Decomposition of Causal Effect with Interference Accounting for Network Change

Siwon Ryu

Abstract

Recent empirical studies emphasize the importance of indirect, or spillover, effects of a program or policy. Most studies assume that the underlying network is exogenous, fixed, or unaffected by the intervention. However, some empirical evidence indicates that the treatment can also have significant network effects. This paper studies the identification and estimation of causal treatment effects while explicitly considering possible causal changes in the network resulting from a program. The main finding is the decomposition of the causal effects into two distinct components: the treatment effect when the network remains unchanged and the effect when the treatment alters only the network structure. This result enhances our understanding of policy/program mechanisms by considering counterfactual scenarios where the network is either altered or remains unchanged due to the treatment. The proposed method applies to both randomized experiments and quasi-experimental designs with parallel trends. A simple estimation procedure for causal effects and their decomposition is proposed, and its performance is evaluated through Monte Carlo simulations, and illustrated using data from a program offering savings accounts. The empirical result shows that the direct effect consists of a positive pure treatment effect and a negative network effect, while the indirect effects are not statistically significant.

Keywords: Causal inference; Network change; Endogenous network; Difference-in-differences

1 Introduction

Evaluating a program is an important topic in empirical economics, typically involving the estimation of causal effects of a program. Many methods are based on the potential outcome framework (e.g., Rubin (1974)), and the baseline assumption is the Stable Unit Treatment Value Assumption (SUTVA) which excludes interference between units or individuals. because economic agents frequently interact with one another. Recent empirical evidence further highlights the potential significance of spillover effects in program evaluations.

When spillover effects on outcomes exist, the potential outcome must be expressed as a function of the entire treatment vector. This situation presents primary challenges for researchers. It is not only difficult to define causal parameters compared to the situations under SUTVA, but it is also challenging to identify meaningful parameters as there are many number of potential outcomes. For example, if N units interact, the possible number of treatment assignments profile, and hence 2^N potentially distinct potential outcomes. To address these challenges, studies commonly employ the constant treatment response (CTR) assumption noted by Manski (2013) or use the exposure mapping to significantly reduce dimensionality (e.g., Leung (2020), Leung (2022), Vazquez-Bare (2022a), Vazquez-Bare (2022b), Aronow and Samii (2017), Bramoullé, Djebbari, and Fortin (2009), Forastiere, Airoldi, and Mealli (2021), Auerbach and Tabord-Meehan (2021)).

Another notable approach is designing experiments to deal with interference. Double randomization, proposed by Hudgens and Halloran (2008), involves first randomizing treatment rates (saturation) to groups and then randomizing treatment according to the set rates. The variation in treatment saturation provides an additional source of identification (e.g., Kang and Imbens (2016), Blackwell (2017), Baird et al. (2018), DiTraglia et al. (2023), Sánchez-Becerra (2021), Imai, Jiang, and Malani (2021), and Hoshino and Yanagi (2023)). Additionally, some studies focus on optimizing experimental design in these settings to maximize social welfare (e.g., Kitagawa and Wang (2023), Ananth (2020), Viviano (2019)).

While exposure mapping and double randomization simplify the analysis, they rely on the assumption that the underlying network structure is either unaffected by the

treatment, exogenous, or remains fixed. For example, exposure often refers to the number of treated friends, which depends on the network structure. As a result, accounting for possible network changes with the exposure mapping approach is not straightforward, since it's difficult to separate changes in exposure from changes in the network without strong assumptions. Similarly, studies using double randomization typically assume that the network remains fixed or is independent of the treatment. This assumption may hold in the short term when there is not enough time for the network structure to change.

However, recent empirical studies suggest that treatment can significantly affect the underlying network structure. For example, Comola and Prina (2021) use experimental data from Nepal and found that providing savings accounts to households led to changes in network degrees. Specifically, the probability of being linked to other households with at least one treated member decreased from 81% to 76%. Banerjee et al. (2024) analyze data from Karnataka, India, and a field experiment in Hyderabad to investigate how exposure to formal financial institutions impact their network density. Similarly, Dupas, Keats, and Robinson (2019) use experimental data from Kenya, where households received free savings accounts, and observe that households became less dependent on distant family members while being more supportive of neighbors and friends in their village. These studies highlight the importance of considering network effects in program evaluations.

This paper introduces a method to identify and estimate the causal effect of a program, taking into account the possibility of causal network changes from the treatment. Additionally, this study decomposes the treatment (or spillover) effect into two parts. The first part examines the impact of the treatment when the network remains unchanged, while the second part focuses on the effect of the treatment when the treatment alters the network structure only. If the treatment does not cause changes in the network, the second component has no effect, and the first component aligns with existing concepts of treatment and spillover effects.

The proposed method relies on dyadic links and a linear potential outcome. It is applied in two experimental settings: first, randomized experiments with information of post-treatment period, and second, quasi-experiment designs with parallel trends, where

the difference-in-differences estimand identifies the average treatment effects on treated. A simple estimation procedure is proposed, and its finite sample performances are evaluated through Monte Carlo simulations. It is also demonstrated using experimental data from Comola and Prina (2020), which involves providing households in Nepal with access to savings accounts. The empirical results show that while the overall direct effect is positive, it consists of a positive treatment effect and a negative network effect. This decomposition and interpretation cannot be achieved using the direct and indirect effect calculations in Comola and Prina (2021).

The proposed method relies on dyadic links and a linear potential outcome model, which corresponds to the linear-in-means (or sums) model for observed outcomes, frequently used in empirical research. The coefficients in this model are identified when both the treatment and the network are exogenous. The exogeneity of the underlying network may hold when the network remains largely unchanged or fixed. However, if the network is not fixed, endogeneity may arise if unobserved factors influencing the outcome are correlated with the potential link formation. In such cases, the standard linear-in-means (or sums) regression is inappropriate for identifying the outcome coefficients. To address this issue, I estimate the distribution of potential links using a dyadic regression model first, and use this information to recover the outcome model's coefficients.

This method is applied in two experimental settings: first, in randomized experiments with post-treatment data, and second, in quasi-experimental designs with parallel trends using both pre- and post-treatment data. A simple estimation procedure is proposed, and its finite sample performance is evaluated through Monte Carlo simulations. The method is further illustrated using experimental data from Comola and Prina (2020), which provided households in Nepal with access to savings accounts. The empirical results show that, while the overall direct effect is positive, it is composed of a positive treatment effect and a negative network effect. This decomposition and interpretation cannot be captured by the direct and indirect effect estimates used in Comola and Prina (2021).

The main contributions of this paper can be summarized in three key aspects. First, it proposes a novel method for analyzing causal effects that accounts for causal net-

work changes within the potential outcome framework. Second, it decomposes the causal effects, allowing a more detailed understanding of the mechanisms behind a program. Third, the methods are applicable to various experimental designs, including both randomized experiments and quasi-experimental designs. Additionally, the proposed methods can be viewed as a generalization of existing approaches. If the network remains unchanged, the method reduces to a linear model of treatment effects with interference, as in Vazquez-Bare (2022b), Leung (2020), or Aronow and Samii (2017). Furthermore, if there is no interference, the method simplifies to the standard potential outcome model under SUTVA.

This study is closely connected to the literature on identifying and estimating causal effects of a program when individuals interact with one another. Various studies address violations of SUTVA by employing exposure mappings in randomized experiments (e.g., Leung (2020), Leung (2022), and Vazquez-Bare (2022b)). When treatment is endogenous due to imperfect compliance, local average treatment effects can be identified (e.g., Vazquez-Bare (2022a), DiTraglia et al. (2023), Hoshino and Yanagi (2023), Kormos, Lieli, and Huber (2023), Kang and Imbens (2016), Blackwell (2017)). In cases where treatment endogeneity arises from selection, such as in quasi-experimental settings, the difference-in-differences approach identifies the average treatment effects on the treated (e.g., Xu (2023) and Butts (2021)). Since treatment assignments under interference can be viewed as multiple treatments, this paper also relates to the challenges of analyzing multiple treatments (e.g., Frölich (2004), Fricke (2017)).

While the studies mentioned above assume a fixed or exogenous network, Comola and Prina (2021) (hereafter CP) explicitly address estimation of treatment effects accounting for network changes. CP assume a two-period linear-in-means model (e.g., Manski (2013), Bramoullé, Djebbari, and Fortin (2009)) to consider network change over time, and defines direct and indirect treatment effects as the partial derivatives of the conditional mean of the reduced-form outcome with respect to the treatment vector. Their main findings suggest that indirect effects may be underestimated if network changes resulting from the treatment are not considered.

This study differs from CP in several key aspects. First, I introduce a potential outcome and potential network framework to provide clear causal interpretations, whereas

CP focuses on estimating a reduced form. Second, I derive the limiting distribution of the estimates for statistical inference. Third, in CP, network changes may reflect both time-varying effects and treatment effects, while I explicitly model causal changes in the network driven by the treatment. Lastly, the most significant difference is that I decompose the causal effects into separate components: the pure treatment effect and the network effect. This decomposition allows researchers to determine how much of the direct and indirect effects are due to network changes, providing a clearer understanding of the underlying mechanisms, which is not addressed by CP or other studies in the literature.

This paper is organized as follows: Section 2 describes the setting, defines the parameters of interest including the decomposition of causal effects, and addresses their identification. Section Section 3 presents the details of the estimation procedure. Section 4 evaluates performance of the proposed method through Monte Carlo simulations. Section 5 provides an empirical illustration of the proposed method. Section 6 concludes.

2 Model and Identification

In this section, I provide an overview of the setting and define the key parameters of interest. First, I discuss the response functions for both potential links and potential outcomes. Then, I introduce the main causal parameters, focusing on the direct/indirect effects. The direct effect captures how an individual's treatment status affects their own outcome, while the indirect effects measure the impact of changes in other individuals' treatment statuses. Next, I propose a decomposition of these effects into two components. The first component considers the impact when the underlying network remains fixed and is referred to as the *treatment effect*, as it aligns with the conventional concept of treatment effects in the literature. The second component reflects the impact that comes exclusively from changes in the network structure, which I refer to as the *network effect*.

2.1 Response Functions for Links and Potential Outcomes

Suppose there are G independent groups with N individuals in each group.[1] If we observe data over two periods, let $t \in \{0, 1\}$ represent the time periods, where some individuals are assigned to a treatment group after $t = 0$. In other words, $t = 0$ represent the pre-treatment period and $t = 1$ is the post-treatment period. Let $D_{ig} \in \{0, 1\}$ be an indicator showing whether individual i receives the treatment. And assume there is no imperfect compliance. To simplify notation, I omit group index, or time index, or both for the rest of this section when there is no risk of confusion.

Each individual interacts with others through an underlying network structure. Specifically, let $A_{ij} \in \{0, 1\}$ represent the link between individuals i and j, where there are no self-links. That is, $A_{ij} = 1$ if two different individuals i and j are linked, and $A_{ii} = 0$ for all i. The network can be either directed or undirected. For each individual i, let $Y_i \in \mathbb{R}$ denote the outcome of interest.

Potential outcomes and potential links are expressed as functions of the entire treatment vector. Let $\boldsymbol{d} \in \{0, 1\}^N$ represent a vector of treatment assignments for N individuals. Corresponding to an assignment $\boldsymbol{d}$, let $A_{ij}(\boldsymbol{d})$ denote the *potential link* between pairs of individuals i and j, and $Y_i(\boldsymbol{d})$ represent the *potential outcome* for individual i.

Since there are 2^N possible potential treatment assignments, defining and analyzing the causal effect of interest becomes challenging, particularly when the number of individuals is large. To address this issue, I first assume that the potential network links are formed based on a dyadic model, as stated in Assumption 1.

Assumption 1 (Dyadic Response on Potential Network Links)**.** *For each pair of individuals (i, j), for any treatment assignments $\boldsymbol{d}, \boldsymbol{d}' \in \{0, 1\}^N$, (i) if $d_i = d'_i$ and $d_j = d'_j$, then $A_{ij}(\boldsymbol{d}) = A_{ij}(\boldsymbol{d}')$ with probability 1. Thus, by abusing notation, the potential link can be expressed as $A_{ij}(\boldsymbol{d}) = A_{ij}(d_i, d_j)$ for any $\boldsymbol{d} = (d_1, ..., d_N)'$; (ii) Additionally, for all (i, j) pairs and for $(d_i, d_j) \in \{0, 1\}^2$, the following condition holds: $E[A_{ij}(d_i, d_j)|\boldsymbol{D}] = E[A_{ij}(d_i, d_j)|D_i, D_j]$, for all (i, j) and $(d_i, d_j) \in \{0, 1\}^2$.*

The first part of Assumption 1 imposes that each pair's potential link is determined

[1] In this section, I assume a fixed group size, but this can be easily extended to allow groups to have different sizes N_g, by considering every moment restrictions conditioning on the group size N_g.

solely by their own treatment statuses (d_i, d_j), and not by the treatment status of other individuals. For example, Assumption 1 is satisfied under a dyadic link formation model, such as Graham (2017):

$$A_{ij}(\boldsymbol{d}) = \mathbb{1}\{\theta_1 d_i + \theta_2 d_j + u_{ij} > 0\}, \tag{1}$$

where u_{ij} is including both individual-specific, and pair-specific unobserved factors influencing link formation. In this model, the potential link A_{ij} depends only on d_i, d_j, but not d_k for $k \notin \{i, j\}$.

The second part of Assumption 1 strengthens the first by assuming that potential links are not only formed through a dyadic model, but are also mean independent of other pairs' treatment statuses, given their own treatments. For example, in the dyadic model above, it implies that: $u_{ij}|\boldsymbol{D} \sim u_{ij}|D_i, D_j$.

The potential outcome is typically expressed as a function of the treatment vector $\boldsymbol{d} \in \{0,1\}^N$. A conventional approach to handling potential outcomes under interference is by using exposure maps or assuming a constant treatment response (e.g., Manski (2013)). If there exists a function f such that for any treatment vectors $\boldsymbol{d}, \boldsymbol{d}' \in \{0,1\}^N$, $f(\boldsymbol{d}) = f(\boldsymbol{d}')$ implies $Y(\boldsymbol{d}) = Y(\boldsymbol{d}')$ with probability 1, then the function f is called an exposure map.

The existence or specific functional form of an exposure mapping is unknown without further restrictions. However, in some cases, it is possible to define an appropriate exposure map for a potential outcome. For instance, if the network is anonymous, only the number of treated and untreated neighbors would be relevant. Leung (2020) demonstrates that assuming (i) local spillover (interference occurs only from neighbors within a network distance of 1 or some fixed number) and (ii) exchangeability is equivalent to having a correctly specified exposure map $(d_i, Q_i(\boldsymbol{d}), R_i(\boldsymbol{d}))$, where d_i is the individual's own treatment status, $Q_i(\boldsymbol{d}) := \sum_j A_{ij}(d_i, d_j)d_j$ is the number of potentially treated neighbors, $R_i(\boldsymbol{d}) := \sum_j A_{ij}(d_i, d_j)(1-d_j)$ is the number of potentially untreated neighbors.[2] Therefore, the potential outcome can be written as $Y(d_i, Q_i(\boldsymbol{d}), R_i(\boldsymbol{d}))$ instead of a function of the entire treatment vector.

[2]If $A_{ij}(d_i, d_j)$ denotes the potential row-normalized link, then $Q_i(\boldsymbol{d})$ and $R_i(\boldsymbol{d})$ represent potential rates at which individual i forms a link with treated and untreated neighbors, respectively.

To focus on identifying causal effects and their decomposition, I assume that the response function for the potential outcome is linear in this exposures $(d_i, Q_i(\boldsymbol{d}), R_i(\boldsymbol{d}))$, as stated in Assumption 2.[3]

Assumption 2 (Linear Response on Potential Outcomes)**.** *For each individual i, let $Y_i(\boldsymbol{d})$ be the potential outcome corresponding to $\boldsymbol{d} \in \{0,1\}^N$. Assume that the potential outcome is determined by the following linear response function:*

$$Y_i(\boldsymbol{d}) = \beta_1 + \beta_2 d_i + \beta_3 Q_i(\boldsymbol{d}) + \beta_4 R_i(\boldsymbol{d}) + \varepsilon_i(d_i),$$

where $Q_i(\boldsymbol{d}) = \sum_{j=1,j\neq i} A_{ij}(d_i, d_j)d_j$, $R_i(\boldsymbol{d}) = \sum_{j=1,j\neq i}^{N} A_{ij}(d_i, d_j)(1-d_j)$, and $\varepsilon_i(d)$ is the potential individual error that satisfies $E[\varepsilon_i(d)|\boldsymbol{D}] = E[\varepsilon_i(d)|D_i]$ for $d \in \{0,1\}$. Without loss of generality, assume that the error term has a mean of zero and does not have an average treatment effect on the treated. (i.e., $E[\varepsilon(1) - \varepsilon(0)|D_i = 1] = 0$).[4]

The parameter β_2 represents the effect of an individual's own treatment d_i when all links and others' treatments remain fixed. Thus, β_2 captures the direct treatment effect. Next, since $Q_i(\boldsymbol{d})$ represents the number of potentially treated neighbors, β_3 captures the spillover (or exposure) effect from one additional treated neighbors. Similarly, β_4 represents the effects from one additional untreated neighbors. These parameters can be interpreted as causal effects under various assumptions, particularly if the network links are unaffected by the treatment.

However, because an individual's own treatment d_i can also influence potential links $\{A_{ij}(d_i, d_j)\}_{i,j}$, it introduces an additional effect on outcomes, driven by these altered links. I define this type of effect as the *network effect*. Changes in the treatment status of others, d_j, can have similar network effects, which are formally defined below. The individual error term $\varepsilon_i(d)$ is assumed to be mean independent of the others' treatment statuses given the individual's own treatment.

Assumption 2 suggests that the observed outcome can be expressed as a linear

[3]In other words, Assumption 2 is equivalent to assuming (i) local spillover, (ii) exchangeability, and (iii) additive separability of potential outcomes with respect to the exposures.

[4]If $E[\varepsilon(d)] = \mu$, then we can rewrite the model by replacing β_1 with $\tilde{\beta}_1 = \beta_1 - \mu$, and $\varepsilon_i(d)$ with $\tilde{\varepsilon}_i(d) = \varepsilon_i(d) - \mu$. If $E[\varepsilon(1) - \varepsilon(0)|D_i = 1] = \tau$, then we can rewrite the model by replacing β_2 with $\tilde{\beta}_2 = \beta_2 + \tau$, and ε_i with $\tilde{\varepsilon}_i(1) = \varepsilon_i(1) - \tau$, $\tilde{\varepsilon}_i(0) = \varepsilon_i(0)$.

network model:

$$Y_i = \beta_1 + \beta_2 D_i + \beta_3 Q_i + \beta_4 R_i + \varepsilon_i, \tag{2}$$

where $Q_i = \sum_{j=1}^{N} A_{ij} D_j = Q_i(\boldsymbol{D})$ and $R_i = \sum_{j=1}^{N} A_{ij}(1 - D_j) = R_i(\boldsymbol{D})$ are the observed numbers of treated and untreated neighbors, respectively. If the network is unaffected by the treatment, i.e., $A_{ij}(d_i, d_j) = A_{ij}$, then the response function reduces a linear response model commonly used in the literature (e.g., Leung (2020), Forastiere, Airoldi, and Mealli (2021)). Moreover, when $\beta_3 = \beta_4 = 0$, which means there is no interaction, the model simplifies to a standard causal model without interference. In this way, the model generalizes existing approaches to account for cases where the individuals interact, and also when network structure is influenced by treatment.

2.2 Causal Parameters and Decomposition

In this subsection, I define the key causal parameters and describe their decomposition. Consider a scenario where each group contains 2 units ($N = 2$). Based on Assumption 2, the potential outcome can be written as a function of an individual's treatment, their neighbor's treatment, and their potential link: $Y_i(\boldsymbol{d}) = y(d_i, d_j, A_{ij}(d_i, d_j))$. The effect of individual i's own treatment (d_i) on their outcome can be decomposed as follows:

$$\begin{aligned} &y(1, 0, A_{ij}(1,0)) - y(0, 0, A_{ij}(0,0)) \\ &= \underbrace{y(1, 0, A_{ij}(1,0)) - y(1, 0, A_{ij}(0,0))}_{\text{=Direct Network Effect}} + \underbrace{y(1, 0, A_{ij}(0,0)) - y(0, 0, A_{ij}(0,0))}_{\text{=Direct Treatment Effect}}. \end{aligned}$$

The first term represents the *direct network effect*, capturing the impact of the treatment on the outcome due to changes in the network links, while treatment status is fixed at $(d_i, d_j) = (1, 0)$. The second term represents the *direct treatment effect*, which denote the effect of changes in treatment status from $(d_i, d_j) = (0, 0)$ to $(1, 0)$, while the link is fixed at $A_{ij}(0, 0)$.

Similarly, the effect of neighbor j's treatment (d_j) on individual i's outcome is

decomposed as:

$$\begin{aligned} & y(0,1,A_{ij}(0,1)) - y(0,0,A_{ij}(0,0)) \\ & = \underbrace{y(0,1,A_{ij}(0,1)) - y(0,1,A_{ij}(0,0))}_{\text{=Indirect Network Effect}} + \underbrace{y(0,1,A_{ij}(0,0)) - y(0,0,A_{ij}(0,0))}_{\text{=Indirect Treatment Effect}}. \end{aligned}$$

The first term captures the *indirect network effect*, reflecting how changes in neighbor j's treatment influence the individual i's outcome by altering the network links from $A_{ij}(0,0)$ to $A_{ij}(0,1)$, while their treatment status is fixed at $(d_i, d_j) = (0,1)$. The second term represents the *indirect treatment effect*, which measures the influence of j's treatment on i's outcome, assuming their link is fixed at $A_{ij}(0,0)$.

The decomposition of direct and indirect effects allows us to separate the pure effects of the treatment from the network effects. The network effects capture the changes in outcomes that are driven by shifts in the network structure, while the treatment effects focus on the changes in outcomes when the network remains fixed at its untreated counterfactual.

Remark 1. There is an alternative way to decompose the effects. For example, the indirect effects can be rewritten as:

$$\begin{aligned} & y(0,1,A_{ij}(0,1)) - y(0,0,A_{ij}(0,0)) \\ & = \underbrace{y(0,0,A_{ij}(0,1)) - y(0,0,A_{ij}(0,0))}_{\text{=Indirect Network Effect}} + \underbrace{y(0,1,A_{ij}(0,1)) - y(0,0,A_{ij}(0,1))}_{\text{=Indirect Treatment Effect}}. \end{aligned}$$

In this decomposition, the *network effect* reflects the causal impact of changes in the network when both units remain untreated, while the *treatment effect* represents the causal impact of the other's treatment, assuming the links are fixed at the untreated counterfactual. The distinction here lies in the baseline counterfactual scenario regarding treatment status and network links. Researchers can choose which decomposition definition best suits their empirical context, depending on the interpretation they find more insightful.

In a more general case with N individuals, each indirect effect could be influenced by the treatment status of all neighbors. However, the potential outcome is influenced by

the treatment of others primarily through the count of treated or untreated neighbors. Therefore, I focus on the marginal impact of a neighbor's treatment, specifically the impact of *one additional* treated or untreated neighbor.

Let $\boldsymbol{e}_1, ..., \boldsymbol{e}_N$ be the standard Euclidean basis, where for each i, $\boldsymbol{e}_i = (e_{i1}, ..., e_{iN})'$, $e_{ii} = 1$, and $e_{ij} = 0$ for all $j \neq i$. Define $m(d, e) = E[A_{ij}|D_i = d, D_j = e]$ and $H(d, e) = E[A_{ij}(d, e) - A_{ij}(0, 0)|D_i = d, D_j = e]$, for $(d, e) \in \{0, 1\}^2$, where $H(d, e)$ represents the average treatment effect on treated (ATT) of links. Then, the direct effect on outcome can be written as:

$$Y_i(\boldsymbol{e}_i) - Y_i(\mathbf{0}) = \beta_2 + \beta_4 \sum_{j \neq i} (A_{ij}(1, 0) - A_{ij}(0, 0)) + \varepsilon_i(1) - \varepsilon_i(0).$$

Assumption 2 provides the causal interpretation for the decomposition. In the counterfactual scenario where no individual is treated, i.e., $\boldsymbol{D} = \mathbf{0}$, the potential outcome is $\beta_1 + \beta_4 \sum_{j \neq i} A_{ij}(0, 0) + \varepsilon_i(0)$. From that scenario, if the links remain fixed at $A_{ij}(0, 0)$ but the individual i is treated, then the outcome becomes $\beta_1 + \beta_2 + \beta_4 \sum_{j \neq i} A_{ij}(0, 0) + \varepsilon_i(1)$. Thus, the causal changes in individual i's outcomes by the own treatment is $\beta_2 + \varepsilon_i(1) - \varepsilon_i(0)$, which defines the direct treatment effect. Next, from this scenario where the links are still fixed at $A_{ij}(0, 0)$, and only individual i is treated, i.e., $\boldsymbol{D} = \boldsymbol{e}_i$, if the links are changed from $A_{ij}(0, 0)$ to $A_{ij}(1, 0)$, then the outcome becomes $\beta_2 + \beta_4 \sum_{j \neq i} A_{ij}(1, 0) + \varepsilon(1)$. Therefore the causal effect on the outcome is $\beta_4 \sum_{j \neq i} (A_{ij}(1, 0) - A_{ij}(0, 0))$. This represent the causal changes from the causal changes in network links, which defines the direct network effect.

Because the individual error term has no ATT by construction, i.e., $E[\varepsilon_i(1) - \varepsilon_i(0)|D_i = 1] = 0$, the *average direct effect* (π^D) on treated is decomposed by:[5]

$$\pi^D := E[Y_i(\boldsymbol{e}_i) - Y_i(\mathbf{0})|\boldsymbol{D} = \boldsymbol{e}_i], = \underbrace{\beta}_{:=\pi^{DT}} + \underbrace{\gamma_2 \sum_{j=1}^{N} H(1, 0)}_{:=\pi^{DN}}.$$

Here, π^{DT} is the *average direct treatment effect* that represents the effect of one's own treatment when the network is fixed. On the other hand, π^{DN} is the *average direct*

network effect that captures the effect of changes in links driven by one's own treatment.

Similarly, the indirect effect on the outcome can be written as:

$$Y_i(\boldsymbol{e}_j) - Y_i(\mathbf{0}) = \beta_3 A_{ij}(0,1) - \beta_4 A_{ij}(0,0).$$

Now, consider a counterfactual situation where no individual is treated, so the potential outcome is given by $\beta_1 + \beta_4 A_{ij}(0,0) + \varepsilon_i(0)$. From this scenario, if individual j is treated, but the links remain fixed at $A_{ij}(0,0)$, the outcome becomes $\beta_1 + \beta_3 A_{ij}(0,0)$, and the difference $(\beta_3 - \beta_4)A_{ij}(0,0)$ defines the indirect treatment effect. Furthermore, from this situation where individual j is treated, but links remain at $A_{ij}(0,0)$, if the link changes to $A_{ij}(0,1)$, the difference in the outcome is $\beta_3(A_{ij}(0,1) - A_{ij}(0,0))$, which defines the indirect network effect.

The average indirect effect (π^I) is decomposed as:

$$\pi^I := E[Y_i(\boldsymbol{e}_j) - Y_i(\mathbf{0}) | \boldsymbol{D} = \boldsymbol{e}_j] = \underbrace{(\beta_3 - \beta_4)m(0,0)}_{:=\pi^{IT}} + \underbrace{\beta_3 H(0,1)}_{:=\pi^{IN}}. \tag{3}$$

Here, π^{IT} and π^{IN} represent the *average indirect treatment effect* and *average indirect network effect*, respectively.

As discussed in Remark 1, there is another expression with a different interpretation for the indirect effects when considering a different comparison of counterfactual scenarios. However, the preceding argument about identification and estimation remains the same for this alternative expression. Therefore, I will use the definition in (3) for the decomposition of indirect effects.

[5]This decomposition remains consistent with another expression of the direct effect. Under the counterfactual scenario where $\boldsymbol{D} = \boldsymbol{e}_i$ (i.e., $D_i = 1$ and $D_j = 0$ for all $j \neq i$), the potential outcome is $\beta_2 + \beta_4 \sum_{j \neq i} A_{ij}(1,0) + \varepsilon_i(1)$. If the links are fixed at $A_{ij}(1,0)$, but individual i is untreated, the outcome becomes $\beta_4 + \sum_{j \neq i} A_{ij}(1,0) + \varepsilon_i(0)$. Therefore, the causal effect of individual i's own treatment on the outcome is again $\beta_2 + \varepsilon_i(1) - \varepsilon_i(0)$, and the remaining term represent the direct network effect.

[5]For the indirect effect, the decomposition varies depending on the counterfactual scenario being compared. Instead of the scenario where individual j is treated and the link is fixed at $A_{ij}(0,0)$, now consider a different counterfactual where individual j is untreated but the link is fixed at $A_{ij}(0,1)$. Comparing the potential outcomes when $\boldsymbol{D} = \boldsymbol{e}_j$ and this new counterfactual, we obtain an indirect treatment effect of $(\beta_3 - \beta_4)A_{ij}(0,1)$ and an indirect network effect of $\beta_4(A_{ij}(0,1) - A_{ij}(0,0))$. Intuitively, the term $\beta_3 - \beta_4$ captures the difference in the marginal effect of individual j's treatment when he is linked to individual i. Thus, when we fix the link at A, for example, $(\beta_3 - \beta_4)A$ represents the

In summary, the parameters of interest are:

$$\begin{aligned}
\pi^{DT} &= \beta_2, \\
\pi^{DN} &= \beta_4(N-1)H(1,0), \\
\pi^{IT} &= (\beta_3 - \beta_4)m(0,0), \\
\pi^{IN} &= \beta_3 H(0,1).
\end{aligned}$$

From the linearity of the response function, any direct or indirect effect (i.e., for any $\boldsymbol{d} \in \{0,1\}^N$ can be expressed as a linear combination of $\boldsymbol{\pi} := (\pi^{DT}, \pi^{DN}, \pi^{IT}, \pi^{IN})$. To separately identify these four effects, we need to estimate the outcome coefficients $\boldsymbol{\beta}$ and the conditional distribution of potential links, specifically $m(0,0)$, $H(1,0)$, and $H(0,1)$.

2.3 Identification

This section discusses the identification of causal effects and their decomposition as defined in Section 2. First, I address the case where only the post-treatment period data is observed, but the treatment is exogenous (i.e., randomized experiments). Then, I discuss the case where data is observed over two periods $t \in \{0,1\}$, with $t = 0$ as the pre-treatment period and $t = 1$ as the post-treatment period, where the treatment satisfies both the parallel trends assumption and no-anticipation.

In each case, we have two types of data: (i) individual-level data and (ii) dyadic-level data. Identification involves using both types. First, by using dyadic-level data that include observed links, the conditional expectations of links are identified as the coefficient $\boldsymbol{\zeta}$ of a dyadic regression on links. Second, the coefficients $\boldsymbol{\beta}$ in the outcome model (2) are identified. If the network is exogenous, the coefficient $\boldsymbol{\beta}$ can be identified through regression of Y_i on $(1, D_i, Q_i, R_i)$ using the information of observed network and treatment. However, I consider possible endogeneity of links due to causal changes in the network driven by the treatment. For instance, if the error term u_{ij} in the link formation model (1) is correlated with the error term ε_i in (2), even if the treatment is randomly

indirect treatment effect. Additionally, when individual j is untreated, treated, the effect of changing A_{ij} from 0 to 1 is β_4, β_3, respectively.

assigned, the network is endogenous. Therefore, I use the estimated distribution of links ($\boldsymbol{\zeta}$) in the dyadic regression to recover the coefficient $\boldsymbol{\beta}$ in the outcome regression. Lastly, both the decomposition and causal effects are identified through the combination of $\boldsymbol{\zeta}$ and $\boldsymbol{\beta}$.

To begin, I list the identifying assumptions about the data distribution as follows:

Assumption 3 (Distributions)**.** *Let $\{\mathcal{V}_{ig}\}_{i,g}$ be individual-level data, and $\{\mathcal{W}_{ijg}\}_{(i,j),g}$ be dyadic-level data. Then, (i) $\mathcal{V}_{ig}$ are identically distributed over i and g, and independent over g; (ii) $\mathcal{W}_{ijg}$ are identically distributed over all (i,j) and g, and independent over g.*

Assumption 4 (Overlap)**.** *For any $(d,e) \in \{0,1\}^2$, $\Pr(D_{ig} = d, D_{jg} = e) \in (0,1)$.*

Assumption 3 is about the distribution of the population, stating that groups are independent while allowing unrestricted dependence between individuals or dyads within each group. Assumption 4 is a standard requirement that ensures the existence of the corresponding conditional distributions. In the following, I explicitly consider two scenarios: (i) when the treatment is randomly assigned and only post-treatment information is observed, and (ii) in a quasi-experimental setting, where both pre- and post-treatment information is observed, assuming parallel trends and no anticipation.

2.3.1 Identification under Randomized Experiment

Consider situations where the program is designed to randomly assign the treatment, and we have dyadic-level data $\{\mathcal{W}_{ij}\} = \{A_{ij}\}$, along with individual-level data $\{\mathcal{V}_i\} = \{D_i, Y_i\}$. In this case, both the individual error term in the outcome response function and potential links are independent of the treatment, as stated in the following:

Assumption 5 (Randomized Experiment)**.** *(i) $E[\varepsilon_i(d)|D_i] = 0$ for $d \in \{0,1\}$; (ii) $E[A_{ij}(d_i,d_j)|D_i,D_j] = E[A_{ij}(d_i,d_j)]$.*

Consider the following conditional expectations: $E[A_{ij}|D_i,D_j] = \boldsymbol{W}_{ij}'\boldsymbol{\zeta}$, where $\boldsymbol{W}_{ij} = (1, D_i, D_j, D_iD_j)' \in \mathbb{R}^4$. The coefficient $\boldsymbol{\zeta}$ is identified provided that $E[\boldsymbol{W}_{ijg}\boldsymbol{W}_{ijg}']$ is nonsingular. Each element in $\boldsymbol{\zeta}$ means $\zeta_1 = m(0,0)$, $\zeta_2 = m(1,0) - m(0,0)$,

$\zeta_3 = m(0,1) - m(0,0)$, and $\zeta_4 = m(1,1) - m(1,0) - m(0,1) + m(0,0)$, where $m(d,e) = E[A_{ij}(d,e)]$.

From (2), the observed outcome is written as: $Y_i = Z_i'\boldsymbol{\beta} + \varepsilon_i$, where $Z_i := (1, D_i, Q_i, R_i)$. Although the treatment is randomly assigned, $E[\varepsilon_i|Z_i] = E[\varepsilon_i|\boldsymbol{A}, \boldsymbol{D}] \neq 0$ because the network is possibly correlated with the individual error term ε_i. Thus, the network can be endogenous because the causal changes in network. However, by exploiting the exogeneity of treatment, we can write: $E[Y_i|\boldsymbol{D}] = E[Z_i|\boldsymbol{D}]'\boldsymbol{\beta}$, and $E[Z_i|\boldsymbol{D}] = (1, D_i, E[Q_i|\boldsymbol{D}], E[R_i|\boldsymbol{D}])$. Here, $E[Q_i|\boldsymbol{D}] = \sum_{j\neq i} E[A_{ij}|D_i, D_j]D_j$ and $E[R_i|\boldsymbol{D}] = \sum_{j\neq i} E[A_{ij}|D_i, D_j](1-D_j)$. Therefore, if $\boldsymbol{\zeta}$ is known, the outcome coefficient $\boldsymbol{\beta}$ can be recovered by the coefficient of a regression $E[Y_i|\boldsymbol{D}] = Z_i(\boldsymbol{\zeta})'\boldsymbol{\beta}$, where $Z_i(\boldsymbol{\beta}) = E[Z_i|\boldsymbol{D}; \boldsymbol{\zeta}]$.

Lastly, since $\zeta_2 = E[A_{ij}(1,0) - A_{ij}(0,0)]$, and $\zeta_3 = E[A_{ij}(0,1) - A_{ij}(0,0)]$, we can recover the decomposition $\boldsymbol{\pi}$ from $\boldsymbol{\beta}$ and $\boldsymbol{\zeta}$. Proposition 1 formally states this identification procedure.

Proposition 1 (Identification With Randomized Experiment)**.** *Suppose Assumptions 1-5 hold. Then:*

(i) In a dyadic regression $E[A_{ij}|\boldsymbol{D}] = \boldsymbol{W}_{ij}'\boldsymbol{\zeta}$, the coefficient $\boldsymbol{\zeta}$ is given by $\boldsymbol{\zeta} = \boldsymbol{R}_{\boldsymbol{W}}^{-1} E[\boldsymbol{W}_{ij} A_{ij}]$ provided that $\boldsymbol{R}_{\boldsymbol{W}} := E[\boldsymbol{W}_{ij}\boldsymbol{W}_{ij}']$ is nonsingular;

(ii) The outcome regression is $E[Y_i|\boldsymbol{D}] = \boldsymbol{Z}_i(\boldsymbol{\zeta})'\boldsymbol{\beta}$, where $\boldsymbol{Z}_i(\boldsymbol{\zeta}) = (1, D_i, Q_i(\boldsymbol{\zeta}), R_i(\boldsymbol{\zeta}))$, $Q_i(\boldsymbol{\zeta}) = \sum_{j\neq i}(\boldsymbol{W}_{ij}'\boldsymbol{\zeta})D_j$, and $R_i(\boldsymbol{\zeta}) = \sum_{j\neq i}(\boldsymbol{W}_{ij}'\boldsymbol{\zeta})(1-D_j)$. The coefficient $\boldsymbol{\beta}$ is given by $\boldsymbol{\beta} = \boldsymbol{R}_{\boldsymbol{Z}}^{-1} E[\boldsymbol{Z}_i(\boldsymbol{\zeta})Y_i]$ provided that $\boldsymbol{R}_{\boldsymbol{Z}} := E[\boldsymbol{Z}_i(\boldsymbol{\zeta})\boldsymbol{Z}_i(\boldsymbol{\zeta})']$ is nonsingular;

(iii) The decomposition of the causal effects is given by $\boldsymbol{\pi} = (\pi^{DT}, \pi^{DN}, \pi^{IT}, \pi^{IN})$ with: $\pi^{DT} = \beta_2$, $\pi^{DN} = (N-1)\beta_4\zeta_2$, $\pi^{IT} = (\beta_3 - \beta_4)\zeta_1$, and $\pi^{IN} = \beta_3\zeta_3$.

Note that $Q_i(\boldsymbol{\zeta}) = \sum_{j\neq i} E[A_{ij}|D_i, D_j]D_j = E\left[\sum_{j\neq i} A_{ij}(D_i, D_j)D_j|\boldsymbol{D}\right] = E[Q_i|\boldsymbol{D}; \boldsymbol{\zeta}]$, and similarly, $R_i(\boldsymbol{\zeta}) = E[R_i|\boldsymbol{D}; \boldsymbol{\zeta}]$.

2.3.2 Identification with Parallel Trend

Next, I consider situations where the treatment is not guaranteed to be random, but pre-treatment information is available. We observe dyadic-level data $\{\mathcal{W}_{ij}\} = \{(A_{ij0}, A_{ij1})\}$, and individual-level data $\{\mathcal{V}_i\} = (D_i, Y_{i0}, Y_{i1})$, where A_{ijt} and Y_{it} denote the observed

link and outcome in period $t \in \{0,1\}$, respectively. Let $\varepsilon_{it}(d)$ represent individual i's error term, as defined in Assumption 2, at period $t \in \{0,1\}$, and $A_{ijt}(d,e)$ be the potential links of pair (i,j) at period $t \in \{0,1\}$. Identification in this situation is based on a difference-in-differences approach.

To this end, let Δ denote the first-difference operator, i.e., for a random variable K_t, $\Delta K = K_1 - K_0$. To distinguish outcome coefficients between the two periods, let $\boldsymbol{\beta}$ represent the coefficients at $t = 1$ and $\boldsymbol{\alpha}$ those at $t = 0$. The following assumptions are required to ensure parallel trend and no-anticipation in individual error term, potential links, and the potential outcome:

Assumption 6 (No Anticipation)**.** *(i)* $E[\varepsilon_{i0}(0)|D_{ig} = 1] = E[\varepsilon_{i0}(1)|D_{ig} = 1]$ *for each individual* i*; (ii)* $E[A_{ij0}(d,e)|D_i = d, D_j = e] = E[A_{ij0}(0,0)|D_i = d, D_j = e]$*, for each pair* (i,j)*, and for all* $(d,e) \in \{0,1\}^2$*. (iii)* $\alpha_2 = 0$ *and* $\alpha_3 = \alpha_4$*.*

Assumption 7 (Parallel Trend)**.** *(i)* $E[\Delta\varepsilon_i(0)|D_i = 1] = E[\Delta\varepsilon_i(0)|D_i = 0]$*, for each individual* i*; (ii)* $E[\Delta A_{ij}(0,0)|D_i = d, D_j = e] = E[\Delta A_{ij}(0,0)|D_i = 0, D_j = 0]$*, for all pairs* (i,j) *and for all* $(d,e) \in \{0,1\}^2$*. (iii)* $\alpha_4 = \beta_4$*.*

Assumption 6 ensures that there is no-anticipation of the treatment at the pre-treatment period. Since no individual is treated at $t = 0$, we can think of $\varepsilon_{i0} = \varepsilon_{i0}(0)$, and thus the first part of Assumption 6 follows. By the same argument, the second part holds when $A_{ij0} = A_{ij0}(0,0)$.

Assumption 7 is the key identifying assumption for a difference-in-differences estimand. For instance, in the identification of $E[A_{ij1}(d,e) - A_{ij1}(0,0)|D_i = d, D_j = e]$, the first term is directly observed, while the second term remains counterfactual. The second part of Assumption 7 recovers this counterfactual term by exploiting the exogenous parallel trend.

The restrictions on coefficient of potential outcome at $t = 0$ in last parts of Assumption 6 and Assumption 7, ensure both no-anticipation and parallel trends of potential outcomes. First, by Assumption 6-(i), we have $E[\varepsilon_{i0}(d_i)|D_i = d_i] = E[\varepsilon_{i0}(0)|D_i = d_i]$ for $d_i \in \{0,1\}$. Therefore,

$$E[Y_{i0}(\boldsymbol{d}) - Y_{i0}(\boldsymbol{0})|\boldsymbol{D} = \boldsymbol{d}] = \alpha_2 d_i + \alpha_3 E[Q_{i0}(\boldsymbol{d})|\boldsymbol{D} = \boldsymbol{d}] + \alpha_4 E[R_{i0}(\boldsymbol{d}) - S_{i0}(\boldsymbol{d})|\boldsymbol{D} = \boldsymbol{d}],$$

where $S_{i0}(\boldsymbol{d}) = Q_{i0}(\boldsymbol{d}) + R_{i0}(\boldsymbol{d}) = \sum_{j \neq i} A_{ij}(d_i, d_j)$. Thus, $\alpha_2 = 0$ and $\alpha_3 = \alpha_4$ implies $E[Y_{i0}(\boldsymbol{d})|\boldsymbol{D} = \boldsymbol{d}] = E[Y_{i0}(\boldsymbol{0})|\boldsymbol{D} = \boldsymbol{d}]$ for all $\boldsymbol{d} \in \{0,1\}^N$. Now, if $\alpha_4 = \beta_4$, then $\Delta Y_i(\boldsymbol{0}) = \beta_1 - \alpha_1 + \beta_4 \sum_{j \neq i} \Delta A_{ij}(0,0) + \Delta\varepsilon_i(0)$. It follows that

$$E[\Delta Y_i(\boldsymbol{0})|\boldsymbol{D} = \boldsymbol{d}] = \beta_1 - \alpha_1 + \beta_4 \sum_{j \neq i} E[\Delta A_{ij}(0,0)|\boldsymbol{D} = \boldsymbol{d}] + E[\Delta\varepsilon_{i1}(0)|D_i = d_i],$$
$$E[\Delta Y_i(\boldsymbol{0})|\boldsymbol{D} = \boldsymbol{0}] = \beta_1 - \alpha_1 + \beta_4 \sum_{j \neq i} E[\Delta A_{ij}(0,0)|\boldsymbol{D} = \boldsymbol{0}] + E[\Delta\varepsilon_{i1}(0)|D_i = 0].$$

Therefore, Assumption 7 implies that the parallel trend holds for the potential outcomes as well.

Similar to the identification procedure under randomized treatment, the coefficient in dyadic regression $E[A_{ijt}|D_i, D_j] = \boldsymbol{W}'_{ij}\boldsymbol{\zeta}_t$ is identified, and contains the conditional averages $m_t(d,e) = E[A_{ijt}(d,e)|D_i = d, D_j = e]$. The difference-in-differences of links is given by $\boldsymbol{\xi} = \boldsymbol{\zeta}_1 - \boldsymbol{\zeta}_2$, which represents the coefficient of the dyadic regression $E[\Delta A_{ij}|D_i, D_j]$.

The outcome coefficients are then identified by the regression of differenced outcome. Note that $\Delta Y_i = \beta_1 - \alpha_1 + \beta_2 D_i + \beta_3 Q_{i1} + \beta_4(R_{i1} - S_{i0}) + \Delta\varepsilon_i$. Thus, it can be written as $E[\Delta Y_i|\boldsymbol{D}] = E[X_i|\boldsymbol{D}]'\boldsymbol{\beta} + E[\Delta\varepsilon_i|\boldsymbol{D}]$. Here $\boldsymbol{\beta} = (\beta_1 - \alpha_1, \beta_2, \beta_3, \beta_4)'$, so the first element is different from that of randomized treatment case. Again, $E[X_i|\boldsymbol{D}] = (1, D_i, E[Q_{i1}|\boldsymbol{D}], E[R_{i1} - S_{i0}|\boldsymbol{D}])$ is estimated using the first-stage estimate of $\boldsymbol{\zeta}_t$ and $\boldsymbol{\xi}$, and $\boldsymbol{\beta}$ is identified subsequently.

Specifically, $\xi_2 = E[A_{ij1}(1,0) - A_{ij1}(0,0)|D_i = d, D_j = e]$ and $\xi_3 = E[A_{ij1}(0,1) - A_{ij1}(0,0)|D_i = 0, D_j = 1]$. Therefore, we can recover the decomposition $\boldsymbol{\pi}$ from $\boldsymbol{\beta}$, $\boldsymbol{\zeta}_1$, and $\boldsymbol{\xi}$. Proposition 4 formally states this identification procedure in this scenario:

Proposition 2. *Suppose Assumptions 1-4, 6, and 7 hold. Then,*

(i) For each $t \in \{0,1\}$, in a dyadic regression $E[A_{ijt}|\boldsymbol{D}] = \boldsymbol{W}'_{ij}\boldsymbol{\zeta}_t$, the coefficient $\boldsymbol{\zeta}_t$ is given by $\boldsymbol{\zeta}_t = \boldsymbol{R}_{\boldsymbol{W}}^{-1} E[\boldsymbol{W}_{ij} A_{ijt}]$ provided that $\boldsymbol{R}_{\boldsymbol{W}} := E[\boldsymbol{W}_{ij}\boldsymbol{W}'_{ij}]$ is nonsingular. Define $\boldsymbol{\zeta} = (\boldsymbol{\zeta}_0, \boldsymbol{\zeta}_1)$;

(ii) Define $\boldsymbol{\xi} = \boldsymbol{\zeta}_1 - \boldsymbol{\zeta}_2$ as the difference-in-differences estimand of links. Then, $\xi_2 = E[A_{ij1}(1,0) - A_{ij1}(0,0)|D_i = 1, D_j = 0]$, and $\xi_3 = E[A_{ij1}(0,1) - A_{ij1}(0,0)|D_i = 0, D_j = 1]$;

(iii) *The outcome regression is* $E[\Delta Y_i|\boldsymbol{D}] = \boldsymbol{X}_i(\boldsymbol{\zeta})'\boldsymbol{\beta}$, *where* $\boldsymbol{\beta} = (\beta_1 - \alpha_1, \beta_2, \beta_3, \beta_4)$, $\boldsymbol{X}_i(\boldsymbol{\zeta}) = (1, D_i, Q_{i1}(\boldsymbol{\zeta}), R_{i1}(\boldsymbol{\zeta}) - S_{i0}(\boldsymbol{\zeta}))$, $Q_{i1}(\boldsymbol{\zeta}) = \sum_{j\neq i}(\boldsymbol{W}'_{ij}\boldsymbol{\zeta}_1)D_j$, $R_{i1}(\boldsymbol{\zeta}) = \sum_{j\neq i}(\boldsymbol{W}'_{ij}\boldsymbol{\zeta}_1)(1 - D_j)$, *and* $S_{i0}(\boldsymbol{\zeta}) = \sum_{j\neq i}(\boldsymbol{W}'_{ij}\boldsymbol{\zeta}_0)$. *The coefficient* $\boldsymbol{\beta}$ *is given by* $\boldsymbol{\beta} = \boldsymbol{R}_{\boldsymbol{X}}^{-1}E[\boldsymbol{X}_i(\boldsymbol{\zeta})Y_i]$ *provided that* $\boldsymbol{R}_{\boldsymbol{X}} := E[\boldsymbol{X}_i(\boldsymbol{\zeta})\boldsymbol{X}_i(\boldsymbol{\zeta})']$ *is nonsingular;*

(iv) *The decomposition of causal effects is given by* $\boldsymbol{\pi} = (\pi^{DT}, \pi^{DN}, \pi^{IT}, \pi^{IN})$ *with* $\pi^{DT} = \beta_2$, $\pi^{DN} = (N-1)\beta_4\xi_2$, $\pi^{IT} = (\beta_3 - \beta_4)\zeta_1$, *and* $\pi^{IN} = \beta_3\xi_3$.

Remark (Identification Under Fixed Network)**.** If links are not affected by the treatment, then $\zeta_2 = \zeta_3 = 0$ in Proposition 1, and $\xi_2 = \xi_3 = 0$ in Proposition 2. Thus, there are no network effects in either direct or indirect effects. The estimation of outcome coefficients and direct and indirect treatment effects remains valid.

Remark (Identification Without Interactions)**.** If $\beta_3 = \beta_4 = 0$ in the outcome response model, then there are no indirect effects or direct network effects. The direct treatment effect is identified by the difference-in-means in Proposition 1, and the canonical difference-in-differences in Proposition 2.

3 Estimation and Inference

In this section, I propose estimators for the parameters identified in Section 2.3 and the decomposition defined in Section 2. Since all identification arguments are constructive, the natural choice for estimators is plug-in estimators. Additionally, because each parameter is defined by a projection coefficient for a conditional expectation, the plug-in estimators are essentially least squares estimators. Hence, the estimation procedure is straightforward but requires 3 stages. For each estimator, clustered standard errors can be employed to conduct inference, taking into account the dependency within groups.

3.1 Estimators

I propose a 3-stage procedure to estimate the parameters and their decomposition. In the first-stage, coefficients ($\boldsymbol{\zeta}$ under randomized experiment setting, or $\boldsymbol{\zeta}_1$, $\boldsymbol{\xi}$ under quasi-experiment setting) of dyadic regression of links are estimated. Subsequently, the

outcome coefficient $\boldsymbol{\beta}$ in Assumption 2 is estimated in the second-stage, by using the first-stage estimate. Finally, the decomposition of causal effects $\boldsymbol{\pi}$ are estimated in the third-stage, by using the estimates in the first-stage and the second-stage. In this Section, I continue to assume each group has N individuals for simplicity.

3.1.1 First-Stage Estimators

Randomized Experiment Setting: We observe the dyadic links $\{A_{ijg}\}_{(i,j),g}$, and recall that $\boldsymbol{W}_{ijg} = (1, D_{ig}, D_{jg}, D_{ig}D_{jg})'$ is the dyadic regressor for a pair (i,j) in group g. The coefficient of a dyadic regression $E[A_{ijg}|\boldsymbol{D}] = \boldsymbol{W}'_{ijg}\boldsymbol{\zeta}$ is estimated by the following least squares estimator:

$$\hat{\boldsymbol{\zeta}} = \left[\frac{1}{G}\sum_{g=1}^{G}\sum_{(i,j):i\neq j}\boldsymbol{W}_{ijg}\boldsymbol{W}'_{ijg}\right]^{-1}\left[\frac{1}{G}\sum_{g=1}^{G}\sum_{(i,j):i\neq j}\boldsymbol{W}_{ijg}A_{ijg}\right].$$

Quasi-Experiment Setting: We observe the dyadic links at pre- and post-treatment periods $\{(A_{ij0g}, A_{ij1g})\}_{(i,j),g}$. The coefficients of a dyadic regressions $E[A_{ijtg}|\boldsymbol{D}] = \boldsymbol{W}'_{ijg}\boldsymbol{\zeta}_t$ and $E[\Delta A_{ijg}|\boldsymbol{D}] = \boldsymbol{W}'_{ijg}\boldsymbol{\xi}$ are estimated by the following least squares estimators:

$$\hat{\boldsymbol{\zeta}}_t = \left[\frac{1}{G}\sum_{g=1}^{G}\sum_{(i,j):i\neq j}\boldsymbol{W}_{ijg}\boldsymbol{W}'_{ijg}\right]^{-1}\left[\frac{1}{G}\sum_{g=1}^{G}\sum_{(i,j):i\neq j}\boldsymbol{W}_{ijg}A_{ijtg}\right],$$
$$\hat{\boldsymbol{\xi}} = \left[\frac{1}{G}\sum_{g=1}^{G}\sum_{(i,j):i\neq j}\boldsymbol{W}_{ijg}\boldsymbol{W}'_{ijg}\right]^{-1}\left[\frac{1}{G}\sum_{g=1}^{G}\sum_{(i,j):i\neq j}\boldsymbol{W}_{ijg}\Delta A_{ijg}\right].$$

3.1.2 Second-Stage Estimators

Randomized Experiment Setting: We observe the individual-level treatment status and outcome $\{D_i, Y_i\}_{i,g}$, and first-stage estimate of conditional mean $E[\widehat{A_{ijg}|D_{ig}}, D_{jg}] = \boldsymbol{W}'_{ij}\hat{\boldsymbol{\zeta}}$. The regressor is constructed by $\boldsymbol{Z}_{ig}(\boldsymbol{\zeta}) = (1, D_{ig}, Q_{ig}(\boldsymbol{\zeta}), R_{ig}(\boldsymbol{\zeta}))$, where $Q_{ig}(\boldsymbol{\zeta}) = \sum_{j\neq i}\boldsymbol{W}'_{ijg}\boldsymbol{\zeta}D_{jg}$ and $R_{ig}(\boldsymbol{\zeta}) = \sum_{j\neq i}\boldsymbol{W}'_{ijg}\boldsymbol{\zeta}(1-D_{jg})$ for each individual i in group g. The coefficient of the outcome regression $E[Y_{ig}|\boldsymbol{D}] = \boldsymbol{Z}_{ig}(\boldsymbol{\zeta})'\boldsymbol{\beta}$ is estimated by the following

least squares estimator:

$$\hat{\boldsymbol{\beta}} = \left[\frac{1}{G}\sum_{g=1}^{G}\sum_{i=1}^{N} \boldsymbol{Z}_{ig}(\hat{\boldsymbol{\zeta}})\boldsymbol{Z}_{ig}(\hat{\boldsymbol{\zeta}})'\right]^{-1}\left[\frac{1}{G}\sum_{g=1}^{G}\sum_{i=1}^{N} \boldsymbol{Z}_{ig}(\hat{\boldsymbol{\zeta}})Y_{ig}\right].$$

Quasi-Experiment Setting: We observe the individual-level treatment status and outcomes at pre and post treatment periods $\{(D_{ig}, Y_{i0g}, Y_{i1g})\}_{i,g}$, and first-stage estimate of conditional mean $E[\widehat{A_{ijtg}|D_{ig}}, D_{jg}] = \boldsymbol{W}_{ij}'\hat{\boldsymbol{\zeta}}_t$, $t \in \{0,1\}$. Denote $\boldsymbol{\zeta} = (\boldsymbol{\zeta}_0', \boldsymbol{\zeta}_1')'$ and $\hat{\boldsymbol{\zeta}} = (\hat{\boldsymbol{\zeta}}_0', \hat{\boldsymbol{\zeta}}_1')'$. The regressor is constructed by $\boldsymbol{X}_{ig}(\boldsymbol{\zeta}) = (1, D_{ig}, Q_{ig}(\boldsymbol{\zeta}_1), R_{ig}(\boldsymbol{\zeta}_1) - S_{ig}(\boldsymbol{\zeta}_0))$, where $Q_{ig}(\boldsymbol{\zeta}_1) = \sum_{j\neq i}\boldsymbol{W}_{ijg}'\boldsymbol{\zeta}_1 D_{jg}$, $R_{ig}(\boldsymbol{\zeta}_1) = \sum_{j\neq i}\boldsymbol{W}_{ijg}'\boldsymbol{\zeta}_1(1-D_{jg})$, and $S_{ig}(\boldsymbol{\zeta}_0) = \sum_{j\neq i}\boldsymbol{W}_{ijg}'\boldsymbol{\zeta}_0$ for each individual i in group g. The coefficient of the outcome regression $E[\Delta Y_{ig}|\boldsymbol{D}] = \boldsymbol{X}_{ig}(\boldsymbol{\zeta})'\boldsymbol{\beta}$ is estimated by the following least squares estimator:

$$\hat{\boldsymbol{\beta}} = \left[\frac{1}{G}\sum_{g=1}^{G}\sum_{i=1}^{N} \boldsymbol{X}_{ig}(\hat{\boldsymbol{\zeta}})\boldsymbol{X}_{ig}(\hat{\boldsymbol{\zeta}})'\right]^{-1}\left[\frac{1}{G}\sum_{g=1}^{G}\sum_{i=1}^{N} \boldsymbol{X}_{ig}(\hat{\boldsymbol{\zeta}})\Delta Y_{ig}\right].$$

3.1.3 Estimator for the decomposition $\boldsymbol{\pi}$

Lastly, the decomposition $\boldsymbol{\pi}$ is estimated using a plug-in estimator.

Randomized Experiment Setting:	$\hat{\boldsymbol{\pi}} = \begin{pmatrix}\hat{\beta}_2 & (N-1)\hat{\beta}_4\hat{\zeta}_2 & (\hat{\beta}_3-\hat{\beta}_4)\hat{\zeta}_1 & \hat{\beta}_3\hat{\zeta}_3\end{pmatrix}$
Quasi-Experiment Setting:	$\hat{\boldsymbol{\pi}} = \begin{pmatrix}\hat{\beta}_2 & (N-1)\hat{\beta}_4\hat{\xi}_2 & (\hat{\beta}_3-\hat{\beta}_4)[\hat{\boldsymbol{\zeta}}_1]_1 & \hat{\beta}_3\hat{\xi}_3\end{pmatrix}$

Remark (Estimation with Covariates)**.** Identification and estimation under conditioning on covariates, or selection-on-observables, can be performed using the inverse probability weighting method proposed by Abadie (2005).

3.2 Inference

Since the proposed estimators are least squares estimators for projection coefficients, standard large sample theory can be applied. The random sample of independent groups plays a key role in the application of asymptotic theory. Let "$\xrightarrow{p}$" and "$\xrightarrow{d}$" denote convergence in probability and in distribution, respectively. First, Proposition 3 summarizes that the T-ratios for the estimators in each stage are asymptotically normal.

Proposition 3. *Suppose Assumptions 1-5 hold, and (i)* $\boldsymbol{R}_{\boldsymbol{W}} := E[\boldsymbol{W}_{ijg}\boldsymbol{W}'_{ijg}]$ *is nonsingular; (ii)* $\boldsymbol{R}_{\boldsymbol{Z}} := E[\boldsymbol{Z}_{ig}(\boldsymbol{\zeta}_0)\boldsymbol{Z}_{ig}(\boldsymbol{\zeta}_0)']$ *is nonsingular; (iii)* $E[Y_{ig}^4] < \infty$. *Let* $\boldsymbol{\zeta}^\star$, $\boldsymbol{\beta}^\star$, *and* $\boldsymbol{\pi}^\star$ *be true values of parameters. Then,* $(\hat{\boldsymbol{\zeta}}, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\pi}}) \xrightarrow{p} (\boldsymbol{\zeta}^\star, \boldsymbol{\beta}^\star, \boldsymbol{\pi}^\star)$ *and*

$$\begin{aligned}
\hat{V}_{\boldsymbol{\zeta}}^{-1/2}\sqrt{G}(\hat{\boldsymbol{\zeta}} - \boldsymbol{\zeta}^\star) &\xrightarrow{d} N(0,1),\\
\hat{V}_{\boldsymbol{\beta}}^{-1/2}\sqrt{G}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^\star) &\xrightarrow{d} N(0,1),\\
\hat{V}_{\boldsymbol{\pi}}^{-1/2}\sqrt{G}(\hat{\boldsymbol{\pi}} - \boldsymbol{\pi}^\star) &\xrightarrow{d} N(0,1).
\end{aligned}$$

$\mathcal{W}_g := \{(A_{ijg}, \boldsymbol{W}_{ijg})\}_{(i,j)}$, $\mathcal{V}_g = \{(D_{ig}, Y_{ig})\}_i$ *are clustered group-level data. The influence functions of* $\boldsymbol{\zeta}$, $\boldsymbol{\beta}$ *are given by*

$$\begin{aligned}
\psi_{\boldsymbol{\zeta}}(\mathcal{W}_g, \boldsymbol{\zeta}) &:= \boldsymbol{R}_{\boldsymbol{W}}^{-1}\frac{1}{N(N-1)}\sum_{(i,j):i\neq j}\boldsymbol{W}_{ijg}(A_{ijg} - \boldsymbol{W}'_{ijg}\boldsymbol{\zeta}),\\
\psi_{\boldsymbol{\beta}}(\mathcal{V}_g, \mathcal{W}_g, \boldsymbol{\zeta}, \boldsymbol{\beta}) &:= \boldsymbol{R}_{\boldsymbol{Z}}^{-1}\left[\boldsymbol{Z}_{ig}(\boldsymbol{\zeta})(Y_{ig} - \boldsymbol{Z}_{ig}(\boldsymbol{\zeta})'\boldsymbol{\beta}) - \boldsymbol{Q}_{\boldsymbol{\zeta}}\psi_{\boldsymbol{\zeta}}(\mathcal{W}_g, \boldsymbol{\zeta})\right],
\end{aligned}$$

where $\boldsymbol{Q}_{\boldsymbol{\zeta}} := E[\boldsymbol{Z}_{ig}(\boldsymbol{\zeta}^\star)\nabla_{\boldsymbol{\zeta}}(\boldsymbol{Z}_{ig}(\boldsymbol{\zeta}^\star)'\boldsymbol{\beta}^\star)]$. *And the influence function* $\psi_{\boldsymbol{\pi}}(\mathcal{V}_g, \mathcal{W}_g, \boldsymbol{\zeta}, \boldsymbol{\beta})$ *is given by*

$$\begin{pmatrix}
[\psi_{\boldsymbol{\beta}}(\mathcal{W}_g, \mathcal{V}_g, \boldsymbol{\zeta}, \boldsymbol{\beta})]_2\\
(N-1)[\psi_{\boldsymbol{\beta}}(\mathcal{W}_g, \mathcal{V}_g, \boldsymbol{\zeta}, \boldsymbol{\beta})]_4[\zeta^\star]_2 + [\boldsymbol{\beta}^\star]_4[\psi_{\boldsymbol{\zeta}}(\mathcal{W}_g, \boldsymbol{\zeta})]_2\\
([\psi_{\boldsymbol{\beta}}(\mathcal{W}_g, \mathcal{V}_g, \boldsymbol{\zeta}, \boldsymbol{\beta})]_3 - [\psi_{\boldsymbol{\beta}}(\mathcal{W}_g, \mathcal{V}_g, \boldsymbol{\zeta}, \boldsymbol{\beta})]_4)\,[\boldsymbol{\zeta}^\star]_1 + ([\boldsymbol{\beta}^\star]_3 - [\boldsymbol{\beta}^\star]_4)[\psi_{\boldsymbol{\zeta}}(\mathcal{W}_g, \boldsymbol{\zeta})]_1\\
[\psi_{\boldsymbol{\beta}}(\mathcal{W}_g, \mathcal{V}_g, \boldsymbol{\zeta}, \boldsymbol{\beta})]_3[\boldsymbol{\zeta}^\star]_3 + [\boldsymbol{\beta}^\star]_3[\psi_{\boldsymbol{\zeta}}(\mathcal{W}_g, \boldsymbol{\zeta})]_3
\end{pmatrix},$$

where $[v]_k$ *denote* k*-th element in vector* v. *Lastly,* $\hat{V}_{\boldsymbol{\zeta}}$, $\hat{V}_{\boldsymbol{\beta}}$, $\hat{V}_{\boldsymbol{\pi}}$ *are sample variance matrices of* $\psi_{\boldsymbol{\zeta}}(\mathcal{W}_g, \hat{\boldsymbol{\zeta}})$, $\psi_{\boldsymbol{\beta}}(\mathcal{V}_g, \hat{\boldsymbol{\zeta}}, \hat{\boldsymbol{\beta}})$, $\psi_{\boldsymbol{\pi}}(\mathcal{V}_g, \mathcal{W}_g, \hat{\boldsymbol{\zeta}}, \hat{\boldsymbol{\beta}})$, *respectively, and* $V^{-1/2}$ *denote a square root matrix of* V^{-1}.

Since A_{ijg}, D_{ig} are indicator variables, the boundedness of the moment $E[Y_{ig}^4]$ is sufficient to apply the law of large numbers and the central limit theorem. Proposition 3 implies that the decomposition $\hat{\boldsymbol{\pi}}$ has an asymptotic normal distribution with zero mean and asymptotic variance $E[\psi_{\boldsymbol{\pi}}(\mathcal{V}_g, \mathcal{W}_g, \boldsymbol{\zeta}^\star, \boldsymbol{\beta}^\star)\psi_{\boldsymbol{\pi}}(\mathcal{V}_g, \mathcal{W}_g, \boldsymbol{\zeta}^\star, \boldsymbol{\beta}^\star)']$. The plug-in standard errors are computed as the square root of the diagonal elements of the variance of the empirical influence function $\psi_{\boldsymbol{\pi}}(\mathcal{V}_g, \mathcal{W}_g, \hat{\boldsymbol{\zeta}}, \hat{\boldsymbol{\beta}})$, which is consistent with the asymptotic variance and therefore valid asymptotically. Furthermore, the asymptotic variances of $\hat{\boldsymbol{\zeta}}$ and $\hat{\boldsymbol{\beta}}$ are given by the variances of $\psi_{\boldsymbol{\zeta}}(\mathcal{W}_g, \boldsymbol{\zeta}^\star)$, and $\psi_{\boldsymbol{\beta}}(\mathcal{V}_g, \mathcal{W}_g, \boldsymbol{\zeta}^\star, \boldsymbol{\beta}^\star)$, respectively. The plug-in standard errors for $\boldsymbol{\beta}$ and $\boldsymbol{\zeta}$ are defined similarly.

The limiting distribution of estimators in quasi-experiment setting with parallel trend can be similarly established, and it is summarized in Proposition 4:

Proposition 4. *Suppose Assumptions 1-4, 6, 7 hold, and (i)* $\boldsymbol{R}_{\boldsymbol{W}} := E[\boldsymbol{W}_{ijg}\boldsymbol{W}_{ijg}']$ *is nonsingular; (ii)* $\boldsymbol{R}_{\boldsymbol{X}} := E[\boldsymbol{X}_{ig}(\boldsymbol{\zeta}^\star)\boldsymbol{X}_{ig}(\boldsymbol{\zeta}^\star)']$ *is nonsingular; (iii)* $E[Y_{itg}^4] < \infty$ *for* $t \in \{0, 1\}$. *Let* $\boldsymbol{\zeta}_t^\star$, $\boldsymbol{\xi}^\star$, $\boldsymbol{\beta}^\star$, *and* $\boldsymbol{\pi}^\star$ *be true values of parameters. Then,* $(\hat{\boldsymbol{\zeta}}_t, \hat{\boldsymbol{\xi}}, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\pi}}) \xrightarrow{p} (\boldsymbol{\zeta}_t^\star, \boldsymbol{\xi}^\star, \boldsymbol{\beta}^\star, \boldsymbol{\pi}^\star)$ *and*

$$
\begin{aligned}
\hat{V}_{\zeta,t}^{-1/2}\sqrt{G}(\hat{\boldsymbol{\zeta}}_t - \boldsymbol{\zeta}_t^\star) &\xrightarrow{d} N(0,1),\\
\hat{V}_{\boldsymbol{\xi}}^{-1/2}\sqrt{G}(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}^\star) &\xrightarrow{d} N(0,1),\\
\hat{V}_{\boldsymbol{\beta}}^{-1/2}\sqrt{G}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^\star) &\xrightarrow{d} N(0,1),\\
\hat{V}_{\boldsymbol{\pi}}^{-1/2}\sqrt{G}(\hat{\boldsymbol{\pi}} - \boldsymbol{\pi}^\star) &\xrightarrow{d} N(0,1),
\end{aligned}
$$

where $\mathcal{W}_g := \{(A_{ij0g}, A_{ij1g}, \boldsymbol{W}_{ijg})\}_{(i,j)}$, $\mathcal{V}_g = \{(D_{ig}, Y_{i0g}, Y_{i1g})\}_i$ *are clustered group-level data. Let* $\boldsymbol{\zeta} = (\boldsymbol{\zeta}_0, \boldsymbol{\zeta}_1)$. *The influence functions of* $\boldsymbol{\zeta}_t$, $\boldsymbol{\xi}$, $\boldsymbol{\beta}$ *are given by*

$$
\begin{aligned}
\psi_{\zeta,t}(\mathcal{W}_g, \boldsymbol{\zeta}_t) &:= \boldsymbol{R}_{\boldsymbol{W}}^{-1}\frac{1}{N(N-1)}\sum_{(i,j):i\neq j}\boldsymbol{W}_{ijg}(A_{ijtg} - \boldsymbol{W}_{ijg}'\boldsymbol{\zeta}_t),\\
\psi_{\boldsymbol{\xi}}(\mathcal{W}_g, \boldsymbol{\xi}) &:= \boldsymbol{R}_{\boldsymbol{W}}^{-1}\frac{1}{N(N-1)}\sum_{(i,j):i\neq j}\boldsymbol{W}_{ijg}(\Delta A_{ijg} - \boldsymbol{W}_{ijg}'\boldsymbol{\xi}),\\
\psi_{\boldsymbol{\beta}}(\mathcal{V}_g, \mathcal{W}_g, \boldsymbol{\zeta}, \boldsymbol{\beta}) &:= \boldsymbol{R}_{\boldsymbol{Z}}^{-1}\left[\boldsymbol{Z}_{ig}(\boldsymbol{\zeta})(\Delta Y_{ig} - \boldsymbol{Z}_{ig}(\boldsymbol{\zeta})'\boldsymbol{\beta}) - \boldsymbol{Q}_{\zeta,1}\psi_{\zeta,1}(\mathcal{W}_g, \boldsymbol{\zeta}_1) - \boldsymbol{Q}_{\zeta,2}\psi_{\zeta,2}(\mathcal{W}_g, \boldsymbol{\zeta}_2)\right],
\end{aligned}
$$

where $Q_{\zeta_t} := E[\boldsymbol{X}_{ig}(\boldsymbol{\zeta}^\star)\nabla_{\zeta_t}(\boldsymbol{X}_{ig}(\boldsymbol{\zeta}^\star)'\boldsymbol{\beta}^\star)]$. *And* $\psi_{\boldsymbol{\pi}}(\mathcal{V}_g, \mathcal{W}_g, \boldsymbol{\zeta}, \boldsymbol{\beta})$ *is defined by*

$$\begin{pmatrix} [\psi_{\boldsymbol{\beta}}(\mathcal{W}_g, \mathcal{V}_g, \boldsymbol{\zeta}, \boldsymbol{\beta})]_2 \\ (N-1)[\psi_{\boldsymbol{\beta}}(\mathcal{W}_g, \mathcal{V}_g, \boldsymbol{\zeta}, \boldsymbol{\beta})]_4[\xi^\star]_2 + [\boldsymbol{\beta}^\star]_4[\psi_{\boldsymbol{\xi}}(\mathcal{W}_g, \boldsymbol{\xi})]_2 \\ ([\psi_{\boldsymbol{\beta}}(\mathcal{W}_g, \mathcal{V}_g, \boldsymbol{\zeta}, \boldsymbol{\beta})]_3 - [\psi_{\boldsymbol{\beta}}(\mathcal{W}_g, \mathcal{V}_g, \boldsymbol{\zeta}, \boldsymbol{\beta})]_4)\,[\boldsymbol{\zeta}_1^\star]_1 + ([\boldsymbol{\beta}^\star]_3 - [\boldsymbol{\beta}^\star]_4)[\psi_{\boldsymbol{\zeta},1}(\mathcal{W}_g, \boldsymbol{\zeta})]_1 \\ [\psi_{\boldsymbol{\beta}}(\mathcal{W}_g, \mathcal{V}_g, \boldsymbol{\zeta}, \boldsymbol{\beta})]_3[\boldsymbol{\xi}^\star]_3 + [\boldsymbol{\beta}^\star]_3[\psi_{\boldsymbol{\xi}}(\mathcal{W}_g, \boldsymbol{\xi})]_3 \end{pmatrix},$$

where $[v]_k$ *denote* k*-th element in vector* v. *Lastly,* $\hat{V}_{\boldsymbol{\zeta},t}, \hat{V}_{\boldsymbol{\xi}}, \hat{V}_{\boldsymbol{\beta}}, \hat{V}_{\boldsymbol{\pi}}$ *are sample variance matrices of* $\psi_{\boldsymbol{\zeta},t}(\mathcal{W}_g, \hat{\boldsymbol{\zeta}}_t)$, $\psi_{\boldsymbol{\xi}}(\mathcal{W}_g, \hat{\boldsymbol{\xi}})$, $\psi_{\boldsymbol{\beta}}(\mathcal{V}_g, \hat{\boldsymbol{\zeta}}, \hat{\boldsymbol{\beta}})$, $\psi_{\boldsymbol{\pi}}(\mathcal{V}_g, \mathcal{W}_g, \hat{\boldsymbol{\zeta}}, \hat{\boldsymbol{\beta}})$, *respectively, and* $V^{-1/2}$ *denote a square root matrix of* V^{-1}.

The baseline argument is the same as in Proposition 3 since all regressors are similar. Note that $\boldsymbol{\xi}$ represents the difference-in-differences coefficient, which captures the average treatment effect on treated (ATT) for links, and is directly used to compute $\boldsymbol{\pi}$. Since $\boldsymbol{\xi}$ and $\boldsymbol{\zeta}_1$ are required to recover the decomposition, estimating $\boldsymbol{\zeta}_0$ is unnecessary. Once again, inference based on plug-in clustered standard errors remains asymptotically valid.

4 A Monte Carlo Study

To investigate the finite sample properties of the estimators introduced in Section 3, I conduct simulations using data generated based on the assumptions outlined in Section 2, across various sample sizes (number of groups).

First, the treatment indicators D_{ig} are generated from a Bernoulli distribution with a success probability of $P_D = 0.5$. Links and outcomes are then generated under two different settings: (i) a randomized experiment, and (ii) a quasi-experiment with parallel trends.

Design 1: Randomized Experiment

First, define $I_{\boldsymbol{\theta}}(d, e) := (1, d, e, de)\boldsymbol{\theta} = \theta_1 + \theta_2 d + \theta_3 e + \theta_4 de$ as potential index. The potential link of pair (i, j), given treatment statuses $(D_i, D_j) = (d, e)$ is generated by a binary response $A_{ij}(d, e) = \mathbb{1}\{I_{\boldsymbol{\theta}}(d, e) \geq \varepsilon_{ij}\}$, where ε_{ij} follows a standard normal

distribution. The mean and the average treatment effect (ATE) on links are given by $m(d,e) := E[A_{ij}(d,e)] = \Phi(I_\theta(d,e))$, and $H(d,e) = E[A_{ij}(d,e) - A_{ij}(0,0)] = \Phi(I_{\boldsymbol{\theta}}(d,e)) - \Phi(\theta_1)$, respectively. Let $\boldsymbol{m} = (m(1,1), m(1,0), m(0,1), m(0,0))'$ be a vector of potential means. The coefficient of dyadic regression for the observed link $A_{ij} = A_{ij}(D_i, D_j)$ on $\boldsymbol{W}_{ij} = (1, D_i, D_j, D_iD_j)$ is then given by $\boldsymbol{\zeta} = \boldsymbol{M}\boldsymbol{m}$, where

$$\boldsymbol{M} = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix}.$$

The outcome if individual i is generated as: $Y_i = \beta_1 + \beta_2 D_i + \beta_3 Q_i + \beta_4 R_i + \nu_i$, where $Q_i = \sum_{j \neq i} A_{ij} D_j$, $R_i = \sum_{j \neq i} A_{ij}(1 - D_j)$, $\nu_i = u_i + \sum_{j \neq i} \varepsilon_{ij}$, and u_i is the standard normal error term. The individual error term ν_i is independent of treatment, but it is correlated with the network. As a result, the coefficient of linear-in-means regression (i.e., regression of Y_i on $(1, D_i, Q_i, R_i)$) does not recover the parameter $\boldsymbol{\beta}$ since $E[\nu_i | \boldsymbol{A}, \boldsymbol{D}] \neq 0$. I assess the bias of this model assuming exogenous network in Table 2. The generated data consists of a dyadic level data $\{A_{ijg}\}_{(i,j):i \neq j,g}$, and individual level data $\{(D_{ig}, Y_{ig})\}_{i,g}$.

Design 2: Quasi-Experiment with Parallel Trend

Let $h_t(\cdot,\cdot) = I_\omega(d,e)$ be a symmetric function for $t \in \{0,1\}$. The potential link for a pair (i,j) at given treatment statuses $(D_i, D_j) = (d,e)$ is generated by a binary response $A_{ij0} = \mathbb{1}\{h_0(D_i, D_j) \geq \varepsilon_{ij0}\}$ at $t = 0$, and $A_{ij1} = \mathbb{1}\{I\boldsymbol{\theta}(d,e) + h_1(D_i, D_j) \geq \varepsilon_{ijt}\}$ at $t = 1$, where ε_{ijt} are standard normal error terms. Thus, the potential links are not independent of the treatment by construction. The conditional means at each period are given by $m_0(d,e) := E[A_{ij0}(d,e) | D_i = d', D_j = e'] = \Phi(h_0(d',e'))$, and $m_1(d,e) := E[A_{ij1}(d,e) | D_i = d', D_j = e'] = \Phi(I\boldsymbol{\theta}(d,e) + h_1(d',e'))$. The average treatment effect on treated (ATT) on links is: $H(d,e) = m_1(d,e) - m_0(d,e)$.

Here, $m_0(d,e)$ does not depend on (d,e), and hence Assumption 6-(ii) holds. Suppose $h_1(d,e) = h_0(d,e) - I_{\boldsymbol{\theta}}(0,0) = h_0(d,e) - \theta_1$. Then, Assumption 7-(ii) is satisfied

by construction since:[6]

$$E[\Delta A_{ij}(0,0)|D_i = d, D_j = e] = \Phi(I_{\boldsymbol{\theta}}(0,0) + h_1(d,e)) - \Phi(h_0(d,e)) = 0.$$

Specifically, I set $h_0(d,e) = I_{\boldsymbol{\omega}}(d,e)$ with a coefficient ω. The coefficient of dyadic regressions of observed link A_{ijt}, and first-difference link ΔA_{ij} on $\boldsymbol{W}_{ij}$ is given by $\boldsymbol{\zeta}_t = \boldsymbol{M}\boldsymbol{m}_t,\ t \in \{0,1\}$ and $\boldsymbol{\xi} = \boldsymbol{\zeta}_1 - \boldsymbol{\zeta}_0$, where $\boldsymbol{m}_t = (m_t(1,1), m_t(1,0), m_t(0,1), m_t(0,0))'$.

At $t = 1$, the outcome is generated by the same way as in the Design 1, and that at $t = 0$, the outcome is generated by $Y_{i0} = \alpha + \beta_4 S_{i0} + \nu_{i0}$, where $S_{i0} = \sum_{j \neq i} A_{ij0}$ is the number of neighbors in the pre-treatment period. Since the individual error term contains dyadic error terms related to potential links, Q_i, R_i, S_i are endogenous regressors. The generated data consists of a dyadic-level data $\{(D_{ig}, D_{jg}, A_{ij0g}, A_{ij1g})\}_{(i,j):i\neq j,g}$, and individual-level data $\{(D_{ig}, Y_{i0g}, Y_{i1g})\}_{i,g}$.

The coefficients are set by $\boldsymbol{\beta} = (2, 1, 0.8, 0.6)$, $\alpha = 1$, $\boldsymbol{\theta} = (-1, 0.1, 0.1, 1)$, and $\boldsymbol{\omega} = (-1.5, 0.3, 0.3, -1)$. The corresponding true value of dyadic coefficients are $\boldsymbol{\zeta} = (0.159, 0.025, 0.025, 0.370)$ for Design 1, and $\boldsymbol{\zeta}_1 = (0.067, 0.069, 0.069, 0.037)$, $\boldsymbol{\zeta}_0 = (0.067, 0.048, 0.048, -0.135)$, and $\boldsymbol{\xi} = (0, 0.021, 0.021, 0.172)$ for Design 2. Lastly, the true value of decomposition is $\boldsymbol{\pi} = (1, 0.290, 0.032, 0.020)$ for Design 1, and $\boldsymbol{\pi} = (1, 0.235, 0.013, 0.016)$ for Design 2.

The estimators proposed in Section 3, along with clustered standard errors for each coefficient, are computed. The coverage rate is calculated as the proportion of cases in which the true value is included in the 95% confidence interval across all simulations. The mean squared error (MSE) is calculated as the average squared deviation between the estimate and its true value over all simulations.

Table 1 presents the results for the decomposition ($\boldsymbol{\pi}$). In the first panel, the average estimates are closely aligned with the true values, even with a relatively small group size in both designs. The second and third panels show that the MSE decreases at a rate of G^{-1}, and the coverage rate of the confidence intervals, based on clustered errors, is near the nominal 95% level. This supports the validity of the proposed asymptotic

[6]This design impose more than parallel trend, since there is no trend. If the trend is given by $E[A_{ij1}(0,0) - A_{ij0}(0,0)|D_i = d', D_j = e'] = T$, then one can define $h_1(d',e')$ as $h_1(d',e') := \Phi^{-1}(T + \Phi(h(d',e'))) - \theta_1$.

theory and confirms that the clustered standard errors perform well. For the outcome coefficient results ($\boldsymbol{\beta}$), see Table 7, and for the dyadic coefficients $\boldsymbol{\zeta}, \boldsymbol{\zeta}_1, \boldsymbol{\xi}$, see Table 6 in Appendix B.

Table 1: Simulation Result of Decomposition

	Design 1				**Design 2**			
G	π^{DT}	π^{DN}	π^{IT}	π^{IN}	π^{DT}	π^{DN}	π^{IT}	π^{IN}
Mean								
100	0.987	0.301	0.032	0.021	1.021	0.244	0.015	0.017
200	0.983	0.299	0.031	0.021	0.976	0.247	0.013	0.017
400	0.995	0.292	0.031	0.02	1.001	0.239	0.013	0.016
800	0.999	0.291	0.032	0.02	0.985	0.239	0.013	0.017
1600	0.998	0.291	0.032	0.02	0.993	0.237	0.013	0.017
True	1	0.29	0.032	0.02	1	0.235	0.013	0.016
MSE								
100	0.308	0.043	1.36e-4	0.05e-4	0.772	0.087	0.98e-4	0.08e-4
200	0.146	0.02	0.65e-4	0.02e-4	0.367	0.039	0.48e-4	0.04e-4
400	0.078	0.01	0.34e-4	0.01e-4	0.187	0.019	0.23e-4	0.02e-4
800	0.035	0.005	0.17e-4	0.01e-4	0.09	0.009	0.12e-4	0.01e-4
1600	0.017	0.002	0.08e-4	0e-4	0.042	0.004	0.06e-4	0e-4
Coverage Rate								
100	0.93	0.894	0.934	0.925	0.934	0.913	0.938	0.904
200	0.932	0.926	0.944	0.928	0.931	0.923	0.945	0.914
400	0.926	0.911	0.938	0.919	0.94	0.929	0.941	0.933
800	0.938	0.931	0.945	0.946	0.941	0.937	0.947	0.94
1600	0.944	0.926	0.942	0.939	0.951	0.955	0.952	0.951

Notes: This table presents the simulation results for $B = 1,000$ replications. Column G denotes the number of independent groups, with each group consisting of $N = 20$ individuals. The first panel shows the mean across all replications, and the row labeled "TRUE" provides the true values for each decomposition. The second and third columns display the mean squared error (MSE) and 95% coverage rates, respectively. $(\pi^{DT}, \pi^{DN}, \pi^{IT}, \pi^{IN})$ represent the direct treatment, direct network, indirect treatment, and indirect network effects, respectively.

It is important to recall that the network is correlated with the individual error term, meaning that the estimation of the linear-in-means regression will be biased. The linear-in-means regression assumes a fixed or exogenous network, ignoring potential causal changes in the network. Table 2 and Table 3 compare the outcome coefficients from the proposed estimation method with those from the linear-in-means regression. These comparisons show that neglecting network changes leads to significantly biased

estimates.

Table 2: Bias When Assuming Exogenous Network (Design 1)

	Considering Network Change				**Exogenous Network**			
G	β_1	β_2	β_3	β_4	β_1	β_2	β_3	β_4
Mean								
100	1.996	0.987	0.802	0.601	4.634	3.785	0.119	-0.358
200	1.982	0.983	0.804	0.609	4.635	3.789	0.119	-0.358
400	1.994	0.995	0.801	0.603	4.635	3.785	0.119	-0.358
800	2.002	0.999	0.8	0.599	4.635	3.787	0.119	-0.359
1600	1.997	0.998	0.8	0.601	4.634	3.786	0.12	-0.359
TRUE	2	1	0.8	0.6	2	1	0.8	0.6
MSE								
100	0.507	0.308	0.015	0.114	6.951	7.777	0.464	0.919
200	0.237	0.146	0.007	0.054	6.948	7.786	0.465	0.919
400	0.125	0.078	0.004	0.028	6.946	7.76	0.463	0.918
800	0.059	0.035	0.002	0.013	6.946	7.77	0.463	0.92
1600	0.03	0.017	0.001	0.007	6.938	7.764	0.463	0.919
Coverage Rate								
100	0.94	0.944	0.946	0.936	0	0	0	0
200	0.957	0.94	0.946	0.951	0	0	0	0
400	0.942	0.938	0.939	0.94	0	0	0	0
800	0.947	0.95	0.959	0.947	0	0	0	0
1600	0.938	0.95	0.951	0.944	0	0	0	0

Notes: This table presents the simulation results for $B = 1,000$ replications. Column G denotes the number of independent groups, with each group consisting of $N = 20$ individuals. The first panel shows the mean across all replications, and the row labeled "TRUE" provides the true values for each decomposition. The second and third columns display the mean squared error (MSE) and 95% coverage rates, respectively. The first 4 columns, labeled "Considering Network Change," show the results from the proposed estimation, while the last 4 columns, labeled "Exogenous Network," show the results from the regression of Y_i on $1, D_i, Q_i, R_i$, which is valid only when $E[\nu_i|\boldsymbol{A}, \boldsymbol{D}] = 0$.

Table 3: Bias When Assuming Exogenous Network (Design 2)

	Considering Network Change				Exogenous Network			
G	$\beta_1 - \alpha$	β_2	β_3	β_4	$\beta_1 - \alpha$	β_2	β_3	β_4
Mean								
100	0.971	1.021	0.801	0.58	0.971	2.84	-0.021	-0.391
200	1.01	0.976	0.807	0.617	0.976	2.832	-0.02	-0.391
400	1.006	1.001	0.797	0.604	0.977	2.838	-0.023	-0.39
800	1.005	0.985	0.804	0.608	0.977	2.836	-0.022	-0.391
1600	1.002	0.993	0.802	0.605	0.978	2.837	-0.022	-0.39
TRUE	1	1	0.8	0.6	1	1	0.8	0.6
MSE								
100	0.27	0.772	0.077	0.392	0.02	3.408	0.677	0.985
200	0.13	0.367	0.038	0.189	0.01	3.367	0.675	0.983
400	0.065	0.187	0.019	0.094	0.005	3.385	0.679	0.981
800	0.033	0.09	0.009	0.046	0.003	3.372	0.676	0.982
1600	0.015	0.042	0.004	0.021	0.002	3.375	0.676	0.98
Coverage Rate								
100	0.943	0.944	0.942	0.943	0.952	0	0	0
200	0.949	0.938	0.946	0.94	0.945	0	0	0
400	0.947	0.944	0.951	0.944	0.948	0	0	0
800	0.951	0.947	0.95	0.949	0.926	0	0	0
1600	0.961	0.958	0.961	0.958	0.893	0	0	0

Notes: This table presents the simulation results for $B = 1,000$ replications. Column G denotes the number of independent groups, with each group consisting of $N = 20$ individuals. The first panel shows the mean across all replications, and the row labeled "TRUE" provides the true values for each decomposition. The second and third columns display the mean squared error (MSE) and 95% coverage rates, respectively. The first 4 columns, labeled "Considering Network Change," show the results from the proposed estimation, while the last 4 columns, labeled "Exogenous Network," show the results from the regression of ΔY_i on $1, D_i, Q_i, R_i - S_i$, which is only valid when $E[\Delta\nu_i|\boldsymbol{A}, \boldsymbol{D}] = 0$.

5 Empirical Illustration

In this section, I apply the proposed estimation method to data from studies conducted by Comola and Prina (2020) and Comola and Prina (2021). The dataset comes from a randomized experiment carried out in villages surrounding Pokhara, Nepal, from 2009 to 2011. The treatment in this experiment involved providing access to savings accounts for households. The pre-treatment survey took place in February 2009, and

the treatment was randomly assigned to half of the households in June 2010 via a public lottery.

According to Prina (2015), formal banking services in Nepal are highly limited, with only 20% of households having a bank account. At the start of the experiment, only 17% of participants had savings accounts, with most keeping their cash at home. The experiment aimed to assess the impact of providing a savings account on economic behaviors such as consumption. Specifically, the treatment offered households the option to open a savings account. The main effects estimated in Comola and Prina (2021) are intent-to-treat (ITT) effects. However, as reported by Prina (2015), the take-up rate was quite high, with 84% of treated households opening an account, and 80% of those actively using it.

The outcome variable of interest in Comola and Prina (2021) is household meat consumption, which may be influenced by peers' consumption behaviors. Using their proposed method, the authors found positive direct and indirect effects on meat consumption in response to opening a savings account. The sample consists of 915 households in 19 villages, including information on their financial networks. The network information is constructed as undirected, where $A_{ij} = 1$ if at least one household i reported having repeated financial exchanges with household j. The network is block-diagonal, as links are based within villages, resulting in a total of 56,308 dyads.

Comola and Prina (2021) estimate a two-period version of the linear-in-means model using an IV estimation strategy similar to that of Bramoullé, Djebbari, and Fortin (2009). They compute the direct and indirect effects as derivatives of the reduced-form outcome. The linear-in-means structure implies specific derivatives representing changes in links in response to the treatment, which correspond to the average treatment effects on treated row-normalized links. The authors estimate a regression of differenced row-normalized links on dyadic treatment, defined as *some treated*, i.e., $\max(D_i, D_j)$.

In contrast, the estimation procedure proposed in this paper estimates the difference in links based on (D_i, D_j, D_iD_j). Table 4 presents the results of these regressions. The first column shows the regression results from Comola and Prina (2021), where the coefficient of the dyadic treatment is estimated at 0.002. This indicates that the average change in row-normalized links (or the probability of being linked) increases by 0.002

percentage points in response to the dyadic treatment.

Table 4: Average Treatment Effects on Treated of Links

Var	Comola, Prina (2021)	Row-Normalized Link	Link
Constant	-0.001	-0.001	-0.003***
	(0.001)	(0.001)	(0.001)
$\max\{D_1, D_2\}$	0.002**		
	(0.001)		
D_1		0.002*	0.004**
		(0.001)	(0.002)
D_2		0.002**	0.004**
		(0.001)	(0.002)
$D_1 \times D_2$		-0.003	-0.003
		(0.002)	(0.002)
Observations	56,308	56,308	56,308

Notes: To compare the results in Comola and Prina (2021), each regression controls for dyadic information such as marital status, children, livestock, and death using a linear model. The dependent variable in the third column is $A_{ij1} - A_{ij0}$, while in the first two columns, it is $A^s_{ij1} - A^s_{ij0}$, where $A^s_{ijt} = A_{ijt} / \sum_{j \neq i} A_{ijt}$ represents the row-normalized links. Standard errors are reported in parentheses. *,** ,*** denote the significance levels at 10%, 5%, and 1%, respectively.

The second and third columns in Table 4 display the first-stage dyadic regression on the differenced link. The second column estimates the regression of differenced row-normalized links for comparison with the first column. Due to the undirected nature of the network, the coefficients for D_1 and D_2 are nearly identical. The actual first-stage regression used in the proposed estimation procedure is shown in the third column. Here, the coefficient estimates for D_1 and D_2 are about double those in the second column, with $H(1,0) = H(0,1)$ estimated at 0.004, because it is not row-normalized. That is, the average degree (i.e., number of neighbors) is about 2, which implies the network is substantially sparse.

Table 5 illustrates how the causal effects are decomposed and estimated, compared to the estimation results in Comola and Prina (2021). In Comola and Prina (2021), direct and indirect effects are computed as the derivative of the reduced-form outcome with respect to the treatment vector, i.e., $\partial E[\mathbf{y}|\mathbf{D}]/\partial D_k$. Thus, the direct effects represent the average partial effect of one's own treatment, while the indirect effects

capture the average partial effect of others' treatments. The linear-in-means structure implies a linear projection of $E[\mathbf{y}|\mathbf{D}]$. As a result, these effects are difficult to interpret causally unless the potential outcome is additively separable with respect to the treatment statuses of others. Furthermore, both direct and indirect effects are mixed effects of treatment and network, i.e., $\pi^{\cdot T}$ and $\pi^{\cdot N}$.

While Comola and Prina (2021) considers changes in the network driven by the intervention, it is challenging to decompose these effects because the changes in the network they consider reflect changes in links over time. This creates a mixture of time-varying effects and causal effects from the treatment.

Table 5: Decomposition of Treatment Effects

		DD				
	CP	G=19	G_{tot}=83	G_{tot}=107	G_{tot}=222	G_{tot}=298
π^{DT}	342.3	478.5***	572.3***	577.6***	462.4***	465.8***
		(174.6)	(113.5)	(116.7)	(128.4)	(161.3)
π^{DN}		-235.3***	-132.3**	-274.3***	-324.3***	-165.4
		(35.4)	(55.6)	(61.7)	(89.8)	(117.1)
π^{IT}	260.9	1.6	2.2	1.7	2.3	3.2
		(4.4)	(6)	(4.7)	(6.6)	(9.2)
π^{IN}		-3.8	-6	-17.9	-58.8	-61.4
		(31)	(49)	(52.8)	(73.7)	(94.4)

Notes: CP represents the estimation of direct and indirect effects in Comola and Prina (2021) (column (3) in Table 2). DD represents the estimation proposed in Section 3. The second column use 19 independent villages. The 3rd-6th columns are the same estimates by dividing each villages into smaller groups. G_{tot} is the total number of independent groups used in each estimation. The standard errors are computed based on plug-in asymptotic variance, and are reported in parentheses. *,**,*** denote the significance levels at 10%, 5%, and 1%, respectively.

The third column named $G = 19$ estimates the direct/indirect treatment/network effects $\boldsymbol{\pi}$. The direct treatment effect π^{DT} is estimated to be 478.5, and the direct network effect is estimated to be -243.2. Therefore, the total direct effect is 243.2, which is smaller than the estimates reported by Comola and Prina (2021). While the overall direct effect is positive, this result suggests that an opposing effect may exist due to changes in the network structure. Specifically, this can be interpreted as the

offer of a savings account tending to increase consumption and strengthen financial networks. However, increased financial links lead to greater savings and consequently reduce consumption. In contrast, both indirect effects are not significant.

There are only 19 villages, which is a small number to ensure the validity of asymptotic properties. To address this issue, I searched for independent subgroups within each village. This is feasible because the underlying network is sparse, with a density of only 2%. Specifically, I used the algorithm proposed by Yan and Sarkar (2021) to divide the network into smaller groups with a block diagonal structure. The number of clusters for each village is chosen as $\lfloor N_v/K \rfloor$ for $K = 10, 8, 4, 3$, resulting in total group numbers G_{tot} of 83, 107, 222, and 298, respectively. Although the results are sensitive to the choice of the number of clusters, the overall pattern remains consistent: direct treatment effects tend to be positive, direct network effects tend to be negative, and indirect effects are generally not significant.

Overall, the results presented in this section suggest that the method proposed in this paper can clearly decompose causal effects into pure treatment effects and those driven by causal changes in network links. In particular, this method is useful for identifying treatment effects that operate in opposite directions, as illustrated in Table 4. However, the data requirements are substantial, as full information on network links and a sufficient number of independent groups are needed.

6 Conclusion

This paper presents a method for identifying and estimating the causal effects of programs, accounting for potential causal network changes induced by treatment. The approach decomposes the treatment effect into two components: the impact when the network remains unchanged and the impact when only the network structure is altered. The effectiveness of the method is demonstrated through a Monte Carlo study and illustrated using data from a study in Nepal by Comola and Prina (2020). This novel approach not only offers a new way to estimate causal effects considering causal network changes, but also provides a decomposition that enhances our understanding of the mechanisms driving the program's impact.

While a linear model for the outcome response is commonly used in practice, future research could explore more flexible functional forms to mitigate the risk of model misspecification. For example, instead of assuming a linear relationship with exposures (own treatment, number of treated, and untreated neighbors), a series approximation could be employed. Additionally, although this study assumes the availability of full network information, future work could relax this requirement by observing only exposure values instead of full network under a different set of assumptions. For instance, if potential exposure distributions are identified instead of potential link distributions, it may still be possible to estimate causal effects and apply a similar decomposition. This approach could be particularly useful in cases where collecting full network data is costly.

References

Abadie, Alberto (2005). "Semiparametric difference-in-differences estimators". In: *The review of economic studies* 72.1, pp. 1–19.

Ananth, Abhishek (2020). *Optimal treatment assignment rules on networked populations*. Tech. rep. working paper.

Aronow, Peter M and Cyrus Samii (2017). "Estimating average causal effects under general interference, with application to a social network experiment". In: *The Annals of Applied Statistics* 11.4, pp. 1912–1947.

Auerbach, Eric and Max Tabord-Meehan (2021). "The local approach to causal inference under network interference". In: *arXiv preprint arXiv:2105.03810*.

Baird, Sarah, J Aislinn Bohren, Craig McIntosh, and Berk Özler (2018). "Optimal design of experiments in the presence of interference". In: *Review of Economics and Statistics* 100.5, pp. 844–860.

Banerjee, Abhijit, Emily Breza, Arun G Chandrasekhar, Esther Duflo, Matthew O Jackson, and Cynthia Kinnan (2024). "Changes in social network structure in response to exposure to formal credit markets". In: *Review of Economic Studies* 91.3, pp. 1331–1372.

Blackwell, Matthew (2017). "Instrumental variable methods for conditional effects and causal interaction in voter mobilization experiments". In: *Journal of the American Statistical Association* 112.518, pp. 590–599.

Bramoullé, Yann, Habiba Djebbari, and Bernard Fortin (2009). "Identification of peer effects through social networks". In: *Journal of econometrics* 150.1, pp. 41–55.

Butts, Kyle (2021). "Difference-in-differences estimation with spatial spillovers". In: *arXiv preprint arXiv:2105.03737*.

Comola, Margherita and Silvia Prina (2020). *Replication data for: Treatment Effect Accounting for Network Changes*. Version V1. DOI: 10.7910/DVN/K6QU2J. URL: https://doi.org/10.7910/DVN/K6QU2J.

— (2021). "Treatment effect accounting for network changes". In: *Review of Economics and Statistics* 103.3, pp. 597–604.

DiTraglia, Francis J, Camilo García-Jimeno, Rossa O'Keeffe-O'Donovan, and Alejandro Sánchez-Becerra (2023). "Identifying causal effects in experiments with spillovers and non-compliance". In: *Journal of Econometrics*.

Dupas, Pascaline, Anthony Keats, and Jonathan Robinson (2019). "The effect of savings accounts on interpersonal financial relationships: Evidence from a field experiment in rural Kenya". In: *The Economic Journal* 129.617, pp. 273–310.

Forastiere, Laura, Edoardo M Airoldi, and Fabrizia Mealli (2021). "Identification and estimation of treatment and interference effects in observational studies on networks". In: *Journal of the American Statistical Association* 116.534, pp. 901–918.

Fricke, Hans (2017). "Identification based on difference-in-differences approaches with multiple treatments". In: *Oxford Bulletin of Economics and Statistics* 79.3, pp. 426–433.

Frölich, Markus (2004). "Programme evaluation with multiple treatments". In: *Journal of Economic Surveys* 18.2, pp. 181–224.

Graham, Bryan S (2017). "An econometric model of network formation with degree heterogeneity". In: *Econometrica* 85.4, pp. 1033–1063.

Hoshino, Tadao and Takahide Yanagi (2023). "Treatment effect models with strategic interaction in treatment decisions". In: *Journal of Econometrics* 236.2, p. 105495.

Hudgens, Michael G and M Elizabeth Halloran (2008). "Toward causal inference with interference". In: *Journal of the American Statistical Association* 103.482, pp. 832–842.

Imai, Kosuke, Zhichao Jiang, and Anup Malani (2021). "Causal inference with interference and noncompliance in two-stage randomized experiments". In: *Journal of the American Statistical Association* 116.534, pp. 632–644.

Kang, Hyunseung and Guido Imbens (2016). "Peer encouragement designs in causal inference with partial interference and identification of local average network effects". In: *arXiv preprint arXiv:1609.04464*.

Kitagawa, Toru and Guanyi Wang (2023). "Who should get vaccinated? Individualized allocation of vaccines over SIR network". In: *Journal of Econometrics* 232.1, pp. 109–131.

Kormos, Mate, Robert P Lieli, and Martin Huber (2023). "Treatment Effect Analysis for Pairs with Endogenous Treatment Takeup". In: *arXiv preprint arXiv:2301.04876*.

Leung, Michael P (2020). "Treatment and spillover effects under network interference". In: *Review of Economics and Statistics* 102.2, pp. 368–380.

— (2022). "Causal inference under approximate neighborhood interference". In: *Econometrica* 90.1, pp. 267–293.

Manski, Charles F (2013). "Identification of treatment response with social interactions". In: *The Econometrics Journal* 16.1, S1–S23.

Newey, Whitney K and Daniel McFadden (1994). "Large sample estimation and hypothesis testing". In: *Handbook of econometrics* 4, pp. 2111–2245.

Prina, Silvia (2015). "Banking the poor via savings accounts: Evidence from a field experiment". In: *Journal of development economics* 115, pp. 16–31.

Rubin, Donald B (1974). "Estimating causal effects of treatments in randomized and nonrandomized studies." In: *Journal of educational Psychology* 66.5, p. 688.

Sánchez-Becerra, Alejandro (2021). "Spillovers, Homophily, and Selection into Treatment: The Network Propensity Score". In: *Job market paper*.

Vazquez-Bare, Gonzalo (2022a). "Causal spillover effects using instrumental variables". In: *Journal of the American Statistical Association*, pp. 1–12.

Vazquez-Bare, Gonzalo (2022b). "Identification and estimation of spillover effects in randomized experiments". In: *Journal of Econometrics*.

Viviano, Davide (2019). "Policy targeting under network interference". In: *arXiv preprint arXiv:1906.10258*.

Xu, Ruonan (2023). "Difference-in-Differences with Interference: A Finite Population Perspective". In: *arXiv preprint arXiv:2306.12003*.

Yan, Bowei and Purnamrita Sarkar (2021). "Covariate regularized community detection in sparse graphs". In: *Journal of the American Statistical Association* 116.534, pp. 734–745.

Appendix

A Proofs

A.1 Proofs for Propositions in Section 2.3

Proof of Proposition 1. Recall that $Q_i = \sum_{j \neq i} A_{ij} D_j$ and $R_i = \sum_{j \neq i} A_{ij}(1 - D_j)$. Then, by Assumption 1, we have

$$\begin{aligned} E[Q_i|\boldsymbol{D}] &= \sum_{j \neq i} E[A_{ij}|D_i, D_j] D_j = \sum_{j \neq i} (\boldsymbol{W}_{ij}' \boldsymbol{\zeta}) D_j =: Q_i(\boldsymbol{\zeta}), \\ E[R_i|\boldsymbol{D}] &= \sum_{j \neq i} E[A_{ij}|D_i, D_j](1 - D_j) = \sum_{j \neq i} (\boldsymbol{W}_{ij}' \boldsymbol{\zeta})(1 - D_j) =: R_i(\boldsymbol{\zeta}). \end{aligned} \tag{4}$$

By taking conditional expectation on the observed outcome equation (2), we have

$$\begin{aligned} E[Y_i|\boldsymbol{D}] &= \beta_1 + \beta_2 D_i + \beta_3 E[Q_i|\boldsymbol{D}] + \beta_4 E[R_i|\boldsymbol{D}] + E[\varepsilon_i|\boldsymbol{D}] \\ &= \beta_1 + \beta_2 D_i + \beta_3 Q_i(\boldsymbol{\zeta}) + \beta_4 R_i(\boldsymbol{\zeta}) + E[\varepsilon_i|\boldsymbol{D}] \\ &= Z_i(\boldsymbol{\zeta})' \boldsymbol{\beta}, \end{aligned}$$

because $E[\varepsilon_i|D_i] = E[\varepsilon_i(0)|D_i = 0] + D_i(E[\varepsilon_i(1)|D_i = 1] - E[\varepsilon_i(0)|D_i = 0]) = E[\varepsilon_i(0)] + D_i E[\varepsilon_i(1) - \varepsilon_i(0)] = 0$ by Assumption 2 and Assumption 5. Therefore, (i) and (ii) follow by the standard identification result of least squares estimator of projection coefficients. Next, by Assumption 5 again, we have

$$\begin{aligned} H(d, e) :&= E[A_{ij}(d, e) - A_{ij}(0, 0)|D_i = d, D_j = e] \\ &= E[A_{ij}|D_i = d, D_j = e] - E[A_{ij}(0, 0)|D_i = 0, D_j = 0] \\ &= E[A_{ij}|D_i = d, D_j = e] - E[A_{ij}|D_i = 0, D_j = 0] = m(d, e) - m(0, 0). \end{aligned}$$

Therefore, $\zeta_2 = m(1, 0) - m(0, 0) = H(1, 0)$ and $\zeta_3 = m(0, 1) - m(0, 0) = H(0, 1)$. As a result, the decomposition defined in Section 2 can be recovered by using $\boldsymbol{\zeta}$ and $\boldsymbol{\beta}$. □

Proof of Proposition 2. Recall that $Q_{i1} = \sum_{j \neq i} A_{ij1} D_j$, $R_{i1} = \sum_{j \neq i} A_{ij1}(1 - D_j)$, and $S_{i0} = \sum_{j \neq i} A_{ij0}$. Then, similar to (4), we have $E[Q_{i1}|\boldsymbol{D}] = Q_{i1}(\boldsymbol{\zeta})$, $E[R_{i1}|\boldsymbol{D}] = R_{i1}(\boldsymbol{\zeta})$, and $E[S_{i0}|\boldsymbol{D}] = S_{i0}(\boldsymbol{\zeta})$. Next, from Assumption 2, Assumption 6, and Assumption 7 the first-differenced observed outcome is given by $\Delta Y_i = (\beta_1 - \alpha_1) + \beta_2 D_i + \beta_3 Q_{i1} + \beta_4(R_{i1} - S_{i0}) + \Delta\varepsilon_i$. Taking conditional expectation on ΔY_i, we have

$$\begin{aligned} E[\Delta Y_i|\boldsymbol{D}] &= (\beta_1 - \alpha_1) + \beta_2 D_i + \beta_3 E[Q_{i1}|\boldsymbol{D}] + \beta_4 E[R_{i1} - S_{i0}|\boldsymbol{D}] + E[\Delta\varepsilon_i|\boldsymbol{D}] \\ &= (\beta_1 - \alpha_1) + \beta_2 D_i + \beta_3 Q_{i1}(\boldsymbol{\zeta}) + \beta_4(R_{i1}(\boldsymbol{\zeta}) - S_{i0}(\boldsymbol{\zeta})) + E[\Delta\varepsilon_i|D_i] \\ &= X_i(\boldsymbol{\zeta})'\boldsymbol{\beta}. \end{aligned}$$

The last term is zero because

$$\begin{aligned} E[\Delta\varepsilon_i|D_i] &= E[\Delta\varepsilon_i(0)|D_i = 0] + D_i\left[E[\Delta\varepsilon_i|D_i = 1] - E[\Delta\varepsilon_i|D_i = 0]\right] \\ &= E[\Delta\varepsilon_i(0)] + D_i E[\varepsilon_{i1}(1) - \varepsilon_{i0}(0)|D_i = 1] \end{aligned}$$

The second equation is because $E[\varepsilon_{i1}(0)|D_i = 1] = E[\varepsilon_{i0}(0)|D_i = 1] + E[\Delta\varepsilon_i(0)|D_i = 0]$ by Assumption 7 and $E[\varepsilon_{i0}(0)|D_i = 1] = E[\varepsilon_{i0}|D_i = 1]$ by Assumption 6, and therefore $E[\varepsilon_{i1}(1) - \varepsilon_{i1}(0)|D_i = 1] = E[\varepsilon_{i1}|D_i = 1] - E[\varepsilon_{i0}|D_i = 1] - E[\Delta\varepsilon_i|D_i = 0]$. Each term in the second equation is zero by Assumption 2. Therefore, (i) and (iii) follow by the standard identification result of least squares estimator of projection coefficients. Next, note that

$$\begin{aligned} &H(d,e) \\ &:= E[A_{ij1}(d,e) - A_{ij1}(0,0)|D_i = d, D_j = e] \\ &= E[A_{ij1}|D_i = d, D_j = e] - E[A_{ij0}(0,0)|D_i = d, D_j = e] - E[\Delta A_{ij}(0,0)|D_i = d, D_j = e] \\ &= E[A_{ij1}|D_i = d, D_j = e] - E[A_{ij0}(d,e)|D_i = d, D_j = e] - E[\Delta A_{ij}(0,0)|D_i = 0, D_j = 0] \\ &= E[\Delta A_{ij}|D_i = d, D_j = e] - E[\Delta A_{ij}|D_i = 0, D_j = 0] = \Delta m(d,e) - \Delta m(0,0), \end{aligned}$$

where $m_t(d,e) = E[A_{ijt}(d,e)|D_i = d, D_j = e]$. The third equation is by Assumption 6-(ii) and Assumption 7-(ii). Therefore, because $\zeta_{2t} = m_t(1,0) - m_t(0,0)$, we have $\xi_2 = H(1,0)$ and $\xi_3 = H(0,1)$. As a result, the decomposition defined in Section 2 can be

recovered by using $\boldsymbol{\zeta} = (\boldsymbol{\zeta}_1, \boldsymbol{\zeta}_2)$ and $\boldsymbol{\beta}$. □

A.2 Proofs for Propositions in Section 3

The following lemma is Lemma 4.3 in Newey and McFadden (1994), and will be used to prove Proposition 3, and Proposition 4.

Lemma 1. *Let $\boldsymbol{V}_g$ be a random vector whose support is $\mathcal{V}$ and $\ell : \mathcal{V} \times \Phi \to \mathbb{R}^M$ be a vector of real valued functions that is integrable with respect to the distribution of $\boldsymbol{V}_g$ at each point $\phi \in \Phi \subset \mathbb{R}^K$. Define followings:*

$$L_G(\phi) = \frac{1}{G}\sum_{g=1}^{G} \ell(\boldsymbol{V}_g, \phi), \quad L(\phi) = E[\ell(\boldsymbol{V}_g, \phi)].$$

Suppose (a) $\{\boldsymbol{V}_g\}$ is independently and identically distributed; (b) $\hat{\phi} \xrightarrow{p} \phi_0$, ϕ_0; (c) $\ell(\boldsymbol{v}, \phi)$ is continuous at ϕ_0 for all $\boldsymbol{v} \in \mathcal{V}$; (d) For some neighborhood $\mathcal{N}$ of ϕ_0, we have $E\left[\sup_{\phi\in\mathcal{N}} \|\ell(\boldsymbol{V}_g, \phi)\|\right] < \infty$. Then, $L(\phi)$ is continuous at ϕ_0 and $L_G(\hat{\phi}) \xrightarrow{p} L(\phi_0)$.

Proof. Consider a sequence $\{\phi_n\} \to \phi_0$. For the neighborhood $\mathcal{N}$ of ϕ_0 satisfying (d), we have $\|\ell(v, \phi_n)\| \leq \sup_{\phi\in\mathcal{N}} \|\ell(v, \phi)\| =: g(v)$, for all but finite number of n, where $g(v)$ is integrable by (d). Thus, by dominated convergence thoerem, we have $\{E[\ell(\boldsymbol{V}_g, \phi_n)]\} \to E[\ell(\boldsymbol{V}_g, \phi_0)]$, which implies continuity of $L(\phi)$ at ϕ_0. See proof of Lemma 4.3 in Newey and McFadden (1994) for $L_G(\hat{\phi}) \xrightarrow{p} L(\phi_0)$. □

Proof of Proposition 3.

A. Consistency, and influence function of $\hat{\boldsymbol{\zeta}}$

Note that $\|\boldsymbol{W}_{ijg}\|^2 = 1 + D_{ig} + D_{jg} + D_{ig}D_{jg} \leq 4$, and $|\boldsymbol{W}'_{ijg}\boldsymbol{\zeta}| = |\zeta_1 + D_{ig}\zeta_2 + D_{jg}\zeta_3 +$

$D_{ig}D_{jg}\zeta_4| \leq |\zeta_1+\zeta_2+\zeta_3+\zeta_4| = m(1,1) \in [0,1]$, because

$$\boldsymbol{\zeta} = \begin{pmatrix} m(0,0) \\ m(1,0) - m(0,0) \\ m(0,1) - m(0,0) \\ m(1,1) - m(1,0) - m(0,1) + m(0,0) \end{pmatrix},$$

where $m(d,e) = E[A_{ij}|D_i = d, D_j = e]$. Let $\omega_{ijg} = A_{ijg} - \boldsymbol{W}'_{ijg}\boldsymbol{\zeta}^\star$. Then, $E[\omega_{ijg}|\boldsymbol{D}_g] = E[\omega_{ijg}|\boldsymbol{W}_{ijg}] = 0$, and $E[\omega_{ijg}^2] = E\left[(A_{ijg} - \boldsymbol{W}'_{ijg}\boldsymbol{\zeta})^2\right] \leq E\left[(|A_{ijg}| + |\boldsymbol{W}'_{ijg}\boldsymbol{\zeta}|)^2\right] \leq 4$. Therefore, by standard asymptotic theory for least squares estimator, we have

$$\begin{aligned}
\frac{1}{G}\sum_{g=1}^{G}\left[\frac{1}{N(N-1)}\sum_{(i,j):i\neq j}\boldsymbol{W}_{ijg}\boldsymbol{W}'_{ijg}\right] &\xrightarrow{p} E\left[\boldsymbol{W}_{ijg}\boldsymbol{W}'_{ijg}\right] =: \boldsymbol{R}_{\boldsymbol{\zeta}}, \\
\frac{1}{G}\sum_{g=1}^{G}\left[\frac{1}{N(N-1)}\sum_{(i,j):i\neq j}\boldsymbol{W}_{ijg}\omega_{ijg}\right] &\xrightarrow{p} E[\boldsymbol{W}_{ijg}\omega_{ijg}] = 0, \\
\frac{1}{\sqrt{G}}\sum_{g=1}^{G}\left[\frac{1}{N(N-1)}\sum_{(i,j):i\neq j}\boldsymbol{W}_{ijg}\omega_{ijg}\right] &\xrightarrow{d} N\left(0, \boldsymbol{\Sigma}_{\boldsymbol{\zeta}}\right),
\end{aligned}$$

as $G \to \infty$, where $\boldsymbol{\Sigma}_{\boldsymbol{\zeta}} := \mathrm{Var}\left(\frac{1}{N(N-1)}\sum_{(i,j):i\neq j}\boldsymbol{W}_{ijg}\omega_{ijg}\right)$. The second probability limit is from the moment condition $E[A_{ijg} - \boldsymbol{W}'_{ijg}\boldsymbol{\zeta}^\star|\boldsymbol{D}_g] = 0$. And the variance $\boldsymbol{\Sigma}_{\boldsymbol{\zeta}}$ of last limiting distribution exists because

$$\begin{aligned}
E\left[\left\|\boldsymbol{W}_{ijg}(A_{ijg} - \boldsymbol{W}'_{ijg}\boldsymbol{\zeta}^\star)\right\|^2\right] &= E\left[\|\boldsymbol{W}_{ijg}\|^4\right]^{\frac{1}{2}} E\left[(A_{ijg} - \boldsymbol{W}'_{ijg}\boldsymbol{\zeta}^\star)^4\right]^{\frac{1}{2}} \\
&\leq E\left[\|\boldsymbol{W}_{ijg}\|^4\right]^{\frac{1}{2}} \left(8E\left[|A_{ijg}|^4\right] + 8E\left[\|\boldsymbol{W}_{ijg}\|^4\right]\|\boldsymbol{\zeta}^\star\|^4\right)^{\frac{1}{2}} \\
&\leq 4\left(8 + 128\,\|\boldsymbol{\zeta}^\star\|^4\right)^{\frac{1}{2}} < \infty.
\end{aligned}$$

by Cauchy-Schwarz inequality and pytagorian rule. Thus, $\hat{\boldsymbol{\zeta}} \xrightarrow{p} \boldsymbol{\zeta}^\star$, and

$$\sqrt{G}(\hat{\boldsymbol{\zeta}} - \boldsymbol{\zeta}^\star) = \frac{1}{\sqrt{G}}\sum_{g=1}^{G}\psi_{\boldsymbol{\zeta}}(\mathcal{W}_g, \boldsymbol{\zeta}^\star) + o_p(1) \xrightarrow{d} N(0, \boldsymbol{R}_{\boldsymbol{\zeta}}^{-1}\boldsymbol{\Sigma}_{\boldsymbol{\zeta}}\boldsymbol{R}_{\boldsymbol{\zeta}}^{-1}), \tag{5}$$

where $\mathcal{W}_g := \{(A_{ijg}, \boldsymbol{W}_{ijg})\}_{(i,j)}$, and $\psi_{\boldsymbol{\zeta}}(\mathcal{W}_g, \boldsymbol{\zeta}) = \boldsymbol{R}_{\boldsymbol{\zeta}}^{-1}\frac{1}{N(N-1)}\sum_{(i,j):i\neq j}\boldsymbol{W}_{ijg}(A_{ijg} - \boldsymbol{W}'_{ijg}\boldsymbol{\zeta})$.

B. Consistency, and influence function of $\hat{\beta}$

Proposition 1 implies a moment condition $E[h(\mathcal{V}_g, \boldsymbol{\zeta}^\star, \boldsymbol{\beta}^\star)] = 0$, where $\mathcal{V}_g = (D_{ig}, Y_{ig})_i$ and $h(\mathcal{V}_g, \boldsymbol{\zeta}, \boldsymbol{\beta}) = \boldsymbol{Z}_{ig}(\boldsymbol{\zeta})(Y_{ig} - \boldsymbol{Z}_{ig}(\boldsymbol{\zeta})'\boldsymbol{\beta}) \in \mathbb{R}^4$. Suppose $(\bar{\boldsymbol{\zeta}}, \bar{\boldsymbol{\beta}}) \xrightarrow{p} (\boldsymbol{\zeta}^\star, \boldsymbol{\zeta}_1)$. Recall that $|\boldsymbol{W}'_{ijg}\boldsymbol{\zeta}| \leq 1$. Thus, $\max\{Q_{ig}(\boldsymbol{\zeta}), R_{ig}(\boldsymbol{\zeta})\} \leq \sum_{j\neq i} \hat{A}_{ijg}(\boldsymbol{\zeta}) \leq N_g(N_g - 1)$. Therefore,

$$\|\boldsymbol{Z}_{ig}(\boldsymbol{\zeta})\|^2 = 1 + D_{ig} + (Q_{ig}(\boldsymbol{\zeta}))^2 + (R_{ig}(\boldsymbol{\zeta}))^2 \leq 2 + 2(N_g(N_g - 1))^2 =: B_1 < \infty.$$

Also, $\nabla_{\boldsymbol{\zeta}} \boldsymbol{Z}_{ig}(\boldsymbol{\zeta}) = (\boldsymbol{0}, \boldsymbol{0}, \sum_{j\neq i} \boldsymbol{W}_{ijg} D_{jg}, \sum_{j\neq i} \boldsymbol{W}_{ijg}(1 - D_{jg}))'$. Thus,

$$\begin{aligned}
\|\nabla_{\boldsymbol{\zeta}} \boldsymbol{Z}_{ig}(\boldsymbol{\zeta})\|^2 &= \operatorname{tr}\left(\sum_{j\neq i} \boldsymbol{W}_{ijg} D_{jg} \sum_{j\neq i} \boldsymbol{W}'_{ijg} D_{jg} + \sum_{j\neq i} \boldsymbol{W}_{ijg}(1 - D_{jg}) \sum_{j\neq i} \boldsymbol{W}'_{ijg}(1 - D_{jg})\right) \\
&= \sum_{j\neq i} \boldsymbol{W}'_{ijg} D_{jg} \sum_{j\neq i} \boldsymbol{W}_{ijg} D_{jg} + \sum_{j\neq i} \boldsymbol{W}'_{ijg}(1 - D_{jg}) \sum_{j\neq i} \boldsymbol{W}_{ijg}(1 - D_{jg}) \\
&\leq 2\left(\sum_{j\neq i} D_{jg}\right)^2 + 2\left(\sum_{j\neq i} D_{ig} D_{jg}\right)^2 + \left\{\sum_{j\neq i}(1 - D_{jg})\right\}^2 + \left\{\sum_{j\neq i} D_{ig}(1 - D_{jg})\right\}^2 \\
&\leq 6(N_g - 1)^2 = B_2 < \infty. \qquad (6)
\end{aligned}$$

The above boundedness imply boundedness of derivative of the moment functions:

$$\begin{aligned}
\|\nabla_{\boldsymbol{\beta}} h(\mathcal{V}_g, \boldsymbol{\zeta}, \boldsymbol{\beta})\| &= \|\boldsymbol{Z}_{ig}(\boldsymbol{\zeta})\boldsymbol{Z}_{ig}(\boldsymbol{\zeta})'\| = \|\boldsymbol{Z}_{ig}(\boldsymbol{\zeta})\|^2 \leq B_1, \\
\|\nabla_{\boldsymbol{\zeta}} h(\mathcal{V}_g, \boldsymbol{\zeta}, \boldsymbol{\beta})\| &= \|h_1\| + \|h_2\| \leq B_3,
\end{aligned}$$

where

$$
\begin{aligned}
\|h_1\| &:= \|\nabla_{\boldsymbol{\zeta}} \boldsymbol{Z}_{ig}(\boldsymbol{\zeta})(Y_{ig} - \boldsymbol{Z}_{ig}(\boldsymbol{\zeta})'\boldsymbol{\beta})\| \\
&= \|\nabla_{\boldsymbol{\zeta}} \boldsymbol{Z}_{ig}(\boldsymbol{\zeta})\| \, |(Y_{ig} - \boldsymbol{Z}_{ig}(\boldsymbol{\zeta})'\boldsymbol{\beta})| \\
&\leq \|\nabla_{\boldsymbol{\zeta}} \boldsymbol{Z}_{ig}(\boldsymbol{\zeta})\| \, (|Y_{ig}| + \|\boldsymbol{Z}_{ig}(\boldsymbol{\zeta})\| \, \|\boldsymbol{\beta}\|) \leq B_2 \left(|Y_{ig}| + \sqrt{B_1} \, \|\boldsymbol{\beta}\| \right), \\
\|h_2\| &:= \|\boldsymbol{Z}_{ig}(\boldsymbol{\zeta}) \nabla_{\boldsymbol{\zeta}} (\boldsymbol{Z}_{ig}(\boldsymbol{\zeta})'\boldsymbol{\beta})\| \\
&\leq \|\boldsymbol{Z}_{ig}(\boldsymbol{\zeta})\boldsymbol{\beta}'\| \, \|\nabla_{\boldsymbol{\zeta}} (\boldsymbol{Z}_{ig}(\boldsymbol{\zeta}))\| \\
&= \|\boldsymbol{Z}_{ig}(\boldsymbol{\zeta})\| \, \|\boldsymbol{\beta}\| \, \|\nabla_{\boldsymbol{\zeta}} (\boldsymbol{Z}_{ig}(\boldsymbol{\zeta}))\| \leq (B_1 B_2)^{\frac{1}{2}} \, \|\beta\|,
\end{aligned}
$$

by Cauchy-Schwarz inequality for Frobenius inner product, and the definition of Frobenius norm and L_2 norm. Also, since $\nabla_{\boldsymbol{\beta}} h(\mathcal{V}_g, \boldsymbol{\zeta}, \boldsymbol{\beta})$ and $\nabla_{\boldsymbol{\zeta}} h(\mathcal{V}_g, \boldsymbol{\zeta}, \boldsymbol{\beta})$ are continuous on $\boldsymbol{\zeta}, \boldsymbol{\beta}$, we can apply Lemma 1 to conclude

$$
\begin{aligned}
\frac{1}{G} \sum_{g=1}^{G} h(\mathcal{V}_g, \bar{\boldsymbol{\zeta}}, \bar{\boldsymbol{\beta}}) &\xrightarrow{p} E[h(\mathcal{V}_g, \boldsymbol{\zeta}^\star, \boldsymbol{\beta}^\star)] = 0 \\
\frac{1}{G} \sum_{g=1}^{G} \nabla_{\boldsymbol{\beta}} h(\mathcal{V}_g, \bar{\boldsymbol{\zeta}}, \bar{\boldsymbol{\beta}}) &\xrightarrow{p} E\left[\nabla_{\boldsymbol{\beta}} h(\mathcal{V}_g, \boldsymbol{\zeta}^\star, \boldsymbol{\beta}^\star)\right] = -E\left[\boldsymbol{Z}_{ig}(\boldsymbol{\zeta}^\star)\boldsymbol{Z}_{ig}(\boldsymbol{\zeta}^\star)'\right] =: -\boldsymbol{R}_{\boldsymbol{Z}} \\
\frac{1}{G} \sum_{g=1}^{G} \nabla_{\boldsymbol{\zeta}} h(\mathcal{V}_g, \bar{\boldsymbol{\zeta}}, \bar{\boldsymbol{\beta}}) &\xrightarrow{p} E\left[\nabla_{\boldsymbol{\zeta}} h(\mathcal{V}_g, \boldsymbol{\zeta}^\star, \boldsymbol{\beta}^\star)\right] = -E[\boldsymbol{Z}_{ig}(\boldsymbol{\zeta}^\star) \nabla_{\boldsymbol{\zeta}} (\boldsymbol{Z}_{ig}(\boldsymbol{\zeta}^\star)'\boldsymbol{\beta}^\star)] =: -\boldsymbol{Q}_{\boldsymbol{\zeta}},
\end{aligned}
\tag{7}
$$

because

$$
E\left[\nabla_{\boldsymbol{\zeta}} h(\mathcal{V}_g, \boldsymbol{\zeta}^\star, \boldsymbol{\beta}^\star)\right] = E\left[\nabla_{\boldsymbol{\zeta}} \boldsymbol{Z}_{ig}(\boldsymbol{\zeta}^\star)(Y_{ig} - \boldsymbol{Z}_{ig}(\boldsymbol{\zeta}^\star)'\boldsymbol{\beta}^\star)\right] - E\left[\boldsymbol{Z}_{ig}(\boldsymbol{\zeta}^\star) \nabla_{\boldsymbol{\zeta}} (\boldsymbol{Z}_{ig}(\boldsymbol{\zeta}^\star)'\boldsymbol{\beta}^\star)\right],
$$

and the second term is zero by the moment condition $E\left[Y_{ig} - \boldsymbol{Z}_{ig}(\boldsymbol{\zeta}^\star)'\boldsymbol{\beta}^\star | \boldsymbol{D}_g\right] = 0$.

In the second-stage estimation, $\hat{\boldsymbol{\beta}}$ solves the following first-order condition:

$$
0 = \frac{1}{G} \sum_{g=1}^{G} h(\mathcal{V}_g, \hat{\boldsymbol{\zeta}}, \hat{\boldsymbol{\beta}}),
\tag{8}
$$

First two convergences in (11) imply

$$\begin{aligned}\hat{\boldsymbol{\beta}} &= \boldsymbol{\beta}^\star + \left[\frac{1}{G}\sum_{g=1}^{G} \boldsymbol{Z}_{ig}(\hat{\boldsymbol{\zeta}})\boldsymbol{Z}_{ig}(\hat{\boldsymbol{\zeta}})'\right]^{-1}\left[\frac{1}{G}\sum_{g=1}^{G} \boldsymbol{Z}_{ig}(\hat{\boldsymbol{\zeta}})(Y_{ig} - \boldsymbol{Z}_{ig}(\hat{\boldsymbol{\zeta}})'\boldsymbol{\beta}^\star)\right] \\ &= \boldsymbol{\beta}^\star + \left[\frac{1}{G}\sum_{g=1}^{G} \nabla_{\boldsymbol{\beta}} h(\mathcal{V}_g, \hat{\boldsymbol{\zeta}}, \boldsymbol{\beta}^\star)\right]^{-1}\frac{1}{G}\sum_{g=1}^{G} h(\mathcal{V}_g, \hat{\boldsymbol{\zeta}}, \boldsymbol{\beta}^\star) \xrightarrow{p} 0.\end{aligned}$$

Next, by applying mean value theorem on (8) and by (5) and (11), we have

$$\begin{aligned}0 &= \frac{1}{\sqrt{G}}\sum_{g=1}^{G} h(\mathcal{V}_g, \boldsymbol{\zeta}^\star, \boldsymbol{\beta}^\star) + \frac{1}{G}\sum_{g=1}^{G} \nabla_{\boldsymbol{\beta}} h(\mathcal{V}_g, \bar{\boldsymbol{\zeta}}, \bar{\boldsymbol{\beta}})\sqrt{G}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^\star) \\ &\quad + \frac{1}{G}\sum_{g=1}^{G} \nabla_{\boldsymbol{\zeta}} h(\mathcal{V}_g, \bar{\boldsymbol{\zeta}}, \bar{\boldsymbol{\beta}})\sqrt{G}(\hat{\boldsymbol{\zeta}} - \boldsymbol{\zeta}^\star) \\ &= \frac{1}{\sqrt{G}}\sum_{g=1}^{G} h(\mathcal{V}_g, \boldsymbol{\zeta}^\star, \boldsymbol{\beta}^\star) - \boldsymbol{R}_{\boldsymbol{\beta}}\sqrt{G}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^\star) - \boldsymbol{Q}_{\boldsymbol{\zeta}}\frac{1}{\sqrt{G}}\sum_{g=1}^{G} \psi_{\boldsymbol{\zeta}}(\mathcal{W}_g, \boldsymbol{\zeta}^\star) + o_p(1), \qquad (9)\end{aligned}$$

Also, by rearranging (9), we have

$$\sqrt{G}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^\star) = \frac{1}{\sqrt{G}}\sum_{g=1}^{G} \psi_{\boldsymbol{\beta}}(\mathcal{V}_g, \mathcal{W}_g, \boldsymbol{\zeta}^\star, \boldsymbol{\beta}^\star) + o_p(1),$$

where $\psi_{\boldsymbol{\beta}}(\mathcal{V}_g, \mathcal{W}_g, \boldsymbol{\zeta}, \boldsymbol{\beta}) := \boldsymbol{R}_{\boldsymbol{\beta}}^{-1}\left[h(\mathcal{V}_g, \boldsymbol{\zeta}, \boldsymbol{\beta}) - \boldsymbol{Q}_{\boldsymbol{\zeta}}\psi_{\boldsymbol{\zeta}}(\mathcal{W}_g, \boldsymbol{\zeta})\right]$.

By Cauchy Schwarz inequality and Pytagorian rule,

$$E\left[\|\psi_{\boldsymbol{\beta}}(\mathcal{V}_g, \mathcal{W}_g, \boldsymbol{\zeta}^\star, \boldsymbol{\beta}^\star)\|^2\right] \le \left\|\boldsymbol{R}_{\boldsymbol{\beta}}^{-1}\right\|^2 E\left[\|h(\mathcal{V}_g, \boldsymbol{\zeta}^\star, \boldsymbol{\beta}^\star)\|^2\right] + \left\|\boldsymbol{Q}_{\boldsymbol{\zeta}}\right\|^2 E\left[\|\psi_{\boldsymbol{\zeta}}(\mathcal{W}_g, \boldsymbol{\zeta}^\star)\|^2\right]$$

we shown $E\left[\|\psi_{\boldsymbol{\zeta}}(\mathcal{W}_g, \boldsymbol{\zeta}^\star)\|^2\right] < \infty$ in Part A.

$$\begin{aligned}E\left[\|h(\mathcal{V}_g, \boldsymbol{\zeta}^\star, \boldsymbol{\beta}^\star)\|^2\right] &= E\left[\|\boldsymbol{Z}_{ig}(\boldsymbol{\zeta}^\star)(Y_{ig} - \boldsymbol{Z}_{ig}(\boldsymbol{\zeta}^\star)'\boldsymbol{\beta}^\star)\|^2\right] \\ &= E\left[\|\boldsymbol{Z}_{ig}(\boldsymbol{\zeta}^\star)\|^2 (Y_{ig} - \boldsymbol{Z}_{ig}(\boldsymbol{\zeta}^\star)'\boldsymbol{\beta}^\star)^2\right] \\ &\le E\left[\|\boldsymbol{Z}_{ig}(\boldsymbol{\zeta}^\star)\|^4\right]^{\frac{1}{2}}\left(8E[Y_{ig}^4] + 8E[\|Z_{ig}(\boldsymbol{\zeta}^\star)\|^4]\,\|\boldsymbol{\beta}^\star\|^4\right)^{\frac{1}{2}} < \infty,\end{aligned}$$

provided that $E[Y_i^4] < \infty$.

C. Consistency, and influence function of $\hat{\boldsymbol{\pi}}$

By applying Delta-method, we have

$$\begin{aligned}\sqrt{G}(\hat{\boldsymbol{\pi}} - \boldsymbol{\pi}^\star) &= \frac{1}{\sqrt{G}} \sum_{g=1}^{G} \psi_{\boldsymbol{\pi}}(\mathcal{W}_g, \mathcal{V}_g, \boldsymbol{\zeta}^\star, \boldsymbol{\beta}^\star) + o_p(1) \\ &\xrightarrow{d} N\left(0, E\left[\psi_{\boldsymbol{\pi}}(\mathcal{W}_g, \mathcal{V}_g, \boldsymbol{\zeta}^\star, \boldsymbol{\beta}^\star)\psi_{\boldsymbol{\pi}}(\mathcal{W}_g, \mathcal{V}_g, \boldsymbol{\zeta}^\star, \boldsymbol{\beta}^\star)'\right]\right),\end{aligned}$$

where $\psi_{\boldsymbol{\pi}}(\mathcal{V}_g, \mathcal{W}_g, \boldsymbol{\zeta}, \boldsymbol{\beta})$ is defined by

$$\begin{pmatrix} [\psi_{\boldsymbol{\beta}}(\mathcal{W}_g, \mathcal{V}_g, \boldsymbol{\zeta}, \boldsymbol{\beta})]_2 \\ (N-1)[\psi_{\boldsymbol{\beta}}(\mathcal{W}_g, \mathcal{V}_g, \boldsymbol{\zeta}, \boldsymbol{\beta})]_4[\boldsymbol{\zeta}]_2 + [\boldsymbol{\beta}]_4[\psi_{\boldsymbol{\zeta}}(\mathcal{W}_g, \boldsymbol{\zeta})]_2 \\ ([\psi_{\boldsymbol{\beta}}(\mathcal{W}_g, \mathcal{V}_g, \boldsymbol{\zeta}, \boldsymbol{\beta})]_3 - [\psi_{\boldsymbol{\beta}}(\mathcal{W}_g, \mathcal{V}_g, \boldsymbol{\zeta}, \boldsymbol{\beta})]_4)\,[\boldsymbol{\zeta}]_1 + ([\boldsymbol{\beta}]_3 - [\boldsymbol{\beta}]_4)[\psi_{\boldsymbol{\zeta}}(\mathcal{W}_g, \boldsymbol{\zeta})]_1 \\ [\psi_{\boldsymbol{\beta}}(\mathcal{W}_g, \mathcal{V}_g, \boldsymbol{\zeta}, \boldsymbol{\beta})]_3[\boldsymbol{\zeta}]_3 + [\boldsymbol{\beta}]_3[\psi_{\boldsymbol{\zeta}}(\mathcal{W}_g, \boldsymbol{\zeta})]_3 \end{pmatrix},$$

where $[v]_k$ denote k-th element in vector v. Recall that $E[\|\psi_{\boldsymbol{\zeta}}(\mathcal{W}_g, \boldsymbol{\zeta}^\star)\|^2]$, and $E[\|\psi_{\boldsymbol{\beta}}(\mathcal{V}_g, \mathcal{W}_g, \boldsymbol{\zeta}^\star, \boldsymbol{\beta}^\star)\|^2]$ are bounded, and therefore $E[\|\psi_{\boldsymbol{\zeta}}(\mathcal{W}_g, \boldsymbol{\zeta}^\star)\psi_{\boldsymbol{\beta}}(\mathcal{V}_g, \mathcal{W}_g, \boldsymbol{\zeta}^\star, \boldsymbol{\beta}^\star)'\|]$ is also bounded. It follows that $E\left[\|\psi_{\boldsymbol{\pi}}(\mathcal{W}_g, \mathcal{V}_g, \boldsymbol{\zeta}^\star, \boldsymbol{\beta}^\star)\|^2\right] < \infty$.

D. Consistency of plug-in standard errors

As shown in Part A, and Part B, we have

$$E\left[\sup_{\boldsymbol{\zeta}} \|\psi_{\boldsymbol{\zeta}}(\mathcal{W}_g, \boldsymbol{\zeta})\|^2\right] < \infty, \text{ and } E\left[\sup_{(\boldsymbol{\zeta},\boldsymbol{\beta})} \|\psi_{\boldsymbol{\beta}}(\mathcal{V}_g, \mathcal{W}_g, \boldsymbol{\zeta}, \boldsymbol{\beta})\|^2\right] < \infty.$$

Therefore, by Cauchy-Schwarz inequality,

$$\begin{aligned}&E\left[\sup_{(\boldsymbol{\zeta},\boldsymbol{\beta})} \|\psi_{\boldsymbol{\zeta}}(\mathcal{W}_g, \boldsymbol{\zeta})\psi_{\boldsymbol{\beta}}(\mathcal{V}_g, \mathcal{W}_g, \boldsymbol{\zeta}, \boldsymbol{\beta})'\|\right] \\ &\leq E\left[\sup_{\boldsymbol{\zeta}} \|\psi_{\boldsymbol{\zeta}}(\mathcal{W}_g, \boldsymbol{\zeta})\|^2\right]^{\frac{1}{2}} E\left[\sup_{(\boldsymbol{\zeta},\boldsymbol{\beta})} \|\psi_{\boldsymbol{\beta}}(\mathcal{V}_g, \mathcal{W}_g, \boldsymbol{\zeta}, \boldsymbol{\beta})\|^2\right]^{\frac{1}{2}} < \infty.\end{aligned}$$

This implies $E[\sup_{(\boldsymbol{\zeta},\boldsymbol{\beta})} \|\psi_{\boldsymbol{\pi}}(\mathcal{W}_g, \mathcal{V}_g, \boldsymbol{\zeta}, \boldsymbol{\beta})\|^2] < \infty$, since $\psi_{\boldsymbol{\pi}}(\mathcal{W}_g, \mathcal{V}_g, \boldsymbol{\zeta}, \boldsymbol{\beta})\psi_{\boldsymbol{\pi}}(\mathcal{W}_g, \mathcal{V}_g, \boldsymbol{\zeta}, \boldsymbol{\beta})'$

is computed by elements in $\psi_{\boldsymbol{\zeta}}(\mathcal{W}_g, \boldsymbol{\zeta})$. Therefore, with consistency of $(\hat{\boldsymbol{\zeta}}, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\pi}})$, and applying Lemma 1, we have

$$\frac{1}{G}\sum_{g=1}^{G} \psi_{\boldsymbol{\zeta}}(\mathcal{W}_g, \hat{\boldsymbol{\zeta}})\psi_{\boldsymbol{\zeta}}(\mathcal{W}_g, \hat{\boldsymbol{\zeta}})' \xrightarrow{p} E[\psi_{\boldsymbol{\zeta}}(\mathcal{W}_g, \boldsymbol{\zeta}^\star)\psi_{\boldsymbol{\zeta}}(\mathcal{W}_g, \boldsymbol{\zeta}^\star)']$$
$$\frac{1}{G}\sum_{g=1}^{G} \psi_{\boldsymbol{\beta}}(\mathcal{V}_g, \mathcal{W}_g, \hat{\boldsymbol{\zeta}}, \hat{\boldsymbol{\beta}})\psi_{\boldsymbol{\beta}}(\mathcal{V}_g, \mathcal{W}_g, \hat{\boldsymbol{\zeta}}, \hat{\boldsymbol{\beta}})' \xrightarrow{p} E[\psi_{\boldsymbol{\beta}}(\mathcal{V}_g, \mathcal{W}_g, \boldsymbol{\zeta}^\star, \boldsymbol{\beta}^\star)\psi_{\boldsymbol{\beta}}(\mathcal{V}_g, \mathcal{W}_g, \boldsymbol{\zeta}^\star, \boldsymbol{\beta}^\star)']$$
$$\frac{1}{G}\sum_{g=1}^{G} \psi_{\boldsymbol{\pi}}(\mathcal{W}_g, \mathcal{V}_g, \hat{\boldsymbol{\zeta}}, \hat{\boldsymbol{\beta}})\psi_{\boldsymbol{\pi}}(\mathcal{W}_g, \mathcal{V}_g, \hat{\boldsymbol{\zeta}}, \hat{\boldsymbol{\beta}})' \xrightarrow{p} E[\psi_{\boldsymbol{\pi}}(\mathcal{W}_g, \mathcal{V}_g, \boldsymbol{\zeta}^\star, \boldsymbol{\beta}^\star)\psi_{\boldsymbol{\pi}}(\mathcal{W}_g, \mathcal{V}_g, \boldsymbol{\zeta}^\star, \boldsymbol{\beta}^\star)'].$$

As a result, the continuous mapping theorem states the desired results. □

Proof or Proposition 4. By applying the same argument of Part A in Proof of Proposition 3 to $\boldsymbol{\zeta}_1, \boldsymbol{\zeta}_2$, and by the fact that $\boldsymbol{\xi} = \boldsymbol{\zeta}_1 - \boldsymbol{\zeta}_0$, we have $\hat{\boldsymbol{\xi}} \xrightarrow{p} \boldsymbol{\xi}^\star$, and

$$\begin{aligned}\sqrt{G}(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}^\star) &= \frac{1}{\sqrt{G}}\sum_{g=1}^{G} \psi_{\boldsymbol{\xi}}(\mathcal{W}_g, \boldsymbol{\xi}^\star) + o_p(1) \xrightarrow{d} N(0, \boldsymbol{R}_{\boldsymbol{W}}^{-1}\boldsymbol{\Sigma}_{\boldsymbol{\xi}}\boldsymbol{R}_{\boldsymbol{W}}^{-1}),\\ \hat{V}_{\boldsymbol{\zeta}} &= \frac{1}{G}\sum_{g=1}^{G} \psi_{\boldsymbol{\xi}}(\mathcal{W}_g, \hat{\boldsymbol{\xi}})\psi_{\boldsymbol{\xi}}(\mathcal{W}_g, \hat{\boldsymbol{\xi}})' \xrightarrow{p} \boldsymbol{R}_{\boldsymbol{W}}^{-1}\boldsymbol{\Sigma}_{\boldsymbol{\xi}}\boldsymbol{R}_{\boldsymbol{W}}^{-1},\end{aligned} \tag{10}$$

where $\boldsymbol{\Sigma}_{\boldsymbol{\xi}} = \text{Var}\left(\frac{1}{N(N-1)}\sum_{(i,j):i\neq j} \boldsymbol{W}_{ijg}(\Delta A_{ijg} - \boldsymbol{W}'_{ijg}\boldsymbol{\xi})\right)$.

The moment function for $\boldsymbol{\beta}$ in this case is given by $\ell(\mathcal{V}_g, \boldsymbol{\zeta}, \boldsymbol{\beta}) = X_{ig}(\boldsymbol{\zeta})(\Delta Y_{ig} - X_{ig}(\boldsymbol{\zeta})'\boldsymbol{\beta})$. Recall that $\max\{Q_{itg}(\boldsymbol{\zeta}), R_{itg}(\boldsymbol{\zeta}), S_{itg}(\boldsymbol{\zeta})\} \leq \sum_{j\neq i} \hat{A}_{ijtg}(\boldsymbol{\zeta}) \leq N_g(N_g - 1)$. Therefore,

$$\begin{aligned}\|\boldsymbol{X}_{ig}(\boldsymbol{\zeta})\|^2 &= 1 + D_{ig} + (Q_{i1g}(\boldsymbol{\zeta}))^2 + (R_{i1g}(\boldsymbol{\zeta}) - S_{i0g}(\boldsymbol{\zeta}))^2\\ &\leq 2 + 5(N_g(N_g - 1))^2 =: B_4 < \infty.\end{aligned}$$

Also, $\nabla_{\zeta_1}\boldsymbol{X}_{ig}(\boldsymbol{\zeta}) = (\boldsymbol{0}, \boldsymbol{0}, \sum_{j\neq i}\boldsymbol{W}_{ijg}D_{jg}, \sum_{j\neq i}\boldsymbol{W}_{ijg}(1 - D_{jg}))'$. Thus, by the same argument in (6), $\left\|\nabla_{\zeta_1}\boldsymbol{X}_{ig}(\boldsymbol{\zeta})\right\|^2 \leq 6(N_g - 1)^2 = B_2 < \infty$. Next, $\nabla_{\zeta_0}\boldsymbol{X}_{ig}(\boldsymbol{\zeta}) =$

$(\mathbf{0}, \mathbf{0}, \mathbf{0}, \sum_{j\neq i} \boldsymbol{W}_{ijg})'$. Thus,

$$\left\|\nabla_{\zeta_0} \boldsymbol{X}_{ig}(\boldsymbol{\zeta})\right\|^2 \leq \sum_{j\neq i} \boldsymbol{W}'_{ijg} \sum_{j\neq i} \boldsymbol{W}_{ijg}$$
$$1 + N_g^2 D_{ig} + \left(\sum_{j\neq i} D_{jg}\right)^2 + \left(\sum_{j\neq i} D_{ig} D_{jg}\right) \leq 4(N_g - 1)^2 = B_5 < \infty.$$

The above boundedness imply boundedness of derivative of the moment functions:

$$\|\nabla_{\boldsymbol{\beta}} \ell(\mathcal{V}_g, \boldsymbol{\zeta}, \boldsymbol{\beta})\| = \|\boldsymbol{X}_{ig}(\boldsymbol{\zeta}) \boldsymbol{X}_{ig}(\boldsymbol{\zeta})'\| = \|\boldsymbol{X}_{ig}(\boldsymbol{\zeta})\|^2 < \infty,$$
$$\left\|\nabla_{\zeta_t} \ell(\mathcal{V}_g, \boldsymbol{\zeta}, \boldsymbol{\beta})\right\| = \|\ell_{1t}\| + \|\ell_{2t}\| < \infty,$$

where

$$\begin{aligned}
\|\ell_{1t}\| &:= \left\|\nabla_{\zeta_t} \boldsymbol{X}_{ig}(\boldsymbol{\zeta})(\Delta Y_{ig} - \boldsymbol{X}_{ig}(\boldsymbol{\zeta})'\boldsymbol{\beta})\right\| \\
&= \left\|\nabla_{\zeta_t} \boldsymbol{X}_{ig}(\boldsymbol{\zeta})\right\| |(\Delta Y_{ig} - \boldsymbol{X}_{ig}(\boldsymbol{\zeta})'\boldsymbol{\beta})| \\
&\leq \left\|\nabla_{\zeta_t} \boldsymbol{X}_{ig}(\boldsymbol{\zeta})\right\| (|\Delta Y_{ig}| + \|\boldsymbol{X}_{ig}(\boldsymbol{\zeta})\| \, \|\boldsymbol{\beta}\|) \leq B_6 \left(|\Delta Y_{ig}| + \sqrt{B_2} \, \|\boldsymbol{\beta}\|\right) < \infty, \\
\|\ell_{2t}\| &:= \left\|\boldsymbol{X}_{ig}(\boldsymbol{\zeta}) \nabla_{\zeta_t} (\boldsymbol{X}_{ig}(\boldsymbol{\zeta})'\boldsymbol{\beta})\right\| \\
&\leq \|\boldsymbol{X}_{ig}(\boldsymbol{\zeta})\boldsymbol{\beta}'\| \left\|\nabla_{\zeta_t} (\boldsymbol{X}_{ig}(\boldsymbol{\zeta}))\right\| \\
&= \|\boldsymbol{X}_{ig}(\boldsymbol{\zeta})\| \, \|\boldsymbol{\beta}\| \left\|\nabla_{\zeta_t} (\boldsymbol{X}_{ig}(\boldsymbol{\zeta}))\right\| < \infty,
\end{aligned}$$

for $t \in \{0, 1\}$, by Cauchy-Schwarz inequality for Frobenius inner product, and the definition of Frobenius norm and L_2 norm. Also, since $\nabla_{\boldsymbol{\beta}} \ell(\mathcal{V}_g, \boldsymbol{\zeta}, \boldsymbol{\beta})$ and $\nabla_{\zeta_t} \ell(\mathcal{V}_g, \boldsymbol{\zeta}, \boldsymbol{\beta})$ are continuous on $\boldsymbol{\zeta}, \boldsymbol{\beta}$, we can apply Lemma 1 to conclude

$$\begin{aligned}
\frac{1}{G} \sum_{g=1}^{G} \ell(\mathcal{V}_g, \bar{\boldsymbol{\zeta}}, \bar{\boldsymbol{\beta}}) &\xrightarrow{p} E[\ell(\mathcal{V}_g, \boldsymbol{\zeta}^\star, \boldsymbol{\beta}^\star)] = 0 \\
\frac{1}{G} \sum_{g=1}^{G} \nabla_{\boldsymbol{\beta}} \ell(\mathcal{V}_g, \bar{\boldsymbol{\zeta}}, \bar{\boldsymbol{\beta}}) &\xrightarrow{p} E\left[\nabla_{\boldsymbol{\beta}} \ell(\mathcal{V}_g, \boldsymbol{\zeta}^\star, \boldsymbol{\beta}^\star)\right] = -E\left[\boldsymbol{X}_{ig}(\boldsymbol{\zeta}^\star) \boldsymbol{X}_{ig}(\boldsymbol{\zeta}^\star)'\right] =: -\boldsymbol{R}_{\boldsymbol{X}} \\
\frac{1}{G} \sum_{g=1}^{G} \nabla_{\zeta_t} \ell(\mathcal{V}_g, \bar{\boldsymbol{\zeta}}, \bar{\boldsymbol{\beta}}) &\xrightarrow{p} E\left[\nabla_{\zeta_t} \ell(\mathcal{V}_g, \boldsymbol{\zeta}^\star, \boldsymbol{\beta}^\star)\right] = -E[\boldsymbol{X}_{ig}(\boldsymbol{\zeta}^\star) \nabla_{\zeta_t} (\boldsymbol{X}_{ig}(\boldsymbol{\zeta}^\star)'\boldsymbol{\beta}^\star)] =: -\boldsymbol{Q}_{\zeta_t},
\end{aligned} \tag{11}$$

Therefore, by the same argument in Part B of proof of Proposition 3, we have $\hat{\boldsymbol{\beta}} \xrightarrow{p} \boldsymbol{\beta}^\star$, and

$$\sqrt{G}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^\star) = \frac{1}{\sqrt{G}} \sum_{g=1}^{G} \psi_{\boldsymbol{\beta}}(\mathcal{V}_g, \mathcal{W}_g, \boldsymbol{\zeta}^\star, \boldsymbol{\beta}^\star) + o_p(1) \xrightarrow{d} N\left(0, V_{\boldsymbol{\beta}}\right)$$
$$\hat{V}_{\boldsymbol{\beta}} = \frac{1}{G} \sum_{g=1}^{G} \psi_{\boldsymbol{\beta}}(\mathcal{V}_g, \mathcal{W}_g, \hat{\boldsymbol{\zeta}}, \hat{\boldsymbol{\beta}}) \psi_{\boldsymbol{\beta}}(\mathcal{V}_g, \mathcal{W}_g, \hat{\boldsymbol{\zeta}}, \hat{\boldsymbol{\beta}})' \xrightarrow{p} V_{\boldsymbol{\beta}},$$

where $\psi_{\boldsymbol{\beta}}(\mathcal{V}_g, \mathcal{W}_g, \boldsymbol{\zeta}, \boldsymbol{\beta}) := \boldsymbol{R}_{\boldsymbol{Z}}^{-1} \left[\ell(\mathcal{V}_g, \boldsymbol{\zeta}, \boldsymbol{\beta}) - \boldsymbol{Q}_{\zeta_1} \psi_{\zeta_1}(\mathcal{W}_g, \boldsymbol{\zeta}) - \boldsymbol{Q}_{\zeta_2} \psi_{\zeta_2}(\mathcal{W}_g, \boldsymbol{\zeta})\right]$, and $V_{\boldsymbol{\beta}} = E\left[\psi_{\boldsymbol{\beta}}(\mathcal{V}_g, \mathcal{W}_g, \boldsymbol{\zeta}^\star, \boldsymbol{\beta}^\star) \psi_{\boldsymbol{\beta}}(\mathcal{V}_g, \mathcal{W}_g, \boldsymbol{\zeta}^\star, \boldsymbol{\beta}^\star)'\right]$. Lastly, by applying Delta-method, we have

$$\sqrt{G}(\hat{\boldsymbol{\pi}} - \boldsymbol{\pi}^\star) = \frac{1}{\sqrt{G}} \sum_{g=1}^{G} \psi_{\boldsymbol{\pi}}(\mathcal{W}_g, \mathcal{V}_g, \boldsymbol{\zeta}^\star, \boldsymbol{\beta}^\star) + o_p(1) \xrightarrow{d} N\left(0, V_{\boldsymbol{\pi}}\right),$$
$$\hat{V}_{\boldsymbol{\pi}} = \frac{1}{G} \sum_{g=1}^{G} \psi_{\boldsymbol{\pi}}(\mathcal{W}_g, \mathcal{V}_g, \hat{\boldsymbol{\zeta}}, \hat{\boldsymbol{\beta}}) \psi_{\boldsymbol{\pi}}(\mathcal{W}_g, \mathcal{V}_g, \hat{\boldsymbol{\zeta}}, \hat{\boldsymbol{\beta}})' \xrightarrow{p} V_{\boldsymbol{\pi}}$$

where $V_{\boldsymbol{\pi}} = E\left[\psi_{\boldsymbol{\pi}}(\mathcal{W}_g, \mathcal{V}_g, \boldsymbol{\zeta}^\star, \boldsymbol{\beta}^\star) \psi_{\boldsymbol{\pi}}(\mathcal{W}_g, \mathcal{V}_g, \boldsymbol{\zeta}^\star, \boldsymbol{\beta}^\star)'\right]$, and $\psi_{\boldsymbol{\pi}}(\mathcal{V}_g, \mathcal{W}_g, \boldsymbol{\zeta}, \boldsymbol{\beta})$ is defined by

$$\begin{pmatrix} [\psi_{\boldsymbol{\beta}}(\mathcal{W}_g, \mathcal{V}_g, \boldsymbol{\zeta}, \boldsymbol{\beta})]_2 \\ (N-1)[\psi_{\boldsymbol{\beta}}(\mathcal{W}_g, \mathcal{V}_g, \boldsymbol{\zeta}, \boldsymbol{\beta})]_4 [\boldsymbol{\xi}]_2 + [\boldsymbol{\beta}]_4 [\psi_{\boldsymbol{\xi}}(\mathcal{W}_g, \boldsymbol{\zeta})]_2 \\ ([\psi_{\boldsymbol{\beta}}(\mathcal{W}_g, \mathcal{V}_g, \boldsymbol{\zeta}, \boldsymbol{\beta})]_3 - [\psi_{\boldsymbol{\beta}}(\mathcal{W}_g, \mathcal{V}_g, \boldsymbol{\zeta}, \boldsymbol{\beta})]_4) [\boldsymbol{\zeta}_1]_1 + ([\boldsymbol{\beta}]_3 - [\boldsymbol{\beta}]_4) [\psi_{\zeta_1}(\mathcal{W}_g, \boldsymbol{\zeta})]_1 \\ [\psi_{\boldsymbol{\beta}}(\mathcal{W}_g, \mathcal{V}_g, \boldsymbol{\zeta}, \boldsymbol{\beta})]_3 [\boldsymbol{\xi}]_3 + [\boldsymbol{\beta}]_3 [\psi_{\boldsymbol{\xi}}(\mathcal{W}_g, \boldsymbol{\zeta})]_3 \end{pmatrix}.$$

□

B Tables

This table shows the simulation results for the first-stage estimation of dyadic regression for Designs 1 and 2, as defined in Section 4.

Table 6: Simulation Result of Dyadic Coefficients

	Design 1				**Design 2, ζ_1**				**Design 2, ξ**			
G	ζ_1	ζ_2	ζ_3	ζ_4	ζ_1	ζ_2	ζ_3	ζ_4	ξ_1	ξ_2	ξ_2	ξ_4
Mean												
100	0.159	0.026	0.026	0.37	0.067	0.069	0.069	0.037	0	0.021	0.021	0.172
200	0.159	0.025	0.025	0.37	0.067	0.069	0.069	0.037	0	0.02	0.02	0.172
400	0.159	0.025	0.025	0.37	0.067	0.069	0.069	0.038	0	0.021	0.021	0.172
800	0.159	0.026	0.026	0.37	0.067	0.069	0.069	0.037	0	0.021	0.021	0.172
1600	0.159	0.025	0.025	0.37	0.067	0.069	0.069	0.037	0	0.021	0.021	0.172
TRUE	0.159	0.025	0.025	0.37	0.067	0.069	0.069	0.037	0	0.021	0.021	0.172
MSE												
100	0.3e-5	0.4e-5	0.4e-5	1.4e-5	0.1e-5	0.3e-5	0.3e-5	1.1e-5	0.3e-5	0.5e-5	0.5e-5	1.7e-5
200	0.1e-5	0.2e-5	0.2e-5	0.7e-5	0.1e-5	0.1e-5	0.1e-5	0.5e-5	0.1e-5	0.3e-5	0.3e-5	0.8e-5
400	0.1e-5	0.1e-5	0.1e-5	0.3e-5	0e-5	0.1e-5	0.1e-5	0.3e-5	0.1e-5	0.1e-5	0.1e-5	0.4e-5
800	0e-5	0.1e-5	0.1e-5	0.2e-5	0e-5	0e-5	0e-5	0.1e-5	0e-5	0.1e-5	0.1e-5	0.2e-5
1600	0e-5	0e-5	0e-5	0.1e-5	0e-5	0e-5	0e-5	0.1e-5	0e-5	0e-5	0e-5	0.1e-5
Coverage Rate												
100	0.947	0.952	0.952	0.95	0.942	0.939	0.939	0.94	0.949	0.934	0.934	0.94
200	0.947	0.938	0.938	0.942	0.946	0.938	0.938	0.961	0.943	0.939	0.939	0.948
400	0.95	0.947	0.947	0.961	0.957	0.95	0.95	0.957	0.954	0.942	0.942	0.948
800	0.954	0.949	0.949	0.946	0.958	0.955	0.955	0.952	0.952	0.952	0.952	0.958
1600	0.948	0.95	0.95	0.953	0.942	0.945	0.945	0.949	0.936	0.947	0.947	0.953

Notes: This table presents the simulation results for $B = 1,000$ replications. Column G denotes the number of independent groups, with each group containing $N = 20$ individuals. The first panel shows the mean across all replications, and the row labeled "TRUE" presents the true values for each decomposition. The second and third columns display the MSE and 95% coverage rates, respectively. $(\pi^{DT}, \pi^{DN}, \pi^{IT}, \pi^{IN})$ denote direct treatment, direct network, indirect treatment, indirect network effects, respectively.

This table shows the simulation results for the second-stage estimation of outcome regression for Designs 1 and 2, as defined in Section 4.

Table 7: Simulation Result of Outcome Coefficients

	Design 1				**Design 2**			
G	π^{DT}	π^{DN}	π^{IT}	π^{IN}	π^{DT}	π^{DN}	π^{IT}	π^{IN}
Mean								
100	1.996	0.987	0.802	0.601	0.971	1.021	0.801	0.58
200	1.982	0.983	0.804	0.609	1.01	0.976	0.807	0.617
400	1.994	0.995	0.801	0.603	1.006	1.001	0.797	0.604
800	2.002	0.999	0.8	0.599	1.005	0.985	0.804	0.608
1600	1.997	0.998	0.8	0.601	1.002	0.993	0.802	0.605
TRUE	2	1	0.8	0.6	1	1	0.8	0.6
MSE								
100	0.507	0.308	0.015	0.114	0.27	0.772	0.077	0.392
200	0.237	0.146	0.007	0.054	0.13	0.367	0.038	0.189
400	0.125	0.078	0.004	0.028	0.065	0.187	0.019	0.094
800	0.059	0.035	0.002	0.013	0.033	0.09	0.009	0.046
1600	0.03	0.017	0.001	0.007	0.015	0.042	0.004	0.021
Coverage Rate								
100	0.94	0.944	0.946	0.936	0.943	0.944	0.942	0.943
200	0.957	0.94	0.946	0.951	0.949	0.938	0.946	0.94
400	0.942	0.938	0.939	0.94	0.947	0.944	0.951	0.944
800	0.947	0.95	0.959	0.947	0.951	0.947	0.95	0.949
1600	0.938	0.95	0.951	0.944	0.961	0.958	0.961	0.958

Notes: This table presents the simulation results for $B = 1,000$ replications. Column G denotes the number of independent groups, with each group containing $N = 20$ individuals. The first panel shows the mean across all replications, and the row labeled "TRUE" presents the true values for each decomposition. The second and third columns display the MSE and 95% coverage rates, respectively. $(\pi^{DT}, \pi^{DN}, \pi^{IT}, \pi^{IN})$ denote direct treatment, direct network, indirect treatment, indirect network effects, respectively.