34

Models and Methods to Identify Peer Effects

Weihua (Edward) An

INTRODUCTION

There have been numerous studies on peer effects. But inadequate attention has been paid to making valid causal inference about peer effects, either due to intellectual negligence or methodological limitations. This paper will review recent advances in statistical modeling and inference on peer effects and point out some directions for future research in this area. In sociology, the literature on methodological issues in studying peer effects can be traced back at least to the classic paper written by Duncan et al. (1968). Other notable studies and reviews include Kandel (1978), Marsden and Friedkin (1993), Doreian (2001), Carrington et al. (2005), Valente (2005), O'Malley and Marsden (2008), Smith and Christakis (2008), and so forth. There also exist some very good reviews on this topic in economics, including Manski (1993, 2000, 2010), Brock and Durlauf (2001a), Blume and Durlauf (2005), Soetevent (2006), Hartmann et al. (2008), Jackson (2008) and Moffitt (forthcoming); in political science, including Fowler, Heaney et al. (2009); and in physics and statistics, including Albert and Barabási (2002), Newman (2003), Goldenberg et al. (2009), and Kolaczyk (2009).

Generally speaking, methodological studies of peer effects can be divided into two different approaches. One is about mathematical modeling of peer effects, like Jackson (2008, esp. Chapter 8), which cares more about the long run behaviors or equilibria of the social interactions of peer groups. The other is the statistical identification and estimation of peer effects. This review will mostly focus on the second approach, especially the identification issues. In addition, this review differs from the above ones in two perspectives. First, it is interdisciplinary, drawing on literature not only from sociology but also from economics, political science, statistics, and so on. Second, it uses potential outcomes framework to unite and elaborate its critiques and emphasizes the conditions under which peer effects can be attributed as causal.

What peer effects are and why they are important

In the literature, there is no consensus on exactly what "peers" mean. They can refer to friends, roommates, classmates, colleagues, neighbors, co-offenders, inmates, even firms making same products or providing same services or operating in same neighborhoods, and so forth, depending on the context. Although sometimes people nominate their spouses or siblings as their best friends, in general, researchers tend not to view social contacts connected through marriage or kinship as peers. Even so, it can still be very hard in practice to clearly define peership. Take "friends" as an example: different subjects may have very different definitions of what friends mean. A general approach to dealing with this kind of ambiguity is to provide an explicit name generator, like "Whom do you usually discuss important affairs with?" But what constitute important affairs are likely to be different across subjects and discussing important affairs is neither the only nor the central element of friendship. Regardless of these conceptualization issues, there are at least eight ways we can categorize peer effects, which may help us better understand the subtleties and varieties of peer effects.

1. Exogenous vs. endogenous peer effects: The difference between exogenous and endogenous peer effects mainly lies in the ultimate cause of peer effects. The former usually refers to the spill-over effects of some exogenous policy intervention on subjects that are not originally targeted by such an intervention but are connected to the target population of such an intervention. For example in Figure 34.1, there is an exogenous intervention (e.g., a smoking prevention or cessation program) that is aimed to change the smoking behavior of subject *j*, Y_j, namely, to prevent or stop subject *j* from smoking. Peer effects in this case refer to the effects of subject *j*'s smoking status or behavioral change on his or her friend subject *i*'s attitude and behavior regarding smoking. In contrast, endogenous peer effects come directly from the attributes of or changes in preferences or behaviors of peers. Still using Figure 34.1 to illustrate, endogenous peer effects refer to situations in which subject *j* directly influences subject *i* without being influenced by any external intervention first. In practice, it could be very difficult to completely separate endogenous peer effects from the exogenous ones, as many of the endogenous peer effects could originally be generated by external forces unobservable to researchers.
2. Positive vs negative peer effects: By desirability, peer effects can be viewed either as positive or negative. For example, peers who smoke cigarettes are usually considered to have negative effects on their contacts while students who have a higher GPA are generally thought to have a positive influence on their classmates.
3. Active vs. passive peer effects: Active peer effects come from connections that people can explicitly recognize while passive peer effects come from peers that a subject does not have an explicit tie with. Friends' effects are examples of the former. Transmission of infectious diseases or market competition can serve as examples of the latter. In practice, peer effects usually refer to active peer effects.
4. Contemporaneous vs. lagged peer effects: The effects due to peers' contemporaneous influence are called contemporaneous peer effects while the effects due to peers' previous influence are called lagged peer effects. Social interactions in a study group generate contemporaneous peer effects while diffusions of infectious diseases represent lagged peer effects.
5. Group vs. individual peer effects: By the size of reference group, peer effects can be either based on the group or the individual. Sometimes people are more likely to be influenced by their peer group while other times only by their best friends or other types of individual social ties. Peer effects within a study group can be an example for the former while obesity, smoking, or monopolistic competition can be examples for the latter.
6. Unidirectional vs. bidirectional peer effects: Unidirectional peer effects occur when peer effects flow only one way, from one subject to another, but not the other way around. Bidirectional peer effects imply that peers influence each other and the peer effects flow reciprocally.
7. Symmetric vs. asymmetric peer effects: When the effects that subject A has on his or her contact B is the same as the effects that subject B has on A, it is called symmetric peer effects. When the two effects are not equal, it is called asymmetric peer effects. For example, religious persons are thought to have stronger effects in converting their contacts into religions than the effects their contacts have in pulling them out of religions. Similar are the peer effects between smokers and nonsmokers.
8. Peer effects on preference, behavior, or outcome: By the content of peer effects, peer effects can be divided into peer effects on preference, behavior, or outcome, or any combination of them.

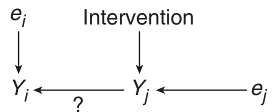

Figure 34.1 Exogenous peer effects from policy intervention to Y_j to Y_i

The important role played by peer effects in mediating social economic outcomes have been documented in the literature repeatedly. Researchers have shown that peers matter on diffusion of innovations (Coleman et al., 1957) and technology adoption (Oster and Thornton, 2009), job seeking and status attainment (Granovetter, 1973, 1974; Williams, 1981; Fernandez and Weinberg, 1997; Lin, 1999; Fernandez et al., 2000), widening of socioeconomic inequality (Finneran and Kelly, 2003; Calvo-Armengol and Jackson, 2004, 2007; Salganik et al., 2006), social spreading of obesity (Christakis and Fowler, 2007; Trogdon et al., 2008; Halliday and Kwak, 2009; Carrell et al., 2010), autism (Liu et al., 2010), cigarette smoking (Ennett and Baumann, 1993; Maxwell, 2002; Christakis and Fowler, 2008), criminal or delinquent behaviors (Baerveldt et al., 2008; Carrington, this volume), and sexually transmitted diseases (Laumann and Youm, 1999; Bearman et al., 2004), flow of mass communication (Katz and Lazarsfeld, 2005), mobilization of social movements and civic participation (Diani and McAdam, 2003; Lim, 2009), patterns of migrations (Garip, 2008) and energy consumption (Ayres et al., 2009), conformity of political opinions (Lazer et al., 2008), spillover of workers' productivity (Moretti, 2004; Greenstone et al., 2008; Mas and Moretti, 2009), and so on.

Understanding peer effects through the lens of social network analysis has important implications for public policies. We can either change or utilize the social network structures of peer groups to improve policy effectiveness. Take the seating arrangement in a classroom, for example; if students with high GPAs have larger positive peer effects on students with low GPAs than on students with high GPAs, then students with high GPAs should be seated with students with low GPAs to maximize the average student GPA. Certainly, the seat arrangement also depends on the objectives of education and the incentives of the teachers. Maybe the teachers do not care about the average GPA but the number of students who have GPAs above a certain threshold. If that is the case, students with high GPAs probably should be seated together. Take smoking prevention for another example. We can choose the popular subjects, for example, those who receive the most friendship nominations in a peer group, as opinion leaders to lead smoking prevention programs and accelerate the diffusion of positive information, attitudes and behaviors regarding smoking (Valente and Davis, 1999).

Understand peer effects using DAGs and counterfactuals

An important issue in studying peer effects is how to identify the observed correlations among peers' attitudes, behaviors, or outcomes as causal. Directed acyclic graphs (DAGs) (Pearl, 2000) provide a very intuitive conceptual tool to present causal paths and can be deployed here to show the causal paths of peer effects. For simplicity, suppose we have data on some outcome of interest (Y) and attributes (X) of a group of subjects and we are interested in whether subjects' outcomes are affected by their peers' outcomes. We can use the following DAG to present the hypothetical causal paths of such endogenous peer effects.

In Figure 34.2, we assume that subject i has nominated j as his or her peer and are interested in if Y_j has any effects on Y_i. We assume that each

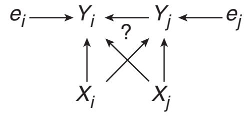

Figure 34.2 The causal paths of endogenous peer effects from Y_j to Y_i

subject's outcome Y is affected not only by their own covariates, X_i, but also by the covariates of their connected peers, X_j. e_i and e_j are just idiosyncratic error terms. This is an extremely simplified version of peer effects. If we have perfect measurement on all the relevant variables and the diagram accurately characterizes how peer effects work in reality, the causal effects from Y_j to Y_i are certainly identifiable, as all the backdoor paths are essentially blocked. But there are several issues to be solved before we can make such an optimistic claim.

The first issue is what we mean by "causal effects". There has been a tremendous amount of philosophical discussion on this, from Aristotle to Hume, to Lewis, and so on (Zalta, 2008). Among them, Donald Rubin's counterfactual model or potential outcomes framework (1974) is increasingly popular and more relevant here. Below is a simple introduction to it. See Morgan and Winship (2007) and Pearl (2009) for more comprehensive reviews.

Suppose there is a binary treatment D and outcome Y we are interested in. Let $Y_i(1)$ denote the potential outcome for unit i under treatment while $Y_i(0)$ is its potential outcome under control. Then the individual treatment effect is defined as the difference between the two potential outcomes, that is, $\tau_i = Y_i(1) - Y_i(0)$. However, since we can observe only one of the potential outcomes for each unit, τ_i is not directly identifiable. To identify each individual treatment effect, τ_i, we need to impute the missing potential outcome for each unit, for example, by matching on covariates or propensity scores, and so on. Below I will use the potential outcomes framework to examine the various difficulties in identifying peer effects.

Ambiguity of the treatment

If we are studying the peer effects of some exogenous policy intervention, then the treatment is well defined, which is the designed policy intervention. But if we are interested in endogenous peer effects, then the treatment is conceptually ambiguous. To fix ideas, let's suppose we are studying the effects of a hypothetical person's best friend's smoking status on his or her own smoking status. What is the treatment in this case? Is it the pure fact that this person's friend is or is not a smoker, or the number of cigarettes smoked per day by the friend, or both? In addition, what is the counterfactual in this case? Is it that this person has no friends at all, or that this person does not know this particular friend, or that this person knows this friend but he or she is not a smoker? In practice, peer effects are usually estimated using dyadic data, which implicitly assumes that the estimated peer effects only apply to populations who have friends but not to those who do not have friends. From this perspective, it can be argued that peer effects are conditional effects, conditioning on that there is at least one friend for the subjects of interest. Hence, in the above example, the counterfactual for a person whose best friend is a smoker would be that this person's best friend is not a smoker. If we are going to use matching methods to estimate peer effects, we would have to adopt a double-matching algorithm, namely, both the person of focus and his/her friend need to be matched with another pair of people who share the similar characteristics with the original pair of people respectively, except that the counterfactual friend is not a smoker. In general, careful thoughts are needed to define the counterfactual clearly in any specific study.

Violation of ignorability

A very important condition for the use of potential outcomes framework to analyze causal effects in observational studies is conditional ignorability, meaning that conditional on covariates, potential outcomes are independent of treatment. In other words, we expect subjects with the same covariates will have the same potential outcomes, $E(Y|D = 1, X = x) = E(Y|D = 0, X = x)$. It is very likely that conditional ignorability does not hold when studying peer effects, for two reasons. One is because of selection bias or homophily. Namely, subjects tend to be friends with other subjects who are similar to them, which may lead assignment of treatment to depend on potential outcomes. For example, overweight people may tend to be friends with other overweight people, corrupt officials tend to befriend other corrupt officials, and so on. Hence it is not your overweight or corrupt friends who make you overweight or corrupt, but you select persons who are overweight or corrupt to be friends from the beginning. Without taking such a selection into account, estimates of peer effects will certainly be biased, often upwardly. The other reason is due to confounding, meaning other factors correlated with the outcomes of both the subject and his or her peers, be they biological or contextual, are omitted from the model. This will lead to omitted variable bias in the estimated peer effects. Below is a diagram showing the causal paths of peer effects with an omitted variable U.

Violation of STUVA

A fundamental assumption of potential outcomes framework is the stable treatment unit value assumption (STUVA), which says that a subject's treatment status should not affect another subject's outcome. There are two features of peer effects that clearly violate such an assumption – one is simultaneity and the other is transitivity. Simultaneity means that Y_i also affects Y_j at the same time Y_j is affecting Y_i. Simultaneity causes ordinary least squares (OLS) estimates of peer effects to be both biased and inconsistent. Transitivity means subject k is a common friend of both subject j and subject i, and Y_k affects both Y_j and Y_i. It also allows Y_i and Y_j to react to Y_k. In some sense, transitivity can be viewed as a special type of simultaneity. Below is a diagram showing peer effects with simultaneity and transitivity.

Special techniques have to be employed to address the simultaneity and transitivity problem. For example, we can check for simultaneity using the Hausman test. If simultaneity does exist and the parameters governing peer effects are not underidentified, either indirect least squares (ILS) or two stage least squares (2SLS) estimators can be applied to estimate the peer effects. See Chapters 18–20 of Gujarati (2002) for an introduction to the simultaneity problem. As to the transitivity problem, a sort of system of equations method is needed.

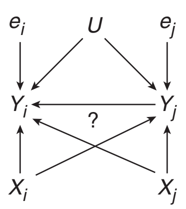

Figure 34.3 Peer effects with omitted variable *U*

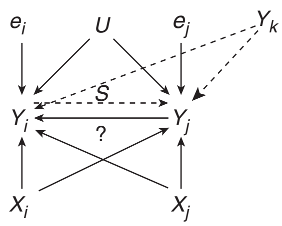

Figure 34.4 Peer effects with omitted variable, simultaneity, and transitivity

The second large issue in modeling and estimating peer effects originates from the static view of network and behavior. We usually assume that social networks and social behaviors embedded in them are fixed. But in many cases, both networks and behaviors keep evolving and we need to model the co-evolution of the networks and behaviors in order to separate peer selection from peer influence and obtain precise estimates of peer effects. There have been only a few studies taking this approach. I will review them more thoroughly below.

The third issue in identifying peer effects comes from missing data and measurement error. They can arise from a variety of reasons. For example, people may forget that some of their social ties are being inquired, there might be coding errors in the data (e.g., because of duplicated names), or a social network is constructed using a convenience sample in which subjects are not appropriately sampled and their links are not fully traced, and so forth.

In summary, some of the biggest problems in studying peer effects arise from (1) unclear definition of peer effects; (2) ambiguity of treatment; (3) violation of ignorability either due to selection or confounding; (4) violation of STUVA either due to simultaneity or transitivity; (5) static view of networks and behaviors; and (6) missing data and measurement error.

The first two problems are more at a conceptual level and have been addressed above. I will provide more thorough review on the rest four issues in the following sections. More specifically, in section 2, I will review the literature on why and how people form peer networks. This will help us better understand the mechanisms of how peer effects work and the importance to separate peer selection from peer influence in order to get valid estimates of peer effects. In section 3, I will review some of the popular models and methods that were developed recently to identify peer effects. This is the central part of this review. Section 4 will focus on missing data and measurement error in social networks and review their consequences on the estimations and inference of peer effects. Lastly, I will summarize and discuss some of the possible breakthroughs in studying peer effects that can be made in the near future.

WHY AND HOW PEER NETWORKS ARE FORMED

Choice, chance, and gene

To fix ideas, I will use the friendship network as an example to illustrate why and how peer networks are formed. In the literature, three factors have been considered as important in friendship formation. To put it simply, they are choice, chance, and gene. Choice means subjects have certain preference and the autonomy to choose whom they want to affiliate with. The underlining assumption is that friendship formation is based on utilitarian consideration. Unsurprisingly, many economists take this point of view and treat friendship as an instrument to obtain material or emotional benefits. For example, Jackson and Wolinsky (1996: 44) assumed that "self-interested individuals choose to form new links or to sever existing links." For another example, Bala and Goyal (2000: 1181) argued that "social networks are formed by individual decisions that trade off the costs of forming and maintaining links against the potential rewards from doing so. We suppose that a link with another agent allows access, in part and in due course, to the benefits available to the latter via his own links." Similar utilitarian argument can be found in Christakis et al. (2010) and Ellison (2010) as well. See Demange and Wooders (2005) for more papers on economic view of group formation. The utilitarian view of friendship formation leads to particular predictions of network structures. According to Bala and Goyal (2000: 1181), for example, "The limiting networks have simple architectures, for example, the wheel, the star, or generalizations of these networks."

Another important factor in friendship formation is chance. By chance I do not mean "randomness" but the structure governing the opportunities for people to meet or know each other. According to Zeng and Xie (2008: 4), opportunity structure refers to "the influence of all factors other than preference on choice" and it can include population composition and organizational structure and activities. Obviously, it is very difficult to operationalize and measure opportunity structure accurately. The basic point is that the likelihood for any pair of subjects to form a friendship is not only determined by their preference and choice, but also the possibility and the number of times that they can meet and interact. In this sense, researchers like Zeng and Xie recognize and emphasize that choices of friends are bounded by the opportunity structure. The following studies present some of the empirical evidence supporting the important role played by opportunity structure. For example, Abu-Ghazzeh (1999: 41) showed that "opportunities to walk around a small group of houses or to sit in small, confined spaces, by contrast, were significantly related to social interaction and friendship formation." For another example, Carrington (2002) demonstrated that the observed amount of sex homogeneity among co-offenders, particularly among males, to a large extent can be attributed to the relatively small number of females involved in crime, rather than signifying a preference for sex homophily. Lastly, Marmaros, and Sacerdote (2004: 1) presented that “two randomly chosen white students interact three times more often than do a black student and a white student. However, placing the black and white student in the same freshman dorm increases their frequency of interaction by a factor of three.”

Many scholars acknowledge that the above two factors may work together in generating social networks. For example, Jackson and Rogers (2007: 1) showed that social networks generated through the combination of both random and local search process “results in a spectrum of features exhibited by large social networks.” Currarini et al. (forthcoming: 1) showed that empirical social networks “can be generated by biases in preferences and biases in meetings.”

Whatever the specific driving force is underneath, social networks present a large degree of homophily “with regard to many sociodemographic, behavioral, and intrapersonal characteristics” (McPherson et al., 2001: 415). Homophily due to selection on the outcomes will bias the estimates of peer effects if not accounted for properly. For example, Arala et al. (2009) found that previous methods overestimated peer influence in product adoption decisions by three to seven folds and that homophily explained more than 50 percent of the perceived behavioral contagion in a global instant messaging network of 27.4 million users.

According to some recent studies, genes also played some part in producing the correlations of friends’ behaviors and some features of the social networks. Fowler et al. (2007) found a significant genetic influence on adolescent friends’ alcohol use (about 30%) and significant correlations of 0.60 and 0.70 between the genetic influences on friends’ alcohol use and adolescents’ own use and their problem use of alcohol. Fowler, Dawesa et al. (2009) showed that at least three social network attributes including in-degree, transitivity, and centrality, were inheritable and could be attributed to genetic factors. Both studies applied the twin pair design. Also see Madden et al. (2002) for a similar study using twin or sibling pairs to assess the respective contributions of peer selection and peer influence to the correlations of peer behaviors.

One limitation of the above analysis is that it ignores the social opportunities, for example, transitivity and reciprocity, produced and delimited by the structures of social networks. Random graph models are invented partly for that purpose.

Random graph models

For brevity, I will review only three classic random graph models here. Readers are encouraged to read other related models like stochastic blockmodels (Holland et al., 1983; Wang and Wong, 1987) and latent space model (Hoff et al., 2002).

Assuming dyadic independence, Holland and Leinhardt (1981) developed the p1 model for friendship formation. Given two randomly chosen subjects i and j, there are four possible relationships between them: (1) there is no tie between them; (2) there is a tie from i to j; (3) there is a tie from j to i; and (4) there is a mutual tie between i and j. The probability of these four types of ties is assumed to be given as follows.

$$\ln(P_{00}) = k_{ij} \tag{34.1}$$

$$\ln(P_{10}) = k_{ij} + \alpha_i + \beta_j + \mu \tag{34.2}$$

$$\ln(P_{01}) = k_{ij} + \alpha_j + \beta_i + \mu \tag{34.3}$$

$$\ln(P_{11}) = k_{ij} + \alpha_i + \beta_j + \alpha_j + \beta_i + 2\mu + \rho_{ij} \tag{34.4}$$

where k_{ij} is a normalizing constant, ensuring probabilities summarized to one, α the sender effect (or “a productivity parameter”), β the receiver effect (or “an attractiveness parameter”), μ a base rate of tie formation or density parameter, and ρ the reciprocity effect (“force of reciprocation”). For identification reasons, it is assumed that ρ_{ij} is the same across all ties as ρ. Then the log likelihood function for observing a network w can be written as

$$\ln P_1(w) \propto \mu L(w) + \sum_i^N \alpha_i w_{i+} + \sum_j^N \beta_j w_{+j} + \rho M(w) \tag{34.5}$$

with the constraint that $\alpha_+ = \beta_+ = 0$. $L(w)$ is the number of ties in the network, w_{i+} the number of outgoing ties, w_{+j} the number of incoming ties, and $M(w)$ the number of mutual ties. Maximum Likelihood method is used to estimate the parameters.

As a random effects version of the p1 model, the p2 model was developed to account for the dependence between ties sharing a same subject and the effects of covariates on tie formation (Van Duijn et al., 2004). The model setup is the same as in the p1 model except for the following:

$$\alpha_i = X_{1i}\gamma_1 + \alpha \tag{34.6}$$

$$\beta_i = X_{2i}\gamma_2 + \beta \tag{34.7}$$

$$\mu_{ij} = \mu + Z_{1ij}\lambda_1 \tag{34.8}$$

$$\rho_{ij} = \rho + Z_{2ij}\lambda_2 \tag{34.9}$$

where α and β are random variables drawn from multivariate distribution with mean zero, and Z's are attributes measured at dyadic level. The estimation is usually done by generalized least squares or by Markov Chain Monte Carlo (MCMC) methods.

p* model can contain more complicated forms of dependence between network structures. The literature on p* model has been growing dramatically in the past few years. Some of the notable papers include Wasserman and Robins (2005), Snijders et al. (2006), Goodreau (2007), Robins, Pattison et al. (2007), Robins, Snijders et al. (2007), and Robins (this volume), to name only a few. In the p* model, the probability of observing any social network is assumed to take the following form:

$$\ln P(W = w|\theta) \propto \sum_k \theta_k S_k(w) \tag{34.10}$$

S_k can be any network statistics of interest, for example, the number of reciprocal ties, the number of triangles, and so on. Note that p1 model is actually a special form of p* model, both belonging to exponential random graph models (ERGMs). Pseudo-likelihood methods or MCMC methods are generally used to estimate the parameters in p* models.

To conclude this section, we can see that peer networks are not formed randomly. On the one hand, people are free to choose their friends. On the other hand, their freedom to choose friends is limited by their biological predisposition and whom they can meet in the physical or social environment in which they are embedded. Taking into account the process of friendship formation is very crucial to address the selection problem in estimating peer effects.

MODELS AND METHODS TO IDENTIFY PEER EFFECTS

Now let's look at how selection, confounding, and simultaneity problems are dealt with in specific models. Depending on the data structure, these models can be roughly divided into four groups: static models, dynamic models, experiments and natural experiments, and simulation studies.

Static models

There are generally two types of static models for studying peer effects, depending on the feature of the dependent variables. Linear-in-means models are suitable for continuous outcomes while binary models aim at dealing with binary outcomes.

Linear-in-means model

The linear-in-means model is very popular in economics (Manski, 1993; Weinberg, 2007; Graham and Hahn, 2009; Bramoullé et al., 2009). It assumes that subject i's outcome is determined by his or her own covariates and the respective mean values of the covariates and outcomes of his or her peer group. Without loss of generality, we can assume there are only two subjects in any peer group. Their outcomes are modeled by the following equations:

$$Y_i = \alpha_1 X_i + \alpha_2 X_j + \beta Y_j + e_i \tag{34.11}$$

$$Y_j = \alpha_1 X_j + \alpha_2 X_i + \beta Y_i + e_j \tag{34.12}$$

First note that the above simultaneous equations cannot be estimated using OLS, because the Y's in the regressors are correlated with the error terms. If we write out the reduced form of the above two equations, we can see that

$$Y_i = \frac{\alpha_1 + \beta\alpha_2}{1-\beta^2} X_i + \frac{\alpha_2 + \beta\alpha_1}{1-\beta^2} X_j + (be_j + e_i) \tag{34.13}$$

Basically what we try to do using the reduced form of equation is to regress the outcomes on all the exogenous covariates and use the estimated coefficients to recover the original parameter values. Since there are only two exogenous covariates but three parameters (α_1, α_2, and β) to be estimated, the above simultaneous equations are not identified without further assumptions. One assumption, which is reasonable in some cases, is that there is an exogenous variable included in one equation but excluded in the other equation. For example, we can think of α_2 is equal to zero in equation (34.11). If this is the case, then (34.11) is identifiable and indirect least squares (ILS) will provide an estimate of the coefficient of peer effects, β. If there are more than one such exogenous variable, the model will be overidentified. In that case, two-stage least squares (2SLS) can be used to estimate the peer effects. Another approach is to use instrumental variables (IVs) for Y_i and Y_j in the regressors so that their correlations with the error terms are blocked. For example, we can use lagged Y's as instruments for Y's in the regressors, but this requires dynamic data. Bramoullé et al. (2009) proposed using the outcomes of intransitive and indirect contacts as IVs to identify peer effects. For example, according to Bramoullé et al. (2009), suppose there is an intransitive triad in which subject i is affected by j and j is affected by k, but i is not affected by k, then k's outcome can be used as an instrumental variable for j's outcome to identify the effects of j on i. Bramoullé et al. (2009) also extended the above technique to deal with situations in which fixed network effects are present. See Chapter 15 of Greene (2002) and Chapters 10–11 of Wooldridge (2002) for more discussion on the general techniques to address the simultaneity problem. If transitivity exists, namely, some of the peer groups overlap with one another, either the method proposed by Bramoullé et al. (2009) or some other method based on the system of equations has to be employed to estimate peer effects. This needs further investigation.

Binary outcome model

If the outcomes of interest are measured as binary, a logit or probit model is usually adopted to model peer effects (Brock and Durlauf, 2001a and 2001b; Bulte and Lilien, 2001; Sorensen, 2006; Krauth, 2009). Below shows the logit model:

$$\text{logit}[P(Y_i = 1)] = \ln\left[\frac{P(Y_i = 1)}{P(Y_i = 0)}\right] = \alpha_1 X_i + \alpha_2 X_j + \beta Y_j \qquad (34.14)$$

$$\text{logit}[P(Y_j = 1)] = \ln\left[\frac{P(Y_j = 1)}{P(Y_j = 0)}\right] = \alpha_1 X_j + \alpha_2 X_i + \beta Y_i \qquad (34.15)$$

The likelihood function can be written as

$$L(Y \mid X, \alpha_1, \alpha_2, \beta) = \prod_{i=1}^{n} \left[\frac{1}{1 + e^{-(\alpha_1 X_i + \alpha_2 X_j + \beta Y_j)}}\right]^{y_i} \left[\frac{1}{1 + e^{(\alpha_1 X_i + \alpha_2 X_j + \beta Y_j)}}\right]^{1-y_i} \qquad (34.16)$$

According to Brock and Durlauf (2001b), the nonlinearity between predictors and outcomes solves the simultaneity problem and facilitate identifying peer effects.

One challenge to static models is that there may be omitted variable bias. In other words, the estimated peer effects may be due to common environment factors rather than peer influence. For example, Bulte and Lilien (2001) showed that when pharmaceuticals' marking efforts were controlled for, the social contagion of physician prescription, originally attributed to peer influence by Coleman et al. (1957), disappeared.

Other static methods to identify peer effects include variance decomposition (Glaeser et al., 1996), comparison of group and individual level of regression coefficients (Glaeser et al., 2002), Rosenbaum's nonparametric methods (2007), strategies based on group size variations (Davezies et al., 2009; Lee, 2009), and so on.

Dynamic models

Dynamic network data are very useful to solve or mitigate the selection and confounding problems. Besides the models being reviewed below, some other useful techniques suitable for analyzing dynamic network data include event history analysis (e.g., Liu et al., 2010), the dynamic matched sample estimation framework (Aral et al., 2009), and so forth.

Dynamic logit model

Christakis and Fowler (2007 and 2008) and Fowler and Christakis (2009a) applied dynamic logit models to analyze peer effects on obesity, smoking and happiness, respectively. The outcomes are assumed to be generated according to the following model:

$$\text{logit}[P(Y_{it} = 1)] = \alpha_1 X_{it} + \gamma Y_{it-1} + \beta_1 Y_{jt} + \beta_2 Y_{jt-1} + e_{it} \qquad (34.17)$$

$$\text{logit}[P(Y_{jt} = 1)] = \alpha_1 X_{jt} + \gamma Y_{jt-1} + \beta_1 Y_{it} + \beta_2 Y_{it-1} + e_{jt} \qquad (34.18)$$

According to Christakis and Fowler (2007: 373), they used "generalized estimating equations to account for multiple observations of the same ego across examinations and across ego-alter pairs. We assumed an independent working correlation structure for the clusters. . . . The use of a time-lagged dependent variable (lagged to the previous examination) eliminated serial correlation in the errors (evaluated with a Lagrange multiplier test) and also substantially controlled for the egos genetic endowment and any intrinsic, stable predisposition to obesity. The use of a lagged independent variable for an alter weight status controlled for homophily."

Cohen-Colea and Fletcherb (2008) raised some critiques of the dynamic logit model used in Christakis and Fowler (2007), claiming that it may overestimate peer influence by ignoring contextual effects. Regardless of the model specification issue, some recent research finds that the magnitudes of peer influence may vary by health behaviors or outcomes of interest. For example, VanderWeele (2010) finds that the contagion effects for obesity and smoking are reasonably robust to possible homophily or contextual effects while those for happiness is less so.

Fixed effects model

Fixed effects can be used to explicitly account for omitted variable bias and selection bias (e.g., Nanda and Sorensen, 2008; Mas and Moretti, 2009). For example, suppose there is a variable measuring the common environmental factor for subject i and j, U_{ij}, and there is another variable measuring the propensity for subject i and j to form a tie, S_{ij}. We can use variable G to represent the effects of time-invariant variables like race, gender, or even genes. Then a dynamic linear peer effects model can be written as:

$$Y_{it} = \alpha_1 X_{it} + \alpha_2 X_{jt} + \beta Y_{jt} + \theta_1 U_{ijt} + \theta_2 S_{ijt} + \theta_3 G_i + e_{it} \qquad (34.19)$$

Using a first-difference estimator and subtracting both sides of the above equation by previous values, we will get

$$\Delta Y_{it} = \alpha_1 \Delta X_{it} + \alpha_2 \Delta X_{jt} + \beta \Delta Y_{jt} + \theta_1 \Delta U_{ijt} + \theta_2 \Delta S_{ijt} + \Delta e_{it} \qquad (34.20)$$

The effects of time-invariant variables are dropped out. If we assume both U_{ijt} and S_{ijt} do not change over time, their effects will be dropped out as well and the above equation can be simplified to

$$\Delta Y_{it} = \alpha_1 \Delta X_{it} + \alpha_2 \Delta X_{jt} + \beta \Delta Y_{jt} + \Delta e_{it} \qquad (34.21)$$

If U_{ijt} and S_{ijt} do change over time and cannot be observed or measured directly, there are generally two ways to solve the problem. One is to use proxy variables to measure the two change variables. For example, ΔU_{ijt}, representing environmental changes, can be approximated by the average change in the outcomes of the neighbors of subjects i and j. It is difficult to find proxy variables for ΔS_{ijt}. So a model for friendship formation is needed to predict the propensity for any pair of subjects to form friendship. For example, we can use the random graph models covered in the previous section to model friendship formation and plug into the outcome model the predicted probabilities of friendship formation to account for friendship selection. The other way is to use IVs. Since it is difficult to come up with any IVs for ΔS_{ijt}, this approach is more suitable to address the unobserved ΔU_{ijt} problem. For example, we can use the outcome of subject j's siblings who do not live in the same neighborhood as does subject j, to instrument for subject j's outcome. In many cases, it is reasonable to assume that siblings' outcomes are correlated and there are no direct effects of subject j's siblings' outcomes on subject i's outcome and so the IVs are valid in this sense. Of course, if there are good measures for environmental changes, we should directly include them into the model. But in practice measures of the environmental changes may not be available, comprehensive, or accurate. So the above approaches may still be useful in many cases.

Stochastic actor-oriented model

The stochastic actor-oriented model was developed by Snijders (2001 and 2005) and also presented in Snijders et al. (2009), Steglich et al. (2010), Snijders (this volume), and so on, to model peer network formation and peer influence (i.e., behavioral changes) jointly. The stochastic actor-oriented model assumes changes in peer networks and behaviors follow two separate continuous Markov processes. The frequency of the two types of changes is determined by two rate functions, one for each: λ_N for network and λ_B for behavior. The waiting time for any change is assumed to follow an exponential distribution, $P(T > t) = e^{-(\lambda N + \lambda B)t}$. Potentially, the λ's can vary across subjects, by incorporating subjects' covariates and network positions into the rate functions. Subjects make both types of changes according to two objective functions, which are assumed to be a linear summation of the effects of network structures and behavioral features.

$$f_i^N(w, w', z) = \sum_k \beta_k^N S_k^N(i, w, w', z), \qquad (34.22)$$

$$f_i^B(w, w', z) = \sum_k \beta_k^B S_k^B(i, w, w', z, z') \qquad (34.23)$$

w and w' represent the network statistics of subject i and his or her peers, and z and z' the covariates (including behavioral measures) of subject i and his or her peers. Thus, this model combines both the random network model with the behavioral model, which allows us to separate peer selection from peer influence. One drawback of this model is that it is too complicated to allow for closed-form estimation of the parameters. Usually stochastic simulation techniques like MCMC are needed to estimate the parameters. Koskinen (2004) and Koskinen and Snijders (2007) have extended the stochastic actor-oriented model using Bayesian methods. Another limitation of this model is that it assumes at any given time, there can only be one change happening in a social network, which might not be realistic in some cases. For example, in faction politics, rupture between leaders of two factions may lead to massive dissolving of ties between the two factions at the same time. Similar examples can be found in international politics.

Experiments and natural experiments

If we say the above models and methods address the various problems in estimating peer effects in data analysis process, experiments and natural experiments are used to solve some of those problems in data generation process. Broadly speaking, there are two types of experiments. One is randomly assigning policy treatment to subjects. This type of experiment is not particularly concerned with estimating peer effects. Instead, it provides an estimate of the total policy effects including peer effects. To separate peer effects, special experimental designs are needed. For example, a partial population design can be used for estimating peer effects under control (PEC), in which treatment is applied to only partial members in each of the treatment groups. The difference between the average outcome of the untreated units in the treatment groups and the average outcome of the control units in the control groups can be viewed as an estimate of PEC. To estimate peer effects under treatment (PET), a group-based treatment design is needed in which treatment is assigned to two groups, with participants in one group being random individuals with no connections between one another while participants in the other group being groups of individuals with internal connections among them, for examples, groups of friends or colleagues. The difference between the average outcome of the former participants and that of the latter participants will provide an estimate of PET. In practice, it needs a large number of experimental groups to account for the variations in social network structures of peer groups when estimating both PEC and PET.

The other type of experiment is randomly assigning friends to subjects. This aims at eliminating the selection problem, but note that it does not necessarily eliminate the confounding problem. For example, in many colleges roommates are randomly assigned. Suppose there is positive correlation of the academic performance of the roommates. Because of the random assignment of roommates, we know such correlation cannot be attributed to selection of roommates. But this does not eliminate the possibility that part of the correlation may be due to roommates' common teaching fellow, common living environment, and so on. Special care has to be taken to mitigate the effects of the common environmental factors in order to obtain good estimates of peer effects.

There have been many studies using the first type of experiment. Here are a few examples for studying PEC. Duo and Saez (2003: 815) "encouraged a random sample of employees in a subset of departments to attend a benefits information fair organized by the university, by promising a monetary reward for attendance. The experiment multiplied by more than five the attendance rate of these treated individuals (relative to controls), and tripled that of untreated individuals within departments where some individuals were treated." They also found that the Tax Deferred Account (TDA) retirement plan "enrollment five and eleven months after the fair was significantly higher in departments where some individuals were treated than in departments where nobody was treated." For another example, Cipollone and Rosolia (2007) provided evidence on the social interaction effect of the schooling achievement of young men on those of young women. The authors "exploit the exemption from compulsory military service (CMS) granted to a few specific cohorts of males living in southern Italy as a result of an earthquake in 1980. The exemption is shown to have increased boys' high school graduation rates by more than 2 percentage points. Graduation rates of girls in the same cohorts increased by about 2 percentage points. Since in Italy women are not subject to military draft, we argue that the change in their schooling achievements is the reaction to the schooling behavior of exempt males" (Cipollone and Rosolia, 2007: 948).

As examples for studying PET, Wing and Jeffrey (1999) showed that in a weight loss program, participants recruited with friends had higher treatment completion rate and greater weight losses than those who were recruited to participate in the program alone; Falk and Ichino (2006) provided experimental evidence showing that subjects working as a pair had higher productivity than those working alone; Babcock and Hartman (2010: 1) found that "subjects who have been incentivized to exercise increase gym usage more if they have more treated friends."

Examples for the second type of experiment are abundant. In Sacerdote (2001: 681), "Freshman year roommates and dormmates are randomly assigned at Dartmouth College. I [Sacerdote] find that peers have an impact on grade point average and on decisions to join social groups such as fraternities." Kremer and Levy (2003: 1) examine "a natural experiment in which students at a large state university are randomly assigned roommates through a lottery system. We find that on average, males assigned to roommates who reported drinking in the year prior to entering college had one quarter-point lower GPA than those assigned to non-drinking roommates." Boisjoly et al. (2006: 2) found that "white students at a large state university who were randomly assigned African-American roommates in their first year are more likely to endorse affirmative action and view a diverse student body as essential for a high-quality education." Camargo et al (2010: 1) also documented that "randomly assigned roommates of different races are as likely to become friends as randomly assigned roommates of the same race," and "in the long-run, white students who are randomly assigned black roommates have a significantly larger proportion of black friends than white students who are randomly assigned white roommates." Rao et al. (2007: 1) found that at a large private university where undergraduates are randomly assigned to residential halls, "a student becomes up to 8.3 percentage points more likely to get immunized if an additional 10 percent of her friends receive flu shots." Another interesting study conducted by Cook et al. (2007) showed that "sixth grade students attending middle schools are much more likely to be cited for discipline problems than those attending elementary schools." The authors explained that this was possibly because the sixth graders attending middle schools were exposed to older peers and the relative freedom from supervision. Using a dataset in which students "are exogenously assigned to peer groups of about 30 students with whom they are required to spend the majority of their time interacting", Carrell et al. (2009: 439) found "academic peer effects of much larger magnitude than found in previous studies." Carrell et al. (2010) reported that when individuals were randomly assigned to peer groups, there were statistically significant peer effects on fitness outcomes and such effects are caused primarily by friends who were the least fit. In a recent experiment on trust games conducted by Fowler and Christakis (2009b: 1), "subjects were randomly assigned to a sequence of different groups." They showed that "focal individuals ('egos') are influenced by fellow group members ('alters') in future interactions with others. Furthermore, this influence persists for multiple periods and spreads up to three degrees of separation (from person to person to person to person)."

Admittedly, there are also some studies that reported only modest peer effects or the magnitudes of peer effects vary by context. For example, Angrist and Lang (2004) found that a school integration program sending minority students from Boston schools to more affluent suburban schools did not significantly affect the scores of white students and only modestly decreased the scores of the minority third graders, both in the host districts. Another study done by Imberman et al. (2009) suggested that the influx of evacuated students because of Hurricanes Katrina and Rita moderately reduced elementary math test scores in the receiving schools in Houston, whereas Katrina evacuees benefited from the relocation, experiencing a .15 standard deviation improvement in scores (Sacerdote, 2008). Jackson (2010: 1) indicated that "While students at schools with higher-achieving peers have better academic achievement, within-school increases in peer achievement improve outcomes only at high-achievement schools." See Boozer and Cacciola (2001), Boruch (2005), Carrell et al. (2008), Ammermueller and Pischke (2009), and Duo et al. (2009) for more experimental studies on educational peer effects.

It should be noted that peer effects are often connected to neighborhood effects. In general, we can say peer effects reflect the impact of the social environment of certain neighborhoods. In this sense, recent studies using experiments (Moving to Opportunity) to detect neighborhood effects are informative to studying peer effects (Kling et al., 2007; Clampet-Lundquist and Massey, 2008; Ludwig et al., 2008; Sampson, 2008). Three comments are in order on these studies. First, if there are no neighborhood effects detected, it means that there might not be peer effects either, as peer effects are often nested in neighborhood effects. Second, one of the disputes in those studies is whether and how to account for the selection bias when participants could move into different neighborhoods. This is in turn determined by the definition of neighborhood effects. If by neighborhood effects we mean "living in a certain type of neighborhood," then we should fix the assignment of neighborhood, for example, we can move subjects from disadvantaged neighborhoods to advantaged neighborhoods and do not allow them to move afterwards to other types of neighbourhoods, especially disadvantaged ones. An analogy may be helpful here. Return to our classic example on the contagion effects of smoking. As said before, the counterfactual is that this subject has a nonsmoker friend. Suppose we allow this subject to freely choose a new friend. Very likely will he or she choose a smoker instead of a nonsmoker as his or her new friend. Then it is obvious to see the estimate of peer effects is biased by such an endogenous selection of friends. Similarly, if we allow individuals to freely choose the neighborhoods they want to reside in, the resulting estimates of the neighborhood effects will be biased as well. Third, a more interesting question is why the subjects from disadvantaged neighborhoods tend to move to other disadvantaged neighborhoods. As Sampson (2008) put it, "selection bias . . . [is] a fundamental social process worthy of study in its own right." Is this due to personal preference, stocks of social capital in different neighborhoods, or segregative policies and actions in the receiving neighborhoods? Similarly, if we are studying peer effects in smoking, we need to ask why smokers tend to be friends with other smokers and where such homophily comes from, and so on. Answers to these questions are very crucial for us to better understand and correctly estimate both neighborhood effects and peer effects.

Simulation studies

Simulations enable researchers to manipulate the features of social networks and policy treatments at their will and so provide them with a flexible tool to detect peer effects and to evaluate model performance, etc. Here are some notable studies using simulations to study peer effects. Christakis and Fowler (2007) argued that the directionality of social ties could be used as an identification strategy for distinguishing peer influence from other social correlation effects. They showed that mutual friends influence each other the most and people who are nominated by others as friends have influence on the nominators while the nominators have no influence on the nominees. Anagnostopoulos et al. (2008) reformulated this as the edge-reversal test, saying that, "Since other forms of social correlation (other than social influence) are only based on the fact that two friends often share common characteristics or are affected by the same external variables and are independent of which of these two individuals has named the other as a friend, we intuitively expect reversing the edges not to change our estimate of the social correlation significantly. On the other hand, social influence spreads in the direction specified by the edges of the graph, and hence reversing the edges should intuitively change the estimate of the correlation" (2008: 11). However, the simulations he conducted indicated that the edge-reversal test might not be an effective tool. One reason for this, I conjecture, is that even though the edge-reversal test could possibly eliminate the bias due to contextual effects, it does not necessarily exclude selection bias. It might be that the nominees just tend to have smaller variations in outcomes of interest. For example, people who are fitter may be more likely to be nominated as friends. In the simplest OLS framework, when you use these friends' weight as predictors of the nominators' weight, it is likely to see significance. But in the reverse regression setting where you use the nominators' weight to predict the nominees' weight, due to the larger variation in the nominators' weight, it is likely you do not obtain significance.

In addition, this might happen just because of random sampling errors. The nominees are usually just a small subset of the subjects in a social network who are repeatedly nominated by others as friends or social contacts. Even if the nominees are randomly chosen from the total subjects, the variation in the nominees' outcomes is likely to be smaller than the variation in the nominators' outcomes due to the repetition of nominations and the variance of the variation in the nominees' outcomes is likely to be larger than the variance of the variation in the nominators' outcomes from trial to trial. Hence, any observed directional peer effects may just be a result of sampling error.

Anagnostopoulos et al. (2008) brought up another type of innovative test of peer effects in dynamic network data: the shuffle test. In brief, they showed that if social influence does not play a role, the estimates of peer effects will be close to each other whether the timing of peers' actions are shuffled or not while in contrast, if social influence indeed plays a role, the estimates of peer effects will generally be different when the timing of peers' actions are shuffled.

Bahr et al. (2009) provided an interesting example on using simulations to study social contagion of obesity. The authors showed that "individuals with similar BMIs will cluster together into groups, and if left unchecked, current social forces will drive these groups toward increasing obesity. . . . The popular strategy for dieting with friends is shown to be an ineffective long-term weight loss strategy, whereas dieting with friends of friends can be somewhat more effective by forcing a shift in cluster boundaries . . . simulations also show that interventions targeting well-connected and/or normal-weight individuals at the edges of a cluster may quickly halt the spread of obesity" (2009: 723). See Fu (2011) for another interesting simulation study on the imitation dynamics of vaccination behaviors on social networks.

MISSING DATA AND MEASUREMENT ERROR

So far we have been assuming that there are neither missing data nor measurement error in the observed social networks. But, in reality, missing data and measurement error are quite prevalent in social network data. First, missing data can arise from several scenarios (Handcock and Gile, 2007). Here are some examples:

1 Nonrandom sampling of subjects: A typical example for this is that the social network data is just based on a convenience sample in which subjects and their contacts are not properly sampled.
2 Missing links: Subjects may not want to release any information about their contacts or just forget to nominate some of their friends. So when there is no tie between subjects, we cannot distinguish if it is because there is really not a tie between them or the tie is missing.
3 Not fully traced links: For example, in a fixed-choice design, subjects are asked to nominate only a fixed number of contacts. As a result, subjects with more than the fixed number of contacts will underreport their real number of friends. This problem may also come from the fact that the boundary of the social network is not clearly defined. For example, what we mean by friendship, peer group, and neighborhoods is

not always clearly or consistently understood by researchers and respondents.

4 Absence or attrition of subjects: These occur when subjects or their contacts are not included in a study, due to absence at the survey time, moving or death, and so on.
5 Missing covariates: Some covariates like income and education may not be readily available when the social network data are collected.

Similarly, measurement error in social network data can arise because of several reasons. First, it might be that what friendship means is different to each respondent—do we allow respondents to cite themselves, their spouses, or relatives as their best friends? Second, although unlikely but still possible, people may misreport their contacts. For example, if you ask middle school students to nominate their close friends, many of them may over-report the number of close friends they have, because of implicit social competition for popularity. In addition, it is well known that due to the sensitivity issue, some of the covariates like income, sexual activities, political orientation, psychological measures, and so on, tend to be reported with a lot of measurement error. Last, inaccurate data input can generate another source of measurement error. For example, people with same names might bring about a lot of confusion to data input, etc.

Missing data and measurement error may have severe consequences on social network analysis of peer effects. Brewer and Webster (1999: 361) noticed that "on average, residents forgot 20 per cent of their friends. Forgetting also influenced the measurement of some social network structural properties, such as density, number of cliques, centralization, and individuals' centralities." Ghani et al. (1998: 2079) found that when there were missing data, "substantial systematic biases are introduced. The direction and magnitude of these biases suggest that, by ignoring them, the risk for the establishment and persistence of infection in a population may be underestimated." Robins et al. (2004: 257) distinguished "ties between respondents from ties that link respondents to non-respondents", and found that "if we assume that the non-respondents are missing at random, . . . treating a sizable proportion of nodes as non-respondents may still result in estimates, and inferences about structural effects, consistent with those for the entire network. . . . If, on the other hand, the principal research focus is on the respondent-only structure, with non-respondents clearly not missing at random, . . . values of parameter estimates may not be directly comparable to those for models that exclude non-respondents." Kossinets (2006: 247) showed that "network boundary specification and fixed choice designs can dramatically alter estimates of network-level statistics. The observed clustering and assortativity coefficients are overestimated via omission of affiliations or fixed choice thereof, and underestimated via actor non-response, which results in inflated measurement error."

There have been some attempts to address the problems due to missing data and measurement error. Butts (2003: 103) developed a family of hierarchical Bayesian models to allow for "the simultaneous inference of informant accuracy and social structure in the presence of measurement error and missing data." Handcock and Gile (2007) developed inference for social networks with missing data based on information from adaptive network mechanisms. Koskinen et al. (2008: 2) discussed "various aspects of fitting exponential family random graph models to networks with missing data and present a Bayesian data augmentation algorithm for the purpose of estimation."

A general procedure to deal with missing data (and potentially measurement error as well) in social network analysis is provided by Butts (2003: 105), which is worth repeating here. "(1) Determining the extent of error in existing data. (2) Determining the mechanisms by which error is produced. (3) Finding means of collecting higher quality data. (4) Minimizing and accounting for the uncertainty associated with missing data in network analyses."

SUMMARY AND FUTURE DIRECTIONS

This chapter reviewed the literature on models and methods to identify peer effects. Below is a summary of what we can learn from the literature. Bear in mind this summary only provides a general outline to deal with the problems in studying peer effects. There must be ad hoc considerations and solutions to the problems in any specific research.

(1) Explicitly define the concept of peers and the boundary of the peer group. (2) Clearly discuss the content, meaning, and directionality of the peer effects. (3) Use dynamic data analysis techniques (e.g., fixed effects models) to model the co-evolution of network and behavior and to control for confounding and selection. (4) Use statistical techniques like instrumental variables, 2SLS, or system of equations to account for simultaneity or transitivity. (5) Adjust for missing data and measurement error through imputation, etc.

Despite the significant progress that has been made in the last decade in modeling and analyzing peer effects, there are certain areas that need further investigation.

1 Experiments: Given the complexity and difficulty in analyzing peer effects using observational data, experiments like those with the partial population design should be applied more often to surpass some of the barriers and to obtain better estimates of peer effects.
2 Front door mechanisms: More research needs to be done to identify and describe the specific mechanisms by which peer effects operate. For example, suppose we are interested in the social spreading of obesity. The potential front door mechanisms may include that friends tend to share similar norms or standards on what constitute a normal weight, that friends tend to eat similar food, that friends tend to do group exercises together, and so on. If we can operationalize some of these mechanisms and get measures on the variables involved, for example, the number of times of physical exercises per week, we can study a specific mechanism by which peers affect each other (e.g., Anderson, 2009; de la Haye et al., 2009). In addition, qualitative studies (e.g., Michell and West, 1996; Stewart-Knox et al., 2005) based on interviews and focus groups can serve for that purpose as well.
3 Model network formation and peer effects jointly: On the one hand, those continuous Markov process models need to be refined to account for more nuanced peer network formation and peer influence processes, for example, to properly distinguish social contagion, social influence, and social learning from each other (Young, 2009). On the other hand, new models and methods, simpler to communicate and compute while still with good approximations to the real world, need to be developed (e.g., Merckena et al., 2009). Given the similarities between spatial models and network models, some techniques developed in spatial analysis may be applicable to social network analysis as well (e.g., O'Malley and Marsden, 2008; Lee, 2009).
4 Computation with large social networks: With large social network data becoming increasingly available and models becoming more complicated, there must be new and better solutions to reduce the mounting computational cost in social network analysis. One possible solution is to use partial samples of social network data to estimate and infer peer effects. The other is more of a brute force approach, just inventing faster computational methods such as parallel computing. Both network sampling and computing for analysis of large social network data demand further research.

Last, I want to close this review with a quote from *The Analects of Confucius* (adopted from the English version by Legge, 2004), which perfectly accords with one of the main purposes of this review and may shed some light on the cultural dynamics of the peer influence process: "When we see men of worth, we should think of equalling them; when we see men of a contrary character, we should turn inwards and examine ourselves."

NOTE

The author wants to thank Professor Nicholas Christakis, Christopher Winship, Filiz Garip and Peter Carrington for their helpful comments and suggestions. Sincere thanks also go to relevant participants of the American Sociological Association Methodology Section Conference at University of Illinois at Urban-Champaign (04/03, 2010), Department of Health Care Policy of Harvard Medical School (06/24, 2010), the SunBelt XXX of International Network for Social Network Analysis at Riva del Garda, Italy (07/04, 2010), the International Sociological Association World Congress in Gothenburg, Sweden (07/12, 2010) and the 105th American Sociological Association Meeting in Atlanta, Georgia, USA (08/15, 2010).

REFERENCES

Abu-Ghazzeh, T.M. (1999) 'Housing layout, social interaction, and the place of contact in Inabu-Nuseir, Jordan'. *Journal of Environmental Psychology,* 19: 41–73.

Albert, R. and Barabási, A.-L. (2002) 'Statistical mechanics of complex networks'. *Reviews of Modern Physics,* 74(1): 47–97.

Ammermueller, A. and Pischke, J.-S. (2009) 'Peer effects in European primary schools: Evidence from the progress in international reading literacy study'. *Journal of Labor Economics,* 27(3): 315–48.

Anagnostopoulos, A., Kumar, R. and Mahdian, M. (2008) 'Influence and correlation in social networks'. In *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.* New York: ACM, pp. 7–15.

Anderson, L.B. (2009) 'The trend in obesity: The effect of social norms on perceived weight and weight goal'. *Working paper*, http://www2.binghamton.edu/economics/graduate/documents/prospectus-by-l-anderson.pdf.

Angrist, J.D. and Lang, K. (2004) 'Does school integration generate peer effects? Evidence from Boston's Metco program'. *American Economic Review,* 94(5): 1613–34.

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

Ayres, I., Raseman, S. and Shih, A. (2009) 'Evidence from two large field experiments that peer comparison feedback can reduce residential energy usage'. *NBER Working Paper* No. 15386, http://www.nber.org/papers/w15386.pdf

Babcock, P.S. and Hartman, J.L. (2010) 'Networks and Workouts: Treatment Size and Status Specific Peer Effects in a Randomized Field Experiment'. *NBER Working Paper* No. 16581, http://www.nber.org/papers/w16581.pdf

Baerveldt, C., Volker, B. and Rossem, R.V. (2008) 'Revisiting selection and influence: An inquiry into the friendship networks of high school students and their association with delinquency'. *Canadian Journal of Criminology and Criminal Justice,* 50(5): 559–87.

Bahr, D.B., Browning, R.C., Wyatt, H.R. and Hill, J.O. (2009) 'Exploiting social networks to mitigate the obesity epidemic'. *Obesity,* 17(4): 723–28.

Bala, V. and Goyal, S. (2000) 'A noncooperative model of network formation'. *Econometrica,* 68(5): 1181–229.

Bearman, P.S., Moody, J. and Stovel, K. (2004) 'Chains of affection: The structure of adolescent romantic and sexual networks'. *American Journal of Sociology,* 110: 33–91.

Blume, L. and Durlauf, S.N. (2005) 'Identifying social interactions: A review'. http://www.ssc.wisc.edu/econ/archive/wp2005–12.pdf.

Boisjoly, J., Duncan, G.J., Kremer, M., Levy, D.M. and Eccles, J. (2006) 'Empathy or antipathy? The impact of diversity'. *American Economic Review,* 96(5): 1890–905.

Boozer, M.A. and Cacciola, S.E. (2001) 'Inside the 'bBlack box' of Project Star: Estimation of peer effects using experimental data'. Yale Economic Growth Center Discussion Paper No. 832, http://www.econ.yale.edu/growth_pdf/cdp832.pdf.

Boruch, R.F. (ed.). (2005) *Place-Based Trials: Experimental Tests of Public Policy.* Thousand Oaks, CA: Sage.

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

Brewer, D. and Webster, C. (1999) 'Forgetting of friends and its effects on measuring friendship networks'. *Social Networks,* 21: 361–73.

Brock, W.A. and Durlauf, S.N. (2001a) 'Interactions-based models'. Chapter 54 of *Handbook of Econometrics V*, edited by J.J. Heckman and E. Leamer. Elsevier Science.

Brock, W.A. and Durlauf, S.N. (2001b) 'Discrete choice with social interactions'. *The Review of Economic Studies,* 68(2): 235–60.

Bulte, C.V. d. and Lilien, G.L. (2001) 'Medical innovation revisited: Social contagion versus marketing effort'. *American Journal of Sociology,* 106(5): 1409–35.

Butts, C. (2003) 'Network inference, error, and informant (in) accuracy: A Bayesian approach'. *Social Networks,* 25: 103–40.

Calvo-Armengol, A. and Jackson, M. (2004) 'The effects of social networks on employment and inequality'. *American Economic Review,* 94 (3): 426–54.

Calvo-Armengol, A. and Jackson, M. (2007) 'Networks in labor markets: Wage and employment dynamics and inequality'. *Journal of Economic Theory,* 132: 27–46.

Camargo, B., Stinebrickner, R. and Stinebrickner, T.R. (2010) 'Interracial Friendships in College'. *NBER Working Paper* No. 15970, http://www.nber.org/papers/w15970.pdf

Carrell, S.E., Fullerton, R.L. and West, J.E. (2009) 'Does your cohort matter? Measuring peer effects in college achievement'. *Journal of Labor Economics,* 27(3): 439–64.

Carrell, S.E., Hoekstra, M. and West, J.E. (2010) 'Is Poor Fitness Contagious? Evidence from Randomly Assigned Friends'. *NBER Working Paper* No. 16518, http://www.nber.org/papers/w16518.pdf.

Carrell, S.E., Malmstrom, F.V. and West, James E. (2008) 'Peer effects in academic cheating'. *Journal of Human Resources,* 43(1): 173–207.

Carrington, P.J. (2002) 'Sex homogeneity in co-offending groups', in J. Hagberg (ed.), *Contributions to Social Network Analysis, Information Theory and Other Topics in Statistics: A Festschrift in Honour of Ove Frank*. Stockholm: Stockholm University. pp. 101–16.

Carrington, P.J., Scott, J. and Wasserman, S. (eds) (2005) *Models and Methods in Social Network Analysis*. New York: Cambridge University Press.

Christakis, Nicholas A., and Fowler, James H. (2007) "The spread of obesity in a large social network over 32 years." *New England Journal of Medicine* 357(4): 370–79.

Christakis, N.A. and Fowler, J.H. (2008) 'The collective dynamics of smoking in a large social network'. *New England Journal of Medicine,* 358(21): 2249–58.

Christakis, N.A., Fowler, J.H., Imbens, G.W. and Kalyanaraman, K. (2010) 'An empirical model for strategic network formation'. *NBER Working Paper* No. 16039, http://www.nber.org/papers/w16039.pdf

Cipollone, P. and Rosolia, A. (2007) 'Social interactions in high school: Lessons from an earthquake'. *The American Economic Review,* 97(3): 948–65.

Clampet-Lundquist, S. and Massey, D.S. (2008) 'Neighborhood effects on economic self-sufficiency: A reconsideration of the moving to opportunity experiment'. *American Journal of Sociology,* 114(1): 107–43.

Cohen-Colea, E. and Fletcherb, J.M. (2008) 'Is obesity contagious? Social networks vs. environmental factors in the obesity epidemic'. *Journal of Health Economics,* 27: 1382–1387.

Coleman, J., Katz, E. and Menzel, H. (1957) 'The diffusion of an innovation among physicians'. *Sociometry,* 20: 253–70.

Cook, P.J., MacCoun, R., Muschkin, C. and Vigdor, J. (2007) 'Should sixth grade be in elementary or middle school? An analysis of school configuration and student behavior'. *NBER Working Paper* No. w12471, http://www.nber.org/papers/w12471.pdf.

Currarini, S., Jackson, M. and Pin, P. (forthcoming) 'An economic model of friendship: Homophily, minorities and segregation'. *Econometrica*.

Davezies, Laurent, D"Haultfoeuille, Xavier Fougre, Deni. (2009) 'Identification of peer effects using group size variation'. *Econometrics Journal,* 12: 397C413.

Demange, G. and Wooders, M. (2005) *Group Formation in Economics: Networks, Clubs, and Coalitions*. New York: Cambridge University Press.

Diani, M. and McAdam, Doug. (eds) (2003) *Social Movements and Networks: Relational Approaches to Collective Action*. New York: Oxford University Press.

Doreian, Pa. (2001) 'Causality in social network analysis'. *Sociological Methods & Research* 30(1): 81–114.

Duo, E., Dupas, P. and Kremer, M. (2009) 'Peer effects, teacher incentives, and the impact of tracking: Evidence from a randomized evaluation in Kenya'. *NBER Working Paper* No. 14475, http://www.nber.org/papers/w14475.pdf.

Duo, E. and Saez, E. (2003) 'The role of information and social interactions in retirement plan decisions: Evidence from a randomized experiment'. *The Quarterly Journal of Economics,* 118(3): 815–42.

Van Duijn, M., Snijders, T. and Zijlstra, B. (2004) 'p2: A random effects model with covariates for directed graphs'. *Statistica Neerlandica,* 58: 234–54.

Duncan, O.D., Haller, A.O. and Portes, A. (1968) 'Peer influences on aspirations: A reinterpretation'. *American Journal of Sociology,* 74(2): 119–37.

Ellison, G., Glaeser, E.L. and Kerr, W.R. (2010) 'What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns'. *American Economic Review,* 100: 1195–1213.

Ennett, S.T. and Baumann, K.E. (1993) 'Peer group structure and adolescent cigarette smoking: A social network analysis'. *Journal of Health and Social Behavior,* 34: 226–36.

Falk, A. and Ichino, A. (2006) 'Clean evidence on peer effects'. *Journal of Labor Economics,* 24(1): 39–57.

Fernandez, R.M., Castilla, E.J. and Moore, P. (2000) 'Social capital at work: Networks and employment at a phone center'. *American Journal of Sociology,* 105(5): 1288–356.

Fernandez, R.M. and Weinberg, N. (1997) 'Sifting and sorting: Personal contacts and hiring in a retail bank'. *American Sociological Review,* 62(6): 883–902.

Finneran, Lisa, and Kelly, Morgan (2003) "Social networks and inequality." *Journal of Urban Economics* 53: 282–99.

Fowler, J.H. and Christakis, N.A. (2009a) "Dynamic spread of happiness in a large social network: Longitudinal analysis over 20 years in the Framingham heart study." *British Medical Journal* 337(42): 2338–47.

Fowler, J.H. and Christakis, N.A. (2009b) 'Cooperative behaviour cascades in human social networks'. http://www.citebase.org/abstract?id=oai:arXiv.org:0908.3497.

Fowler, J., Dawesa, C. and Christakis, N. (2009) 'Model of genetic variation in human social networks'. *Proceedings of the National Academy of Sciences of the USA,* 106(6): 1720–24.

Fowler, J.H., Heaney, M.T., Nickerson, D.W., Padgett, J.F. and Sinclair, Betsy. (2009) 'Causality in political networks'. *Political Networks Paper Archive Working Papers*, Southern Illinois University Carbondale, http://opensiuc.lib.siu.edu/cgi/viewcontent.cgi?article=1034&context=pn_wp.

Fowler, T., Shelton, K., Lifford, K., Rice1, F., McBride, A., Nikolov, I., Neale, M.C., Harold, G., Thapar, A. and van den Bree, M.B.M. (2007) 'Genetic and environmental influences on the relationship between peer alcohol use and own alcohol use in adolescents'. *Addiction,* 102: 894–903.

Fu, F., Rosenbloom, D.I., Wang, L. and Nowak, M.A. (2011) 'Imitation dynamics of vaccination behaviour on social networks'. *Proceedings of Royal Society B*, 278 (1702): 42–49.

Garip, F. (2008) 'Social capital and migration: How do similar resources lead to divergent outcomes?' *Demography,* 45(3): 591–617.

Ghani, A.C., Donnelly, C.A. and Garnett, G.P. (1998) 'Sampling biases and missing data in explorations of sexual partner networks for the spread of sexually transmitted diseases'. *Statistics in Medicine,* 17(18): 2079–97.

Glaeser, E.L., Sacerdote, B.I. and Scheinkman, J.A. (1996) 'Crime and social interactions'. *Quarterly Journal of Economics,* 111(2): 507–48.

Glaeser, E.L., Sacerdote, B.I. and Scheinkman, J.A. (2002) 'The social multiplier'. *NBER Working Paper* 9153, http://www.nber.org/papers/w9153.pdf.

Goldenberg, A., Zheng, A.X., Fieberg, S.E., and Airoldi, E.M. (2009) 'A survey of statistical network models'. *Foundations and Trends in Machine Learning*.

Goodreau, S. (2007) 'Advances in exponential random graph (p*) models applied to a large social network'. *Social Networks,* 29: 231–48.

Graham, B.S. and Hahn, J. (2009) 'Identification and estimation of the linear-in-means model of social interactions'. *Economics Letters,* 88(1): 1–6.

Granovetter, Mark S. (1973) "The strength of weak ties." *American Journal of Sociology* 78(6): 1360–80.

Granovetter, M.S. (1974) *Getting a Job: A Study of Contacts and Careers*. Chicago: University of Chicago Press.

Greene, W.H. (2002) *Econometric Analysis*. 5th ed. Prentice Hall.

Greenstone, M., Hornbeck, R. and Moretti, E. (2008) 'Identifying agglomeration spillovers: Evidence from million dollar plants'. *NBER Working Paper* 13833, http://www.nber.org/papers/w13833.pdf.

Gujarati, D. (2002) *Basic Econometrics*. 4th ed. McGraw-Hill/Irwin.

Halliday, T.J. and Kwak, S. (2009) 'Weight gain in adolescents and their peers'. *Economics and Human Biology,* 7: 181–90.

Handcock, M. and Gile, K. (2007) 'Modeling social networks with sampled or missing data'. *Working Paper* No. 75, Center for Statistics and the Social Sciences, University of Washington, http://www.csss.washington.edu/Papers/wp75.pdf.

Hartmann, W.R., Manchanda, P., Nair, H., Bothner, M., Dodds, P., Godes, D., Hosanagar, K. and Tucker, C. (2008) 'Modeling social interactions: Identification, empirical methods and policy implications'. *Marketing Letters,* 19: 287–304.

de la Haye, K., Robins, G., Mohrd, P. and Wilsone, C. (2009) 'Obesity-related behaviors in adolescent friendship networks'. *Social Networks*.

Hoff, P.D., Raftery, A.E. and Handcock, M.S. (2002) 'Latent space approaches to social network analysis'. *Journal of the American Statistical Association,* 97: 1090–98.

Holland, P.W., Laskey, K.B. and Leinhardt, S. (1983) 'Stochastic blockmodels: First steps'. *Social Networks,* 5(2): 109–37.

Holland, P.W. and Leinhardt, S. (1981) 'An exponential family of probability distributions for directed graphs (with discussion)'. *Journal of the American Statistical Association,* 76: 33–65.

Imberman, S, Kugler, A.D. and Sacerdote, B. (2009) 'Katrina's children: Evidence on the structure of peer effects from Hurricane evacuees'. *NBER Working Paper* No. 15291, http://papers.nber.org/papers/w15291.pdf.

Jackson, C.K. (2010) 'Peer Quality or Input Quality?: Evidence from Trinidad and Tobago'. *NBER Working Paper* No. 16598, http://www.nber.org/papers/w16598.pdf

Jackson, M.O. (2008) *Social and Economic Networks*. Princeton: Princeton University Press.

Jackson, M.O. and Rogers, B.W. (2007) 'Meeting strangers and friends of friends: How random are socially generated networks?' *American Economic Review*, http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.84.8381&rep=rep1&type=pdf.

Jackson, M.O. and Wolinsky, A. (1996) 'A strategic model of social and economic networks'. *Journal of Economic Theory,* 71: 44–74.

Kandel, D.B. (1978) 'Homophily, selection, and socialization in adolescent friendships'. *American Journal of Sociology,* 84(2): 427–36.

Katz, E. and Lazarsfeld, P.F. (2005) *Personal Influence: The Part Played by People in the Flow of Mass Communications*. New Brunswick, NJ: Transaction.

Kling, J.R., Liebman, J.B. and Katz, L.F. (2007) 'Experimental analysis of neighborhood effects'. *Econometrica,* 75(1): 83–119.

Kolaczyk, E. (2009) *Statistical Analysis of Network Data*. Springer.

Koskinen, J. (2004) 'Bayesian inference for longitudinal social networks'. Research Report 2004:4, Department of Statistics, Stockholm University, http://gauss.stat.su.se/site/pdfer/RR2004_4.pdf.

Koskinen, J.H., Robins, G.L. and Pattison, P.E. (2008) 'Analysing exponential random graph (p-star) models with missing data using Bayesian data augmentation'. MelNet Social Networks Laboratory Technical Report 08–04, Department of Psychology, School of Behavioural Science, University of Melbourne, Australia, http://www.sna.unimelb.edu.au/publications/MelNet_Techreport_08_04.pdf.

Koskinen, J. and Snijders, T. (2007) 'Bayesian inference for dynamic social network data." *Journal of Statistical Planning and Inference,* 137: 3930–38.

Kossinets, G. (2006) 'Effects of missing data in social networks'. *Social Networks* 28: 247–68.

Krauth, B.V. (2009) 'Simulation-based estimation of peer effects'. *Journal of Econometrics,* 1: 243–71.

Kremer, M. and Levy, D.M. (2003) 'Peer effects and alcohol use among college students'. *Working Paper*, http://www.povertyactionlab.org/papers/12_Kremer_Peer_Effects_and_Alcohol_Use.pdf.

Laumann, E.O. and Youm, Y. (1999) 'Racial/ethnic group differences in the prevalence of sexually transmitted diseases in the United States: A network explanation'. *Sexually Transmitted Diseases,* 26(5): 250–61.

Lazer, D., Rubineau, B., Katz, N., Chetkovich, C. and Neblo, M.A. (2008) 'Networks and political attitudes: Structure, influence, and co-evolution'. *Harvard Kennedy School Faculty Research Working Paper Series* RWP08-044, http://web.hks.harvard.edu/publications/getFile.aspx?Id=310

Lee, L.-f. (2009) 'Identification and estimation of spatial econometric models with group interactions, contextual factors and fixed effects'. *Journal of Econometrics,* 140(2): 333–74.

Legge, J. (Trans.) (2004) *The Analects of Confucius*, http://ebooks.adelaide.edu.au/c/confucius/c748a/part4.html.

Lim, C. (2009) 'Mobilizing on the margin: How does interpersonal recruitment affect citizen participation in politics?' *Social Science Research*.

Lin, N. (1999) 'Social networks and status attainment'. *Annual Review of Sociology,* 25: 467–87.

Liu, K.-Y., King, M. and Bearman, P.S. (2010) 'Social influence and the autism epidemic'. *American Journal of Sociology,* 115(5): 1387–1434.

Ludwig, J., Liebman, J.B., Kling, J.R., Duncan, G.J., Katz, L.F., Kessler, R.C. and Sanbonmatsu, L. (2008) 'What can we learn about neighborhood effects from the moving to opportunity experiment?' *American Journal of Sociology,* 114(1): 144–88.

Madden, Pamela A. F., Bucholz, Kathleen K., Todorov, Alexandre A., Grant, Julia D., and Heath, Andrew C. (2002) "The assessment of peer selection and peer environmental influences on behavior using pairs of siblings or twins'. *Twin Research,* 5(1): 38–43.

Manski, C.F. (1993) 'Identification of endogenous social effects: The reflection problem'. *Review of Economic Studies,* 60: 531–42.

Manski, C.F. (2000) 'Economic analysis of social interactions'. *NBER Working Paper* No. 7580, http://papers.nber.org/papers/w7580.pdf.

Manski, C.F. (2010) 'Identification of treatment response with social interactions'. *Working Paper*, http://faculty.wcas.northwestern.edu/~cfm754/treatment_with_interactions.pdf.

Marmaros, D. and Sacerdote, B. (2004) 'How do friendships form?' *NBER Working Paper* 11530, http://www.nber.org/papers/w11530.pdf.

Marsden, P.V. and Friedkin, N.E. (1993) 'Network studies of social influence'. *Sociological Methods Research,* (22): 127–51.

Mas, A. and Moretti, E. (2009) 'Peers at work'. *American Economic Review,* 99(1): 112–45.

Maxwell, K.A. (2002) 'Friends: The role of peer influence across adolescent risk behaviors'. *Journal of Youth and Adolescence,* 31(4): 267–77.

McPherson, M., Smith-Lovin, L. and Cook, J.M. (2001) 'Birds of a feather: Homophily in social networks'. *Annual Review of Sociology,* 27: 415–44.

Merckena, L., Candela, M., Willemsd, P. and de Vries, H. (2009) 'Social influence and selection effects in the context of smoking behavior: Changes during early and mid adolescence'. *Health Psychology,* 28(1): 73–82.

Michell, L., and West, P. (1996) 'Peer pressure to smoke: The meaning depends on the method'. *Health Education Research,* 11(1): 39–49.

Moffitt, R.A. (forthcoming) 'Policy interventions, low-level equilibria, and social interactions'. In Steven Durlauf and Peyton Young (eds), *Social Dynamics.* MIT Press.

Moretti, E. (2004) 'Workers' education, spillovers and productivity: Evidence from plant-level production functions'. *American Economic Review,* 94(3): 656–90.

Morgan, S.L. and Winship, C. (2007) *Counterfactuals and Causal Inference: Methods and Principles for Social Research*. New York: Cambridge University Press.

Newman, M.E.J. (2003) 'The structure and function of complex networks'. *SIAM Review,* 45: 167–256.

Nanda, R. and Sorensen, J.B. (2008) 'Peer effects and entrepreneurship'. *Harvard Business School Working Paper,* pp. 08–051, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1084874.

O'Malley, J.A. and Marsden, P.V. (2008) 'The analysis of social networks'. *Health Services and Outcomes Research Methodology,* 8(4): 222–69.

Oster, E. and Thornton, R. (2009) 'Determinants of Technology Adoption: Private Value and Peer Effects in Menstrual Cup Take-Up'. *NBER Working Paper* No. 14828, http://www.nber.org/papers/w14828.pdf.

Pearl, J. (2000) *Causality: Models, Reasoning, and Inference*. New York: Cambridge University Press.

Pearl, J. (2009) 'Causal inference in statistics: An overview'. *Statistics Surveys,* 3: 96–146.

Rao, N., Mobius, Markus M. and Rosenblat, T. (2007) 'Social networks and vaccination decisions'. *FRB of Boston Working Paper* No. 07–12, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1073143.

Robins, G., Pattison, P., Kalish, Y. and Lusher, D. (2007) 'An introduction to exponential random graph (p*) models for social networks'. *Social Networks,* 29: 173–91.

Robins, G., Pattison, P. and Woolcock, J. (2004) 'Missing data in networks: Exponential random graph (p*) models for networks with non-respondents'. *Social Networks,* 26: 257–83.

Robins, G., Snijders, T. Wang, P., Handcock, M. and Pattison, P. (2007) 'Recent developments in exponential random graph (p*) models for social networks'. *Social Networks,* 29: 192–215.

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

Sacerdote, B. (2001) 'Peer effects with random assignment: Results for Dartmouth roommates'. *Quarterly Journal of Economics,* 116: 681–704.

Sacerdote, B. (2008) 'When the saints come marching in: Effects of Hurricanes Katrina and Rita on student evacuees'. *NBER Working Paper* No. 14385, http://www.nber.org/papers/w14385.pdf.

Salganik, M.J., Dodds, P.S. and Watts, D.J. (2006) 'Experimental study of inequality and unpredictability in an artificial cultural market'. *Science,* 311: 854–56.

Sampson, R.J. (2008) 'Moving to inequality: Neighborhood effects and experiments meet social structure'. *American Journal of Sociolog,y* 114(1): 189–231.

Smith, K.P. and Christakis, N.A. (2008) 'Social networks and health'. *Annual Review of Sociology,* 34: 405–29.

Snijders, T.A.B. (2001) 'The statistical evaluation of social network dynamics'. In M. Sobel and M. Becker (eds), *Sociological Methodology*. Boston: Basil Blackwell. 361–95.

Snijders, T.A.B. (2005) 'Models for longitudinal network data'. In Peter Carrington, John Scott, and Stanley Wasserman (eds), *Models and Methods in Social Network Analysis*. New York: Cambridge University Press. Chapter 11.

Snijders, T.A.B., van de Buntz, G.G. and SteglichIn, C.E.G. (2009) 'Introduction to stochastic actor-based models for network dynamics'. Draft article for special issue of *Social Networks on Dynamics of Social Networks*, http://www.stats.ox.ac.uk/~snijders/SnijdersSteglichVdBunt2009.pdf.

Snijders, T.A.B., Pattison, P., Robins, G. and Handcock, M. (2006) 'New specifications for exponential random graph models'. *Sociological Methodology,* 36: 99–153.

Soetevent, A.R. (2006) 'Empirics of the identification of social interactions: An evaluation of the approaches and their results'. *Journal of Economic Surveys,* 20(2): 193–228.

Sorensen, A.T. (2006) 'Social learning and health plan choices'. *RAND Journal of Economics,* 37(4): 929–45.

Steglich, C., Snijders, T. and Pearson, M. (2010) 'Dynamic networks and behavior: Separating selection from influence'. *Sociological Methodology*, http://www.stats.ox.ac.uk/~snijders/siena/SteglichSnijdersPearson2009.pdf.

Stewart-Knox, B.J., Sittlington, J., Rugkasa, J., Harrisson, S., Treacy, M. and Abaunza, P.S. (2005) 'Smoking and peer groups: Results from a longitudinal qualitative study of young people in Northern Ireland'. *British Journal of Social Psychology,* 44: 397–414.

Trogdon, J.G., Nonnemaker, J. and Pais, J. (2008) 'Peer effects in adolescent overweight'. *Journal of Health Economics,* 27: 1388–99.

Valente, T.W. (2005) 'Network models and methods for studying the diffusion of innovations'. In Peter Carrington, John Scott, and Stanley Wasserman (eds) *Models and Methods in Social Network Analysis*. New York: Cambridge University Press. pp. 98–116.

Valente, T.W. and Davis, R.L. (1999) 'Accelerating the diffusion of innovations using opinion leaders'. *The ANNALS of the American Academy of Political and Social Science,* 566: 55–67.

VanderWeele, T.J. (2010) 'Sensitivity analysis for contagion effects in social networks', *Unpublished Manuscript*.

Wang, Y.J. and Wong, G.Y. (1987) 'Stochastic blockmodels for directed graphs'. *Journal of the American Statistical Association,* 82(397): 8–19.

Wasserman, S. and Robins, G. (2005) 'An introduction to random graphs, dependence graphs, and p*'. In Peter Carrington, John Scott, and Stanley Wasserman (eds), *Models and Methods in Social Network Analysis*. New York: Cambridge University Press. Chapter 8.

Weinberg, B.A. (2007) 'Social interactions with endogenous associations'. *NBER Working Paper* No. 13038, http://www.nber.org/papers/w13038.pdf.

Williams, R.A. (1981) 'Peer influence vs. peer selection: An attempted separation'. *CDE Working Paper* 81–20, http://www.ssc.wisc.edu/cde/cdewp/8120.pdf.

Wing, R.R. and Jeffery, R.W. (1999) 'Benefits of Recruiting Participants With Friends and Increasing Social Support for Weight Loss and Maintenance'. *Journal of Consulting and Clinical Psychology,* 67(l): 132–138.

Wooldridge, J. (2002) *Econometric Analysis of Cross Section and Panel Data*. Cambridge: MIT Press.

Young, H. P. (2009) 'Innovation diffusion in heterogeneous populations: Contagion, social influence, and social learning'. *American Economic Review,* 99(5): 1899–924.

Zalta, E.N. (2008) *The Stanford Encyclopedia of Philosophy*. The Metaphysics Research Lab Center for the Study of Language and Information, Stanford University, http://plato.stanford.edu/. Especially the entries on 'Aristotle on Causality', 'David Hume', 'Causal Processes' amd 'Counterfactual Theories of Causation'.

Zeng, Z. and Xie, Y. (2008) 'A preference-opportunity-choice framework with applications to intergroup friendship'. *American Journal of Sociology,* 114(3): 615–48.