

A DECOMPOSITION ANALYSIS OF DIFFUSION OVER A LARGE NETWORK

KYUNGCHUL SONG
University of British Columbia

Diffusion over a network refers to the phenomenon of a change of state of a cross-sectional unit in one period leading to a change of state of its neighbors in the network in the next period. One may estimate or test for diffusion by estimating a cross-sectionally aggregated correlation between neighbors over time from data. However, the estimated diffusion can be misleading if the diffusion is confounded by omitted covariates. This paper focuses on the measure of diffusion proposed by He and Song (2022, Preprint, arXiv:1812.04195v4 [stat.ME]), provides a method of decomposition analysis to measure the role of the covariates on the estimated diffusion, and develops an asymptotic inference procedure for the decomposition analysis in such a situation. This paper also presents results from a Monte Carlo study on the small sample performance of the inference procedure.

1. INTRODUCTION

Diffusion of people's or firms' choices over a social or an industrial network has drawn attention in economics, sociology, and marketing. Examples include diffusion of technology or product recommendations over social or industrial networks. (See, e.g., Leskovec, Adamic, and Huberman (2007), Conley and Udry (2010), Banerjee et al. (2013), and de Matos, Ferreira, and Krackhardt (2014) to name but a few.)¹

Disentangling the role of covariates from the true causal effects has been a primary concern in almost every study of causal inference. For example, the propensity score method in program evaluations attempts to measure the effect of a social program after "eliminating the confounding effect" of covariates. (See Rosenbaum and Rubin (1983). See also Imbens and Wooldridge (2009) for a literature review on program evaluations.) In such situations, the role of a covariate

I thank Mahdi Ebrahimi Kahou for his valuable comments at the beginning of this research, and Yige Duan for excellent assistance in this research, including numerous helpful comments on this work. I also thank the Co-Editor and two anonymous referees for criticisms and suggestions. I acknowledge that this research was supported by the Social Sciences and Humanities Research Council of Canada. Address correspondence to Kyungchul Song, Vancouver School of Economics, University of British Columbia, 6000 Iona Drive, Vancouver, BC V6T 1L4, Canada; e-mail: kysong@mail.ubc.ca.

¹In this paper, we focus only on the diffusion of observed state-switches such as the purchase of a product or the adoption of a technology. Hence, we do not consider in this paper the diffusion of information or news over a network where the spread of the information or news does not manifest itself through certain observable choices.

is determined by its influence on the program participation by *the individual to which the covariate belongs*. However, in studies of social interactions or social networks, what matters for causal inference is the relation between covariates and outcomes not only for the sample unit that the covariates belong to, but also of their *neighboring units*. Such a relation arises when the network is formed based on homophily on the covariates. For example, suppose a network is formed among students roughly based on their parents' income, so that a student from a high-income family is more likely to be a friend of another student from a high-income family than from a low-income family. When one observes the purchase of an expensive smartphone of a particular brand by students over two periods, the correlation of purchases between friends over time does not necessarily indicate diffusion of purchases over the network; this can merely be due to the fact that the purchases mostly come from students from high-income families.

Spurious diffusion caused by covariates has received attention in the literature. For example, Aral, Muchnik, and Sundararajan (2009) attempt to distinguish influence-based contagion and homophily-based diffusion. They find that the peer influence is generally overestimated when homophily effect is ignored. Shalizi and Thomas (2011) point out challenges arising from the confounding of social contagion, homophily, and the influence of individual traits. It is not hard to see that failing to condition on covariates that play a crucial role in network formation through homophily and individual's decisions would lead to bias in the measurement of diffusion.

This paper's goal is to develop a method for a decomposition analysis that can be used to gauge the significance of covariates in measuring diffusion. The main idea is analogous to the idea of using a Hausman test to check the omitted variable bias in a linear regression model. Suppose that the parameter of interest is a coefficient in the linear regression model, and one would like to see whether there is any impact of omitting a subset of regressors on the estimated parameter of interest. For this, one can compare the two estimated parameters, one with all the regressors included and the other with a subset of regressors omitted, to see the role of the omitted regressors.

In the same spirit, in this paper, we define *relational diffusion* Δ_S as follows:²

$\Delta_S = \text{Identified Diffusion with Omitted Covariates} - \text{True Diffusion}$,

that is, Δ_S represents the difference between the identified diffusion with omitted covariates and the true diffusion.³ The relational diffusion, Δ_S , gauges the impact of omitting a covariate upon the measurement of diffusion. This will reveal whether a specific covariate is a significant source of relational diffusion. However, it

²I thank Peter Phillips for suggesting this terminology. Relational diffusion captures the cross-sectional dependence of outcomes that is either due to their diffusion caused by covariates which are related to each other causally or due to the noncausal cross-sectional dependence of covariates which influence the outcomes.

³ Δ_S will be clear later, the subscript S denotes the index set of omitted covariates, i.e., S is such that $X_{i,s}, s \in S$, is omitted, where $X_{i,k}$ is the k th entry of the covariate vector $X_i \in \mathbf{R}^d$.

is not immediately clear how to disentangle the role of covariates in this way. Here, unlike the omitted variable bias in a linear regression model, the magnitude of relational diffusion is related to the cross-sectional dependence structure of covariates. (For example, students from high-income families are friends to each other.) One idea would be to compare two conditional covariances, both between observed outcomes and previous-period outcomes of their neighbors, where one is conditioned on the full set of covariates and the other on the set of covariates with the covariate of interest omitted. The main difficulty with this approach is that inference requires knowledge of the cross-sectional dependence structure among the covariates of different cross-sectional units, but this dependence structure is rarely known in practice. In many applications, there is no reason to believe that this dependence structure coincides with the network over which the diffusion arises.

In order to overcome this difficulty, we adopt the approach of Kuersteiner and Prucha (2013) and use conditional probabilities for inference, where we condition on the entire cross section of covariates so that our inference is robust to the unknown cross-sectional dependence structure of covariates. For the concreteness of the procedure, we focus on the measure of diffusion called average diffusion at the margin (ADM), which was recently introduced by He and Song (2022) and shown to be identified by a spatiotemporal dependence measure. For the analysis of relational diffusion, we decompose the spatiotemporal dependence measure constructed with some covariates omitted into the ADM and the gap (denoted by Δ_S). If the omission of the covariates causes no relational diffusion, we must have $\Delta_S = 0$. Thus, the role of covariates is determined by whether Δ_S is zero or not. This paper develops asymptotic inference on Δ_S for each index set S of omitted covariates, and shows that it is asymptotically valid under regularity conditions. We also provide a multiple testing procedure that selects the covariates such that $\Delta_S \neq 0$ with the asymptotic control of the familywise error rate (FWER). (See Section 9.1 of Lehmann and Romano (2005) for the definition of FWER.) This framework of decomposition analysis is carefully designed so that all the quantities are defined conditional on the covariates so that the unknown cross-sectional dependence structure of covariates does not affect the asymptotic validity of inference.

This paper provides results from a small-scale Monte Carlo simulation study. The study investigates the finite sample performance of asymptotic confidence intervals using networks generated according to the preferential attachment random graph generation model of Barabási and Albert. (See Jackson (2008, Sect. 5.2).) The results show a reasonably stable behavior of finite sample coverage probabilities. The simulation studies also show that the more aligned the cross-sectional dependence structure of covariates is to the contact network over which diffusion arises, the larger the relational diffusion becomes.

The literature of epidemiology, sociology, and economics studied diffusion of various phenomena such as disease, information, and technology. (See Chapter 17

of Newman (2010) for a review of the models and the literature.) Recent contributions in economics that study diffusion over a network include Banerjee et al. (2019), Akbarpour, Malladi, and Saberi (2020), Beaman et al. (2021), and Sadler (2020). This paper's study of diffusion as a causal parameter is closely related to the recent literature on causal inference with network interference. See van der Laan (2014), Aronow and Samii (2017), and Leung (2020). (We refer the readers to He and Song (2022) for a more extensive literature review in this area.) This paper's causal inference framework basically follows He and Song (2022), but departs from the paper by developing a formal way of quantifying the role of covariates in causing the spuriousness of diffusion. This requires a substantial modification of their procedure.

The paper is organized as follows. The next section explains the causal framework for analysis of diffusion, introduces a spatiotemporal dependence measure for each set of covariates, and provides a decomposition of the measure into a component due to the covariates and a residual. The section then concludes by establishing identification of diffusion and explaining the role of cross-sectional dependence of covariates in creating relational diffusion. Section 3 focuses on inference on diffusion decomposition. The section offers asymptotic inference on the component that is due to the covariates and provides conditions for its asymptotic validity. Section 4 presents and discusses results from a Monte Carlo simulation study. Section 5 concludes. Mathematical proofs are collected in the Appendix.

2. DIFFUSION OVER A NETWORK AND IDENTIFICATION OF CAUSAL EFFECTS

2.1. Diffusion over a Contact Network

Let us consider a generic model of diffusion of binary actions over a large network of people as follows. There are two states, 0 and 1, and everybody starts with the default state of 0. For example, the diffusion may be about that of a certain farming technology over a network of farmers, where the state of 0 represents the nonadoption of the technology, and 1 represents its adoption. Each person's binary action $A_{i,t} = 1$ at time t records a *switch of the state* at time t from state 0 to state 1. We assume that the switched state is *irreversible* in the sense that the switch of the state can happen only once. Hence, if $A_{i,t} = 1$ at some time $t \geq 0$, we have $A_{i,s} = 0$, for all $s > t$. This is the case, especially when the switch of the state is defined to be the switch of a state *for the first time*. The binary actions spread over a network over time.

To formalize this process, suppose that there is a directed network called a *contact network* over a set $N = \{1, \dots, n\}$ of people, where each neighborhood of a person represents the set of people whose influence person is directly exposed to. More specifically, we denote the contact network by $G_{\text{ctt}} = (N, E_{\text{ctt}})$ (with subscript "ctt" mnemonic for "contact"). The edge set E_{ctt} consists of edges ij , where the presence of an edge ij in E_{ctt} means that person i is exposed to the direct influence

from j .⁴ We denote the in-neighborhood of person i by⁵

$$N_{\text{ctt}}(i) = \{j \in N : ij \in E_{\text{ctt}}\}, \tag{2.1}$$

which represents the set of people whose influence person i is directly exposed to. The contact network describes whose actions in one period potentially affect whose actions in the next period. Thus, each person i 's binary action $A_{i,t} \in \{0, 1\}$ is a function of $(A_{j,t-1})_{j \in N_{\text{ctt}}(i)}$ for a set $N_{\text{ctt}}(i)$ of neighbors in G_{ctt} , and her own state vector $U_{i,t} \in \mathbf{R}^d$: for $t = 1, 2, 3, \dots, T$,

$$A_{i,t} = \begin{cases} \rho_{i,t}((A_{j,t-1})_{j \in N_{\text{ctt}}(i)}, U_{i,t}), & \text{if } A_{i,s} = 0, \text{ for all } s = 0, 1, \dots, t-1, \\ 0, & \text{otherwise,} \end{cases} \tag{2.2}$$

for some map $\rho_{i,t}$. Here, $A_{i,t} = 1$ represents the switch of the state of person i (i.e., an ‘‘action’’ by person i) at time t . Suppose that person i has switched the state at some time $s = 0, 1, \dots, t-1$, so that $A_{i,s} = 1$. Since this switched state is irreversible, no further switch of the state is allowed for this person after time $t-1$. Hence, this person should have $A_{i,t} = 0$.

The diffusion process in (2.2) is a generalized version of a threshold model of diffusion studied in the literature. (See Granovetter (1978). See also a recent contribution by Acemoglu, Ozdaglar, and Yildiz (2011) for an example.) A special case of this model is a linear threshold model where the map $\rho_{i,t}$ is given by

$$\rho_{i,t}((A_{j,t-1})_{j \in N_{\text{ctt}}(i)}, U_{i,t}) = 1 \left\{ \sum_{j \in N_{\text{ctt}}(i)} p_{ij} A_{j,t-1} \geq U_{i,t} \right\}, \tag{2.3}$$

where p_{ij} is a weight that individual i gives to j .

2.2. The Researcher’s Observation

The researcher observes each person’s state at time $t = 0$, which is denoted by $Y_{i,0}$, and her state at time $t = t_1$, which is denoted by $Y_{i,1}$. These observed binary states are related to the state-switches as follows:

$$Y_{i,0} = A_{i,0}, \text{ and } Y_{i,1} = \sum_{s=0}^{t_1} A_{i,s}. \tag{2.4}$$

Recall that $Y_{i,0} = A_{i,0} \in \{0, 1\}$. By the irreversibility of state-switches, each person can switch the state at most once, which implies that $Y_{i,1} \in \{0, 1\}$. Hence, $Y_{i,1} = 1$ if and only if person i is in state 1 at time t_1 . When $Y_{i,1} = 0$, this means that the person i has never switched the state including the initial period. The researcher

⁴The notion of a directed edge ij from j to i is taken from the notation in Newman (2010) where the graph is represented by an adjacency matrix A and its (i, j) th entry A_{ij} is 1 if and only if there is an edge from j to i .

⁵The in-neighborhood in a directed graph refers to a set of neighbors whose edges with the person i are from the neighbors to the person i .

does *not* observe the diffusion process in real time. The researcher observes the states of people at two time periods $t = 0$ and $t_1 = 1$.

Our setting accommodates information diffusion where $A_{j,0}$ represents the indicator of a person j who receives information first in the network, and $A_{i,t}$, $t \geq 1$, the indicator of certain binary action (such as purchase of a good) by person i in time t . However, we require that both $Y_{i,0}$ and $Y_{i,1}$ as defined in (2.4) are observed at some time t_1 for each person i . We exclude the situation where there is information diffusion and we do not know who the initial receivers of the information are.

Let us introduce a graph that represents the causal connections between *observed actions* $Y_{i,0}$'s and $Y_{i,1}$'s. We can trace the actions at a given time t_1 back to the initial actions at time 0. To see this, first let $N_{\text{cau}}(i)$ denote the set of people such that each $j \in N_{\text{cau}}(i)$ is connected to i along the contact network G_{ctt} , i.e., there exist $i_1, i_2, \dots, i_{t-1}, i_t = i$ such that $j \in N_{\text{ctt}}(i_1)$ and

$$i_{s-1} \in N_{\text{ctt}}(i_s), \text{ for } s = 2, \dots, t_1. \tag{2.5}$$

The sets $N_{\text{cau}}(i)$, $i \in N$, define a network, say, $G_{\text{cau}} = (N, E_{\text{cau}})$, where $ij \in E_{\text{cau}}$ if and only if $j \in N_{\text{cau}}(i)$. The set $N_{\text{cau}}(i)$ represents all the people whose initial actions potentially have influenced person i 's decision at time t_1 indirectly through the influences of neighbors in the contact network.⁶ We call the graph G_{cau} the *causal graph* for (Y_0, Y_1) , where $Y_0 = (Y_{0,i})_{i \in N}$ and $Y_1 = (Y_{1,i})_{i \in N}$.

The researcher, however, does not observe the causal graph (or the contact network). Instead, she observes a graph G_{obs} . With regard to the relation between G_{cau} and G_{obs} , we make the following assumption:

Assumption 2.1. The observed graph G_{obs} contains G_{cau} as a subgraph.

This assumption does not require that the observed graph G_{obs} ‘‘approximates’’ the causal graph G_{cau} in any sense. Neither is it required to contain the directional information in the causal relations in $Y_{j,0}$ and $Y_{i,1}$. In fact, the assumption is satisfied if G_{obs} contains an undirected supergraph of G_{cau} as a subgraph. This is convenient, because the observed graph may not capture the direction of causality accurately in practice. The essence of Assumption 2.1 is that it requires the observed graph to capture the cross-sectional dependence among $Y_{i,1}$'s. This assumption is substantially weaker than the assumption for the networks, for example, used in linear-in-means models (e.g., Manski, 1993; Bramoullé, Djebbari, and Fortin, 2009). In these models, it is not enough to assume that the observed network

⁶Suppose that \mathbf{A}_{ctt} is the adjacency matrix of the contact network G_{ctt} such that its (i,j) th entry is given by $[\mathbf{A}_{\text{ctt}}]_{ij} = 1 \{ij \in E_{\text{ctt}}\}$, for each $i, j \in N$, i.e., there is an edge from j to i in G_{ctt} if and only if the (i,j) th entry of \mathbf{A}_{ctt} is one. Then, for $i \neq j$, we have $j \in N_{\text{cau}}(i)$ if and only if

$$\sum_{m=1}^{t_1} [\mathbf{A}_{\text{ctt}}^m]_{ij} > 0.$$

Recall that the (i,j) th entry of $\mathbf{A}_{\text{ctt}}^m$, denoted by $[\mathbf{A}_{\text{ctt}}^m]_{ij}$ here, counts the number of the walks of length m from j to i . Hence, for $i \neq j$, we have $j \in N_{\text{cau}}(i)$ if and only if there is a walk of length less than or equal to m from j to i in the contact network.

contains the true network as a subgraph. (See de Paula, Rasul, and Souza (2020) and Lewbel, Qu, and Tang (2021) for approaches that do not require network data at all.)⁷

Let \mathcal{F} be the σ -field generated by X and the adjacency matrices of G_{ctt} and G_{obs} , where $X = (X_i)_{i \in N}$ is the collection of covariate vectors, $X_i \in \mathbb{R}^p$. Throughout the paper, we assume that the covariates, and the graphs G_{obs} and G_{ctt} are stochastic. We also allow the graphs G_{obs} and G_{ctt} to be a large connected graph, where every pair of people is connected directly or indirectly. However, for asymptotic inference, we require the graphs to be not too dense. We make this assumption precise later.

2.3. Average Diffusion at the Margin

We introduce the causal parameter of interest following the potential outcome approach in program evaluations. By recursively applying the equation in (2.2), we can rewrite $Y_{i,1}$ as a function of $(Y_{k,0})_{k \in N_{\text{cau}}(i)}$ and unobserved heterogeneities:

$$Y_{i,1} = \begin{cases} \tilde{\rho}_i((Y_{k,0})_{k \in \bar{N}_{\text{cau}}(i)}, V_i), & \text{if } A_{i,s} = 0, \text{ for all } s = 0, 1, \dots, t_1 - 1, \\ 0, & \text{otherwise,} \end{cases} \tag{2.6}$$

where the vector $V_{i,t}$ consists of components $U_{j,s}$ with $j \in \bar{N}_{\text{cau}}(i) \equiv N_{\text{cau}}(i) \cup \{i\}$ and $s \leq t$, and $\tilde{\rho}_{i,t}$ is determined as the compositions of maps $\rho_{j,s}$ with $j \in \bar{N}_{\text{cau}}(i)$ and $s \leq t$.

For $d \in \{0, 1\}$, $i \in N$, and $j \in N_{\text{cau}}(i)$, we introduce a potential outcome $Y_{ij}^*(d)$ which is the same as $Y_{i,1}$ except that $Y_{j,0}$ in $\tilde{\rho}_i$ on the right-hand side of (2.6) is replaced by d . (If $j \notin N_{\text{cau}}(i)$, then $Y_{ij}^*(d)$ is simply taken to be $Y_{i,1}$.) This is the state of person i in period t_1 , when the initial action of person j is *counterfactually* fixed to be $d \in \{0, 1\}$. Our focus is on the ADM at $t = t_1$:

$$\text{ADM} = \frac{1}{n} \sum_{j \in N} \sum_{i \in N} \mathbf{E} [Y_{ij}^*(1) - Y_{ij}^*(0) \mid \mathcal{F}]. \tag{2.7}$$

The ADM was introduced by He and Song (2022). It measures the expected increase in the number of switchers when one additional randomly chosen individual j switches her state in the initial period, whereas other people choose their initial actions according to the randomness of the event $A_{i,0} = 1$. The impact of a randomly selected person j changing the initial action from 0 to 1 on the number of the total switchers until time t_1 is measured after integrating out the conditional distribution of other people’s initial actions $A_{i,0}$, $i \neq j$, given \mathcal{F} . Hence, the “average” in the ADM is twofold. The first “average” refers to the expectation over the conditional distribution of $Y_{k,0}$, $k \neq j$, given \mathcal{F} , and then the second average is over the random selection of j .

⁷It is possible to relax this assumption into He and Song (2022) called *Dependency Causal Graphs*. However, we do not pursue this more general framework here.

Suppose that there is no diffusion in the sense that the map $\rho_{i,t}$ in (2.2) does not depend on $(A_{j,t-1})_{j \in N_{\text{ctt}}(i)}$, that is,

$$A_{i,t} = \begin{cases} \rho_{i,t}(U_{i,t}), & \text{if } A_{i,s} = 0, \text{ for all } s = 0, 1, \dots, t-1, \\ 0, & \text{otherwise.} \end{cases} \tag{2.8}$$

Then, $Y_{ij}^*(1) - Y_{ij}^*(0) = 0$, for all $i \in N$ and $j \in N_{\text{cau}}(i)$. Hence, in this case, $\text{ADM} = 0$.

When the conditional probability of an initial switch, $P\{A_{j,0} = 1 \mid \mathcal{F}\}$, is very small, this does not necessarily make the ADM small, because the ADM compares the expected number of switches between two counterfactual scenarios (one with a randomly chosen j being an initial switcher and the other not), and the two scenarios use the same conditional distribution $A_{j,0}$ given \mathcal{F} . However, if $P\{A_{j,0} = 1 \mid \mathcal{F}\}$ is very small, it may affect the quality of the asymptotic inference that we introduce later.

2.4. Identification of the ADM

As mentioned before, we assume that the researcher observes the initial actions, each person’s states by time $t = t_1$, covariates and observed graph G_{obs} . That is, the researcher observes $(Y_{i,0}, Y_{i,1}, G_{\text{obs}}, X)$ for each $i \in N$.

For the initial actions $A_{i,0}$ and unobserved heterogeneity $U_i = (U_{i,1}, U_{i,2}, \dots, U_{i,T})$, we make the following assumption that describes the conditional cross-sectional independence given \mathcal{F} .

Assumption 2.2. $(U_i, Y_{i,0})$ ’s are conditionally independent across i ’s given \mathcal{F} .

The assumption requires that the cross-sectional dependence among $(U_i, Y_{i,0})$ ’s comes solely from the cross-sectional dependence of X_i ’s or characteristics of networks G_{ctt} and G_{obs} . For example, this condition is satisfied if, at each period, the action $A_{i,t}$ is determined by the neighbor’s actions $A_{j,t-1}$ in the previous period, and idiosyncratic unobserved heterogeneities $U_{i,t}$ that are cross-sectionally independent once one condition on the whole covariate vector X and the graphs G_{ctt} and G_{obs} . The covariate X_i can include network characteristics of G_{obs} such as average degrees of agent i or of her neighbors. It can also include an average of the characteristics of the neighbors.

We also assume an analog of an unconfoundedness condition in program evaluations as follows.

Assumption 2.3. For all $i, j \in N$, $i \neq j$, $(Y_{ij}^*(1), Y_{ij}^*(0), G_{\text{ctt}})$ is conditionally independent of $Y_{j,0}$ given (G_{obs}, X) .

This condition is satisfied, for example, if the initial actions are determined solely by (G_{obs}, X) and some other random events that are independent of all other components. See He and Song (2022) for a detailed discussion on this assumption.

For the purpose of the decomposition analysis in this paper, we generalize the notion of the above unconfoundedness condition to accommodate the situation where one omits some covariates. First, let us introduce notation for subvectors

of covariates. Recall that $X_i \in \mathbf{R}^p$, for each $i \in N$, and let $X_{i,k}$ be the k th entry of X_i . Let $\mathbb{S} = \{1, \dots, p\}$. For each $S \subset \mathbb{S}$, let $X_{i,S} = (X_{i,k})_{k \in S}$, $X_{i,-S} = (X_{i,k})_{k \in \mathbb{S} \setminus S}$, $X_S = (X_{i,S})_{i \in N}$, and $X_{-S} = (X_{i,-S})_{i \in N}$. Let us introduce the following notion of unconfoundedness condition.

Definition 2.1. For each $S \subset \mathbb{S}$, we say that S -unconfoundedness holds, if, for all $i, j \in N$, $i \neq j$, $(Y_{ij}^*(1), Y_{ij}^*(0), G_{\text{ctt}}, X_S)$ is conditionally independent of $Y_{j,0}$ given (G_{obs}, X_{-S}) .

The S -unconfoundedness condition is stronger than the unconfoundedness condition in Assumption 2.3. In fact, the S -unconfoundedness satisfies a monotonicity property: if $S \subset S'$, the S' -unconfoundedness implies the S -unconfoundedness.⁸ Hence, the larger the set S is, the S -unconfoundedness condition becomes stronger. Especially, the \emptyset -unconfoundedness corresponds to the unconfoundedness in Assumption 2.3 used by He and Song (2022). At the other extreme, the \mathbb{S} -unconfoundedness corresponds to the randomized control trial where the covariates are entirely irrelevant in the treatment assignment (i.e., the variables $Y_{j,0}$ here).

The rest of the section is devoted to presenting the result that the ADM is identified using only $(Y_0, Y_1, G_{\text{obs}}, X_{-S})$, $Y_0 = (Y_{i,0})_{i \in N}$ and $Y_1 = (Y_{i,1})_{i \in N}$, under the S -unconfoundedness condition. In other words, if the S -unconfoundedness condition holds, then one can identify the ADM with covariates X_S omitted. Hence, omitting X_S does not cause any relational diffusion. Later we develop a decomposition method to quantify the magnitude of relational diffusion, which can be used to test whether the S -unconfoundedness holds or not.

To facilitate the identification analysis, let us make the following assumption on the initial actions $Y_{j,0}$, and covariates.

Assumption 2.4. (i) There exist a known distribution function $F : \mathbf{R} \rightarrow [0, 1]$ and unknown parameter $\gamma_0 \in \mathbf{R}^p$ such that for all $j \in N$,

$$E[Y_{j,0} | G_{\text{obs}}, X] = F(X_j^\top \gamma_0), \tag{2.9}$$

F is nonconstant and has density bounded away from zero.

(ii) For all $j \in N$, the support of X_j is not contained in any proper linear subspace of \mathbf{R}^p , and for any proper subset $S \subset \mathbb{S}$, there exists $c_0 > 0$ such that for all $n \geq 1$,

$$\frac{1}{n} \sum_{j \in N} \lambda_{\min}(X_{j,-S} X_{j,-S}^\top) \geq c_0,$$

where $\lambda_{\min}(A)$ for a symmetric matrix A denotes the minimum eigenvalue of A .

(iii) There exists $c_1 \in (0, 1/2)$ such that $c_1 < F(X_j^\top \gamma_0) < 1 - c_1$ for all $j \in N$.

Assumption 2.4(i) requires that the initial action $Y_{j,0}$ be conditionally independent of X_{-j} given X_j for each $j \in N$. As in the literature of program evaluations

⁸This monotonicity does not hold if one considers instead an alternative, weaker notion of S -unconfoundedness: for all $i, j \in N$, $i \neq j$, $(Y_{ij}^*(1), Y_{ij}^*(0), G_{\text{ctt}})$ is conditionally independent of $Y_{j,0}$ given (G_{obs}, X_{-S}) . The failure of nonmonotonicity in sets S follows from the results in Phillips (1988).

(e.g., Imbens and Wooldridge, 2009), one can view $F(X_j^\top \gamma_0)$ as the parameterized propensity score of person j , i.e., the propensity of the person j to switch the state at time 0. As we explained before, this assumption is not as strong as it appears in our context because one can include other people’s covariates as part of X_j , such as the average of the characteristics of the neighbors in the observed graph G_{obs} . While it is possible to extend our framework to other forms of parametric or semiparametric specifications, the specification (2.9) is most commonly used in practice, and simplifies the proposal of this paper in terms of both exposition and implementation. Assumption 2.4(ii) is typical in the literature of index models, often invoked for identification of γ_0 . (See, e.g., 2.1 of Horowitz (2009).) Assumption 2.4(iii) is analogous to the overlap condition used in the literature of program evaluations, which requires that the probability of the initial switch of actions is bounded away from zero and one. (See, e.g., Imbens and Wooldridge, 2009.) This condition can be violated, when X_j has unbounded support or the dimension of X_j is large. In our context, the assumption is not plausible especially when the diffusion starts with only a very small number of “seed people.” The analysis in Khan and Tamer (2010) who focused on i.i.d. observations can potentially be extended to this case. However, a full development in this direction is outside of the scope of this paper.

He and Song (2022) introduced a spatiotemporal dependence measure of (Y_0, Y_1) as follows:

$$C = \frac{1}{n} \sum_{j \in N} \sum_{i \in N: ij \in E_{\text{obs}}} \frac{\text{Cov}(Y_{j,0}, Y_{i,1} \mid \mathcal{F})}{\text{Var}(Y_{j,0} \mid \mathcal{F})}, \tag{2.10}$$

and showed that

$$\text{ADM} = C,$$

under Assumptions 2.1–2.3. They developed asymptotic inference for C under the parametric propensity score assumption in Assumption 2.4(i). In this paper, we analyze the consequence of omitting covariates in the construction of C and develop ways to measure the impact of the omission.

Let us introduce an analog of C with X_S omitted. For each $S \subset \mathbb{S}$, we let

$$\mu_{j,0,-S} = F(X_{j,-S}^\top \gamma_{-S}^*),$$

where

$$\gamma_{-S}^* = \arg \max_{\gamma_{-S} \in \mathbb{R}^{p-|S|}} Q_S(\gamma_{-S})$$

and

$$Q_S(\gamma_{-S}) = \sum_{j \in N} \mathbf{E} [Y_{j,0} \log F(X_{j,-S}^\top \gamma_{-S}) + (1 - Y_{j,0}) \log (1 - F(X_{j,-S}^\top \gamma_{-S})) \mid \mathcal{F}].$$

It is not hard to see that under the S -unconfoundedness condition, $\gamma_{-S}^* = \gamma_{0,-S}$, where $\gamma_{0,-S}$ is equal to γ_0 except that the entries with indices in S are eliminated.

We have defined $\mu_{j,0,-S}$ using γ_{-S}^* instead of $\gamma_{0,-S}$ so that it is well defined regardless of whether the S -unconfoundedness condition holds or not. Then, the analog of the measure C with covariates X_S omitted can be written as follows:

$$C_{-S} = \frac{1}{n} \sum_{j \in N} \sum_{i \in N: ij \in E_{\text{obs}}} \frac{\mathbf{E}[(Y_{j,0} - \mu_{j,0,-S})Y_{i,1} \mid \mathcal{F}]}{\text{Var}(Y_{j,0} \mid \mathcal{F})}. \tag{2.11}$$

The quantity C_{-S} captures the covariation between $Y_{i,1}$ and the “residuals” from projecting the local weighted average of the period 0 actions $Y_{j,0}$ over in-neighbors on the covariates X_{-S} . The quantity C_{-S} is different from C , due to the subtraction by $\mu_{j,0,-S}$ in the conditional covariance. When $S = \emptyset$, i.e., no covariate is omitted from the vector X_i , C_{-S} is reduced to C .

The theorem below shows that under the S -unconfoundedness condition, ADM is identified as $C_{-S'}$ for all $S' \subset S$.

THEOREM 2.1. *Suppose that Assumptions 2.2–2.4 hold, and the S -unconfoundedness is satisfied for some $S \subset \mathbb{S}$. Then, for all $S' \subset S$,*

$$\text{ADM} = C_{-S'}. \tag{2.12}$$

Suppose that there is no diffusion (i.e., (2.8)) and the S -unconfoundedness is satisfied for some $S \subset \mathbb{S}$. Then, $\text{ADM} = C_{-S'} = 0$ for all $S' \subset S$. In practice, the S -unconfoundedness condition can be too strong. If the condition fails, equation (2.12) is not guaranteed to hold. In other words, the estimated ADM with X_S omitted can be away from zero significantly, even when the true ADM is zero. The omission of X_S creates relational diffusion in this case. By measuring the discrepancy between the ADM and C_{-S} , one can check whether omitting the covariates X_S causes relational diffusion, and quantify its magnitude. In the next subsection, we elaborate this idea.

2.5. Relational Diffusion and Decomposition Analysis

Suppose that the researcher omits X_S from the covariate vectors, and identifies the ADM by C_{-S} . When the S -unconfoundedness fails, the omitted covariates X_S may create what seems like a diffusion phenomenon even when there is no diffusion in reality. To see this more explicitly, let us define

$$\Delta_S \equiv C_{-S} - \text{ADM}.$$

Then, we can write

$$\Delta_S = \frac{1}{n} \sum_{j \in N} \sum_{i \in N: ij \in E_{\text{obs}}} \mathbf{E}[Y_{i,1} \delta_{j,S} \mid \mathcal{F}] \tag{2.13}$$

with

$$\delta_{j,S} = \frac{\mu_{j,0} - \mu_{j,0,-S}}{\mu_{j,0}(1 - \mu_{j,0})}, \tag{2.14}$$

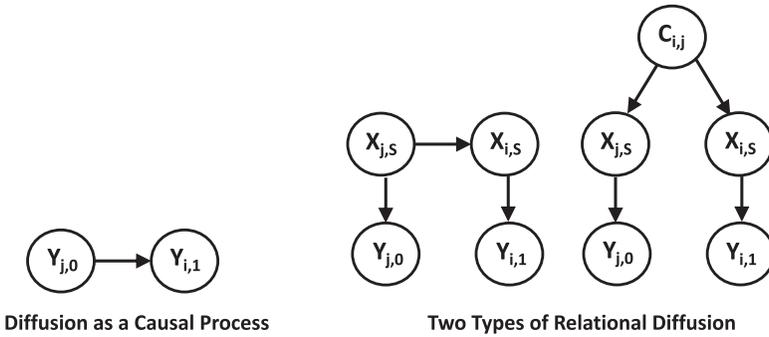


FIGURE 1. Illustration of relational diffusion: The figure illustrates relational diffusion that arises due to the cross-sectional dependence of the covariates. In both figures, the absence of an arrow between a pair of vertices represents the absence of causal relations. Cross-sectional dependence between $X_{j,S}$ and $X_{i,S}$ can arise due to the causal relation between them, or because there is a common factor (denoted $C_{i,j}$ here) which affects both random variables. In either case, nonzero correlation between $Y_{j,0}$ and $Y_{i,1}$ can arise, and contribute to the estimated diffusion, even though there is no causal relation between $Y_{j,0}$ and $Y_{i,1}$.

and $\mu_{j,0} = F(X_j^\top \gamma_0)$. Hence, Δ_S constitutes the remainder term in the decomposition as follows:

$$C_{-S} = ADM + \Delta_S. \tag{2.15}$$

By Theorem 2.1, we have $\Delta_S = 0$ under the S -unconfoundedness condition. However, when the S -unconfoundedness condition fails, the estimated version of C_{-S} can be nonzero, even when ADM is zero. This relational diffusion can be measured by Δ_S .

Omitted variable bias in a linear regression model arises when omitted variables are correlated with other regressors. This correlation is the correlation within the same sample unit. In contrast, relational diffusion arises from the cross-sectional dependence of covariates. (This is illustrated in Figure 1.) Our Monte Carlo simulation results show that the cross-sectional dependence of covariates can play a significant role in determining the magnitude of relational diffusion.

3. INFERENCE ON THE RELATIONAL DIFFUSION

3.1. Estimation of Δ_S

Let us consider estimating the magnitude of the relational diffusion, Δ_S . We use a sample analog of Δ_S as an estimator. First, define

$$\hat{\mu}_{j,0} = F(X_j^\top \hat{\gamma}), \text{ and } \hat{\mu}_{j,0,-S} = F(X_{j,-S}^\top \hat{\gamma}_{-S}), \tag{3.1}$$

where $\hat{\gamma}$ is estimated using maximum likelihood estimator (MLE), i.e.,

$$\hat{\gamma} = \arg \max_{\gamma} \hat{Q}(\gamma), \tag{3.2}$$

and

$$\hat{Q}(\gamma) = \sum_{j \in N} (Y_{j,0} \log F(X_j^\top \gamma) + (1 - Y_{j,0}) \log(1 - F(X_j^\top \gamma))).$$

Similarly, we obtain $\hat{\gamma}_{-S}$ after removing $X_{j,S}$ from the index $X_j^\top \gamma$ in the above maximization.

Then, we construct an estimator of Δ_S as follows:

$$\hat{\Delta}_S = \frac{1}{n} \sum_{j \in N} \sum_{i \in N: ij \in E_{\text{obs}}} Y_{i,1} \hat{\delta}_{j,S}, \tag{3.3}$$

where

$$\hat{\delta}_{j,S} = \frac{\hat{\mu}_{j,0} - \hat{\mu}_{j,0,-S}}{\hat{\mu}_{j,0}(1 - \hat{\mu}_{j,0})}.$$

3.2. Asymptotic Inference on Δ_S

We first establish the asymptotic linear representation of $\sqrt{n}(\hat{\Delta}_S - \Delta_S)$. Let us introduce some notation to simplify the expression of the representation. Let

$$f_j = f(X_j^\top \gamma_0), \quad f_{j,-S} = f(X_{j,-S}^\top \gamma_{-S}^*), \quad \text{and} \quad f'_{j,-S} = f'(X_{j,-S}^\top \gamma_{-S}^*),$$

where f' denotes the first-order derivative of f and f is the density of F that appears in Assumption 2.4. Define

$$\Gamma_j = \frac{f_j}{\sigma_{j,0}^2} X_j, \quad \text{and} \quad \Gamma_{j,S} = \frac{f_{j,-S}}{\sigma_{j,0}^2} X_{j,-S}, \tag{3.4}$$

where $\sigma_{j,0}^2 = \mu_{j,0}(1 - \mu_{j,0})$, and

$$\tilde{U}_j = -H^{-1} \varepsilon_{j,0} \Gamma_j, \quad \text{and} \quad \tilde{U}_{j,S} = -H_S^{-1} \varepsilon_{j,0,-S} \Gamma_{j,S}, \tag{3.5}$$

with $\varepsilon_{j,0} = Y_{j,0} - \mu_{j,0}$, $\varepsilon_{j,0,-S} = Y_{j,0} - \mu_{j,0,-S}$,

$$H = -\frac{1}{n} \sum_{i \in N} \frac{f_i^2 X_i X_i^\top}{\mu_{i,0}(1 - \mu_{i,0})}, \quad \text{and}$$

$$H_S = \frac{1}{n} \sum_{i \in N} \left(\frac{\mu_{i,0}(f'_{i,-S} \mu_{i,0,-S} - f_{i,-S}^2)}{\mu_{i,0,-S}^2} - \frac{(1 - \mu_{i,0})(f'_{i,-S}(1 - \mu_{i,0,-S}) + f_{i,-S}^2)}{(1 - \mu_{i,0,-S})^2} \right) X_{i,-S} X_{i,-S}^\top.$$

(Note that H_S is analogous to the hessian matrix in the misspecified MLE.) Later, we show that under regularity conditions,

$$\sigma_S^{-1} \sqrt{n}(\hat{\Delta}_S - \Delta_S) \rightarrow_d N(0, 1), \tag{3.6}$$

where

$$\sigma_S^2 = \text{Var} \left(\frac{1}{\sqrt{n}} \sum_{i \in N} q_{i,S} \mid \mathcal{F} \right),$$

and (with $\delta_{j,S}$ defined in (2.14))

$$q_{i,S} = \kappa_{1,S} \tilde{U}_i - \kappa_{2,S} \tilde{U}_{i,S} + \left(\sum_{j \in N_{\text{Obs}}(i)} \delta_{j,S} \right) Y_{i,1}, \tag{3.7}$$

with $\mu_{i,1} = \mathbf{E}[Y_{i,1} \mid \mathcal{F}]$,

$$\kappa_{1,S} = \frac{1}{n} \sum_{i \in N} \left(\sum_{j \in N_{\text{Obs}}(i)} (\delta_{j,S}(2\mu_{j,0} - 1) + 1) \Gamma_j^\top \right) \mu_{i,1}, \text{ and}$$

$$\kappa_{2,S} = \frac{1}{n} \sum_{i \in N} \left(\sum_{j \in N_{\text{Obs}}(i)} \Gamma_{j,S}^\top \right) \mu_{i,1}.$$

However, consistent estimation of σ_S^2 is not feasible in our context. To see this, we rewrite

$$\sigma_S^2 = \frac{1}{n} \sum_{i \in N} \sum_{j \in N_{\text{Obs}}(i)} \mathbf{E} \left[(q_{i,S} - \mathbf{E}[q_{i,S} \mid \mathcal{F}]) (q_{i,S} - \mathbf{E}[q_{i,S} \mid \mathcal{F}]) \mid \mathcal{F} \right]. \tag{3.8}$$

In order to estimate this quantity consistently, we should be able to consistently estimate $\mathbf{E}[q_{i,S} \mid \mathcal{F}]$. However, this latter term involves X and is heterogeneous across i 's. Furthermore, we cannot simply model this as a parametric function given X_i , because $Y_{i,1}$ involved in $q_{i,S}$ potentially depends on X in a complex form due to the latent contact network in the diffusion process. Instead, we adopt a conservative inference procedure by using the linear projection of the sample version of $[q_{1,S}, \dots, q_{n,S}]^\top$ onto the range space of $[X_1, \dots, X_n]^\top$ in the euclidean space \mathbf{R}^n . (See (3.9) and (3.10).)

First, we define a sample analog of $q_{i,S}$. Let

$$\hat{f}_j = f(X_j^\top \hat{\gamma}), \quad \hat{f}_{j,-S} = f(X_{j,-S}^\top \hat{\gamma}_{-S}), \text{ and } \hat{f}'_{j,-S} = f'(X_{j,-S}^\top \hat{\gamma}_{-S}),$$

and define

$$\hat{\Gamma}_j = \frac{\hat{f}_j}{\hat{\sigma}_{j,0}^2} X_j, \text{ and } \hat{\Gamma}_{j,S} = \frac{\hat{f}_{j,-S}}{\hat{\sigma}_{j,0}^2} X_{j,-S},$$

where f is the density of F that appears in Assumption 2.4, and $\hat{\sigma}_{j,0}^2 = \hat{\mu}_{j,0}(1 - \hat{\mu}_{j,0})$, and

$$\hat{U}_j = \hat{H}^{-1} \hat{\varepsilon}_{j,0} \hat{\Gamma}_j, \text{ and } \hat{U}_{j,S} = \hat{H}_S^{-1} \hat{\varepsilon}_{j,0,-S} \hat{\Gamma}_{j,S},$$

with $\hat{\varepsilon}_{j,0} = Y_{j,0} - \hat{\mu}_{j,0}$, $\hat{\varepsilon}_{j,0,-S} = Y_{j,0} - \hat{\mu}_{j,0,-S}$,

$$\hat{H} = -\frac{1}{n} \sum_{i \in N} \frac{\hat{f}_i^2 X_i X_i^\top}{\hat{\mu}_{i,0}(1 - \hat{\mu}_{i,0})}, \text{ and}$$

$$\hat{H}_S = \frac{1}{n} \sum_{i \in N} \left(\frac{\hat{\mu}_{i,0}(\hat{f}'_{i,-S} \hat{\mu}_{i,0,-S} - \hat{f}_{i,-S}^2)}{\hat{\mu}_{i,0,-S}^2} - \frac{(1 - \hat{\mu}_{i,0})(\hat{f}'_{i,-S}(1 - \hat{\mu}_{i,0,-S}) + \hat{f}_{i,-S}^2)}{(1 - \hat{\mu}_{i,0,-S})^2} \right) X_{i,-S} X_{i,-S}^\top.$$

Then, the sample analog of $q_{i,S}$ is given by

$$\hat{q}_{i,S} = \hat{\kappa}_{1,S} \hat{U}_i - \hat{\kappa}_{2,S} \hat{U}_{i,S} + \left(\sum_{j \in N_{\text{obs}}(i)} \hat{\delta}_{j,S} \right) Y_{i,1},$$

with

$$\hat{\kappa}_{1,S} = \frac{1}{n} \sum_{i \in N} \left(\sum_{j \in N_{\text{obs}}(i)} (\hat{\delta}_{j,S}(2\hat{\mu}_{j,0} - 1) + 1) \hat{\Gamma}_j^\top \right) Y_{i,1}, \text{ and}$$

$$\hat{\kappa}_{2,S} = \frac{1}{n} \sum_{i \in N} \left(\sum_{j \in N_{\text{obs}}(i)} \hat{\Gamma}_{j,S}^\top \right) Y_{i,1}.$$

To construct a confidence interval, we take the square of the standard error to be

$$\hat{\sigma}_S^2 = \frac{1}{n} \sum_{i_1, i_2 \in N: \bar{N}_{\text{obs}}(i_1) \cap \bar{N}_{\text{obs}}(i_2) \neq \emptyset} (\hat{q}_{i_1,S} - \hat{h}_{i_1,S})(\hat{q}_{i_2,S} - \hat{h}_{i_2,S}), \tag{3.9}$$

where $\bar{N}_{\text{obs}}(i) = N_{\text{obs}}(i) \cup \{i\}$ with $N_{\text{obs}}(i) = \{j \in N : ij \in E_{\text{obs}}\}$,

$$\hat{h}_{i,S} = X_i^\top \hat{\lambda}_S, \text{ and } \hat{\lambda}_S = \left(\frac{1}{n} \sum_{i \in N} X_i X_i^\top \right)^{-1} \frac{1}{n} \sum_{i \in N} X_i \hat{q}_{i,S}. \tag{3.10}$$

Then, the $(1 - \alpha)$ -level confidence interval for Δ_S is given by

$$\mathbb{C}_{1-\alpha} = \left[\hat{\Delta}_S - \frac{z_{1-\alpha/2} \hat{\sigma}_S}{\sqrt{n}}, \hat{\Delta}_S + \frac{z_{1-\alpha/2} \hat{\sigma}_S}{\sqrt{n}} \right], \tag{3.11}$$

where $z_{1-\alpha/2}$ is the $1 - (\alpha/2)$ percentile of $N(0, 1)$.

3.3. Asymptotic Theory

For the asymptotic validity of the confidence interval $C_{1-\alpha}$, we use the following set of assumptions.

Assumption 3.1. (Nondegeneracy) There exists a small $c_2 > 0$ such that the following is satisfied for all $n \geq 1$ and all $S \subset \mathbb{S}$:

$$\sigma_S^2 \equiv \text{Var} \left(\frac{1}{\sqrt{n}} \sum_{i \in N} q_{i,S} \mid \mathcal{F} \right) > c_2. \tag{3.12}$$

Assumption 3.1 requires the nondegeneracy of the distribution of the test statistics. This condition requires that the randomness of q_i (conditional on \mathcal{F}) does not disappear as $n \rightarrow \infty$. Since it is unlikely in practice that the finite sample conditional distribution (given \mathcal{F}) of

$$\frac{1}{\sqrt{n}} \sum_{i \in N} (q_{i,S} - \mathbf{E}[q_{i,S} \mid \mathcal{F}])$$

is degenerate, it appears to be reasonable to use Assumption 3.1 in deriving its asymptotic approximation.

We require conditions for the observed graph G_{obs} as follows.

Assumption 3.2. There exists $k > 0$ such that

$$\max_{i \in N} |N_{\text{obs}}(i)| = O_P((\log(n))^k),$$

as $n \rightarrow \infty$.

Assumption 3.2 requires that the observed network is not too dense. See He and Song (2022) for conditions for a generic network formation model such that Assumption 3.2 is satisfied.

The next set of conditions are regularity conditions used to deal with the estimation error of the MLE $\hat{\gamma}$ and the quasi-MLE $\hat{\gamma}_{-S}$. Define

$$H_0(\gamma) = -\frac{1}{n} \sum_{i \in N} \frac{f^2(X_i^\top \gamma) X_i X_i^\top}{F(X_i^\top \gamma)(1 - F(X_i^\top \gamma))}, \text{ and}$$

$$H_{0,S}(\gamma_{-S}) = \frac{1}{n} \sum_{i \in N} \left(\frac{\mu_{i,0} (f'(X_{i,-S}^\top \gamma_{-S}) F(X_{i,-S}^\top \gamma_{-S}) - f^2(X_{i,-S}^\top \gamma_{-S}))}{F^2(X_{i,-S}^\top \gamma_{-S})} \right) X_{i,-S} X_{i,-S}^\top$$

$$- \frac{1}{n} \sum_{i \in N} \left(\frac{(1 - \mu_{i,0}) (f'(X_{i,-S}^\top \gamma_{-S}) (1 - F(X_{i,-S}^\top \gamma_{-S})) + f^2(X_{i,-S}^\top \gamma_{-S}))}{(1 - F(X_{i,-S}^\top \gamma_{-S}))^2} \right) X_{i,-S} X_{i,-S}^\top.$$

The quantities $H_0(\gamma)$ and $H_{0,S}(\gamma_{-S})$ are the ‘‘hessians’’ of population MLE and quasi-MLE objective functions conditional on X . The following conditions are similar to conditions used in the literature of MLE or MLE under misspecification.

Assumption 3.3. For each $S \subset \mathbb{S}$, the following conditions are satisfied.

- (i) The parameter space Γ for γ_0 is compact, and γ_0 and γ_{-S}^* lie in the interior of Γ .
- (ii) There exists $c_3 > 0$ such that for all $n \geq 1$,

$$\lambda_{\min}(-H_0(\gamma_0)) \geq c_3, \text{ and } \lambda_{\min}(-H_{0,S}(\gamma_{-S}^*)) \geq c_3, \tag{3.13}$$

with probability one.

- (iii) The density f of F is log-concave.
- (iv) F is three times continuously differentiable with bounded derivatives, and for any compact set $K \subset \mathbf{R}$, there exists a constant $c_K > 0$ that depends only on K such that

$$\inf_{z \in K} F(z)(1 - F(z)) > c_K.$$

The assumption below puts a condition on the covariate vector X_i . As our object of interest Δ_S is defined in terms of conditional probability given X , we do not require any condition on the cross-sectional dependence structure of X_i 's.

Assumption 3.4. There exists constant $C > 0$ such that for all $n \geq 1$,

$$\max_{j \in N} \|X_j\| \leq C. \tag{3.14}$$

The bounded support condition on X_i has been used in the literature. (See, e.g., Hirano, Imbens, and Ridder (2003).) The following theorem establishes that the confidence interval $\mathbb{C}_{1-\alpha}$ defined in (3.11) is asymptotically valid.

THEOREM 3.1. *Suppose that Assumptions 2.2–3.4 hold. Then, for each $S \subset \mathbb{S}$,*

$$\liminf_{n \rightarrow \infty} P\{\Delta_S \in \mathbb{C}_{1-\alpha}\} \geq 1 - \alpha.$$

The central part of the asymptotic validity result in Theorem 3.1 comes from the asymptotic normality result in (3.6). To see how this asymptotic normality arises, first, note that we have the following asymptotic linear representation (see Theorem A.1 in the Appendix):

$$\sqrt{n}(\hat{\Delta}_S - \Delta_S) = \frac{1}{\sqrt{n}} \sum_{i \in N} (q_{i,S} - \mathbf{E}[q_{i,S} | \mathcal{F}]) + o_P(1), \tag{3.15}$$

where $q_{i,S}$ is as defined in (3.7). Let us define the graph $G_{\text{obs}}^* = (N, E_{\text{obs}}^*)$ and $E_{\text{obs}}^* = \{i_1 i_2 : N_{\text{obs}}(i_1) \cap N_{\text{obs}}(i_2) \neq \emptyset, i_1 \neq i_2\}$. Under Assumptions 2.1 and 2.2, the quantities $q_{i,S}$ can be shown to have graph G_{obs}^* as a conditional dependency graph given \mathcal{F} .⁹ Then, we can apply the central limit theorem to the right-hand side of (3.15), as long as the observed graph G_{obs} is not too dense, using Theorem 2.4

⁹A triangular array $\xi_{i,n}, i \in N$, is said to have a graph $G_{\text{obs}}^* = (N, E_{\text{obs}}^*)$ as a *conditional dependency graph* given \mathcal{F} , if, for any two subsets A and B of N such that no two nodes $i \in A$ and $j \in B$ are adjacent in G_{obs}^* , $(\xi_{i,n})_{i \in A}$ and $(\xi_{j,n})_{j \in B}$ are conditionally independent given \mathcal{F} .

of Penrose (2003) or Corollary 3.1 of Lee and Song (2019). The required condition for the observed graph G_{obs} is fulfilled by Assumption 3.2.

3.4. A Step-Down Procedure for Detecting the Sources of Relational Diffusion

One might be interested in detecting which set of covariates cause relational diffusion. In this section, we develop a multiple testing procedure that detects the set of such covariates with asymptotic control of FWER. First, let us introduce an individual hypothesis for each covariate index $s \in \mathbb{S}$:

$$H_{0,s} : \Delta_s = 0, \text{ and } H_{1,s} : \Delta_s \neq 0. \tag{3.16}$$

Define

$$S_P = \{s \in \mathbb{S} : \Delta_s \neq 0\}, \tag{3.17}$$

where the subscript P is placed as a reminder that this quantity depends on the conditional distribution of $Y_{i,1}$ given \mathcal{F} . Then, we would like to find a data-dependent random set $\hat{S} \subset \mathbb{S}$ such that

$$\limsup_{n \rightarrow \infty} P \left\{ \hat{S} \not\subset S_P \right\} \leq \alpha. \tag{3.18}$$

We declare the set \hat{S} to be the set of covariates which causes relational diffusion, i.e., $\Delta_s \neq 0$. The probability on the left-hand side of (3.18) is the FWER, which is the probability that there is at least one covariate X_s , with $s \in \hat{S}$, which is falsely declared to be causing a relational diffusion.

Let us consider the following step-down procedure inspired by Romano and Shaikh (2010). For each subset $S \subset \mathbb{S}$, let $\mathbb{Z}_{S,b}$, $b = 1, \dots, B$, be i.i.d. random vectors in $\mathbf{R}^{|S|}$, drawn from $N(0, I_{|S|})$, and let $\hat{\Omega}_S$ be an $|S| \times |S|$ matrix whose entries are given by

$$\hat{\sigma}_{s_1, s_2} = \frac{1}{n} \sum_{i_1, i_2 \in N: \bar{N}_{\text{obs}}(i_1) \cap \bar{N}_{\text{obs}}(i_2) \neq \emptyset} (\hat{q}_{i_1, s_1} - \hat{h}_{i_1, s_1})(\hat{q}_{i_2, s_2} - \hat{h}_{i_2, s_2}), \quad s_1, s_2 \in S, \tag{3.19}$$

and $\hat{q}_{i,s}$ and $\hat{h}_{i,s}$ are entries of $\hat{q}_{i,\mathbb{S}}$ and $\hat{h}_{i,\mathbb{S}}$ corresponding to the covariate index s . Then, we construct $c_{1-\alpha}(S)$ to be the $1 - \alpha$ percentile of $\{\hat{W}_{S,b} : b = 1, \dots, B\}$, where

$$\hat{W}_{S,b} = \max_{s \in S} \frac{\left[|\hat{\Omega}_S^{1/2} \mathbb{Z}_{S,b}| \right]_s}{\hat{\sigma}_s} \tag{3.20}$$

and $\left[|\hat{\Omega}_S^{1/2} \mathbb{Z}_{S,b}| \right]_s$ denotes the s -entry of the vector whose elements are equal to the absolute value of the elements of $\hat{\Omega}_S^{1/2} \mathbb{Z}_{S,b}$.

Setting $\hat{R}_1 = \mathbb{S}$, we recursively define

$$\hat{R}_{t+1} = \left\{ s \in \mathbb{S} : \frac{\sqrt{n}|\hat{\Delta}_s|}{\hat{\sigma}_s} \leq c_{1-\alpha}(\hat{R}_t) \right\}, \quad t = 1, 2, \dots,$$

and we stop when $\hat{R}_{t+1} = \hat{R}_t$, and take $\hat{S} = \mathbb{S} \setminus \hat{R}_t$.

Let us present our result that shows asymptotic control of FWER. Let $\mathbf{q}_{i,S}$ be the $|S|$ -dimensional vector whose entries are given by $q_{i,s}, s \in S$. Define

$$\Omega_{S,i} = \mathbf{E}[(\mathbf{q}_{i,S} - \mathbf{E}[\mathbf{q}_{i,S} | \mathcal{F}])(\mathbf{q}_{i,S} - \mathbf{E}[\mathbf{q}_{i,S} | \mathcal{F}])^\top | \mathcal{F}]. \tag{3.21}$$

We introduce a condition under which the conditional distribution of $\mathbf{q}_{i,S}$ given \mathcal{F} is not degenerate uniformly over $1 \leq i \leq n$ and over $n \geq 1$.

Assumption 3.5. There exists $c_4 > 0$ such that $\min_{1 \leq i \leq n} \lambda_{\min}(\Omega_{S,i}) > c_4$ for all $n \geq 1$.

The following theorem shows that this set \hat{S} controls the FWER asymptotically.

THEOREM 3.2. *Suppose that Assumptions 2.2–3.5 hold. Then,*

$$\limsup_{n \rightarrow \infty} P \left\{ \hat{S} \not\subset S_p \right\} \leq \alpha. \tag{3.22}$$

4. MONTE CARLO SIMULATIONS

4.1. Data Generating Process

Let us first explain the data generating process we use for our Monte Carlo simulation study. First, we generate the contact network G_{ctt} . For this, we choose the adjacency matrix of the contact network as a block diagonal matrix and each block matrix is generated by the Barabási–Albert model which starts with the 20 households per village with an Erdős–Rényi random graph. We treat each block matrix as a village and each node as a household. In total, we have 30 villages and each village has 50 or 200 households. Thus, the total number of the households is either $30 \times 50 = 1,500$ or $30 \times 200 = 6,000$.

We generate the observed graph G_{obs} as follows. The adjacency matrix of the observed graph is set to be a block diagonal matrix constructed as follows. For each block matrix B in the adjacency matrix of the contact network G_{ctt} , we form each block matrix by taking each of its entries to be 1 if and only if the corresponding entry of the matrix $B + B^2 + \dots + B^t$ is nonzero. The graph statistics are presented in Table 1. We fix the realized contact network and the observed graphs, and generate outcomes using the same networks across Monte Carlo simulations. As we are not considering the randomness of the networks in our simulation study, what matters for our purpose is the shape of the realized networks in finite samples, rather than the stochastic property of the random graph models that are used to obtain the realizations.

TABLE 1. Characteristics of the networks.

B–A graph	Contact network		Observed graph	
	$n = 1,500$	$n = 6,000$	$n = 1,500$	$n = 6,000$
Max. deg.	14	33	27	95
Ave. deg.	1.0667	1.2633	2.9373	4.3488
Cluster	0.0275	0.0129	0.1110	0.0991

Notes: This table compares the network characteristics of the contact networks and the observed graphs. Both of them are constructed from 30 independent Barabási–Albert graphs in a block diagonal way. Each block graph is based on the Erdős–Rényi graph of size 20 with $p = 1/19$, and contains 50 or 200 nodes. Thus, the size of the contact networks and the observed graphs is either $30 \times 50 = 1,500$ or $30 \times 200 = 6,000$.

We generate the binary actions $A_t = (A_{i,t})_{i \in N}$ as follows. For each $j \in N$, we specify

$$A_{j,0} = 1 \{F(X_j^\top \gamma_0) \geq U_{j,0}\}, \tag{4.1}$$

where $U_{j,0}$'s are i.i.d. and follow the uniform distribution on $[0, 1]$, and F is the distribution function of $N(0, 1)$. We set $p = 4$ and $\gamma_0 = [0.6, -0.1, -0.3, 0.3]^\top$.

The covariates X_j 's constitute an $n \times p$ matrix $\mathbf{X} = [\mathbf{1}; \mathbf{X}_o]$, where $\mathbf{1}$ is a vector of ones and \mathbf{X}_o is an $n \times (p - 1)$ matrix which is generated as

$$\mathbf{X}_o = (\alpha \mathbf{E}_c + (1 - \alpha)\tilde{\mathbf{E}} + I)\tilde{\mathbf{X}}_o, \quad \alpha \in [0, 1],$$

where $\tilde{\mathbf{X}}_o$ is also an $n \times (p - 1)$ matrix with i.i.d. entries from the uniform distribution on $[0, 1]$, I is an $n \times n$ identity matrix, \mathbf{E}_c is the adjacency matrix of the contact network G_{ctt} , and $\tilde{\mathbf{E}}$ is that of an independently generated Erdős–Rényi graph with the same scale and average degree as \mathbf{E}_c 's.¹⁰ The scalar $\alpha \in [0, 1]$ captures to what extent the cross-sectional dependence of X_j 's is aligned with the contact network G_{ctt} . As α gets closer to 1, the cross-sectional dependence structure of X_j 's is more aligned with the contact network G_{ctt} . When $\alpha = 1$, X_j and X_i are correlated if and only if j and i are adjacent in G_{ctt} . When $\alpha = 0$, the cross-sectional dependence of X_j 's is determined by an independently generated Erdős–Rényi graph. We choose $\alpha \in \{0, 0.5, 1\}$ and see how the choice affects relational diffusion.

In the simulation, we consider a variant of the linear threshold diffusion model in (2.3) as follows: for $t = 1, 2, \dots$,

$$A_{i,t} = \begin{cases} 1\{\delta_0 \bar{A}_{i,t-1} + X_i^\top \beta_0 - U_{i,t} > 0\}, & \text{if } A_{i,t-1} = A_{i,t-2} = \dots = A_{i,0} = 0, \\ 0, & \text{otherwise,} \end{cases} \tag{4.2}$$

¹⁰We calculate the average degree of \mathbf{E}_c by firstly adding up the in-degrees and out-degrees of each node, and then taking average of the total degrees and dividing by 2. In this way, the two adjacent matrices will be approximately equally dense so that varying α will not affect the dispersion of X_j 's or the extensiveness of the cross-sectional dependence substantially.

TABLE 2. The true values of ADM and Δ_S .

		$\delta_0 = 0$			$\delta_0 = 0.5$		
		$\alpha = 1.0$	$\alpha = 0.5$	$\alpha = 0.0$	$\alpha = 1.0$	$\alpha = 0.5$	$\alpha = 0.0$
ADM	$n = 1,500$	0	0	0	0.0374	0.0362	0.0336
	$n = 6,000$	0	0	0	0.0394	0.0373	0.0333
Δ_S	$n = 1,500$	0.2222	0.1625	0.0486	0.2387	0.1765	0.0549
	$n = 6,000$	0.3282	0.2817	0.0603	0.3684	0.3151	0.0701

Notes: We considered 30 villages, where each village is populated by 50 households or 200 households. When $\alpha = 1$, the cross-sectional dependence of covariates is shaped by the contact network, and when $\alpha = 0$, it is entirely unrelated to the contact network. Recall that when $\delta_0 = 0$, there is no diffusion. However, Δ_S is not zero, exhibiting relational diffusion. Relational diffusion is larger when $\alpha = 1, 0.5$ than when $\alpha = 0.0$. This confirms that the magnitude of relational diffusion is related to how similar the cross-sectional dependence structure of the covariates is to the contact network.

where $U_{i,t}$'s are i.i.d. and have the distribution function of $N(0, 1)$, the covariates X_i are the same for the same cross-sectional unit across the short period, and

$$\bar{A}_{i,t-1} = \frac{1}{|N_{\text{ctt}}(i)|} \sum_{j \in N_{\text{ctt}}(i)} A_{j,t-1},$$

where $N_{\text{ctt}}(i)$ is the in-neighborhood of i in the contact network we have generated before. In addition, we choose $\delta_0 \in \{0, 0.5\}$ and set $\beta_0 = [-1, 0.3, -0.4, -0.1]^\top$. For the simulations, we have set $t_1 = 2$, and the Monte Carlo simulation number to be 10,000.

As for the specification of the diffusion model, it is important to note that although we assume that the researcher knows the specification for (4.1), she does not know that $A_{i,t}$'s are generated as in (4.2). In other words, we allow her to be entirely agnostic about the specification of $A_{i,t}$, except that it is generated from a generalized diffusion model of the form in (2.2), and hence the estimation and inference on relational diffusion proposed in this paper does not rely on any information of this particular specification in (4.2).

As for the omitted covariates, we considered $S = \{4\}$, so that we omitted the last entry of the p -dimensional covariate vector X_i . The true values of ADM and Δ_S are presented in Table 2. We computed the true values by simulations using 100,000 simulation draws. Recall that when $\alpha = 1$, the cross-sectional dependence of covariates is shaped by the contact network, and when $\alpha = 0$, it is entirely unrelated to the contact network. When $\delta_0 = 0$, there is no diffusion. However, we see that Δ_S is not zero, exhibiting relational diffusion. The relational diffusion is larger when $\alpha = 1.0, 0.5$ than when $\alpha = 0.0$. This confirms that the magnitude of relational diffusion is related to how similar the cross-sectional dependence structure of the covariates is to the contact network.

TABLE 3. Empirical coverage probabilities.

		$\delta_0 = 0$			$\delta_0 = 0.5$		
		$\alpha = 1.0$	$\alpha = 0.5$	$\alpha = 0.0$	$\alpha = 1.0$	$\alpha = 0.5$	$\alpha = 0.0$
Cov. prob. at 99%	$n = 1,500$	0.9738	0.9777	0.9566	0.9798	0.9771	0.9573
	$n = 6,000$	0.9957	0.9891	0.9867	0.9963	0.9903	0.9849
Cov. prob. at 95%	$n = 1,500$	0.9455	0.9308	0.9035	0.9435	0.9286	0.9038
	$n = 6,000$	0.9834	0.9570	0.9476	0.9842	0.9584	0.9431
Cov. prob. at 90%	$n = 1,500$	0.9001	0.8764	0.8424	0.8991	0.8783	0.8422
	$n = 6,000$	0.9670	0.9150	0.8994	0.9696	0.9149	0.8892
Median CI length	$n = 1,500$	0.3099	0.2440	0.0915	0.3287	0.2335	0.1001
	$n = 6,000$	0.3836	0.2102	0.0672	0.4270	0.2281	0.0753

Notes: We considered 30 villages, where each village is populated by 50 households or 200 households. When $\alpha = 1$, the cross-sectional dependence of covariates is shaped by the observed graph, and when $\alpha = 0$, it is entirely unrelated to the observed graph. The Monte Carlo number was equal to 10,000.

4.2. Estimation and Results

For the Monte Carlo simulations, we have estimated $\hat{\Delta}_S$ as in (3.3). Recall that F is chosen to be the distribution function of $N(0, 1)$. The results of the finite sample coverage probabilities for the confidence intervals are shown in Table 3. When we use 50 households per village, the coverage probability exhibits slight under coverage. However, this coverage probability improves when the number of households is increased to 200 households.

Interestingly, the effect of increase in the number of the households per village depends on α , i.e., whether the cross-sectional dependence structure of the covariates is similar to the contact network or not. When it is similar to the contact network ($\alpha = 1$), the increase in the number of the households increases the length of the confidence intervals. On the other hand, when it is very different from the contact network ($\alpha = 0$), the increase leads to a shorter length of the confidence intervals. Thus, it appears that when the cross-sectional dependence structure of the covariates is aligned with the contact network, the increase in the sample size seems to magnify the standard error in the confidence interval.

Part of this effect should also be coming from the increased neighborhood sizes as the number of households increases. For example, note that as we increase the number of households from 50 to 200, the maximum degree and the average degree of the causal graph increase from 27 and 95 to 2.9373 and 4.3488, respectively. Hence, as the number of households grows, the cross-sectional dependence also becomes more extensive.

5. CONCLUSION

In this paper, we develop a method of quantifying the role of the covariates contributing to relational diffusion. This paper's proposal can be useful in practice

especially when there is a concern about potential bias in the estimated diffusion due to missing covariates. In this situation, one may want to quantify the role of covariates in the estimated diffusion and see whether the role is statistically significant. This paper provides a statistical method that is potentially useful in such a situation.

There are multiple extensions of the paper’s proposal. First, it would be interesting to consider a situation with multiple networks and to measure relational diffusion along each network. Note that in the context of linear spatial models, Drukker, Egger, and Prucha (2022) studied situations with multiple networks and provided asymptotic inference. Second, it could be interesting to investigate whether there exists inference based on permutation on the diffusion decomposition. Conditional on X and G_{obs} , observations are all heterogeneously distributed. Hence, standard nonparametric bootstrap does not work. (See Kojevnikov (2021) for a bootstrap method for network dependence processes.) However, there could be a permutation-based approach that exhibits better finite sample performance than asymptotic inference. This was shown in the Monte Carlo study of Song (2018) in estimating the graph concordance. It would be interesting to see if such a phenomenon extends to this decomposition analysis studied in this paper.

APPENDIX: Mathematical Proofs

Proof of Theorem 2.1. First, we show that $\text{ADM} = \text{C}$. Since $Y_{j,0} \in \{0, 1\}$, we write

$$Y_{i,1}Y_{j,0} = Y_{ij}^*(1)Y_{j,0}, \text{ and } Y_{i,1}(1 - Y_{j,0}) = Y_{ij}^*(0)(1 - Y_{j,0}).$$

Hence, taking conditional expectations given \mathcal{F} , and using Assumption 2.3,

$$\frac{\mathbf{E}[Y_{i,1}Y_{j,0} \mid \mathcal{F}]}{\mu_{j,0}^*} - \frac{\mathbf{E}[Y_{i,1}(1 - Y_{j,0}) \mid \mathcal{F}]}{1 - \mu_{j,0}^*} = \mathbf{E}[Y_{ij}^*(1) - Y_{ij}^*(0) \mid \mathcal{F}],$$

where $\mu_{j,0}^* = \mathbf{E}[Y_{j,0} \mid \mathcal{F}]$. By rearranging terms, we find that the left-hand side is equal to

$$\frac{\text{Cov}(Y_{i,1}, Y_{j,0} \mid \mathcal{F})}{\mu_{j,0}^*(1 - \mu_{j,0}^*)},$$

proving that $\text{ADM} = \text{C}$.

If the S -unconfoundedness holds, we have

$$\mathbf{E}[Y_{j,0} \mid G_{\text{ctt}}, G_{\text{obs}}, X_{-S}] = \mathbf{E}[Y_{j,0} \mid \mathcal{F}].$$

(See, e.g., Lemma 4.2(ii) of Dawid (1979).) Since F is not constant on the support of $X_j^\top \gamma_0$, and the support of X_j is not contained in a proper linear subspace of \mathbf{R}^d by Assumption 2.4, we have $\gamma_{0,S} = 0$, where $\gamma_{0,S}$ is the vector consisting of entries in γ_0 with indexes in S . Hence, $\mathbf{E}[Y_{j,0} \mid \mathcal{F}] = F(X_{j,-S}^\top \gamma_{0,-S})$. Since both $\gamma_{0,-S}$ and γ_{-S}^* maximize $Q_S(\gamma_{-S})$ over γ_{-S} uniquely, we must have $\gamma_{0,S} = \gamma_{-S}^*$. Therefore, $\mu_{j,0,-S} = \mathbf{E}[Y_{j,0} \mid \mathcal{F}]$. This means that $\text{C}_{-S} = \text{C}$. Finally, the S -unconfoundedness implies the S' -unconfoundedness for all $S' \subset S$, yielding the desired result. \square

The rest of the proofs are devoted to proving Theorems 3.1 and 3.2. Throughout the auxiliary results below, we assume that the conditions of Theorem 3.2 hold. (In fact, Assumption 3.5 is used only for the proof of Theorem 3.2.)

LEMMA A.1. For each $S \subset \mathbb{S}$, $\max_{i \in N} \mathbf{E}[q_{i,S}^4 | \mathcal{F}] = O_P(d_{mx}^4)$, where $q_{i,S}$ is as defined in (3.7), and

$$d_{mx} = \max_{i \in N} |N_{\text{obs}}(i)|.$$

Proof. The results follows because $q_{i,S}$ involves a sum over $j \in N_{\text{obs}}(i)$ and this sum is bounded by Cd_{mx} for some constant $C > 0$ that does not depend on n . □

The following lemma gives an asymptotic linear representation of the estimators $\hat{\gamma}$ and $\hat{\gamma}_{0,-S}$.

LEMMA A.2.

$$\begin{aligned} \sqrt{n}(\hat{\gamma} - \gamma_0) &= \frac{1}{\sqrt{n}} \sum_{i \in N} \tilde{U}_i + o_P(1), \text{ and} & \text{(A.1)} \\ \sqrt{n}(\hat{\gamma}_{-S} - \gamma_{-S}^*) &= \frac{1}{\sqrt{n}} \sum_{i \in N} (\tilde{U}_{i,S} - \mathbf{E}[\tilde{U}_{i,S} | \mathcal{F}]) + o_P(1). \end{aligned}$$

Furthermore,

$$\begin{aligned} \max_{j \in N} |\hat{\mu}_{j,0} - \mu_{j,0}| &= O_P(n^{-1/2}), \text{ and} & \text{(A.2)} \\ \max_{j \in N} |\hat{\mu}_{j,0,-S} - \mu_{j,0,-S}| &= O_P(n^{-1/2}). \end{aligned}$$

Proof. For both statements of (A.1) and (A.2), the proof can proceed in the same way as in the proof of Lemmas C.5 and C.6 of He and Song (2022). □

We are prepared to present the asymptotic linear representation of $\sqrt{n}(\hat{\Delta}_S - \Delta_S)$.

THEOREM A.1.

$$\sqrt{n}(\hat{\Delta}_S - \Delta_S) = \frac{1}{\sqrt{n}} \sum_{i \in N} (q_{i,S} - \mathbf{E}[q_{i,S} | \mathcal{F}]) + o_P(1),$$

where $q_{i,S}$ is as defined in (3.7).

Proof. First, let

$$\hat{\Delta}_1 = \frac{1}{n} \sum_{i \in N} \left(\sum_{j \in N_{\text{obs}}(i)} \frac{\hat{\mu}_{j,0}}{\hat{\sigma}_{j,0}^2} \right) Y_{i,1}, \text{ and } \Delta_1 = \frac{1}{n} \sum_{i \in N} \left(\sum_{j \in N_{\text{obs}}(i)} \frac{\mu_{j,0}}{\sigma_{j,0}^2} \right) Y_{i,1}.$$

Furthermore, we define for $S \subset \mathbb{S}$,

$$\hat{\Delta}_{S,2} = \frac{1}{n} \sum_{i \in N} \left(\sum_{j \in N_{\text{obs}}(i)} \frac{\hat{\mu}_{j,0,-S}}{\hat{\sigma}_{j,0}^2} \right) Y_{i,1}, \text{ and } \Delta_{S,2} = \frac{1}{n} \sum_{i \in N} \left(\sum_{j \in N_{\text{obs}}(i)} \frac{\mu_{j,0,-S}}{\sigma_{j,0}^2} \right) Y_{i,1}.$$

Let us write

$$\sqrt{n}(\hat{\Delta}_S - \Delta_S) = \sqrt{n}(\hat{\Delta}_1 - \Delta_1) - \sqrt{n}(\hat{\Delta}_{S,2} - \Delta_{S,2}) + \sqrt{n}\Delta_{S,3},$$

where

$$\Delta_{S,3} = \frac{1}{n} \sum_{i \in N} \left(\sum_{j \in N_{\text{obs}}(i)} \delta_{j,S} \right) \varepsilon_{i,1},$$

where $\varepsilon_{i,1} = Y_{i,1} - \mathbf{E}[Y_{i,1} | \mathcal{F}]$.

First, let us analyze $\sqrt{n}(\hat{\Delta}_1 - \Delta_1)$. We write this as $A_{n1} + A_{n2}$, where

$$A_{n1} = \frac{1}{\sqrt{n}} \sum_{i \in N} \left(\sum_{j \in N_{\text{obs}}(i)} \frac{\hat{\mu}_{j,0} - \mu_{j,0}}{\sigma_{j,0}^2} \right) Y_{i,1} + o_P(1), \text{ and}$$

$$A_{n2} = \frac{1}{\sqrt{n}} \sum_{i \in N} \left(\sum_{j \in N_{\text{obs}}(i)} \mu_{j,0} \left(\frac{1}{\hat{\sigma}_{j,0}^2} - \frac{1}{\sigma_{j,0}^2} \right) \right) Y_{i,1}.$$

The term $o_P(1)$ in A_{n1} is due to (A.2), Assumption 3.2, and the assumption that $\mu_{j,0} \in (c_1, 1 - c_1)$ by Assumption 2.4(iii). Using the first-order Taylor expansion $F(X_j^\top \hat{\gamma})$ around $F(X_j^\top \gamma_0)$, and using Lemma A.2, we obtain that

$$A_{n1} = \frac{1}{n} \sum_{i \in N} \left(\sum_{j \in N_{\text{obs}}(i)} \Gamma_j^\top \right) \mu_{i,1} \frac{1}{\sqrt{n}} \sum_{j \in N} \tilde{U}_j + o_P(1).$$

(Recall the definitions of Γ_j and \tilde{U}_j in (3.4) and (3.5).) Similarly, as for A_{n2} , we obtain that

$$\begin{aligned} A_{n2} &= \frac{1}{\sqrt{n}} \sum_{i \in N} \left(\sum_{j \in N_{\text{obs}}(i)} \frac{\mu_{j,0}(\sigma_{j,0}^2 - \hat{\sigma}_{j,0}^2)}{\sigma_{j,0}^4(1 + o_P(1))} \right) Y_{i,1} \\ &= \frac{1}{\sqrt{n}} \sum_{i \in N} \left(\sum_{j \in N_{\text{obs}}(i)} \frac{\mu_{j,0}(\hat{\mu}_{j,0} - \mu_{j,0})(2\mu_{j,0} - 1)}{\sigma_{j,0}^4(1 + o_P(1))} \right) Y_{i,1} + o_P(1) \\ &= \frac{1}{n} \sum_{i \in N} \left(\sum_{j \in N_{\text{obs}}(i)} \frac{\mu_{j,0} f(X_j^\top \gamma_0) X_j^\top (2\mu_{j,0} - 1)}{\sigma_{j,0}^4(1 + o_P(1))} \right) Y_{i,1} \sqrt{n}(\hat{\gamma} - \gamma_0) + o_P(1). \end{aligned}$$

Again, using Lemma A.2, we conclude that

$$\begin{aligned}
 A_{n2} &= \frac{1}{n} \sum_{i \in N} \left(\sum_{j \in N_{\text{obs}}(i)} \frac{\mu_{j,0} f(X_j^\top \gamma_0) X_j^\top (2\mu_{j,0} - 1)}{\sigma_{j,0}^4} \right) \mu_{i,1} \frac{1}{\sqrt{n}} \sum_{j \in N} \tilde{U}_j + o_P(1) \\
 &= \frac{1}{n} \sum_{i \in N} \left(\sum_{j \in N_{\text{obs}}(i)} \frac{\mu_{j,0} (2\mu_{j,0} - 1)}{\sigma_{j,0}^2} \Gamma_j^\top \right) \mu_{i,1} \frac{1}{\sqrt{n}} \sum_{j \in N} \tilde{U}_j + o_P(1).
 \end{aligned}$$

Hence, we find that

$$\begin{aligned}
 \sqrt{n}(\hat{\Delta}_1 - \Delta_1) &= \frac{1}{n} \sum_{i \in N} \left(\sum_{j \in N_{\text{obs}}(i)} \Gamma_j^\top \right) \mu_{i,1} \frac{1}{\sqrt{n}} \sum_{j \in N} \tilde{U}_j \tag{A.3} \\
 &\quad + \frac{1}{n} \sum_{i \in N} \left(\sum_{j \in N_{\text{obs}}(i)} \frac{\mu_{j,0} (2\mu_{j,0} - 1)}{\sigma_{j,0}^2} \Gamma_j^\top \right) \mu_{i,1} \frac{1}{\sqrt{n}} \sum_{j \in N} \tilde{U}_j + o_P(1).
 \end{aligned}$$

Let us turn to $\sqrt{n}(\hat{\Delta}_{S,2} - \Delta_{S,2})$. We write this as $A_{n1,S} + A_{n2,S}$, where

$$\begin{aligned}
 A_{n1,S} &= \frac{1}{\sqrt{n}} \left(\sum_{j \in N_{\text{obs}}(i)} \frac{\hat{\mu}_{j,0,-S} - \mu_{j,0,-S}}{\sigma_{j,0}^2} \right) Y_{i,1} + o_P(1), \text{ and} \\
 A_{n2,S} &= \frac{1}{\sqrt{n}} \sum_{i \in N} \left(\sum_{j \in N_{\text{obs}}(i)} \mu_{j,0,-S} \left(\frac{1}{\hat{\sigma}_{j,0}^2} - \frac{1}{\sigma_{j,0}^2} \right) \right) Y_{i,1}.
 \end{aligned}$$

Similarly as before, the term $o_P(1)$ in $A_{n1,S}$ is due to (A.2) and the assumption that $\mu_{j,0} \in (c_1, 1 - c_1)$ by Assumption 2.4(iii). Using the same arguments as before, we find that

$$\begin{aligned}
 A_{n1,S} &= \frac{1}{n} \sum_{i \in N} \left(\sum_{j \in N_{\text{obs}}(i)} \Gamma_{j,S}^\top \right) \mu_{i,1} \frac{1}{\sqrt{n}} \sum_{j \in N} (\tilde{U}_{j,S} - \mathbf{E}[\tilde{U}_{j,S} | \mathcal{F}]) + o_P(1), \text{ and} \\
 A_{n2,S} &= \frac{1}{n} \sum_{i \in N} \left(\sum_{j \in N_{\text{obs}}(i)} \frac{\mu_{j,0,-S} (2\mu_{j,0} - 1)}{\sigma_{j,0}^2} \Gamma_j^\top \right) \mu_{i,1} \frac{1}{\sqrt{n}} \sum_{j \in N} \tilde{U}_j + o_P(1).
 \end{aligned}$$

(Recall that $\mathbf{E}[\tilde{U}_j | \mathcal{F}] = 0$ because $\mathbf{E}[\varepsilon_{j,0} | \mathcal{F}] = 0$.) Combining these results with (A.3), we obtain the desired result. \square

Let $\hat{\Delta}_S$ be the $|S|$ -dimensional vector whose entries are given by $\hat{\Delta}_s, s \in S$.

LEMMA A.3.

$$\Omega_S^{-1/2} \sqrt{n}(\hat{\Delta}_S - \Delta_S) \rightarrow_d N(0, I_{|S|}), \tag{A.4}$$

where Ω_S is as defined in (3.21).

Proof. By Theorem A.1, we first write

$$\sqrt{n}(\hat{\Delta}_S - \Delta_S) = \frac{1}{\sqrt{n}} \sum_{i \in N} (\mathbf{q}_{i,S} - \mathbf{E}[\mathbf{q}_{i,S} | \mathcal{F}]) + o_P(1). \tag{A.5}$$

Take $b \in \mathbf{R}^{|S|}$ such that $b^\top b = 1$. Recall the definition of G_{obs}^* after Theorem 3.1. By Assumptions 2.1 and 2.2, $\{q_i\}_{i \in N}$ has G_{obs}^* as a conditional dependency graph given \mathcal{F} , which is a special case of conditional neighborhood dependency introduced in Lee and Song (2019). Let $\sigma^2(b) = b^\top \Omega_S b$, and we apply their Corollary 3.1 and Assumption 3.1 to deduce that

$$\begin{aligned} \sup_{u \in \mathbf{R}} & \left| P \left\{ \frac{1}{\sigma(b)\sqrt{n}} \sum_{i \in N} b^\top (\mathbf{q}_{i,S} - \mathbf{E}[\mathbf{q}_{i,S} | \mathcal{F}]) \leq u \mid \mathcal{F} \right\} - \Phi(u) \right| \\ & \leq C \left(\frac{\sqrt{d_{mx}^* d_{av}^* \mu_3^3}}{n^{1/4}} - \log \left(\frac{d_{mx}^* d_{av}^* \mu_3^3}{\sqrt{n}} \right) \frac{\sqrt{(d_{mx}^*)^2 d_{av}^* \mu_4^4}}{\sqrt{n}} \right), \end{aligned}$$

for some constant $C > 0$ that does not depend on n , where

$$\mu_p^p = \max_{i \in N} \mathbf{E} \left[\left| \frac{b^\top \mathbf{q}_{i,S} - \mathbf{E}[b^\top \mathbf{q}_{i,S} | \mathcal{F}]}{\sigma(b)} \right|^p \mid \mathcal{F} \right].$$

Thus, the desired result follows from this and Assumption 3.2 and the Cramér–Wold device. □

Let $h_{i,S}$ denote the population version of $\hat{h}_{i,S}$ which is defined as follows:

$$h_{i,S} = X_i^\top \lambda_S, \text{ and } \lambda_S = \left(\frac{1}{n} \sum_{i \in N} X_i X_i^\top \right)^{-1} \frac{1}{n} \sum_{i \in N} X_i \mathbf{E}[q_{i,S} | \mathcal{F}].$$

LEMMA A.4.

$$\begin{aligned} \frac{1}{n} \sum_{i \in N} (\hat{q}_{i,S} - q_{i,S})^2 &= O_P(n^{-1} d_{mx}^4), \text{ and} \\ \frac{1}{n} \sum_{i \in N} (\hat{h}_{i,S} - h_{i,S})^2 &= O_P(n^{-1} d_{mx}^4). \end{aligned}$$

Proof. Inspecting the terms in $\hat{q}_{i,S}$, we find that the estimation error of $\hat{q}_{i,S}$ comes from the estimation errors of $\hat{\gamma}$ and $\hat{\gamma}_{-S}$. It is not hard to see from Lemma A.2 that

$$\hat{\kappa}_{1,S} - \kappa_{1,S} = O_P(n^{-1/2} d_{mx}^2), \text{ and } \hat{\kappa}_{2,S} - \kappa_{2,S} = O_P(n^{-1/2} d_{mx}^2).$$

Furthermore,

$$\max_{i \in N} |\hat{U}_i - \tilde{U}_i| = O_P(n^{-1/2}), \text{ and } \max_{i \in N} |\hat{U}_{i,S} - \tilde{U}_{i,S}| = O_P(n^{-1/2}),$$

and

$$\max_{i \in N} \left| \sum_{j \in N_{\text{obs}}(i)} (\hat{\delta}_{j,S} - \delta_{j,S}) \right| = O_P(n^{-1/2} d_{mx}).$$

Collecting these rate results, we find that

$$\frac{1}{n} \sum_{i \in N} (\hat{q}_{i,S} - q_{i,S})^2 = O_P(n^{-1} d_{mx}^4).$$

Thus, from Assumption 3.2, we obtain the first statement of the lemma.

The second statement immediately follows because

$$\sum_{i \in N} (\hat{h}_{i,S} - h_{i,S})^2 \leq \sum_{i \in N} (\hat{q}_{i,S} - q_{i,S})^2.$$

(See the proof of Lemma B.13 of He and Song (2022).) □

Define

$$\eta_{i,S} = \mathbf{q}_{i,S} - \mathbf{h}_{i,S}, \text{ and } \hat{\eta}_{i,S} = \hat{\mathbf{q}}_{i,S} - \hat{\mathbf{h}}_{i,S}, \tag{A.6}$$

where $\hat{\mathbf{q}}_{i,S}$ and $\hat{\mathbf{h}}_{i,S}$ are the vectors having entries $\hat{q}_{i,s}$, and $\hat{h}_{i,s}$, $s \in S$, respectively, and similarly with $\mathbf{h}_{i,S}$. Let $\tilde{\Omega}_S$ be the $|S| \times |S|$ matrix whose (m, ℓ) th entry for $m, \ell = 1, \dots, |S|$ is given by

$$\frac{1}{n} \sum_{i_1, i_2 \in \bar{E}_{\text{obs}}^*} \mathbf{E}[\eta_{i_1,m} \eta_{i_2,\ell} \mid \mathcal{F}],$$

where $\eta_{i_1,m}$ denotes the m th entry of $\eta_{i_1,S}$, and $\bar{E}_{\text{obs}}^* = \{i_1 i_2 : N_{\text{obs}}(i_1) \cap N_{\text{obs}}(i_2) \neq \emptyset\} \cup \{ii : i \in N\}$. Let the s th diagonal entry of $\tilde{\Omega}_S$ be denoted by $\tilde{\sigma}_s^2$.

LEMMA A.5. $\hat{\Omega}_S = \tilde{\Omega}_S + o_P(1)$.

Proof. For $m, \ell = 1, \dots, |S|$, define

$$V_{m\ell} = \frac{1}{n} \sum_{i_1, i_2 \in \bar{E}_{\text{obs}}^*} (\hat{\eta}_{i_1,m} \hat{\eta}_{i_2,\ell} - \eta_{i_1,m} \eta_{i_2,\ell}), \text{ and}$$

$$W_{m\ell} = \frac{1}{n} \sum_{i_1, i_2 \in \bar{E}_{\text{obs}}^*} (\eta_{i_1,m} \eta_{i_2,\ell} - \mathbf{E}[\eta_{i_1,m} \eta_{i_2,\ell} \mid \mathcal{F}]),$$

where $\hat{\eta}_{i_1,m}$ denotes the m th entry of $\hat{\eta}_{i_1,S}$. Then, using Assumption 3.2 and Lemma A.4, and following the same argument as in the proofs of Lemmas B.14 and B.15 of He and Song (2022), we find that

$$V_{m\ell} + W_{m\ell} = o_P(1).$$

Since $V_{m\ell} + W_{m\ell}$ is the (m, ℓ) th element of $\hat{\Omega}_S - \tilde{\Omega}_S$, we obtain the desired result. □

LEMMA A.6.

$$\sup_{\tau \in \mathbf{R}} \left| P \left\{ \hat{W}_{S,b} \leq \tau \mid \mathcal{H} \right\} - P \left\{ \tilde{W}_S \leq \tau \mid \mathcal{F} \right\} \right| = o_P(1), \tag{A.7}$$

where \mathcal{H} denotes the σ -field generated by $(G_{\text{ctt}}, G_{\text{obs}}, X, Y_0, Y_1)$, and

$$\tilde{W}_S = \max_{s \in S} \frac{|\tilde{\Omega}_S^{1/2} \mathbb{Z}_S|}{\tilde{\sigma}_s}. \tag{A.8}$$

Proof. First, for each $i = 1, \dots, n$ and $S \subset \mathbb{S}$, we define

$$\tilde{\Omega}_{S,i} = \mathbf{E} \left[\boldsymbol{\eta}_{i,S} \boldsymbol{\eta}_{i,S}^\top \mid \mathcal{F} \right].$$

By the same arguments in the proof of Lemma B.9 of He and Song (2022), we can see that $\tilde{\Omega}_{S,i} - \Omega_{S,i}$ is positive semidefinite for all $1 \leq i \leq n$. As in (3.14) of Kojevnikov and Song (2022), we find that for any $\nu > 0$,

$$\begin{aligned} & P \left\{ \sup_{\tau \in \mathbf{R}} \left| P \left\{ \hat{W}_{S,b} \leq \tau \mid \mathcal{H} \right\} - P \left\{ \tilde{W}_S \leq \tau \mid \mathcal{F} \right\} \right| > \nu \right\} \\ & \leq \frac{C}{\nu} \left((c_2^2 n)^{-1/6} + n^{-1/2} + P \left\{ \min_{1 \leq i \leq n} \lambda_{\min}(\tilde{\Omega}_{S,i}) < c_4 \right\} \right), \end{aligned}$$

for some constant $C > 0$ that does not depend on n , where $c_4 > 0$ is the constant in Assumption 3.5. Hence, by Assumption 3.5, the last probability in the above display vanishes as $n \rightarrow \infty$. This gives the desired result. □

Define

$$\hat{T}_S = \max_{s \in S} \left| \frac{\sqrt{n}(\hat{\Delta}_s - \Delta_s)}{\hat{\sigma}_s} \right|. \tag{A.9}$$

LEMMA A.7. For any $S \subset \mathbb{S}$,

$$P \left\{ \hat{T}_S \leq \hat{c}_{1-\alpha}(S) \right\} \geq 1 - \alpha + o_P(1). \tag{A.10}$$

Proof. Let $c_{1-\alpha}(S)$ be the $1 - \alpha$ percentile of the conditional distribution of

$$W_S = \max_{s \in S} \frac{|\Omega_S^{1/2} \mathbb{Z}_S|}{\tilde{\sigma}_s}, \tag{A.11}$$

given \mathcal{F} .

First, we show that for all $S \subset \mathbb{S}$,

$$P \left\{ \hat{c}_{1-\alpha}(S) < c_{1-\alpha}(S) \right\} = o(1). \tag{A.12}$$

To see this, note that by Assumption 3.2 and Lemma A.5, for each $\tau \in \mathbf{R}$,

$$P \left\{ \hat{T}_S \leq \tau \mid \mathcal{F} \right\} = P \left\{ \tilde{W}_S \leq \tau \mid \mathcal{F} \right\} + o_P(1). \tag{A.13}$$

Since the conditional density of W_S given \mathcal{F} is bounded uniformly over n , by Assumption 3.5, the term $o_P(1)$ is uniform over $\tau \in \mathbf{R}$. Hence,

$$\sup_{\tau \in \mathbf{R}} \left| P \left\{ \hat{T}_S \leq \tau | \mathcal{F} \right\} - P \left\{ \tilde{W}_S \leq \tau | \mathcal{F} \right\} \right| = o_P(1). \tag{A.14}$$

Since $\tilde{\Omega}_S - \Omega_S$ is positive semidefinite, by Theorem 1 of Jensen (1984), we find that for each $\tau \in \mathbf{R}$,

$$P \{ W_S \leq \tau \} \leq P \left\{ \tilde{W}_S \leq \tau \right\}. \tag{A.15}$$

Take any $\epsilon > 0$. On the event that

$$\sup_{\tau \in \mathbf{R}} \left| P \left\{ \tilde{W}_S \leq \tau | \mathcal{F} \right\} - P \left\{ \hat{W}_{S,b} \leq \tau | \mathcal{H} \right\} \right| \leq \epsilon,$$

we have

$$\begin{aligned} P \left\{ \tilde{W}_S \leq \hat{c}_{1-\alpha}(S) | \mathcal{F} \right\} &\geq P \left\{ \hat{W}_{S,b} \leq \hat{c}_{1-\alpha}(S) | \mathcal{H} \right\} - \epsilon \\ &= 1 - \alpha - \epsilon \\ &= P \left\{ W_S \leq c_{1-\alpha-\epsilon}(S) \right\} \\ &\geq P \left\{ \tilde{W}_S \leq c_{1-\alpha-\epsilon}(S) \right\}, \end{aligned}$$

by (A.15). Hence,

$$P \left\{ \hat{c}_{1-\alpha}(S) < c_{1-\alpha-\epsilon}(S) \right\} \leq P \left\{ \sup_{\tau \in \mathbf{R}} \left| P \left\{ \hat{W}_{S,b} \leq \tau | \mathcal{H} \right\} - P \left\{ \tilde{W}_S \leq \tau | \mathcal{F} \right\} \right| > \epsilon \right\} = o(1), \tag{A.16}$$

by Lemma A.6. Since the choice of ϵ was arbitrary, we obtain the desired result of (A.12). Hence,

$$\begin{aligned} P \left\{ \hat{T}_S \leq \hat{c}_{1-\alpha}(S) \right\} &\geq P \left\{ \hat{T}_S \leq c_{1-\alpha}(S) \right\} + o(1) \\ &\geq P \left\{ W_S \leq c_{1-\alpha}(S) \right\} + o(1) = 1 - \alpha + o(1), \end{aligned} \tag{A.17}$$

by (A.13) and (A.15). □

Proof of Theorem 3.1. Using the same arguments in the proof of Lemmas A.3 and A.5, we find that

$$\hat{\sigma}_S^2 = \tilde{\sigma}_S^2 + o_P(1), \tag{A.18}$$

and

$$\sigma_S^{-1/2} \sqrt{n}(\hat{\Delta}_S - \Delta_S) \rightarrow_d N(0, 1). \tag{A.19}$$

The desired result follows from these two results. □

Proof of Theorem 3.2. Note that $\hat{c}_{1-\alpha}(S)$ is increasing in S . Hence, the desired result follows from Lemma A.7 and Theorem 2.1 of Romano and Shaikh (2010). □

REFERENCES

- Acemoglu, D., A. Ozdaglar, & E. Yildiz (2011) Diffusion of innovations in social networks. In *Proceedings of the 50th IEEE Conference on Decision and Control and European Control Conference (CDC-ECC), 12/2011*, pp. 2329–2334 Institute of Electrical and Electronics Engineers (IEEE).
- Akbarpour, M., S. Malladi, & A. Saberi (2020) Just a Few Seeds More: Value of Network Information for Diffusion. Working paper.
- Aral, S., L. Muchnik, & A. Sundararajan (2009) Distinguishing influence-based contagion from Homophily-driven diffusion in dynamic networks. *Proceedings of the National Academy of Sciences of the United States of America* 106(51), 21544–21549.
- Aronow, P. & C. Samii (2017) Estimating average causal effects under general interference, with application to a social network experiment. *Annals of Applied Statistics* 11, 1912–1947.
- Banerjee, A., A.G. Chandrasekhar, E. Duflo, & M.O. Jackson (2013) The diffusion of microfinance. *Science* 341(6144), 1236498.
- Banerjee, A., A.G. Chandrasekhar, E. Duflo, & M.O. Jackson (2019) Using gossips to spread information: Theory and evidence from two randomized controlled trials. *Review of Economic Studies* 86, 2453–2490.
- Beaman, L., A. BenYishay, J. Magruder, & A.M. Mobarak (2021) Can network theory-based targeting increase technology adoption? *American Economic Review* 111(6), 1918–1943.
- Bramoullé, Y., H. Djebbari, & B. Fortin (2009) Identification of peer effects through social networks. *Journal of Econometrics* 150, 41–55.
- Conley, T.G. & C.R. Udry (2010) Learning about a new technology: Pineapple in Ghana. *The American Economic Review* 100(1), 35–69.
- Dawid, P.A. (1979) Conditional Independence in statistical theory. *Journal of the Royal Statistical Society. Series B* 41, 1–31.
- de Matos, M.G., P. Ferreira, & D. Krackhardt (2014) Peer influence in the diffusion of the iPhone 3G over a large social network. *Management Information System Quarterly* 38, 1103–1133.
- de Paula, A., I. Rasul, & P.C. Souza (2020) Identifying Network Ties from Panel Data: Theory and an Application to Tax Competition. Preprint, [arxiv:1910.07452v2](https://arxiv.org/abs/1910.07452v2).
- Drukker, D.M., P.H. Egger, & I.R. Prucha (2022) Simultaneous Equation Models with Higher Order Spatial or Social Network Interactions. *Econometric Theory*, in press.
- Granovetter, M. (1978) Threshold models of collective behavior. *American Journal of Sociology* 83, 1420–1443.
- He, X. & K. Song (2022) Measuring Diffusion over a Large Network. Preprint, [arXiv:1812.04195v4](https://arxiv.org/abs/1812.04195v4) [stat.ME].
- Hirano, K., G. Imbens, & H. Ridder (2003) Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica* 71, 1161–1189.
- Horowitz, J.L. (2009) *Semiparametric and Nonparametric Methods in Econometrics*. Springer.
- Imbens, G.W. & J.M. Wooldridge (2009) Recent developments in the econometrics of program evaluation. *Journal of Economic Literature* 47, 5–86.
- Jackson, M.O. (2008) *Social and Economic Networks*. Princeton University Press.
- Jensen, D.R. (1984) Ordering ellipsoidal measures: Scale and peakness ordering. *SIAM Journal on Applied Mathematics* 44(6), 1226–1231.
- Khan, S. & E. Tamer (2010) Irregular identification, support conditions, and inverse weight estimation. *Econometrica* 78, 2021–2042.
- Kojevnikov, D. (2021) The Bootstrap for Network Dependent Processes. Preprint, [arXiv:2101.12312v1](https://arxiv.org/abs/2101.12312v1) [econ.EM].
- Kojevnikov, D. & K. Song (2022) Econometric Inference on Large Bayesian Games with Heterogeneous Beliefs. [arXiv:1404.2015v3](https://arxiv.org/abs/1404.2015v3) [stat.AP].
- Kuersteiner, G.M. & I.R. Prucha (2013) Limit theory for panel data models with cross sectional dependence and sequential exogeneity. *Journal of Econometrics* 174(2), 107–126.

- Lee, J.H. & K. Song (2019) Stable limit theorems for empirical processes under conditional neighborhood dependence. *Bernoulli* 25, 1189–1224.
- Lehmann, E.L. & J.P. Romano (2005) *Testing Statistical Hypotheses*. Springer.
- Leskovec, J., L.A. Adamic, & B.A. Huberman (2007) The dynamics of viral marketing. *ACM Transactions on the Web* 1(1), 5.
- Leung, M.P. (2020) Treatment and spillover effects under network interference. *Review of Economics and Statistics* 102, 368–380.
- Lewbel, A., X. Qu, & X. Tang (2021) Social Networks with Unobserved Links. *Working paper*.
- Manski, C.F. (1993) Identification of endogenous social effects: The reflection problem. *Review of Economic Studies* 60(3), 531–542.
- Newman, M.E.J. (2010) *Networks: An Introduction*. Oxford University Press.
- Penrose, M. (2003) *Random Geometric Graphs*. Oxford University Press.
- Phillips, P.C.B. (1988) Conditional and unconditional statistical Independence. *Journal of Econometrics* 38, 341–348.
- Romano, J.P. & A.M. Shaikh (2010) Inference for the identified set in partially identified econometric models. *Econometrica* 78, 169–211.
- Rosenbaum, P.R. & D.B. Rubin (1983) The central role of the propensity score in observational studies for causal effects. *Biometrika* 70(1), 41–55.
- Sadler, E. (2020) Diffusion games. *American Economic Review* 110, 225–270.
- Shalizi, C.R. & A.C. Thomas (2011) Homophily and contagion are generically confounded in observational social network studies. *Sociological Methods and Research* 40, 211–239.
- Song, K. (2018) Measuring the graph concordance of locally dependent observations. *Review of Economics and Statistics* 100, 535–549.
- van der Laan, M. (2014) Causal inference for a population of causally connected units. *Journal of Causal Inference* 2, 13–74.