Computational and Structural Biotechnology Journal 18 (2020) 1754–1760

COMPUTATIONAL AND STRUCTURAL BIOTECHNOLOGY JOURNAL

journal homepage: www.elsevier.com/locate/csbj

Review

Statistical methods for the estimation of contagion effects in human disease and health networks

Ran Xu

Department of Allied Health Sciences, University of Connecticut, Storrs, CT, United States

ARTICLE INFO

Article history:
Received 24 February 2020
Received in revised form 17 June 2020
Accepted 19 June 2020
Available online 25 June 2020

Keywords:
Contagion effects
Networks
Omitted variable bias
Stochastic actor-oriented models
Instrumental variables
Latent-space models

ABSTRACT

Contagion effects, sometimes referred to as spillover or influence effects, have long been central to the study of human disease and health networks. Accurate estimation and identification of contagion effects are important in terms of understanding the spread of human disease and health behavior, and they also have various implications for designing effective public health interventions. However, many challenges remain in estimating contagion effects and it is often unclear when it is difficult to correctly estimate contagion effects, or why a particular method would need to be applied. In this review I explain the challenges in estimating contagion effects, and how they can be framed as an omitted variable bias problem. I then discuss how such challenges have been addressed in randomized experiments and traditional statistical analyses, as well as several state-of-the-art statistical methods. Finally, I conclude by summarizing recent advancements and noting remaining challenges, as well as appropriate next steps.

Contents

1. Introduction

Contagion effects, sometimes referred to as spillover or influence effects, have long been central to the study of human disease and health networks. Such effects are defined as the propensity for an individual's behavior (or disease state) to vary along with the prevalence of that behavior in a reference group [1] such as one's social contacts. Contagion effects have received much attention and have been widely studied in various phenomena such as the spread of health knowledge and behavior (e.g. smoking and registering in health forums) [2,3], health status and disease (e.g. obesity and acquired immunodeficiency syndrome (AIDS)) [4,5], and psychological states (e.g. depression) [6,7]. Accurate estimation and identification of contagion effects are important in terms of understanding the spread of human disease and health behavior, and they also have various implications for designing effective public health interventions.

However, many challenges remain in estimating contagion effects, especially from observational network data, because it is difficult to separate the effect of contagion or influences from other

https://doi.org/10.1016/j.csbj.2020.06.027

processes that operate simultaneously.[1] That is, when we observe that people in close relationships or interactions tend to be similar in health behavior or a disease state, it is difficult to identify the underlying mechanisms that generate these patterns. One mechanism could be influence and contagion [8–10] whereby individuals assimilate the behavior of their network partners. Another mechanism could be selection — in particular, homophily [11,12], in which individuals seek to interact with similar others. Furthermore, there could be some common social-environmental factors -- individuals with previous similarities select themselves into the same social settings (e.g. hospital or alcoholics anonymous (AA) support group), and actual network formation just reflects the opportunities of meeting in this social setting [13–15].[2]

Entanglement among these different mechanisms unavoidably induces bias when we estimate contagion effects [16]. Various statistical methods and recent advancements in the field of social network analysis have attempted to reduce the bias in estimating contagion effects, such as instrumental variable (IV) methods [17], propensity score methods [18], stochastic actor-oriented models (SAOMs) [19], and the latent space-adjusted approach [20]. Although each method potentially leverages extra information in the data to reduce bias, they all have individual strengths and weaknesses, and none can claim to eliminate all sources of bias. Furthermore, considerable misconception remains regarding when it is difficult to correctly estimate contagion effects, or why a particular method would need to be applied.

In the following sections, I first explain the challenges in estimating contagion effects and how they can be framed as an omitted variable bias problem. I then discuss how such challenges have been addressed in randomized experiments and traditional statistical analyses. Several state-of-the-art statistical methods for estimating contagion effects follow, including instrumental variable methods and stochastic actor-oriented models and methods that create proxies for the omitted variables. Finally, we close by summarizing recent advancements and noting remaining challenges, as well as appropriate next steps.

2. Challenges in estimating contagion Effects: An omitted variable bias problem

Similarities of health behavior, disease state, and characteristics of two individuals in a network relationship can be caused by three primary mechanisms—contagion/influence, homophilous selection, or common social or environmental factors [21]. While it is possible to rule out some mechanisms through random treatment assignment or networks in experiments, entanglement among these different mechanisms makes it difficult to correctly estimate the contagion effect from observational data. The challenges in estimation caused by entanglement among contagion effects and common social-environmental factors can be easily framed as an omitted variable bias problem (e.g., ignoring the group or environment individuals belong to when estimating the contagion/influence model). What is less obvious is that entanglement between the contagion/influence and the homophilous selection can also essentially be framed as an omitted variable bias problem [20]. Shalizi and Thomas have shown that when there is an unobserved trait that co-determines both behavior and network choice, contagion effects are generally unidentifiable mainly because contagion/influence and homophily (selection) are generically confounded through this unobserved trait [16].

For example, assuming that the frequency of individual i's (ego) drug use at time t, $drug\_use_{it}$, is the outcome of interest. It is a function of his/her previous drug use, $drug\_use_{it-1}$, his/her friend j's (alter) previous drug use, $drug\_use_{jt-1}$ (ego's network exposure, i.e. contagion/influence), and an unobserved tendency for substance abuse (arrow D in Fig. 1). At the same time, there is a homophilous selection based on this unobserved substance-abuse tendency in the network -- individuals with similar levels of substance-abuse tendency are more likely to be friends (arrow A in Fig. 1). As a result, person j's drug use, which is a function of person j's substance-abuse tendency (arrow B_j), will be correlated with person i's substance-abuse tendency through homophilous selection (arrow C in Fig. 1). However, as the substance-abuse tendency is unobserved, it violates the key assumption of most estimation methods (i.e., the omitted variable should not correlate with the independent variables) such that the estimates of the contagion effects will be biased and inconsistent. Simulation evidence from Xu showed a substantial upward bias in the estimates of contagion effects when an omitted variable is present in both the behavioral and network selection model and a downward bias in the estimates of contagion effects when the omitted variable only presents in the behavioral model while the network is static [20].

3. Randomized network experiments

As other scientific fields, one of the cornerstone approaches to minimizing bias in the estimation of the contagion effect is the use of randomized experiments. There are primarily two types of randomized experiments designed to estimate contagion effects. The first type randomly assigns individuals' networks. In this type of experiment, as one's network partner is randomized, there are no unobserved factors that drive network selection, and behavior and selection are no longer entangled. For example, Centola randomly assigned participants to specific network structures and found that individuals are much more likely to adopt health behaviors when social reinforcements exist in the networks [3]. Sacerdote randomly assigned roommates to students and found that roommates socially influenced subjects' grade point averages as well as decisions to join social groups such as fraternities [22]. Similarly, Kremer and Levy found that males randomly assigned to roommates that drank alcohol prior to college had a lower grade point average (GPA) than those assigned to nondrinking roommates [23].

The second type of randomization preserves subjects' pre-existing social contacts but randomly assigns subjects to treatment (intervention) conditions. By randomizing alters' (those who have network connections with the ego/focal user) behavior, the ego's network exposure is no longer correlated with the omitted variable that affects the ego's behavior or network, and contagion effects can thus be identified. For example Aral and Walker randomly manipulated whether Facebook users received notifications that their friends had adopted a particular product to study peers' influence on product adoption and found that younger and married individuals were less susceptible to influence [24]. In another example, Kramer et al. randomly assigned positive and negative expressions that Facebook users received from their friends and found that emotions are contagious via social networks [25]. For a more thorough review, see [21].

[1] Assuming contagion is simultaneous rather than lagged poses additional challenges. Manski concluded that identification is difficult if not impossible when simultaneous influence exists, requiring strong assumptions to make statistical and causal inferences. This is not the focus of this paper, and we assume there are some lags in transmission of contagion effects.

[2] There are also structural constraints such as transitivity and preferential attachment that could cause people to become friends. However, these mechanisms alone do not entangle influence (e.g., one befriends another due to high popularity but they have different behaviors). In these cases, another mechanism must be present to induce similarity between these friends (e.g., selection of common friends based on similarity in attributes), and thus the entanglement goes back to the original three mechanisms: influence, selection based on homophily, and social-environmental factors.

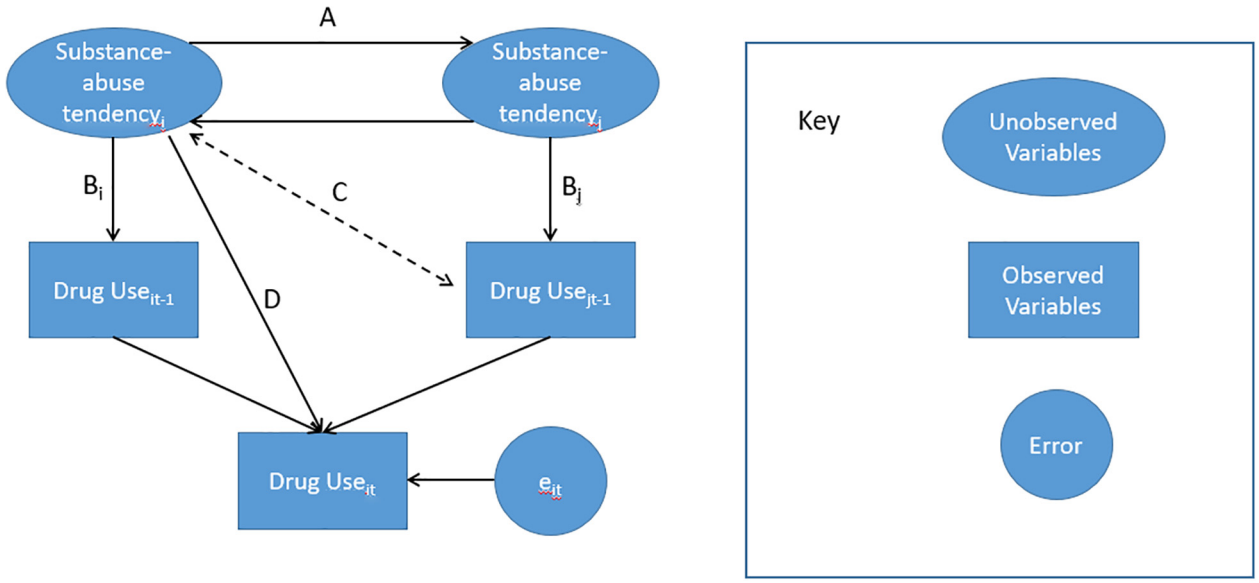

Fig. 1. Contagion effects are unidentifiable due to the omitted variable bias.

While randomized experiments have provided a sound basis for inference, there are several limitations to conducting randomized experiments to study network processes [26]. First, it is often challenging to conduct randomized experiments due to ethical concerns or logistical constraints [27–29]. Second, concerns have often been raised about violating the stable treatment unit value assumption (SUTVA) or the no-spillover effect in network studies, which presumably focus on behavior or disease states transferred from one person to another (e.g., [30]). A third concern is that while most experiments assume that network relations are static, most network relations are actually continuously dynamic in real world settings and people constantly make decisions regarding new interactions and terminating existing ones. For example Carell et al. found a negative treatment effect for the students they intended to help because these students avoided the peers with whom the designed-group intended them to interact and instead formed more homogeneous sub-groups [31]. Finally, while randomized assignments do apply to most human experiments in question, virtually all of them have been conducted on non-random populations. Thus, although employing randomization may represent gains in internal validity, such gains might be outweighed by a loss in external validity or representativeness by selecting a specific sample [26,32].

4. Conventional statistical methods

A conventional statistical model that is widely adopted in estimating contagion/influence effect has origins in Friedkin & Johnsen [33]:

$$Y_{it} = \beta_0 + \beta_1 Y_{it-1} + \beta_2 A_{ijt-1} Y_{jt-1} + \beta_3 X_{it} + e_{it} \quad (1)$$

where Y_{it} is the behavior of i at time t, Y_{it-1} is the behavior of *i at time t-1*, and Y_{jt-1} is the behavior of *alter j at time t-1*. A_{ijt-1} is a weighting matrix based on the observed social network at time t-1. Together $A_{ijt-1}Y_{jt-1}$ represent the network exposure of individual i. $A_{ijt-1}Y_{jt-1}$ can take many different forms. Two popular forms include the average of network partners' behavior or the sum of network partners' behavior, and X_{it} represents other time varying or invariant concurrent variables that might affect behavioral outcome Y. Correspondingly, β_2 represents the contagion effect of interest.

This relatively simple yet effective model can take many different forms of specifications to accommodate a binary outcome, a count variable, and time-to-event data [34,35]. Although this model does not directly model the selection process—and there could still be strong homophily in the selection process—this model controls for the lagged dependent variable Y_{it-1} and other covariates X_{it} that are likely to drive both the behavior and selection model. As long as all of the variables affecting the behavioral outcome or both the behavior and selection are observed and controlled using this relatively simple model, contagion effects can be correctly estimated [16,20]. However, it is rare that researchers can obtain all relevant variables in the real world, especially with observational data.

We also note here this model assumes a lagged, rather than simultaneous, contagion effect. That is, the model assumes contagion does not occur in the same period in the discrete time data. To model contagion effects as simultaneous, Christakis et al. represent the contagion model on a dyadic level [5] as:

$$Y_{it} = \beta_0 + \beta_1 Y_{it-1} + \beta_2 Y_{jt-1} + \beta_3 Y_{jt} + \beta_4 X_{it} + e_{it} \quad (2)$$

Specifically, they use the lagged measure of an alter's behavior Y_{jt-1} to control for homophily and simultaneous measure of the alter's behavior Y_{jt} to represent contagion. However, this method is potentially problematic for two reasons. First, the inclusion of both lagged and simultaneous measures of the alter's behavior can cause difficulties in both estimation and interpretation [36]. For example, an unbiased estimation of simultaneous contagion effects requires specific structural constraints or a maximum likelihood estimation based on an equilibrium model [1,37]. Second, it remains likely that this model omits other variables that drive both behavior and network selection.

Another similar but more sophisticated approach is through propensity score matching. Specifically Aral et al. estimated the contagion effect of adoption of a mobile service-application in a large global instant-messaging network [18]. Individuals with friends adopting the mobile service application were considered the treatment group while those without were considered control group. They modeled the probability of receiving treatment as

$$P(T_{it} = 1 | X_{it}) = \frac{\exp[\alpha_{it} + \beta_{it} X_{it} + e_{it}]}{1 + \exp[\alpha_{it} + \beta_{it} X_{it} + e_{it}]} \quad (3)$$

where T_{it} is the treatment status of individual i on day t and X_{it} represents the vector of demographic and behavioral covariates of i. Each treated individual is then matched to an untreated individual with the most similar predicted probability of receiving treatment from model (3). Matched pairs are thus "equally likely to have a certain number of adopter friends because of observed and correlated latent homophily, contrasting them on the sole dimension of their neighbors' actual adoption status" [18]. Contagion is estimated as the difference in the probability of adoption among matched pairs. This method is appealing as it deliberately controls for the selection process by modeling the choice of friends (adopt vs. not adopt) as a function of various demographic and behavioral variables. However, as with other propensity score-matching methods, this model also requires a "strong ignorability assumption" [38], meaning there are no omitted variables that affect the treatment status—in this case, the network selection. Hence this method suffers from the same limitations as the previous methods.

5. Stochastic actor-oriented models

Given the limitations of traditional statistical methods, many new estimation methods and procedures have been developed over time. One of the models that has received much attention in recent years is called stochastic actor-oriented model (SAOM) [19]. SAOM is a type of micro-simulation model that characterizes how actors change both their network relations and behavior over time. In principle, it is similar to an agent-based model [39], but the SAOM uses statistical inference to estimate the network dynamics and contagion/influence effects from observational data. Specifically, in the simulation process, an SAOM assumes the underlying time is continuous and that actors control their behavior and outgoing ties. At a given moment, one probabilistically selected actor has the opportunity to change one outgoing tie or small step in his or her behavior. The change follows a Markov process in which small changes in networks and behavior are accumulated in each micro step, and large differences can then be observed between initial and final networks [19]. For statistical inference, the parameter values of the simulation algorithms are selected such that the simulated and observed data resemble each other most closely [40]. Specifically the parameters can be estimated by matching key statistics of the simulated and observed networks via method of moments, generalized method of moments, or likelihood-based methods [41].

Each micro-step change for each actor has two parts: a change opportunity process and a change determination process [19]. The change opportunity process decides the rate of actors to make changes, and an actor's waiting time until the next micro step of either kind is exponentially distributed with parameter

$$\lambda_{total} = \sum_i \lambda_i^{network} + \lambda_i^{behavior} \tag{4}$$

where $\lambda_i^{network}$ decides actor i's rate to change networks and $\lambda_i^{behavior}$ decides actor i's rate to change behavior. Both parameters can be functions of an actor's network positions (e.g., centrality) as well as individual characteristics (e.g., age and sex).

The change determination process decides an actor's choice of tie or behavior when the actor has an opportunity to make a change. Specifically, the choice probability of network ties follows a multinomial logit shape and can be expressed as

$$\exp\left[\sum_k \beta_k s_{ki}(x)\right] / \sum_{x'} \exp\left[\sum_k \beta_k s_{ki}(x')\right] \tag{5}$$

here, x represents the state of the network and $s_{ik}(x)$ represents the various effects based on network x for actor i, such as reciprocity, transitivity, centrality, and homophily based on various characteristics. The sum in the denominator extends over all possible next network states x' [40]. Similarly, the probability of actor i choosing a specific behavior can be expressed as

$$\exp\left[\sum_k \beta_k s_{ki}(z)\right] / \sum_{z'} \exp\left[\sum_k \beta_k s_{ki}(z')\right] \tag{6}$$

where z represents the behavior state and the sum in the denominator extends over all possible next behavior states z'. The behavior effects are represented by $s_{ki}(z)$, such as the similarity between the ego's behavior and the alter's average behavior, which represents the behavior assimilation process, and the corresponding parameter estimate represents the contagion or influence effect.

SAOM is appealing as it intuitively incorporates both the influence- and network-selection process from an individual-level perspective, such that the network-selection effects are adjusted for in estimation of contagion effects. SAOM has been applied to study various contagion phenomena such as the spread of smoking, marijuana use, and disease infection [42–44]. However, as the model is based on a simulation algorithm, computation can be very time consuming, especially for relatively large networks. The model is also relatively inflexible as researchers can only use pre-defined network and behavioral terms. More importantly, although SAOM simultaneously models contagion and network-selection processes, it can still suffer from the aforementioned omitted variable bias problem, and contagion estimates from SAOMs are not more conservative relative to other conventional methods [45]. As Steglich et al. pointed out, estimates from SAOMs are still biased when "non-observed variables co-determine the probabilities of change in network and/or behavior [40]."

6. Instrumental variable methods

Neither conventional statistical methods nor SAOMs can account for the bias in the estimation of contagion effects induced by the omitted variable in the behavioral (and network selection) model. Instrumental variable (IV) methods also have the potential to address the omitted variable-bias problem, and have been widely used across many different fields in the social sciences, including Mendelian randomization in epidemiology and bioinformatics [46]. IV methods are used in situations in which explanatory variables are correlated with unobserved error terms that can be caused by simultaneity, omitted variables, measurement errors, and so forth. IV methods work through identifying a set of new variables (instruments) that only correlate with endogenous explanatory variables, but not directly with outcome variables or the unobserved error terms. As such, the methods achieve consistent estimation by "blocking out" the correlation between the endogenous variable and the unobserved errors [47,48].

A handful of studies have used IV methods to estimate contagion effects. The instrumental variables used in these studies can be broadly grouped into two categories: substantive and structural. Substantive IVs are usually theoretical or empirical constructs (e.g., individual characteristics) that only correlate with an alter's behavior, but not directly with an ego's behavior of interest or unobserved variables that affect an ego's behavior/network selection. For example, Duncan et al. used a friend's intelligence as an instrument for the friend's occupational and educational aspirations to estimate peer effects on an individual's aspirations [49]. O'Malley et al. used genetic alleles as IVs for friends' BMI to estimate social contagion effects on weight status [50]. An used friends' family smoking status as an IV for friends' smoking status to estimate peer effects on smoking [51]. However, all of these methods require a strong theoretical argument of the validity of the instrumental variables—that is, the IVs only correlate with an

alter's behavior but not with an ego's behavior or unobserved variables through other routes. In many cases, these assumptions are essentially untestable. In addition, weak instrument problems often occur when the IV is only weakly correlated with an alter's behavior, which induces inconsistency problems and large standard errors of estimates, especially with small sample sizes [48,52].

On the other hand, structural IVs exploit structural properties of networks or the data to identify instrumental variables. For example, Bramoullé et al. argued that if there are intransitive triads in the observed network—for example i– > j– > k, but i and k are not connected—then i's outcome can be used as an instrument for j to estimate contagion effects for k's outcome, since k is not directly influenced by i [17]. In another example, Xu utilized the dynamic nature of the longitudinal data and argued that if the omitted variable that determines behavior/selection is constant over time, the omitted variable can be removed by first-differencing the model [20]. In such a case, contagion effects can then be correctly estimated by using all the past outcome values as IVs, as proposed by Arellano and Bond [53]. However, identification of the contagion effects in these models would also require somewhat strong assumptions and data requirements. In Bramoullé et al. [17], the validity of the IV depends on the assumption that i does not influence k through any alternative path, and simulations have shown that the quality of IV estimates decreases with denser networks and complex functions of intransitivity. In Xu [20], the validity of the IV depends on the assumption that the unobserved variable is constant over time; that errors in the future are independent of past values of the outcome; and that errors are serial-independent, which make them difficult to test. In addition, simulation evidence also shows substantial instability in the contagion estimates when the number of time points is small.

7. Incorporating the omitted variable in the estimation procedure

Finally, another school of thought is to directly deal with the omitted variable that drives the behavior/network selection. By directly incorporating the omitted variable in the estimation process of contagion effects, it is likely that the bias in the contagion-effects estimates will be substantially reduced. There are two possible approaches here. The first is to incorporate the omitted variable in the estimation procedure. Although the omitted variable is unobserved, it can be modeled as a latent construct in a structural equation model (SEM) framework, a widely used method often employed as an alternative approach to deal with unobserved variables [54]. Xu adapted a SEM model from Bollen & Brand and modeled the contagion process in a SEM framework with the omitted variable represented as a latent construct of the ego's behavior over time [20,55]. In principle, this method is appealing as it accounts for the entanglement among contagion/influence and other processes though incorporating the correlation between the omitted variable and the alter's behavior in the estimation procedure. Simulation evidence has shown that this method can produce a small bias in contagion-effects estimates, but the performance is unstable when the number of time points is small or the true coefficient is large.

Another approach is to create a proxy variable for the unobserved variable that drives both the behavior and network selection. Intuitively, if there is any information about the unobserved variable from the selection process, it can be extracted and used in the estimation of the contagion/influence process, which will reduce the bias in the contagion-effects estimates. Liu and Chen employed network- embedding algorithms from the machine learning approach and represented the network topology structure as actor attributes in a low-dimensional space [56]. Specifically, they employed a factorization- based algorithm, GraRep; a random walk-based algorithm, node2vec; and a deep learning-based algorithm, Structural Deep Network Embedding (SDNE), to extract the low-dimensional embedding vector (representing network topology) as actors' attributes and included these attributes as additional covariates in the contagion-effects estimation using a conventional statistical model as in model (1). Simulation results show that although none of these methods completely removed the bias in the estimates of contagion effects, controlling for these embedding vectors significantly improved the estimation when compared with no control at all. However, a potential problem with this machine-learning approach is that the embedding vectors represent the network topology generated by all possible mechanisms, including patterns generated by factors already accounted for in the contagion estimation, such as homophily based on the observed variables. As a result, the embedding vectors can often be a poor measurement of the unobserved variables that drive both behavior and network selection.

A more careful differentiation of the network topology generated by observed variables vs. unobserved variables has also been proposed. Specifically, Shalizi & McFowland and Xu built on the theoretical logic of latent space models as applied to social-network data [20,57,58]. Latent space models assume that each individual has a "latent position" that lies in an unobserved n-dimensional social space, and the probability of interaction between any two actors depends on the latent positions of these two actors. Specifically, they take a logistic form and specify the selection model as

$$logodds(Z_{ij} = 1 | c_i, c_j, x_{ij}, \alpha, \beta) = \alpha + \beta' x_{ij} - D(c_i, c_j) \quad (7)$$

here, Z_{ij} indicates whether there is a network tie from i to j, x_{ij} is a vector of observed covariates (at the dyadic or node level), c indicates the latent social position of i and j, and $D(c_i, c_j)$ represents the distance between i and j's latent position such that a smaller distance indicates a higher likelihood of having a tie. For any pair of i and j, a smaller distance between the latent social position and the unobserved variable (homophily) results in a higher likelihood of a network tie and vice versa. Thus, by accounting for all the observed covariates in the latent space model, when two individuals are close to each other in terms of the unobserved variable, they are more likely to have a network tie and they should also be close to each other in terms of latent social positions. Székely et al. have shown that for two one-dimensional variables X and Y (e.g., latent social position and the unobserved variable that drives both behavior and network selection), if the distance correlation (e.g., correlation between them (e.g., correlation between |Xi-Xj| and |Yi-Yj|) is 1, then one can be written as a linear function of the other [59]. Thus the estimated latent social positions from the latent space model can be included as a proxy for the unobserved variable in the estimation of contagion effects.

Shalizi & McFowland showed that if the network grows according to a continuous latent space model, latent homophilous attributes can then be consistently estimated and controlling for these latent attributes allows for unbiased and consistent estimation of contagion/influence effects in additive- influence models [57]. Simulation evidence from Xu showed that when there is a time-invariant unobserved variable that co-determines behavior and network selection, the estimated latent social positions can be good proxies for the unobserved variable [20]. Moreover, the latent space-adjusted approach outperforms other state-of-art estimation approaches in producing the smallest bias and standard error of the contagion effect in a dynamic linear-in-mean influence model as in model (1). The results are also robust to the inclusion of additional covariates, structural properties (e.g., transitivity) in

networks, different scaling of the latent space model, or even misspecifications [20].

The latent space-adjusted approach is appealing, as carefully differentiating the unobserved variables from the observed ones in the selection process, as well as the inclusion of latent position as additional covariates, is flexible and can be applied to any specification of the contagion/influence model. However, the approach also has several limitations. One limitation is that the latent space-adjusted approach requires the same unobserved variable in both the behavioral and selection models. It does not account for the unobserved variables that are only present in one of the processes but not the other (though such variables are usually less of a concern). A second limitation is that the choice of dimensions of latent social space in the latent space model is not clear. Although Xu [20] chose one-dimensional latent social positions, this does not need to be the case and there is no clear rule on how many dimensions users should use. A third limitation is that computation of latent social position is very time consuming, and the time increases significantly with the increase of data or the number of dimensions in the latent social position.

8. Summary and outlook

Studying the spread of disease and the contagion/social influence of health phenomena is at the center of social network studies. The correct estimation of contagion effects is crucial in terms of advancing scientific knowledge, informing public policy, and designing effective health interventions. In this review, we describe how estimation of contagion effects can be biased due to its entanglement with homophilous selection or common social or environmental factors, and how this challenge in estimation can be framed as an omitted variable bias problem. We then introduced various methods with the potential to correctly estimate contagion effects, including adjusting for observed covariates through conventional statistical models, isolating contagion from other confounding factors through randomized experiments or instrumental variables, modeling the contagion and selection process at the same time through stochastic actor-oriented models, and incorporating the omitted variable in the estimation procedure through machine learning or a latent space model. Each method has unique strengths and weaknesses and researchers should carefully employ these methods based on the characteristics of the data collected. For example, if we have a rich set of covariates and longitudinal data, conventional methods and an SAOM can often yield good results. The instrumental variable method is a viable option if observed data is limited but theoretically sound instrumental variables are present. Alternatively, the latent space-adjusted approach can be employed if none of the previous conditions apply and the primary concern is entanglement between the contagion and selection process.

There are also many new challenges, especially with the development of technology and the growth of "big data." All of the methods in this review were originally designed to deal with small-to-midsize infrequent human interaction data. With the development of new bio sensors and social sensors, we can collect much more biological, physiological, social and network information around humans, and we can collect this information at a much higher frequency. Similarly, with the popularization of social media data and other technologies, the size of the network in question is also expanding, often by tens of millions. This causes at least two unique problems. First, the scaling problem poses various challenges to an estimation, especially for computationally intensive methods such as SAOMs and the latent space-adjusted approach. New algorithms are required to handle big social network data more efficiently. Additionally, as individuals interact differently in large networks, new models must be developed to account for network characteristics such as nesting and clustering. Second, with big network and biological/physiological variables of interest, new theories must be developed to account for how humans interact differently in big network, as well as how the contagion process operates for biological and physiological traits. Note that we did not specify a contagion/influence model in this review, as there are various specifications of contagion/influence models and the "correct" specification is critically dependent upon the theories of the contagion/influence process. For example, the contagion of sexually transmitted disease is fundamentally different from learning and engaging in weight loss behavior through role modeling or peer support, and we need different theories to specify contagion models of the two processes. Thus the development of new theories of contagion is at the heart of correctly specifying and estimating an appropriate contagion model.

With the improvement of capacity to harness data, the development of scientific theories, and appropriate statistical methods, the scientific community will continuously improve its understanding of the contagion and influence process in human disease and health networks.

CRediT authorship contribution statement

Ran Xu: Conceptualization, Writing - original draft, Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Manski CF. Identification of endogenous social effects: The reflection problem. Rev Econ Stud 1993;60:531–42.
[2] Christakis NA, Fowler JH. The collective dynamics of smoking in a large social network. N Engl J Med 2008;358:2249–58.
[3] Centola D. The spread of behavior in an online social network experiment. Science 2010;329:1194–7.
[4] Klovdahl AS. Social networks and the spread of infectious diseases: the AIDS example. Soc Sci Med 1985;21:1203–16.
[5] Christakis NA, Fowler JH. The spread of obesity in a large social network over 32 years. N Engl J Med 2007;357:370–9.
[6] Cacioppo JT, Fowler JH, Christakis NA. Alone in the crowd: the structure and spread of loneliness in a large social network. J Pers Soc Psychol 2009;97:977–91.
[7] German D, Sutcliffe CG, Sirirojn B, Sherman SG, Latkin CA, Aramrattana A, et al. Unanticipated effect of a randomized peer network intervention on depressive symptoms among young methamphetamine users in Thailand. J Community Psychol 2012;40:799–813.
[8] Friedkin NE. Choice shift and group polarization. Am Sociol Rev 1999;1:856–75.
[9] Friedkin NE. Norm formation in social influence networks. Soc Networks 2001;23:167–89.
[10] Oetting ER, Donnermeyer JF. Primary socialization theory: the etiology of drug use and deviance. I. Subst Use Misuse 1998;33:995–1026.
[11] McPherson JM, Smith-Lovin L. Homophily in voluntary organizations: Status distance and the composition of face-to-face groups. Am Sociol Rev 1987;1:370–9.
[12] McPherson M, Smith-Lovin L, Cook JM. Birds of a feather: Homophily in social networks. Annu Rev Sociol 2001;27:415–44.
[13] Feld SL. Social structural determinants of similarity among associates. Am Sociol Rev 1982;1:797–801.
[14] Feld SL. The focused organization of social ties. Am J Sociol 1981;86:1015–35.
[15] Kalmijn M, Flap H. Assortative meeting and mating: Unintended consequences of organized settings for partner choices. Soc Forces 2001;79:1289–312.
[16] Shalizi CR, Thomas AC. Homophily and contagion are generically confounded in observational social network studies. Sociol Method Res 2011;40:211–39.
[17] Bramoullé Y, Djebbari H, Fortin B. Identification of peer effects through social networks. J Econometrics 2009;150:41–55.
[18] Aral S, Muchnik L, Sundararajan A. Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. PNAS 2009;106:21544–9.

[19] Snijders TA, Van de Bunt GG, Steglich CE. Introduction to stochastic actor-based models for network dynamics. Soc Networks 2010;32:44–60.
[20] Xu R. Alternative estimation methods for identifying contagion effects in dynamic social networks: A latent-space adjusted approach. Soc Networks 2018;54:101–17.
[21] VanderWeele TJ, An W. Social networks and causal inference. In: Morgan SL, editor. Handbook of causal analysis for social research. Dordrecht: Springer; 2013. p. 353–74.
[22] Sacerdote B. Peer effects with random assignment: Results for Dartmouth roommates. Q J Econ 2001;116:681–704.
[23] Kremer M, Levy D. Peer effects and alcohol use among college students. J Econ Perspect 2008;22:189–206.
[24] Aral S, Walker D. Identifying influential and susceptible members of social networks. Science 2012;337:337–41.
[25] Kramer AD, Guillory JE, Hancock JT. Experimental evidence of massive-scale emotional contagion through social networks. PNAS 2014;111:8788–90.
[26] Frank KA, Xu R (in press) Causal Inference for Social Network Analysis. Oxford Handbook of Social Network Analysis. Oxford: Oxford University Press.
[27] Cook TD. Randomized experiments in educational policy research: A critical examination of the reasons the educational evaluation community has offered for not doing them. Educ Eval Policy 2002;24:175–99.
[28] Cook TD. Why have educational evaluators chosen not to do randomized experiments? Ann Am Acad Polit Sci 2003;589:114–49.
[29] Rubin DB. Estimating causal effects of treatments in randomized and nonrandomized studies. J Educ Psychol 1974;66:688.
[30] Sun M, Penuel WR, Frank KA, Gallagher HA, Youngs P. Shaping professional development to promote the diffusion of instructional expertise among teachers. Educ Eval Policy 2013;35:344–69.
[31] Carrell SE, Sacerdote BI, West JE. From natural variation to optimal policy? The importance of endogenous peer group formation. Econometrica 2013;81:855–82.
[32] Frank KA, Maroulis S, Duong M, Kelcey B. What would it take to change an inference? using rubin's causal model to interpret the robustness of causal inferences. Educ Eval Policy 2013;35:437–60.
[33] Friedkin NE, Johnsen EC. Social influence and opinions. J Math Sociol 1990;15:193–206.
[34] Hall JA, Valente TW. Adolescent smoking networks: the effects of influence and selection on future smoking. Addict Behav 2007;32:3054–9.
[35] Iyengar R, Van den Bulte C, Valente TW. Opinion leadership and social contagion in new product diffusion. Market Sci 2011;30:195–212.
[36] Lyons R. The spread of evidence-poor medicine via flawed social-network analysis. Statistics Polit Policy 2011;2(1).
[37] Doreian P. Linear models with spatially distributed data: Spatial disturbances or spatial effects?. Sociol Method Res 1980;9:29–60.
[38] Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. Biometrika 1983;70:41–55.
[39] Macy MW, Willer R. From factors to actors: Computational sociology and agent-based modeling. Annu Rev Sociol 2002;28:143–66.
[40] Steglich C, Snijders TA, Pearson M. Dynamic networks and behavior: separating selection from influence. Sociol Methodol 2010;40:329–93.
[41] Snijders T, Steglich C, Schweinberger M. Modeling the coevolution of networks and behavior. In: van Montfort K, Oud J, Satorra A, editors. Longitudinal models in the behavioral and related sciences. Abingdon: Routledge; 2017. p. 41–71.
[42] Schaefer DR, Haas SA, Bishop NJ. A dynamic model of US adolescents' smoking and friendship networks. Am J Public Health 2012;102:e12–8.
[43] Tucker JS, de la Haye K, Kennedy DP, Green HD, Pollard MS. Peer influence on marijuana use in different types of friendships. J Adolesc Health 2014;54:67–73.
[44] Silk MJ, Croft DP, Delahay RJ, Hodgson DJ, Weber N, Boots M, et al. The application of statistical network models in disease research. Methods Ecol Evol 2017;8:1026–41.
[45] Ragan DT, Osgood DW, Ramirez NG, Moody J, Gest SD. A comparison of peer influence estimates from SIENA stochastic actor–based models and from conventional regression approaches. Sociol Method Res 2019;0049124119852369.
[46] Katan MB. Apolipoprotein E isoforms, serum cholesterol, and cancer. Lancet 1986;1:507–8.
[47] An W. Models and methods to identify peer effects. In: Scott J, Carrington PJ, editors. The Sage handbook of social network analysis. London: Sage; 2011. p. 515–32.
[48] Wooldridge JM. Econometric analysis of cross section and panel data. Boston: MIT press; 2010. p. 1045.
[49] Duncan OD, Haller AO, Portes A. Peer influences on aspirations: A reinterpretation. Am J Sociol 1968;74:119–37.
[50] O'Malley AJ, Elwert F, Rosenquist JN, Zaslavsky AM, Christakis NA. Estimating peer effects in longitudinal dyadic data using instrumental variables. Biometrics 2014;70:506–15.
[51] An W. Instrumental variables estimates of peer effects in social networks. Soc Sci Res 2015;50:382–94.
[52] Bound J, Jaeger DA, Baker RM. Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. J Am Stat Assoc 1995;90:443–50.
[53] Arellano M, Bond S. Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. Rev Econ Stud 1991;158:277–97.
[54] Kaplan D. Structural equation modeling: Foundations and extensions. London: Sage Publications; 2008. p. 255.
[55] Bollen KA, Brand JE. A general panel model with random and fixed effects: A structural equations approach. Soc Forces 2010;89:1–34.
[56] Liu Y, Chen X. Estimation of peer influence effect in online games using machine learning approaches. International Conference on Information Resources Management Proceedings, 2019.
[57] Shalizi CR, McFowland III E (2018) Estimating causal peer influence in homophilous social networks by inferring latent locations. arXiv preprint arXiv:1607.06565.
[58] Hoff PD, Raftery AE, Handcock MS. Latent space approaches to social network analysis. J Am Stat Assoc 2002;97:1090–8.
[59] Székely GJ, Rizzo ML, Bakirov NK. Measuring and testing dependence by correlation of distances. Ann Stat 2007;35:2769–94.