(5):1015–35
Feld SL. 1982. Social structural determinants of similarity among associates. *Am. Sociol. Rev.* 47(6):797–801
Fernandez RM, Weinberg N. 1997. Sifting and sorting: personal contacts and hiring in a retail bank. *Am. Sociol. Rev.* 62(6):883–902

Fisher JC. 2019. Social space diffusion: applications of a latent space model to diffusion with uncertain ties. *Sociol. Methodol.* 49(1):258–94
Fortin B, Yazbeck M. 2015. Peer effects, fast food consumption and adolescent weight gain. *J. Health Econ.* 42:125–38
Fowler JH, Dawes CT, Christakis NA. 2009. Model of genetic variation in human social networks. *PNAS* 106(6):1720–24
Freeman LC. 2004. *The Development of Social Network Analysis. A Study in the Sociology of Science*. Vancouver, BC: Empirical
Friedkin NE, Cook KS. 1990. Peer group influence. *Sociol. Methods Res.* 19(1):122–43
Friedkin NE, Johnsen EC. 2011. *Social Influence Network Theory*. Cambridge, UK: Cambridge Univ. Press
Gest SD, Davidson AJ, Rulison KL, Moody J, Welsh JA. 2007. Features of groups and status hierarchies in girls' and boys' early adolescent peer networks. *New Dir. Child Adolesc. Dev.* 2007(118):43–60
Goldsmith-Pinkham P, Imbens GW. 2013. Social networks and the identification of peer effects. *J. Bus. Econ. Stat.* 31(3):253–64
Goodreau SM. 2007. Advances in exponential random graph (p^*) models applied to a large social network. *Soc. Netw.* 29(2):231–48
Goodreau SM, Kitts JA, Morris M. 2009. Birds of a feather, or friend of a friend? Using exponential random graph models to investigate adolescent social networks. *Demography* 46(1):103–25
Gould RV, Fernandez RM. 1989. Structures of mediation: a formal approach to brokerage in transaction networks. *Sociol. Methodol.* 19:89–126
Graham BS. 2008. Identifying social interactions through conditional variance restrictions. *Econometrica* 76(3):643–60
Granovetter MS. 1973. The strength of weak ties. *Am. J. Sociol.* 78(6):1360–80
Greenberg J. 2021. Social network positions, peer effects, and evaluation updating: an experimental test in the entrepreneurial context. *Organ. Sci.* 32(5):1174–92
Handcock MS, Gile K. 2007. Modeling social networks with sampled or missing data. arXiv:1010.0891 [stat.AP]
Handcock MS, Hunter DR, Butts CT, Goodreau SM, Morris M. 2008. statnet: Software tools for the representation, visualization, analysis and simulation of network data. *J. Stat. Softw.* 24:1–11
He R, Zheng T. 2015. GLMLE: graph-limited enabled fast computation for fitting exponential random graph models to large social networks. *Soc. Netw. Anal. Min.* 5:8
Hernán MA, Robins JM. 2020. *Causal Inference: What If*. Boca Raton, FL: Chapman and Hall/CRC
Hsieh CS, Lee L. 2016. A social interactions model with endogenous friendship formation and selectivity. *J. Appl. Econom.* 31(2):301–19
Hsieh CS, Lee L, Boucher V. 2020. Specification and estimation of network formation and network interaction models with the exponential probability distribution. *Quant. Econ.* 11(4):1349–90
Hunter DR, Handcock MS. 2006. Inference in curved exponential family models for networks. *J. Comput. Graph. Stat.* 15(3):565–83
Hunter DR, Handcock MS, Butts CT, Goodreau SM, Morris M. 2008. ergm: A package to fit, simulate and diagnose exponential-family models for networks. *J. Stat. Softw.* 24(3):1–29
Imbens GW, Rubin DB. 2015. *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. Cambridge, UK: Cambridge Univ. Press
Jiang S, Gao Q, Chen H, Roco MC. 2014. The roles of sharing, transfer, and public funding in nanotechnology knowledge-diffusion networks. *J. Assoc. Inf. Sci. Technol.* 66(5):1017–29
Kitts JA, Leal DF. 2021. What is(n't) a friend? Dimensions of the friendship concept among adolescents. *Soc. Netw.* 66:161–70
Kitts JA, Quintane E. 2020. Rethinking social networks in the era of computational social science. In *Oxford Handbook of Social Networks*, ed. R Light, J Moody, pp. 71–97. Oxford, UK: Oxford Univ. Press
Kossinets G, Watts DJ. 2009. Origins of homophily in an evolving social network. *Am. J. Sociol.* 115(2):405–50
Krackhardt D. 1987. Cognitive social structures. *Soc. Netw.* 9(2):109–34
Krivitsky PN, Morris M. 2017. Inference for social network models from egocentrically sampled data, with application to understanding persistent racial disparities in HIV prevalence in the US. *Ann. Appl. Stat.* 11(1):427–55

Laumann EO. 1965. Subjective social distance and urban occupational stratification. *Am. J. Sociol.* 71(1):26–36
Lazarsfeld PF, Merton RK. 1954. Friendship as a social process: a substantive and methodological analysis. In *Freedom and Control in Modern Society*, ed. M Berger, T Abel, CH Page, pp. 18–66. New York: Van Nostrand
Lee L. 2007. Identification and estimation of econometric models with group interactions, contextual factors and fixed effects. *J. Econom.* 140(2):333–74
Lee Y, Ogburn EL. 2020. Network dependence can lead to spurious associations and invalid inference. arXiv:1908.00520 [stat.AP]
Leifeld P, Cranmer SJ. 2019. A theoretical and empirical comparison of the temporal exponential random graph model and the stochastic actor-oriented model. *Netw. Sci.* 7(1):20–51
Leszczensky L, Pink S. 2015. Ethnic segregation of friendship networks in school: testing a rational-choice argument of differences in ethnic homophily between classroom- and grade-level networks. *Soc. Netw.* 42:18–26
Leszczensky L, Pink S. 2017. Intra- and inter-group friendship choices of Christian, Muslim, and non-religious youth in Germany. *Eur. Sociol. Rev.* 33(1):72–83
Lin N. 2001. Building a network theory of social capital. In *Social Capital*, ed. N Lin, K Cook, RS Burt, pp. 3–30. London: Routledge
Lizardo O. 2006. How cultural tastes shape personal networks. *Am. Sociol. Rev.* 71(5):778–807
Lusher D, Koskinen J, Robins G. 2013. *Exponential Random Graph Models for Social Networks: Theory, Methods, and Applications*. Cambridge, UK: Cambridge Univ. Press
Malacarne T. 2017. Rich friends, poor friends: inter-socioeconomic status friendships in secondary school. *Socius* 3:237802311773699
Manski CF. 1993. Identification of endogenous social effects: the reflection problem. *Rev. Econ. Stud.* 60(3):531–42
Marsden PV. 1988. Homogeneity in confiding relations. *Soc. Netw.* 10(1):57–76
Marsden PV. 2002. Egocentric and sociocentric measures of network centrality. *Soc. Netw.* 24(4): 407–22
Marsden PV. 2005. Recent developments in network measurements. In *Models and Methods in Social Network Analysis*, ed. PJ Carrington, J Scott, S Wasserman, pp. 8–30. Cambridge, UK: Cambridge Univ. Press
Marsden PV, Friedkin NE. 1993. Network studies of social influence. *Sociol. Methods Res.* 22(1):127–51
McDermott R, Fowler JH, Christakis NA. 2013. Breaking up is hard to do, unless everyone else is doing it too: social network effects on divorce in a longitudinal sample. *Soc. Forces* 92(2):491–519
McFarland DA, Moody J, Diehl D, Smith JA, Thomas RJ. 2014. Network ecology and adolescent social structure. *Am. Sociol. Rev.* 79(6):1088–121
McPherson M, Smith-Lovin L, Cook JM. 2001. Birds of a feather: homophily in social networks. *Annu. Rev. Sociol.* 27:415–44
Micklewright J, Schnepf SV, Silva PN. 2012. Peer effects and measurement error: the impact of sampling variation in school survey data (evidence from PISA). *Econ. Educ. Rev.* 31(6):1136–42
Mohnen M. 2021. Stars and brokers: knowledge spillovers among medical scientists. *Manag. Sci.* 68(4):2513–32
Molina JL, García-Macías A, Lubbers MJ, Valenzuela-Garcia H. 2020. The embeddedness of social capital in personal networks. *Netw. Sci.* 8(2):189–203
Moody J, Mucha PJ. 2013. Portrait of political party polarization. *Netw. Sci.* 1(1):119–21
Moody J, White DR. 2003. Structural cohesion and embeddedness: a hierarchical concept of social groups. *Am. Sociol. Rev.* 68(1):103–27
Moreno JL. 1934. *Who Shall Survive? A New Approach to the Problem of Human Interrelations*. Washington, DC: Nerv. Ment. Dis. Publ.
Morgan SL, Winship C. 2014. *Counterfactuals and Causal Inference: Methods and Principles for Social Research*. Cambridge, UK: Cambridge Univ. Press
Mouw T. 2006. Estimating the causal effect of social capital: a review of recent research. *Annu. Rev. Sociol.* 32:79–102
Nickell S. 1981. Biases in dynamic models with fixed effects. *Econometrica* 49(6):1417–26
Ogburn EL, Shpitser I, Lee Y. 2020a. Causal inference, social networks and chain graphs. *J. R. Stat. Soc. Ser. A* 183(4):1659–76

Ogburn EL, Sofrygin O, Diaz I, van der Laan MJ. 2020b. Causal inference for social network data. arXiv:1705.08527 [stat.ME]
O'Malley AJ, Elwert F, Rosenquist JN, Zaslavsky AM, Christakis NA. 2014. Estimating peer effects in longitudinal dyadic data using instrumental variables. *Biometrics* 70(3):506–15
O'Malley AJ, Marsden PV. 2008. The analysis of social networks. *Health Servic. Outcomes Res. Methodol.* 8(4):222–69
Padgett JF, Ansell CK. 1993. Robust action and the rise of the Medici, 1400–1434. *Am. J. Sociol.* 98(6):1259–319
Perry BL, Pescosolido BA. 2015. Social network activation: the role of health discussion partners in recovery from mental illness. *Soc. Sci. Med.* 125:116–28
Perry BL, Pescosolido BA, Borgatti SP. 2018. *Egocentric Network Analysis: Foundations, Methods, and Models*. Cambridge, UK: Cambridge Univ. Press
Pinheiro DL, Dowd TJ. 2009. All that jazz: the success of jazz musicians in three metropolitan areas. *Poetics* 37(5–6):490–506
Reza S, Manchanda P, Chong JK. 2021. Identification and estimation of endogenous peer effects using partial network data from multiple reference groups. *Manag. Sci.* 67(8):5070–105
Robins G, Snijders TAB, Wang P, Handcock MS, Pattison P. 2007. Recent developments in exponential random graph (p^*) models for social networks. *Soc. Netw.* 29(2):192–215
Rohrer JM, Keller T, Elwert F. 2021. Proximity can induce diverse friendships: a large randomized classroom experiment. *PLOS ONE* 16(8):e0255097
Rose CD. 2017. Identification of peer effects through social networks using variance restrictions. *Econom. J.* 20(3):S47–60
Sacerdote B. 2001. Peer effects with random assignment: results for Dartmouth roommates. *Q. J. Econ.* 116(2):681–704
Schaefer DR, Haas SA, Bishop NJ. 2012. A dynamic model of US adolescents' smoking and friendship networks. *Am. J. Public Health* 102(6):e12–18
Schaefer DR, Light JM, Fabes RA, Hanish LD, Martin CL. 2010. Fundamental principles of network formation among preschool children. *Soc. Netw.* 32(1):61–71
Scott L. 1991. *Social Network Analysis: A Handbook*. London, UK: SAGE
Shalizi CR, Thomas AC. 2011. Homophily and contagion are generically confounded in observational social network studies. *Sociol. Methods Res.* 40(2):211–39
Small ML. 2007. Racial differences in networks: Do neighborhood conditions matter? *Soc. Sci. Q.* 88(2):320–43
Small ML. 2013. Weak ties and the core discussion network: why people regularly discuss important matters with unimportant alters. *Soc. Netw.* 35(3):470–83
Small ML, Adler L. 2019. The role of space in the formation of social ties. *Annu. Rev. Sociol.* 45:111–32
Small ML, Sukhu C. 2016. Because they were there: access, deliberation, and the mobilization of networks for support. *Soc. Netw.* 47:73–84
Smith JA, Gauthier GR. 2020. Estimating contextual effects from ego network data. *Sociol. Methodol.* 50(1):215–75
Snijders TAB. 2001. The statistical evaluation of social network dynamics. *Sociol. Methodol.* 31(1):361–95
Snijders TAB. 2011. Statistical models for social networks. *Annu. Rev. Sociol.* 37:131–53
Snijders TAB, Bosker RJ. 1999. *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*. London: SAGE
Snijders TAB, Pattison PE, Robins GL, Handcock MS. 2006. New specifications for exponential random graph models. *Sociol. Methodol.* 36(1):99–153
Song L. 2011. Social capital and psychological distress. *J. Health Soc. Behav.* 52(4):478–92
Steglich C, Snijders TAB, Pearson M. 2010. Dynamic networks and behavior: separating selection from influence. *Sociol. Methodol.* 40(1):329–93
Thoits PA. 2011. Mechanisms linking social ties and support to physical and mental health. *J. Health Soc. Behav.* 52(2):145–61
Uzzi B. 1999. Embeddedness in the making of financial capital: how social relations and networks benefit firms seeking financing. *Am. Sociol. Rev.* 64(4):481–505

van Duijn MAJ, Gile KJ, Handcock MS. 2009. A framework for the comparison of maximum pseudo-likelihood and maximum likelihood estimation of exponential family random graph models. *Soc. Netw.* 31(1):52–62
VanderWeele TJ. 2011. Sensitivity analysis for contagion effects in social networks. *Sociol. Methods Res.* 40(2):240–55
VanderWeele TJ, An W. 2013. Social networks and causal inference. In *Handbook of Causal Analysis for Social Research*, ed. S Morgan, pp. 353–74. New York: Springer
VanderWeele TJ, Ogburn EL, Tchetgen Tchetgen EJ. 2012. Why and when "flawed" social network analyses still yield valid tests of no contagion. *Stat. Politics Policy* 3(1):2151–60
Waldinger F. 2012. Peer effects in science: evidence from the dismissal of scientists in Nazi Germany. *Rev. Econ. Stud.* 79:838–61
Wasserman S, Faust K. 1994. *Social Network Analysis: Methods and Applications*. Cambridge, UK: Cambridge Univ. Press
Wasserman S, Pattison P. 1996. Logit models and logistic regressions for social networks: I. An introduction to Markov graphs and p^*. *Psychometrika* 61(3):401–25
Whyte WF. 1943. *Street Corner Society: The Social Structure of an Italian Slum*. Chicago: Univ. Chicago Press
Wimmer A, Lewis K. 2010. Beyond and below racial homophily: ERG models of a friendship network documented on Facebook. *Am. J. Sociol.* 116(2):583–642
Yang S, Keller FB, Zheng L. 2016. *Social Network Analysis: Methods and Examples*. London: SAGE
Yeung R, Nguyen-Hoang P. 2016. Endogenous peer effects: fact or fiction? *J. Educ. Res.* 109(1):37–49
Zeng Z, Xie Y. 2008. A preference-opportunity-choice framework with applications to intergroup friendship. *Am. J. Sociol.* 114(3):615–48