Causal Effects of Treatments with Network Changes

Siwon Ryu*

Latest Version of Draft: **Here**

Abstract

Recent empirical studies emphasize the importance of indirect, or spillover effects in program evaluation. Most studies assume that the underlying network is exogenous, fixed, or unaffected by the intervention. However, 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 (network effect). 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 estimation procedure for causal effects and their decomposition is proposed, and its performance is evaluated through Monte Carlo simulations. The methodology is illustrated using data from a program offering savings accounts. The empirical results show that total direct effect is small due to offsetting positive treatment effects and negative network effects.

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

*Department of Economics, Rice University. Email: Siwon.Ryu@rice.edu

1 Introduction

Program evaluation is an important topic in empirical economics, typically involving the estimation of causal effects of programs. Methods are often based on the potential outcome framework (e.g., Rubin (1974)) to define causal effects, with the baseline assumption of the *Stable Unit Treatment Value Assumption* (SUTVA), which excludes interference between units or individuals. However, as economic agents frequently interact with one another, recent empirical evidence 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 for all individuals. This creates significant challenges for researchers, as defining causal parameters becomes more complex compared to SUTVA settings. Furthermore, identifying meaningful parameters becomes increasingly difficult due to the exponential growth in the number of potential outcomes as the number of individuals increases. For example, if N units interact, the possible number of treatment assignments profile and the potential outcomes is 2^N. To address these challenges, studies often adopt the constant treatment response (CTR) framework proposed by Manski (2013). This approach assumes that an individual's potential outcome depends on specific summary statistics derived from the treatment statuses of others, rather than the full treatment vector. For example, an individual's outcome may be determined by their own treatment status and the number of treated friends, rather than the complete profile of treatments. Those summary statistics are called exposures, and a spillover (or exposure) effect can be defined as the impact of changes in exposure, which indirectly reflects changes in the treatment statuses of others. This exposure mapping approach significantly reduces the dimensionality of the potential outcomes framework (e.g., Leung (2020, 2022), Vazquez-Bare (2023a, 2023b), Aronow and Samii (2017), Bramoullé, Djebbari, and Fortin (2009), Forastiere, Airoldi, and Mealli (2021), Auerbach and Tabord-Meehan (2021)).

Another notable approach to addressing interference is the design of experiments utilizing double randomization, as proposed by Hudgens and Halloran (2008). This method involves first randomizing treatment rates (saturation) across groups and then randomizing treatment assignments within each group according to the specified 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)). Furthermore, some studies focus on optimizing experimental designs in such settings to maximize social welfare (e.g., Kitagawa and Wang (2023), Ananth (2021), Viviano (2024)).

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, let D_i denote the treatment indicator for individuals $i = 1, ..., N$, and let A_{ij} represent the indicator of link between individuals i and j. In this setting, exposure often refers to the number of treated friends, which can be expressed as $\sum_{j=1, j\neq i}^{N} A_{ij} D_j$. Consequently, exposure generally depends on the network structure. If we assume that the network links A_{ij} are also influenced by the treatment vector $\boldsymbol{D} = (D_1, ..., D_N)$, changes in exposure would reflect both changes in others' treatment statuses (D_j for $j \neq i$) and changes in the network links A_{ij} driven by $\boldsymbol{D}$. As a result, the exposure effect becomes confounded by these two factors: others' treatments and changes in the network structure. Thus, addressing possible network changes within the exposure mapping framework is challenging without additional assumptions.

The assumption of fixed network may hold in the short term when there is insufficient time for the network structure to change. However, recent empirical studies suggest that interventions can significantly influence the underlying network structure. For example, Comola and Prina (2021) use experimental data from Nepal and find that providing savings accounts to households leads 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 become 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 presents a method to identify and estimate the causal effect of a program, accounting for the possibility that the treatment may induce changes in the network structure. Additionally, this study decomposes the treatment (or spillover) effect into two components.

The first component examines the impact of the treatment when the network structure remains fixed, while the second component assesses the effect of the treatment when it alters the network structure only. If the treatment does not result in changes to the network, the second component has no effect, and the first component corresponds to the conventional concept of the treatment (or spillover) effect. In this framework, I define the second component as the *network effect* and refer to the first component as the treatment (or spillover) effect, following established conventions.

Causal changes in the network structure can be evaluated using the potential network. For instance, let $A_{ij}(\boldsymbol{d})$ represent potential link indicator for individuals i and j, according to the treatment vector $\boldsymbol{d} = (d_1, ..., d_N)$. To describe the decomposition, consider an intervention that shifts the treatment statuses of all individuals from $\boldsymbol{d}$ to $\boldsymbol{d'}$. The overall treatment (or spillover) effect measures the causal changes in outcomes resulting from both changes in the treatment vector ($\boldsymbol{d} \to \boldsymbol{d'}$), and the underlying network ($A_{ij}(\boldsymbol{d}) \to A_{ij}(\boldsymbol{d'})$). In this paper, the (pure) treatment (or spillover) effect is defined as the causal effect of changes in the treatment statuses ($\boldsymbol{d} \to \boldsymbol{d'}$), while the network is fixed at $A_{ij}(\boldsymbol{d})$. The network effect, on the other hand, is defined as the causal effect of changes in the network ($A_{ij}(\boldsymbol{d}) \to A_{ij}(\boldsymbol{d'})$), while treatment statuses are fixed at $\boldsymbol{d}$.

The proposed method is based on dyadic link formation and a linear potential outcome assumption. Specifically, potential network links are determined solely by the treatment statuses of each pair, i.e., $A_{ij}(\boldsymbol{d}) = A_{ij}(d_i, d_j)$, and the potential outcome is modeled as a linear function of the own treatment status and the number of treated and untreated friends, with an individual error term $\varepsilon_i(d_i)$ that depends only on the individual's own treatment status. Further assumptions about potential link formation ($A_{ij}(d_i, d_j)$) and the distribution of the potential outcome ($\varepsilon_i(d_i)$) are specified by the experimental design of interest. This study explicitly considers two experimental settings. First, randomized experiments with post-treatment period information, where both potential links and outcome errors are assumed to be independent of the treatment (i.e., unconfounded). Second, quasi-experimental designs with parallel trends, incorporating pre- and post-treatment information, where both potential links and outcome errors are assumed to satisfy the no-anticipation and parallel trends assumptions.

An estimation procedure is proposed, and its finite sample performance is evaluated through Monte Carlo simulations. The approach is also demonstrated using experimental

data from Comola and Prina (2021), which involves a program providing savings accounts to households in Nepal. The empirical results indicate positive direct and indirect treatment effects on consumption. However, the total direct effect is small due to the negative network effect offsetting the positive treatment effect. This suggests that while opening a savings account directly increases consumption, it may also reduce consumption by altering the network structure, such as increasing the number of friends in this context. This decomposition and interpretation go beyond the direct and indirect effect calculations presented in Comola and Prina (2021).

The main contributions of this paper can be summarized in three key aspects. First, I propose a novel method for analyzing causal effects that accounts for causal network changes within the potential outcome framework. Second, the method decomposes the causal effects, providing a more detailed understanding of the mechanisms driving a program's impact. Third, the method is applicable to various experimental designs, including both randomized and quasi-experimental approaches, and it can be easily extended to other experimental settings with additional conditions. Moreover, 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 (2023b), Leung (2020), or Aronow and Samii (2017). Furthermore, if there is no interference, the method simplifies to the standard potential outcome model under SUTVA.

Related Literature

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 using the exposure mapping approach under randomized experiments (e.g., Leung (2020, 2022), and Vazquez-Bare (2023b)). When treatment is endogenous due to imperfect compliance, local average treatment effects can be identified (e.g., Vazquez-Bare (2023a), DiTraglia et al. (2023), Hoshino and Yanagi (2023), Kormos, Lieli, and Huber (2023), Kang and Imbens (2016), Blackwell (2017)). In cases where treatment is not exogenous such as a quasi-experimental situation, studies use difference-in-differences approach (e.g., Xu (2023) and Butts (2021)) or regression discontinuity design (e.g., Auerbach, Cai, and Rafi (2024)). 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 previous studies generally assume a fixed or exogenous network, Comola and Prina (2021) explicitly address estimation of treatment effects accounting for network changes. The authors propose a two-period linear-in-means model:

$$Y_{i1} = \beta_1 \sum_{j \neq i} \tilde{A}_{ij0} Y_{j1} + \beta_2 \sum_{j \neq i} \Delta\tilde{A}_{ij} Y_{j1} + \gamma D_i + \delta_1 \sum_{j \neq i} \tilde{A}_{ij0} D_j + \delta_2 \sum_{j \neq i} \Delta\tilde{A}_{ij} D_j + \varepsilon_{i1}, \quad (1)$$

where $\tilde{A}_{ijt} = A_{ijt} / \sum_{j \neq i} A_{ijt}$ represents the row-normalized link at time $t \in \{0, 1\}$, $\Delta A_{ij} = A_{ij1} - A_{ij0}$ is the first-difference of row-normalized network link, Y_{it} is the observed outcome at time t, ε_{it} is the individual error term, and D_i is the treatment indicator. In this model, both treatment and network structure are assumed to be exogenous in the sense that $E[\varepsilon_{i1} | \boldsymbol{A}_1, \boldsymbol{A}_0, \boldsymbol{D}] = 0$. The authors estimate the coefficients using an instrumental variable (IV) estimation strategy similar to that of Bramoullé, Djebbari, and Fortin (2009). And then, direct and indirect effects are defined as the partial derivatives of the conditional mean of the reduced-form outcome with respect to the treatment vector: $\partial E[Y_{i1} | \boldsymbol{D}] / \partial \boldsymbol{D}'$. Their main findings using an experimental data in Nepal suggest that indirect effects may be underestimated if network changes resulting from the treatment are not considered.

This study differs from their paper in several key aspects. First, I introduce a potential outcome and potential network framework to provide clear causal interpretations, whereas their approach defines treatment effects as derivatives of a reduced-form outcome. Their treatment effects are difficult to interpret as causal effects when the network is not exogenous (i.e., when $E[\varepsilon_{i1} | \boldsymbol{A}_1, \boldsymbol{A}_0, \boldsymbol{D}] \neq 0$), whereas I consider a more general situation that includes the potential endogeneity of the network. Second, while their structural equation (1) suggests that network changes may reflect time-varying effects, I explicitly model causal changes in the network driven by treatment using the potential network framework. Lastly, the most significant difference is that I decompose causal effects into two distinct components: the treatment effect when the network is fixed, and the network effect, which captures the causal impact of changes in the network alone. This decomposition allows researchers to determine the extent to which causal effects are driven by network changes, providing a clearer understanding of the underlying mechanisms that have yet to be explored in the existing 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 3 presents the estimation procedure. Section 4 evaluates the performance of the proposed method through Monte Carlo simulations, and Section 5 provides an empirical illustration. Section 6 concludes.

2 Model and Identification

In this section, I provide an overview of the model setting and define the key parameters of interest. First, I discuss the response functions for both potential link and potential outcome. Then, I introduce the main causal parameters, focusing on direct and indirect effects. The direct effects capture 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, which I refer 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 Network Links and the Potential Outcome

Suppose there are G independent groups with N individuals in each group.[1] If we observe data over two periods, let $t \in \{0,1\}$ denote 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 $\boldsymbol{D}_g = (D_{1g}, ..., D_{Ng}) \in \{0,1\}^N$ be the treatment vector. We assume that there is no imperfect compliance. To simplify notations, 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 individuals i and j are linked, and $A_{ii} = 0$ for all i. Denote $\boldsymbol{A}$ as the $N \times N$ adjacency matrix with $[\boldsymbol{A}]_{ij} = A_{ij}$. The network can be either directed or

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

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 next 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 pairs (i,j) and for $(d_i, d_j) \in \{0,1\}^2$, the following condition holds: $E[A_{ij}(d_i,d_j)|\boldsymbol{D}] = E[A_{ij}(d_i,d_j)|D_i, D_j]$.*

Assumption 1-(i) states that each pair's potential link is determined 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):

$$A_{ij}(\boldsymbol{d}) = \mathbb{1}\{\theta_0 + \theta_1 d_i + \theta_2 d_j + u_{ij} > 0\}, \tag{2}$$

where u_{ij} includes both individual-specific, and pair-specific unobserved factors influencing the link formation. In this model, the potential link A_{ij} depends only on d_i, d_j, but not d_k for $k \notin \{i,j\}$. Assumption 1-(ii) strengthens (i) by assuming that each pair's observed treatment statuses are sufficient for determining their potential link distribution. 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 $f(\cdot)$ 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 $f(\cdot)$ 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, i.e., 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, and $R_i(\boldsymbol{d}) := \sum_j A_{ij}(d_i, d_j)(1-d_j)$ 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.

To identify causal effects and their decomposition, I assume that the response function for the potential outcome is linear in these exposures $(d_i, Q_i(\boldsymbol{d}), R_i(\boldsymbol{d}))$, as stated next in Assumption 2.

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:*

$$Y_i(\boldsymbol{d}) = \beta_0 + \beta_I d_i + \beta_T Q_i(\boldsymbol{d}) + \beta_U R_i(\boldsymbol{d}) + \varepsilon_i(d_i),$$

where $Q_i(\boldsymbol{d}) = \sum_{j=1, j\neq i}^N A_{ij}(d_i, d_j)d_j$, $R_i(\boldsymbol{d}) = \sum_{j=1, j\neq i}^N A_{ij}(d_i, d_j)(1-d_j)$, 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., $E[\varepsilon_i(1) - \varepsilon_i(0)|D_i = 1] = 0$).[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.

The parameter β_I represents the effect of an individual's own treatment d_i when all links and others' treatments remain fixed. Thus, β_I captures the direct effect of the own treatment.

[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 individual i's potential fraction of treated and untreated neighbors, respectively.

[3] If $E[\varepsilon(d)] = \mu$, then we can rewrite the model by replacing β_0 with $\tilde{\beta}_0 = \beta_0 - \mu$, and $\varepsilon_i(d)$ with $\tilde{\varepsilon}_i(d) = \varepsilon_i(d) - \mu$. If $E[\varepsilon_i(1) - \varepsilon_i(0)|D_i = 1] = \tau$, then we can rewrite the model by replacing β_I with $\tilde{\beta}_I = \beta_I + \tau$, and ε_i with $\tilde{\varepsilon}_i(1) = \varepsilon_i(1) - \tau$, $\tilde{\varepsilon}_i(0) = \varepsilon_i(0)$.

Next, since $Q_i(\boldsymbol{d})$ represents the number of potentially treated neighbors, β_T captures the spillover (or exposure) effect from one additional treated neighbor. Similarly, β_U represents the effects from one additional untreated neighbor. These parameters can be interpreted as causal effects under various assumptions, particularly if the network links are unaffected by the treatment.

However, an individual's own treatment d_i can also influence potential links $\{A_{ij}(d_i, d_j)\}_{i,j}$, introducing 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 other individual, 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 network model:

$$Y_i = \beta_0 + \beta_I D_i + \beta_T Q_i + \beta_U R_i + \varepsilon_i, \tag{3}$$

where $Q_i = Q_i(\boldsymbol{D})$ and $R_i = R_i(\boldsymbol{D})$ 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 to a linear response model commonly used in the literature (e.g., Cai, Janvry, and Sadoulet (2015), Leung (2020), Forastiere, Airoldi, and Mealli (2021)). Moreover, when $\beta_T = \beta_U = 0$, which means there is no interaction, the model simplifies to a standard causal model without interference. In particular, the model generalizes existing approaches to account for cases where the individuals interact, and also when the network structure is influenced by the treatment.

2.2 Data

The observed outcome and treatment consist of individual-level data $\{\mathcal{V}_{ig}\}_{i,g}$, and the observed network links consist of dyadic-level data $\{\mathcal{W}_{ijg}\}_{(i,j),g}$. For example, if we observe only a single period of data, $\mathcal{V}_{ig} = (Y_{ig}, D_{ig})$, and $\mathcal{W}_{ijg} = A_{ijg}$. The following assumptions are made about the data distributions:

Assumption 3 (Distribution)**.** *(i) individual-level data $\mathcal{V}_{ig}$ are identically distributed over i*

and g, and independent over across g; (ii) dyadic-level data W_{ijg} are identically distributed over all pairs (i,j) and g, and independent across g.

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

Assumption 3 states that groups are identical and independent, but it allows unrestricted dependence between individuals and pairs within each group. Moreover, this implies that potential links and potential outcomes are also identically distributed. Consequently Assumption 3-(ii) impose additional symmetry on the network links, i.e., $A_{ij}(d,e)$ and $A_{ji}(d,e)$ are identically distributed for all $(d,e) \in \{0,1\}^2$. Furthermore, if the network is undirected, i.e., when $A_{ij} = A_{ji}$, then we have $A_{ij}(d,e) \sim A_{ji}(e,d) \sim A_{ij}(e,d)$ for all $(d,e) \in \{0,1\}^2$. Therefore, the potential link between two individuals depends on the number of treated individuals between them, i.e., A_{ij} is determined by $d_i + d_j$. Assumption 4 is a standard requirement that ensures the existence of the corresponding conditional distributions.

2.3 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 decom-

posed 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 treatment 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 counterpart.

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* represents 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 general case with N individuals, indirect effect can 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

neighbor.

Let $\{\boldsymbol{e}_1, ..., \boldsymbol{e}_N\}$ be the standard Euclidean basis in $\mathbb{R}^N$, where for each i, $\boldsymbol{e}_i = (e_{i1}, ..., e_{iN})'$, $e_{ii} = 1$, and $e_{ij} = 0$ for all $j \neq i$. Define $m(d,e) = E[A_{ij}(0,0)|D_i = d, D_j = e]$ and $H(d,e) = E[A_{ij}(d,e) - A_{ij}(0,0)|D_i = d, D_j = e]$, for $(d,e) \in \{0,1\}^2$.[4] Here, $m(d,e)$ represents the conditional probability of forming a link between i and j when both individuals are untreated, while $H(d,e)$ denotes the average treatment effect of treated (ATT) on links. The *direct effect* on outcome is then expressed as $Y_i(\boldsymbol{e}_i) - Y_i(\mathbf{0})$, where $\mathbf{0} = (0, ..., 0) \in \mathbb{R}^N$. 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 $\beta_0 + \beta_U \sum_j A_{ij}(0,0) + \varepsilon_i(0)$. In this scenario, if the links remain at $A_{ij}(0,0)$ and only individual i is treated, then the outcome becomes $\beta_0 + \beta_I + \beta_U \sum_j A_{ij}(0,0) + \varepsilon_i(1)$. Therefore, the causal change in individual i's outcome due to their own treatment is the difference $\beta_I + \varepsilon_i(1) - \varepsilon_i(0)$, which defines the *direct treatment effect*. Next, in the 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 $\beta_0 + \beta_I + \beta_U \sum_j A_{ij}(1,0) + \varepsilon(1)$. Therefore the causal effect on the outcome from changes in links is the difference $\beta_U \sum_j \big(A_{ij}(1,0) - A_{ij}(0,0)\big)$, and this defines the *direct network effect*. Consequently, the direct effect is the sum of direct treatment, and direct network effects as follows:

$$\underbrace{Y_i(\boldsymbol{e}_i) - Y_i(\mathbf{0})}_{\text{Direct Effect}} = \underbrace{\beta_I + \varepsilon_i(1) - \varepsilon_i(0)}_{\text{Direct Treatment Effect}} + \underbrace{\beta_U \sum_{j=1, j\neq i}^{N} \big(A_{ij}(1,0) - A_{ij}(0,0)\big)}_{\text{Direct Network Effect}} \tag{4}$$

The *average direct effect* (π^D) is defined as the conditional expectation of direct effect ($Y_i(\boldsymbol{e}_i) - Y_i(\mathbf{0})$) given $\boldsymbol{D} = \boldsymbol{e}_i$. From (4), the average direct effect π^D is given by the sum of the conditional expectation of direct treatment effect, and the conditional expectation of direct network effect. Note that the individual error term has no ATT without loss of generality, i.e., $E[\varepsilon_i(1) - \varepsilon_i(0)|\boldsymbol{D} = \boldsymbol{e}_i] = E[\varepsilon_i(1) - \varepsilon_i(0)|D_i = 1] = 0$, as stated by Assumption 2, and therefore the average direct treatment effect is β_I. Next, in the conditional expectation of direct network effect in (4), $E[A_{ij}(1,0) - A_{ij}(0,0)|\boldsymbol{D} = \boldsymbol{e}_i] = E[A_{ij}(1,0) - A_{ij}(0,0)|D_i = 1, D_j = 0] = H(1,0)$ by Assumption 1. Therefore, we have the decomposition of average

[4] If the network is undirected, then $A_{ij}(d,e) \sim A_{ji}(e,d)$ for all $(d,e) \in \{0,1\}^2$, resulting in $H(1,0) = E[A_{ij}(1,0) - A_{ij}(0,0)|D_i = 1, D_j = 0] = E[A_{ji}(0,1) - A_{ji}(0,0)|D_j = 0, D_i = 1] = H(0,1)$.

direct effect π^D as follows:[5]

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

Here, π^{DT} refers to the *average direct treatment effect* representing the impact of one's own treatment when the network is fixed. On the other hand, π^{DN} denotes the *average direct network effect*, capturing the effect of changes in links driven by one's own treatment. Note that the average direct network effect π^{DN} depends on the number of individuals in the group. This is because an individual's treatment affects all potential links associated with them. $H(1,0)$ represents the causal effect on the probability of forming a link when the individual is treated, and thus, $(N-1)H(1,0)$ measures the expected increase in links due to the individual's own treatment. Therefore, when N is large, the individual is more likely to gain more connections as a result of the treatment.

Similarly, the *indirect effect* on the outcome is expressed as $Y_i(\boldsymbol{e}_j) - Y_i(\mathbf{0})$. To interpret this, consider a counterfactual situation where no individual is treated, i.e., $\boldsymbol{D} = \mathbf{0}$, where the potential outcome is given by $\beta_0 + \beta_U A_{ij}(0,0) + \varepsilon_i(0)$. In this scenario, if the link remains at $A_{ij}(0,0)$, and only individual j is treated, the outcome becomes $\beta_0 + \beta_T A_{ij}(0,0) + \varepsilon_i(0)$. The difference $(\beta_T - \beta_U)A_{ij}(0,0)$ defines the *indirect treatment effect*. Furthermore, in the situation where individual j is treated, but the link remains at $A_{ij}(0,0)$, if the link changes to $A_{ij}(0,1)$, the potential outcome becomes $\beta_0 + \beta_T A_{ij}(0,1) + \varepsilon_i(0)$, and the difference $\beta_T(A_{ij}(0,1) - A_{ij}(0,0))$ defines the *indirect network effect*. Thus, the indirect effect is the sum of the indirect treatment effect and the indirect network effect, expressed as follows:

$$\underbrace{Y_i(\boldsymbol{e}_j) - Y_i(\mathbf{0})}_{\text{Indirect Effect}} \quad = \quad \underbrace{(\beta_T - \beta_U)A_{ij}(0,0)}_{\text{Indirect Treatment Effect}} \quad + \quad \underbrace{\beta_T(A_{ij}(0,1) - A_{ij}(0,0))}_{\text{Indirect Network Effect}} \tag{5}$$

The *average indirect effect* (π^I) is defined as the conditional expectation of the indirect effect ($Y_i(\boldsymbol{e}_j) - Y_i(\mathbf{0})$) given $\boldsymbol{D} = \boldsymbol{e}_j$. From (5), the average indirect effect π^I is the

[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 $\beta_I + \beta_U \sum_{j\neq i} A_{ij}(1,0) + \varepsilon_i(1)$. If the links are fixed at $A_{ij}(1,0)$, but individual i is untreated, the outcome becomes $\beta_U + \sum_{j\neq i} A_{ij}(1,0) + \varepsilon_i(0)$. Therefore, the causal effect of individual i's own treatment on the outcome is again the difference $\beta_I + \varepsilon_i(1) - \varepsilon_i(0)$, and the remaining term represent the direct network effect.

sum of the conditional expectations of the indirect treatment effect and the indirect network effect. Specifically, $E[A_{ij}(0,0)|\boldsymbol{D}=\boldsymbol{e}_j] = E[A_{ij}(0,0)|D_i=0,D_j=1] = m(0,1)$, and $E[A_{ij}(0,1)-A_{ij}(0,0)|\boldsymbol{D}=\boldsymbol{e}_j] = E[A_{ij}(0,1)-A_{ij}(0,0)|D_i=0,D_j=1] = H(0,1)$ by Assumption 1. Consequently, we can decompose the average indirect effect π^I as follows:

$$\pi^I := E[Y_i(\boldsymbol{e}_j)-Y_i(\boldsymbol{0})|\boldsymbol{D}=\boldsymbol{e}_j] \quad = \quad \underbrace{(\beta_T-\beta_U)m(0,1)}_{:=\pi^{IT}} \quad + \quad \underbrace{\beta_T H(0,1)}_{:=\pi^{IN}}. \tag{6}$$

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 an alternative expression with a different interpretation for the indirect effects when considering a different comparison of counterfactual scenarios.[6] However, the preceding argument regarding identification and estimation remains the same for this alternative expression. Therefore, we use the definition in (6) for the decomposition of indirect effects.

In summary, the parameters of interest are:

$$\pi^{DT}=\beta_I, \quad \pi^{DN}=\beta_U(N-1)H(1,0), \quad \pi^{IT}=(\beta_T-\beta_U)m(0,1), \quad \pi^{IN}=\beta_T H(0,1).$$

Based on the linearity of the outcome response function (Assumption 2), any direct or indirect effect, i.e., for any $\boldsymbol{d},\boldsymbol{d}'\in\{0,1\}^N$, can be similarly defined. For example, if $\boldsymbol{d}$ represents the situation where individual i is untreated, while N_T other individuals are treated, and $\boldsymbol{d}'$ is the same situation except that individual i is treated, then $E[Y_i(\boldsymbol{d}')-Y_i(\boldsymbol{d})|\boldsymbol{D}=\boldsymbol{d}']$ defines an average direct effect π^D, and in this comparison, we have $\pi^{DT}=\beta_I$, and $\pi^{DN}=(\beta_T N_T+\beta_U(N-1-N_T))H(1,0)$. The definition of the decomposition can vary depending on which comparison is most relevant in a given empirical context. However, it is important to note that once we identify the outcome coefficients $\boldsymbol{\beta}$ and the conditional distribution of potential links,

[6] For the indirect effect, the decomposition varies depending on the counterfactual scenarios 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 scenario, we obtain an indirect treatment effect of $(\beta_T-\beta_U)A_{ij}(0,1)$ and an indirect network effect of $\beta_U(A_{ij}(0,1)-A_{ij}(0,0))$. Intuitively, the term $\beta_T-\beta_U$ 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, $(\beta_T-\beta_U)A$ represents the indirect treatment effect. Additionally, when individual j is untreated, treated, the effect of changing A_{ij} from 0 to 1 is β_U,β_T, respectively.

specifically $m(0,1)$, $H(1,0)$, and $H(0,1)$, the decomposition can be recovered regardless of which situations are being compared.

2.4 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 are observed, but the treatment is exogenous (i.e., randomized experiments). Then, I discuss the case where data are 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 and no-anticipation assumptions.

In both cases, 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 $\boldsymbol{\zeta}$ from a dyadic regression on links. Second, the coefficients $\beta_I, \beta_T, \beta_U$ in the outcome model (3) are identified using individual-level data. If the network is exogenous, the coefficient $\boldsymbol{\beta}$ can be identified using the information of observed network and treatment. Specifically, in (3), if $E[\varepsilon_i|\boldsymbol{A},\boldsymbol{D}] = 0$, then all regressors in (3) are exogenous, and the coefficient from regression of Y_i on $(1, D_i, Q_i, R_i)$ recovers $\boldsymbol{\beta}$. However, I consider possible endogeneity of network links. For instance, if the links are formed by the dyadic model (2), and the dyadic error term u_{ij} in (2) is correlated with the individual error term ε_i in (3), then even if the treatment is randomly assigned, the network is endogenous and the aforementioned regression using the observed network will not recover the outcome coefficient $\boldsymbol{\beta}$. Instead, I use predicted network $E[A_{ij}|\boldsymbol{D}]$ from the first-step dyadic regression to recover the coefficient $\boldsymbol{\beta}$ in the outcome regression.[7] Lastly, causal effects and the decomposition are recovered by the combination of $\boldsymbol{\zeta}$ and $\boldsymbol{\beta}$.

2.4.1 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

[7] This strategy is often employed in the literature to deal with endogenous network. See, for example, Kelejian and Piras (2014), König, Liu, and Zenou (2019), Lee et al. (2021)

this case, both the individual error term in the outcome response function and potential links are independent of the treatment, as stated as follows:

Assumption 5 (Randomized Experiment)**.** *The treatment is exogenous: (i)* $E[\varepsilon_i(d)|D_i] = 0$ *for all* $d \in \{0,1\}$*; (ii)* $E[A_{ij}(d_i,d_j)|D_i,D_j] = E[A_{ij}(d_i,d_j)]$ *for all* $(d_i,d_j) \in \{0,1\}^2$.

The identification process consists of three steps. In the first step, define the dyadic regressor vector as $\boldsymbol{W}_{ij} = (1, D_i, D_j, D_iD_j)' \in \mathbb{R}^4$. The conditional exepctation of links A_{ij} given D_i, D_j is then given by a saturated regression $E[A_{ij}|D_i,D_j] = \boldsymbol{W}_{ij}'\boldsymbol{\zeta}$. Notice that if $\Pr(D_i = D_j) < 1$, then $E[\boldsymbol{W}_{ijg}\boldsymbol{W}_{ijg}']$ is nonsingular and the coefficient $\boldsymbol{\zeta}$ is identified via least squares estimand. Define $\bar{A}(d,e) = E[A_{ij}|D_i = d, D_j = e]$. Then, the coefficient $\boldsymbol{\zeta}$ recovers the following:

$$\boldsymbol{\zeta} = \begin{pmatrix} \zeta_1 \\ \zeta_2 \\ \zeta_3 \\ \zeta_4 \end{pmatrix} = \begin{pmatrix} \bar{A}(0,0) \\ \bar{A}(1,0) - \bar{A}(0,0) \\ \bar{A}(0,1) - \bar{A}(0,0) \\ \bar{A}(1,1) - \bar{A}(1,0) - \bar{A}(0,1) + \bar{A}(0,0) \end{pmatrix}. \tag{7}$$

Moreover, by Assumption 5, $\bar{A}(d,e) = E[A_{ij}|D_i = d, D_j = e] = E[A_{ij}(d,e)]$. Therefore, $\zeta_2 = H(1,0)$, $\zeta_3 = H(0,1)$, and $\zeta_1 = m(0,1)$.

In the second step, recall that from (3), the conditional expectation of observed outcome is given as $E[Y_i|\boldsymbol{A},\boldsymbol{D}] = \boldsymbol{Z}_i'\boldsymbol{\beta} + E[\varepsilon_i|\boldsymbol{A},\boldsymbol{D}]$, where $\boldsymbol{Z}_i := (1, D_i, Q_i, R_i)$ and $\boldsymbol{\beta} = (\beta_0, \beta_I, \beta_T, \beta_U)$. Here, even though the treatment is randomly assigned, it can be $E[\varepsilon_i|\boldsymbol{A},\boldsymbol{D}] \neq 0$ because the network is possibly correlated with the individual error term ε_i. Instead, by Assumption 5, we can write: $E[Y_i|\boldsymbol{D}] = E[\boldsymbol{Z}_i|\boldsymbol{D}]'\boldsymbol{\beta}$. Here, in the conditional expectation $E[\boldsymbol{Z}_i|\boldsymbol{D}]$, we have $E[Q_i|\boldsymbol{D}] = \sum_j E[A_{ij}|\boldsymbol{D}]D_j$, and $E[R_i|\boldsymbol{D}] = \sum_j E[A_{ij}|\boldsymbol{D}](1 - D_j)$. Note that by Assumption 1 and from the first-step dyadic regression, $E[A_{ij}|\boldsymbol{D}] = E[A_{ij}|D_i,D_j] = \boldsymbol{W}_{ij}'\boldsymbol{\zeta}$. Therefore, we can write $E[\boldsymbol{Z}_i|\boldsymbol{D}] = \boldsymbol{Z}_i(\boldsymbol{\zeta}) := (1, D_i, Q_i(\boldsymbol{\zeta}), R_i(\boldsymbol{\zeta}))$, where $Q_i(\boldsymbol{\zeta}) := \sum_j (\boldsymbol{W}_{ij}'\boldsymbol{\zeta})D_j$ and $R_i(\boldsymbol{\zeta}) := \sum_j (\boldsymbol{W}_{ij}'\boldsymbol{\zeta})(1 - D_j)$. The outcome coefficient $\boldsymbol{\beta}$ can be recovered by the coefficient of a regression $E[Y_i|\boldsymbol{D}] = \boldsymbol{Z}_i(\boldsymbol{\zeta})'\boldsymbol{\beta}$, once $\boldsymbol{\zeta}$ is known.

Then, in the last step, we can recover the decomposition $\boldsymbol{\pi}$ from $\beta_I, \beta_T, \beta_U$, $\zeta_2 = H(1,0)$, $\zeta_3 = H(0,1)$, $\zeta_1 = m(0,1)$. Proposition 1 formally states this identification procedure.

Proposition 1 (Identification with Randomized Experiment)**.** *Suppose Assumptions 1-5 hold.*

Define $\boldsymbol{W}_{ij} = (1, D_i, D_j, D_i D_j) \in \mathbb{R}^4$, *and* $\boldsymbol{Z}_i(\boldsymbol{\zeta}) = (1, D_i, Q_i(\boldsymbol{\zeta}), R_i(\boldsymbol{\zeta})) \in \mathbb{R}^4$, *where* $Q_i(\boldsymbol{\zeta}) = \sum_j (\boldsymbol{W}'_{ij}\boldsymbol{\zeta}) D_j$, *and* $R_i(\boldsymbol{\zeta}) = \sum_j (\boldsymbol{W}'_{ij}\boldsymbol{\zeta})(1 - D_j)$. *Then:*

(i) In a dyadic regression $E[A_{ij}|\boldsymbol{D}] = \boldsymbol{W}'_{ij}\boldsymbol{\zeta}$, *the coefficient* $\boldsymbol{\zeta}$ *is given by* $\boldsymbol{\zeta} = \boldsymbol{B}_{\boldsymbol{W}}^{-1} E[\boldsymbol{W}_{ij} A_{ij}]$ *provided that* $\boldsymbol{B}_{\boldsymbol{W}} := E[\boldsymbol{W}_{ij}\boldsymbol{W}'_{ij}]$ *is nonsingular (i.e.,* $D_i \neq D_j$ *with probability 1);*

(ii) The conditional expectation of outcome is given by $E[Y_i|\boldsymbol{D}] = \boldsymbol{Z}_i(\boldsymbol{\zeta})'\boldsymbol{\beta}$, *and the coefficient* $\boldsymbol{\beta} = (\beta_0, \beta_I, \beta_T, \beta_U)$ *is given by* $\boldsymbol{\beta} = \boldsymbol{B}_{\boldsymbol{Z}}^{-1} E[\boldsymbol{Z}_i(\boldsymbol{\zeta}) Y_i]$, *provided that* $\boldsymbol{B}_{\boldsymbol{Z}} := E[\boldsymbol{Z}_i(\boldsymbol{\zeta})\boldsymbol{Z}_i(\boldsymbol{\zeta})']$ *is nonsingular;*

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

In this case, because the treatment is independent of the potential outcome, the causal effects and their decomposition in this case are interpreted as the average direct/indirect treatment/network effects.

2.4.2 Identification with Parallel Trend

Next, I examine cases where the treatment may not 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 $t \in \{0, 1\}$. Additionally, let $\boldsymbol{\beta}_t = (\beta_{0t}, \beta_{It}, \beta_{Tt}, \beta_{Ut})$ represent the outcome coefficient at $t \in \{0, 1\}$. Similarly denote $Q_{it}, R_{it}, Q_{it}(\boldsymbol{d}), R_{it}(\boldsymbol{d})$ as their respective values at $t \in \{0, 1\}$.

In this case, identification relies on a difference-in-differences approach. Let Δ denote the first-difference operator, i.e., for a random variable K_t, $\Delta K = K_1 - K_0$. The following assumptions are required to ensure parallel trend and no-anticipation in potential links and the potential outcome:

Assumption 6 (No Anticipation)**.** *(i)* $E[\varepsilon_{i0}(0)|D_i = 1] = E[\varepsilon_{i0}(1)|D_i = 1]$ *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)* $\beta_{I0} = 0$ *and* $\beta_{T0} = \beta_{U0}$.

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)* $\beta_{U0}=\beta_{U1}$.

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 $\varepsilon_{i0}=\varepsilon_{i0}(0)$ with probability 1, and thus Assumption 6-(i) holds.[8] By the same argument, Assumption 6-(ii) holds when $A_{ij0}=A_{ij0}(0,0)$ with probability 1.

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

Furthermore, Assumptions 6 and 7, ensure both no-anticipation and parallel trends in the potential outcome. 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, for any $\boldsymbol{d}\in\{0,1\}^N$, we can write:

$$E[Y_{i0}(\boldsymbol{d})-Y_{i0}(\mathbf{0})|\boldsymbol{D}=\boldsymbol{d}]=\beta_{I0}d_i+\beta_{T0}E[Q_{i0}(\boldsymbol{d})|\boldsymbol{D}=\boldsymbol{d}]+\beta_{U0}E[R_{i0}(\boldsymbol{d})-R_{i0}(\mathbf{0})|\boldsymbol{D}=\boldsymbol{d}].$$

By Assumption 6-(ii), we can show that $E[R_{i0}(\mathbf{0})|\boldsymbol{D}=\boldsymbol{d}]=E[R_{i0}(\boldsymbol{d})+Q_{i0}(\boldsymbol{d})|\boldsymbol{D}=\boldsymbol{d}]$. Therefore, the parametric restriction in Assumption 6-(iii) guarantees no-anticipation of the potential outcomes, i.e., $E[Y_{i0}(\boldsymbol{d})|\boldsymbol{D}=\boldsymbol{d}]=E[Y_{i0}(\mathbf{0})|\boldsymbol{D}=\boldsymbol{d}]$ for all $\boldsymbol{d}\in\{0,1\}^N$. Next, Assumption 7-(iii) implies $\Delta Y_i(\mathbf{0})=\Delta\beta_0+\beta_U\sum_j\Delta A_{ij}(0,0)+\Delta\varepsilon_i(0)$. It follows that:

$$E[\Delta Y_i(\mathbf{0})|\boldsymbol{D}=\boldsymbol{d}]=\Delta\beta_0+\beta_U\sum_{j=1,j\neq i}^{N}E[\Delta A_{ij}(0,0)|\boldsymbol{D}=\boldsymbol{d}]+E[\Delta\varepsilon_{i1}(0)|D_i=d_i],$$
$$E[\Delta Y_i(\mathbf{0})|\boldsymbol{D}=\mathbf{0}]=\Delta\beta_0+\beta_U\sum_{j=1,j\neq i}^{N}E[\Delta A_{ij}(0,0)|\boldsymbol{D}=\mathbf{0}]+E[\Delta\varepsilon_{i1}(0)|D_i=0].$$

Therefore, Assumption 7-(i), (ii) imply that the parallel trend holds for the potential outcomes as well.

Similar to the identification procedure under exogenous treatment, in the first step, the

[8] Note that Assumption 6-(i) holds by construction of individual error term that it does not have ATT.

coefficients $\boldsymbol{\zeta}_t = (\zeta_{1t}, \zeta_{2t}, \zeta_{3t}, \zeta_{4t})$ and $\boldsymbol{\xi}$ in dyadic regressions $E[A_{ijt}|D_i, D_j] = \boldsymbol{W}'_{ij}\boldsymbol{\zeta}_t$, and $E[\Delta A_{ij}|D_i, D_j] = \boldsymbol{W}'_{ij}\boldsymbol{\xi}$ are identified when $D_i \neq D_j$ with positive probability. Here, $\boldsymbol{\xi} = \Delta\boldsymbol{\zeta}$ is the difference-in-differences coefficient, and in particular, the second and the third elements identify:

$$\begin{pmatrix}\xi_2\\ \xi_3\end{pmatrix} = \begin{pmatrix}E[\Delta A_{ij}|D_i = 1, D_j = 0] - E[\Delta A_{ij}|D_i = 0, D_j = 0]\\ E[\Delta A_{ij}|D_i = 0, D_j = 1] - E[\Delta A_{ij}|D_i = 0, D_j = 0]\end{pmatrix} = \begin{pmatrix}H(1,0)\\ H(0,1)\end{pmatrix}.$$

The last equation is by Assumptions 6-(ii) and 7-(ii). Moreover, by Assumption 7-(ii), $m(0,1) := E[A_{ij1}(0,0)|D_i = 0, D_j = 1] = E[A_{ij0}(0,0)|D_i = 0, D_j = 1] + E[\Delta A_{ij}(0,0)|D_i = 0, D_j = 0]$, where the first term is $E[A_{ij0}|D_i = 0, D_j = 1] = \zeta_{30} + \zeta_{10}$ by Assumption 6-(ii), and the second term is identified by $\xi_1 = \zeta_{11} - \zeta_{10}$. Therefore, $m(0,1)$ is recovered by $\zeta_{30} + \zeta_{11}$.

Under Assumptions 1-4, 6, and 7, the first-differenced observed outcome is given by

$$\Delta Y_i = \Delta\beta_0 + \beta_{I1}D_i + \beta_{T1}Q_{i1} + \beta_{U1}(R_{i1} - S_{i0}) + \Delta\varepsilon_i = \boldsymbol{X}'_i\boldsymbol{\beta} + \Delta\varepsilon_i, \tag{8}$$

where $S_{i0} = \sum_j A_{ij0}$, $\boldsymbol{X}_i = (1, D_i, Q_{i1}, R_{i1} - S_{i0})$, and $\boldsymbol{\beta} = (\Delta\beta_0, \beta_{I1}, \beta_{T1}, \beta_{U1})$. In this case, because the treatment is not exogenous, the conditional expectation of the last term is not trivially zero. By Assumption 2, $E[\Delta\varepsilon_i|\boldsymbol{D}] = E[\Delta\varepsilon_i|D_i]$. Using Assumptions 6-(i) and 7-(ii), we have $E[\Delta\varepsilon_i|D_i = 1] - E[\Delta\varepsilon_i|D_i = 0] = E[\varepsilon_{i1}(1) - \varepsilon_{i1}(0)|D_i = 1] = 0$, where the last equality is by construction of individual error term. This implies that $E[\Delta\varepsilon_i|D_i] = E[\Delta\varepsilon_i] = 0$. Thus the conditional expectation of (8) can be written as $E[\Delta Y_i|\boldsymbol{D}] = E[\boldsymbol{X}_i|\boldsymbol{D}]'\boldsymbol{\beta}$. Again, $E[\boldsymbol{X}_i|\boldsymbol{D}]$ is estimated by using the first-step parameters $\boldsymbol{\zeta}_t$ and $\boldsymbol{\xi}$. For simplicity, let $\boldsymbol{\zeta} = (\boldsymbol{\zeta}'_1, \boldsymbol{\zeta}'_2)'$. Then, define $E[\boldsymbol{X}_i|\boldsymbol{D}] = \boldsymbol{X}_i(\boldsymbol{\zeta}) := (1, D_i, Q_{i1}(\zeta_1), R_{i1}(\zeta_1) - S_{i0}(\zeta_0))$, where $Q_{i1}(\zeta_1), R_{i1}(\zeta_1), S_{i0}(\zeta_0)$ are those values replacing $E[A_{ijt}|\boldsymbol{D}]$ as $\boldsymbol{W}'_{ij}\boldsymbol{\zeta}_t$. Subsequently, $\boldsymbol{\beta}$ is identified as a coefficient of $E[\Delta Y_i|\boldsymbol{D}] = \boldsymbol{X}_i(\boldsymbol{\zeta})'\boldsymbol{\beta}$.

In the last step, the decomposition $\boldsymbol{\pi}$ is recovered from $\beta_{I1}, \beta_{T1}, \beta_{U1}$, $\xi_2 = H(1,0)$, $\xi_3 = H(0,1)$, and $\zeta_{30} + \zeta_{11} = m(0,1)$. Proposition 2 formally establish identification in this case:

Proposition 2 (Identification with Parallel Trends)**.** *Suppose Assumptions 1-4, 6, and 7 hold. Define* $\boldsymbol{W}_{ij} = (1, D_i, D_j, D_iD_j) \in \mathbb{R}^4$*, and* $\boldsymbol{X}_i(\boldsymbol{\zeta}) = (1, D_i, Q_{i1}(\boldsymbol{\zeta}_1), R_{i1}(\boldsymbol{\zeta}_1) - S_{i0}(\boldsymbol{\zeta}_0)) \in \mathbb{R}^4$*, where* $\boldsymbol{\zeta} = (\boldsymbol{\zeta}'_0, \boldsymbol{\zeta}'_1)'$*,* $Q_{i1}(\boldsymbol{\zeta}_1) = \sum_j(\boldsymbol{W}'_{ij}\boldsymbol{\zeta}_1)D_j$*,* $R_{i1}(\boldsymbol{\zeta}_1) = \sum_j(\boldsymbol{W}'_{ij}\boldsymbol{\zeta}_1)(1 - D_j)$*, and* $S_{i0}(\boldsymbol{\zeta}_0) = \sum_j(\boldsymbol{W}'_{ij}\boldsymbol{\zeta}_0)$*. Then:*

(i) In dyadic regressions $E[A_{ijt}|\boldsymbol{D}] = \boldsymbol{W}'_{ij}\boldsymbol{\zeta}_t$ *for* $t \in \{0,1\}$*, and* $E[\Delta A_{ij}|\boldsymbol{D}] = \boldsymbol{W}'_{ij}\boldsymbol{\xi}$*, the coefficients* $\boldsymbol{\zeta}_t$ *and* $\boldsymbol{\xi}$ *are given by* $\boldsymbol{\zeta}_t = \boldsymbol{B}_{\boldsymbol{W}}^{-1} E[\boldsymbol{W}_{ij} A_{ijt}]$ *and* $\boldsymbol{\xi} = \Delta\boldsymbol{\zeta}$*, provided that* $\boldsymbol{B}_{\boldsymbol{W}} := E[\boldsymbol{W}_{ij}\boldsymbol{W}'_{ij}]$ *is nonsingular (i.e.,* $D_i \neq D_j$ *with probability 1);*

(ii) The conditional expectation of differenced outcome is given by $E[\Delta Y_i|\boldsymbol{D}] = \boldsymbol{X}_i(\boldsymbol{\zeta})'\boldsymbol{\beta}$*, and the coefficient* $\boldsymbol{\beta} = (\Delta\beta_0, \beta_{I1}, \beta_{T1}, \beta_{U1})$ *is given by* $\boldsymbol{\beta} = \boldsymbol{B}_{\boldsymbol{X}}^{-1} E[\boldsymbol{X}_i(\boldsymbol{\zeta})\Delta Y_i]$*, provided that* $\boldsymbol{B}_{\boldsymbol{X}} := E[\boldsymbol{X}_i(\boldsymbol{\zeta})\boldsymbol{X}_i(\boldsymbol{\zeta})']$ *is nonsingular;*

(iii) The decomposition of causal effects is given by $\boldsymbol{\pi} = (\pi^{DT}, \pi^{DN}, \pi^{IT}, \pi^{IN})$ *with* $\pi^{DT} = \beta_{I1}$*,* $\pi^{DN} = (N-1)\beta_{U1}\xi_2$*,* $\pi^{IT} = (\beta_{T1} - \beta_{U1})(\zeta_{30} + \zeta_{11})$*, and* $\pi^{IN} = \beta_{T1}\xi_3$*.*

In this case, the causal effects and their decomposition are interpreted as the average direct/indirect treatment/network effects of treated (ATTs).

Remark 2 (Identification with a Fixed Network)**.** If links are not affected by the treatment, then $H(1,0) = H(0,1) = 0$. 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 3 (Identification without Interactions)**.** If $\beta_T = \beta_U = 0$ in the outcome response model, then there are no indirect effects and the direct network effect, i.e., $\pi^{DN} = \pi^{IT} = \pi^{IN} = 0$. In this scenario, the direct treatment effect is identified by the difference-in-means in the settings of Proposition 1, and the canonical difference-in-differences in the settings of Proposition 2. □

3 Estimation and Inference

In this section, I propose estimators for the parameters identified in Section 2.4 and the decomposition defined in Section 2. Since all identification arguments are constructive, by using conditional expectation, the linear parameters are estimated by least-squares. Hence, the estimation procedure is straightforward but requires three steps. For each estimator, clustered standard errors can be used to conduct inference, taking into account the dependency within groups.

3.1 Estimators

In the first-step, coefficients ($\boldsymbol{\zeta}$ under randomized experiment setting, or $\boldsymbol{\zeta}_1$, $\boldsymbol{\zeta}_2$ under quasi-experiment setting) in dyadic regressions of links are estimated. Subsequently, the outcome coefficient $\boldsymbol{\beta}$ is estimated in the second-step, by using the first-step estimates. Finally, the decomposition of causal effects $\boldsymbol{\pi}$ are estimated in the third-step, by using the estimates in the first two steps.

3.1.1 First-Step 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 both 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-Step Estimators

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

The coefficient of the outcome regression $E[Y_{ig}|\boldsymbol{D}] = \boldsymbol{Z}_{ig}(\hat{\boldsymbol{\zeta}})'\boldsymbol{\beta}$ 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-step 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 $\boldsymbol{X}_{ig}(\hat{\boldsymbol{\zeta}}) = (1, D_{i1g}, Q_{ig}(\hat{\boldsymbol{\zeta}}_1), R_{i1g}(\hat{\boldsymbol{\zeta}}_1) - S_{i0g}(\hat{\boldsymbol{\zeta}}_0))$ for each individual i in group g. The coefficient of the outcome regression $E[\Delta Y_{ig}|\boldsymbol{D}] = \boldsymbol{X}_{ig}(\hat{\boldsymbol{\zeta}})'\boldsymbol{\beta}$ 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: $\hat{\boldsymbol{\pi}} = \begin{pmatrix}\hat{\beta}_I & (N-1)\hat{\beta}_4\hat{\xi}_2 & (\hat{\beta}_T - \hat{\beta}_U)\hat{\xi}_1 & \hat{\beta}_T\hat{\xi}_3\end{pmatrix}$

Quasi-Experiment Setting: $\hat{\boldsymbol{\pi}} = \begin{pmatrix}\hat{\beta}_I & (N-1)\hat{\beta}_U\hat{\xi}_2 & (\hat{\beta}_T - \hat{\beta}_U)(\hat{\xi}_{30} + \hat{\xi}_{11}) & \hat{\beta}_T\hat{\xi}_3\end{pmatrix}$

Remark 4 (Estimation with Covariates)**.** If all the identifying assumptions hold conditional on a set of covariates, the inverse probability weighting method proposed by Abadie (2005) can be applied in identification arguments. In this case, estimators using the corresponding propensity score approach can also be applied.□

3.2 Inference

Since the proposed estimators are least squares estimators for projection coefficients, standard large sample theory can be applied. In this section, I consider large number of independent groups for the asymptotic properties. However, when the underlying network is sparse, we

can obtain the same result considering both large number of groups or individuals with additional restrictions. 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 step are asymptotically normal.

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

$$\hat{V}_{\boldsymbol{b}}^{-1/2}\sqrt{G}(\hat{\boldsymbol{b}} - \boldsymbol{b}^\star) \xrightarrow{d} N(\boldsymbol{0}, \boldsymbol{I}),$$

for $\boldsymbol{b} \in \{\boldsymbol{\zeta}, \boldsymbol{\beta}, \boldsymbol{\pi}\}$, where $\hat{V}_{\boldsymbol{b}}$ is a plug-in estimator of the asymptotic variance of $\boldsymbol{b}$. Let $\mathcal{V}_g = \{(D_{ig}, Y_{ig}), (A_{ijg}, \boldsymbol{W}_{ijg}) : \forall i, \forall (i,j)\}$ be group-level data. The influence functions of $\boldsymbol{\zeta}$, $\boldsymbol{\beta}$ are given by

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

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

$$\begin{pmatrix}
\psi_{\boldsymbol{\beta},2}(\mathcal{V}_g, \boldsymbol{\zeta}, \boldsymbol{\beta}) \\
(N-1)\psi_{\boldsymbol{\beta},4}(\mathcal{V}_g, \boldsymbol{\zeta}, \boldsymbol{\beta})\zeta_2^\star + \beta_4^\star\psi_{\boldsymbol{\zeta},2}(\mathcal{V}_g, \boldsymbol{\zeta}) \\
\left(\psi_{\boldsymbol{\beta},3}(\mathcal{V}_g, \boldsymbol{\zeta}, \boldsymbol{\beta}) - \psi_{\boldsymbol{\beta},4}(\mathcal{V}_g, \boldsymbol{\zeta}, \boldsymbol{\beta})\right)\zeta_1^\star + (\beta_T^\star - \beta_U^\star)\psi_{\boldsymbol{\zeta},1}(\mathcal{V}_g, \boldsymbol{\zeta}) \\
\psi_{\boldsymbol{\beta},3}(\mathcal{V}_g, \boldsymbol{\zeta}, \boldsymbol{\beta})\zeta_3^\star + \beta_3^\star\psi_{\boldsymbol{\zeta},3}(\mathcal{W}_g, \boldsymbol{\zeta})
\end{pmatrix},$$

where $\psi_{\boldsymbol{b},k}(\cdot)$ denote k-th element in vector $\psi_{\boldsymbol{b}}(\cdot)$ for $\boldsymbol{b} \in \{\boldsymbol{\zeta}, \boldsymbol{\beta}, \boldsymbol{\pi}\}$. Lastly, $\hat{V}_{\boldsymbol{\zeta}}$, $\hat{V}_{\boldsymbol{\beta}}$, $\hat{V}_{\boldsymbol{\pi}}$ are sample variance-covariance matrices of $\psi_{\boldsymbol{\zeta}}(\mathcal{V}_g, \hat{\boldsymbol{\zeta}})$, $\psi_{\boldsymbol{\beta}}(\mathcal{V}_g, \hat{\boldsymbol{\zeta}}, \hat{\boldsymbol{\beta}})$, $\psi_{\boldsymbol{\pi}}(\mathcal{V}_g, \hat{\boldsymbol{\zeta}}, \hat{\boldsymbol{\beta}})$, 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 asymp-

totic variance $E[\psi_{\boldsymbol{\pi}}(\mathcal{V}_g, \boldsymbol{\zeta}^\star, \boldsymbol{\beta}^\star)\psi_{\boldsymbol{\pi}}(\mathcal{V}_g, \boldsymbol{\zeta}^\star, \boldsymbol{\beta}^\star)']$. The plug-in standard errors are computed as the square root of the diagonal elements of the variance of the empirical influence function $\psi_{\boldsymbol{\pi}}(\mathcal{V}_g, \hat{\boldsymbol{\zeta}}, \hat{\boldsymbol{\beta}})$, 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 $\psi_{\boldsymbol{\zeta}}(\mathcal{V}_g, \boldsymbol{\zeta}^\star)$, and $\psi_{\boldsymbol{\beta}}(\mathcal{V}_g, \boldsymbol{\zeta}^\star, \boldsymbol{\beta}^\star)$, respectively. The plug-in standard errors for $\boldsymbol{\beta}$ and $\boldsymbol{\zeta}$ are defined similarly.

The limiting distribution of estimators in a quasi-experiment setting with parallel trends can be similarly established, and as summarized in Proposition 4:

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

$$\hat{V}_{\boldsymbol{b}}^{-1/2}\sqrt{G}(\hat{\boldsymbol{b}} - \boldsymbol{b}^\star) \xrightarrow{d} N(0,1),$$

where $\boldsymbol{b} \in \{\boldsymbol{\zeta}_0, \boldsymbol{\zeta}_1, \boldsymbol{\xi}, \boldsymbol{\beta}, \boldsymbol{\pi}\}$. Let $\mathcal{W}_g := \{(A_{ij0g}, A_{ij1g}, \boldsymbol{W}_{ijg}), (D_{ig}, Y_{i0g}, Y_{i1g}) : \forall i, \forall (i,j)\}$ be group-level data, and denote $\boldsymbol{\zeta} = (\boldsymbol{\zeta}_0', \boldsymbol{\zeta}_1')'$. The influence functions of $\boldsymbol{\zeta}_t$, $\boldsymbol{\xi}$, $\boldsymbol{\beta}$ are given by

$$\begin{aligned}
\psi_{\boldsymbol{\zeta}_t}(\mathcal{V}_g, \boldsymbol{\zeta}_t) &:= \boldsymbol{B}_{\boldsymbol{W}}^{-1} \frac{1}{N(N-1)} \sum_{(i,j):i\neq j} \boldsymbol{W}_{ijg}(A_{ijtg} - \boldsymbol{W}_{ijg}'\boldsymbol{\zeta}_t), \\
\psi_{\boldsymbol{\xi}}(\mathcal{V}_g, \boldsymbol{\xi}) &:= \boldsymbol{B}_{\boldsymbol{W}}^{-1} \frac{1}{N(N-1)} \sum_{(i,j):i\neq j} \boldsymbol{W}_{ijg}(\Delta A_{ijg} - \boldsymbol{W}_{ijg}'\boldsymbol{\xi}), \\
\psi_{\boldsymbol{\beta}}(\mathcal{V}_g, \boldsymbol{\zeta}, \boldsymbol{\beta}) &:= \boldsymbol{B}_{\boldsymbol{Z}}^{-1} \left[\boldsymbol{Z}_{ig}(\boldsymbol{\zeta})(\Delta Y_{ig} - \boldsymbol{Z}_{ig}(\boldsymbol{\zeta})'\boldsymbol{\beta}) - \boldsymbol{C}_{\boldsymbol{\zeta},1}\psi_{\boldsymbol{\zeta},1}(\mathcal{V}_g, \boldsymbol{\zeta}_1) - \boldsymbol{C}_{\boldsymbol{\zeta},2}\psi_{\boldsymbol{\zeta},2}(\mathcal{V}_g, \boldsymbol{\zeta}_2)\right],
\end{aligned}$$

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

$$\begin{pmatrix}
\psi_{\boldsymbol{\beta},2}(\mathcal{V}_g, \boldsymbol{\zeta}, \boldsymbol{\beta}) \\
(N-1)\psi_{\boldsymbol{\beta},4}(\mathcal{V}_g, \boldsymbol{\zeta}, \boldsymbol{\beta})\xi_2^\star + \beta_4^\star\psi_{\boldsymbol{\xi},2}(\mathcal{V}_g, \boldsymbol{\xi}) \\
\left(\psi_{\boldsymbol{\beta},3}(\mathcal{V}_g, \boldsymbol{\zeta}, \boldsymbol{\beta}) - \psi_{\boldsymbol{\beta},4}(\mathcal{V}_g, \boldsymbol{\zeta}, \boldsymbol{\beta})\right)(\zeta_{30}^\star + \zeta_{11}^\star) + (\beta_3^\star - \beta_4^\star)\left(\psi_{\boldsymbol{\zeta}_0,3}(\mathcal{V}_g, \boldsymbol{\zeta}) + \psi_{\boldsymbol{\zeta}_1,1}(\mathcal{V}_g, \boldsymbol{\zeta})\right) \\
\psi_{\boldsymbol{\beta},3}(\mathcal{V}_g, \boldsymbol{\zeta}, \boldsymbol{\beta})\xi_3^\star + \beta_3^\star\psi_{\boldsymbol{\xi},3}(\mathcal{V}_g, \boldsymbol{\xi})
\end{pmatrix},$$

where $\psi_{\boldsymbol{b},k}(\cdot)$ denote k-th element in vector $\psi_{\boldsymbol{b}}(\cdot)$ for $\boldsymbol{b} \in \{\boldsymbol{\zeta}, \boldsymbol{\beta}, \boldsymbol{\pi}\}$. Lastly, $\hat{V}_{\boldsymbol{\zeta},t}, \hat{V}_{\boldsymbol{\xi}}, \hat{V}_{\boldsymbol{\beta}}, \hat{V}_{\boldsymbol{\pi}}$ are sample variance matrices of $\psi_{\boldsymbol{\zeta}_t}(\mathcal{V}_g, \hat{\boldsymbol{\zeta}}_t)$, $\psi_{\boldsymbol{\xi}}(\mathcal{V}_g, \hat{\boldsymbol{\xi}})$, $\psi_{\boldsymbol{\beta}}(\mathcal{V}_g, \hat{\boldsymbol{\zeta}}, \hat{\boldsymbol{\beta}})$, $\psi_{\boldsymbol{\pi}}(\mathcal{V}_g, \hat{\boldsymbol{\zeta}}, \hat{\boldsymbol{\beta}})$, 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 $\boldsymbol{\pi}$. Once again, inference based on plug-in clustered standard errors remains asymptotically valid.

4 A Monte Carlo Study

To examine the finite sample performance 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 single index. The potential link of pair (i,j), given treatment statuses $(D_i, D_j) = (d,e)$ is generated by a binary response $A_{ij}(d,e) = \mathbb{1}\{I_{\boldsymbol{\theta}}(d,e) \geq u_{ij}\}$, where u_{ij} follows a standard normal distribution. Let $\Phi(\cdot)$ be the cdf of standard normal distribution. The mean and the average treatment effect (ATE) on links are given by $\bar{A}(d,e) := E[A_{ij}(d,e)] = \Phi(I_\theta(d,e))$, $m(0,1) := E[A_{ij}(0,0)] = \bar{A}(0,0)$, and $H(d,e) = E[A_{ij}(d,e) - A_{ij}(0,0)] = \Phi(I_{\boldsymbol{\theta}}(d,e)) - \Phi(\theta_1)$, respectively. Let $\bar{\boldsymbol{A}} = (\bar{A}(1,1), \bar{A}(1,0), \bar{A}(0,1), \bar{A}(0,0))'$ be a vector of potential means. The coefficient of dyadic regression for the observed link $A_{ij} = A_{ij}(D_i, D_j)$ on $\boldsymbol{W}_{ij} = (1, D_i, D_j, D_i D_j)$ is then given by $\boldsymbol{\zeta} = \boldsymbol{M}\bar{\boldsymbol{A}}$, 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: $Y_i = \beta_0 + \beta_I D_i + \beta_T Q_i + \beta_U R_i + \varepsilon_i$, where

$Q_i = \sum_{j \neq i} A_{ij} D_j$, $R_i = \sum_{j \neq i} A_{ij}(1 - D_j)$, $\varepsilon_i = e_i + \sum_{j \neq i} u_{ij}$, and e_i is the standard normal error term. The individual error term ε_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 $E[\varepsilon_i | \boldsymbol{A}, \boldsymbol{D}] \neq 0$. I assess the bias of this model assuming exogenous network in Table 3. 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 $h_t(d, e)$ be a symmetric index 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 $A_{ij0} = \mathbb{1}\{h_0(D_i, D_j) \geq u_{ij0}\}$ at $t = 0$, and $A_{ij1} = \mathbb{1}\{I_{\boldsymbol{\theta}}(d, e) + h_1(D_i, D_j) \geq u_{ij1}\}$ at $t = 1$, where u_{ijt} are standard normal error terms. Thus, the potential links are not independent of the treatment by construction. The conditional means at $t = 0$ is given by $E[A_{ij0}(d, e) | D_i = d', D_j = e'] = \Phi(h_0(d', e'))$, and hence Assumption 6-(ii) holds. And those at $t = 0$ is $E[A_{ij1}(d, e) | D_i = d', D_j = e'] = \Phi(I_{\boldsymbol{\theta}}(d, e) + h_1(d', e'))$. Therefore, we have $m(0, 1) = \Phi(I_{\boldsymbol{\theta}}(0, 0) + h_1(0, 1))$, and the average treatment effect on treated (ATT) on links is: $H(d, e) = \Phi(I_{\boldsymbol{\theta}}(d, e) + h_1(d, e)) - \Phi(I_{\boldsymbol{\theta}}(0, 0) + h_1(d, e))$.

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:[9]

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

Specifically, I set $h_0(d, e) = I_\omega(d, e)$ 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 $\bar{\boldsymbol{A}}_t = (\bar{A}_t(1, 1), \bar{A}_t(1, 0), \bar{A}_t(0, 1), \bar{A}_t(0, 0))'$, and $\bar{A}_t(d, e) = E[A_{ijt} | D_i = d, D_j = e]$.

At $t = 1$, the outcome is generated by the same way as in the Design 1, and at $t = 0$, the outcome is generated by $Y_{i0} = \alpha + \beta_U S_{i0} + \varepsilon_{i0}$, where $S_{i0} = \sum_{j \neq i} A_{ij0}$ is the number of neighbors in the pre-treatment period, $\varepsilon_{i0} = e_{i0} + \sum_{j \neq i} u_{ij0}$, and $e_{i0} \sim N(0, 1)$. Since the individual error term contains dyadic error terms related to potential links, Q_{i1}, R_{i1}, S_{i0} are endogenous regres-

[9] 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 $h_1(d', e') := \Phi^{-1}(T + \Phi(h(d', e'))) - \theta_1$.

sors. 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 $\{(D_{ig}, Y_{i0g}, Y_{i1g})\}_{i,g}$. Table 1 shows the true parameter values and corresponding true values of decomposition.

Table 1: True Parameter Values

	Design 1	Design 2
Outcome coefficients	$\beta = (2, 1, 0.8, 0.6)$	$\beta_1 = (2, 1, 0.8, 0.6)$ $\beta_0 = (-1, 0, 0.6, 0.6)$
Link Formation	$\theta = (-1, 0.1, 0.1, 1)$	$\theta = (-1, 0.1, 0.1, 1)$ $\omega = (-1.5, 0.3, 0.3, -1)$
Distribution of Links	$\zeta = (2, 1, 0.8, 0.6)$	$\boldsymbol{\zeta}_1 = (0.067, 0.069, 0.069, 0.037)$ $\boldsymbol{\zeta}_0 = (0.067, 0.048, 0.048, -0.135)$ $\boldsymbol{\xi} = (0, 0.021, 0.021, 0.172)$
Decomposition	$\boldsymbol{\pi} = (1, 0.290, 0.032, 0.020)$	$\boldsymbol{\pi} = (1, 0.235, 0.013, 0.016)$

The estimators proposed in Section 3 are computed with clustered standard errors. 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 2 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 theory and confirms that the clustered standard errors perform well. For the outcome coefficient $\boldsymbol{\beta}$, see Table B.2, and for the dyadic coefficients $\boldsymbol{\zeta}, \boldsymbol{\zeta}_1, \boldsymbol{\xi}$, see Table B.1 in Appendix B.

As discussed in Section 2, if the network is exogenously given, the outcome coefficient can be consistently estimated using an outcome regression based on the observed network. However, because the data-generating process introduces a correlation between links and the outcome, assuming an exogenous network leads to bias. Table 3 compares the outcome coefficients from the proposed estimation method with those assuming an exogenous network, as well as estimates that ignore interference. For each design in Table 3, the second column

(Exo.) presents the simulation results of the coefficient estimates from regressing the outcome on regressors based on the observed network (i.e., Q_{i1}, R_{i1}, or $R_{i1} - S_{i0}$), which are biased. The third column (Indep.) shows the simulation results when we ignore all spillover effects, i.e., estimates from regressing the outcome on the intercept and the individual treatment D_i only. These comparisons demonstrate that ignoring potential endogeneity from the causal impact on the network, as well as spillover effects, leads to significantly biased estimates.

Table 2: 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}
Median								
50	1.0011	0.2597	0.0312	0.0197	1.0052	0.1794	0.0136	0.0148
100	1.0026	0.2725	0.032	0.0199	1.0045	0.2116	0.0129	0.0156
200	0.9998	0.2814	0.0311	0.0202	0.9969	0.2214	0.0132	0.016
400	0.9995	0.2851	0.0319	0.0203	1.0058	0.226	0.0136	0.0163
800	1.0033	0.2862	0.032	0.0203	0.9992	0.2324	0.0132	0.0164
TRUE	1	0.2896	0.0317	0.0203	1	0.2348	0.0134	0.0165
MSE								
50	0.6166	0.0903	0.0027	0.0001	1.6011	0.2159	0.002	0.0002
100	0.3012	0.0413	0.0014	0.0001	0.7761	0.0908	0.001	0.0001
200	0.1463	0.0192	0.0007	<0.0001	0.384	0.0406	0.0005	<0.0001
400	0.073	0.0092	0.0003	<0.0001	0.1869	0.0187	0.0002	<0.0001
800	0.0358	0.0046	0.0002	<0.0001	0.0897	0.0089	0.0001	<0.0001
Coverage Rate								
50	0.9363	0.9042	0.9311	0.932	0.9436	0.9335	0.9342	0.8982
100	0.9394	0.9154	0.9365	0.9391	0.9429	0.9243	0.9342	0.9209
200	0.9433	0.9334	0.9443	0.9433	0.943	0.9346	0.9429	0.9304
400	0.9428	0.9383	0.9469	0.9415	0.9443	0.941	0.9465	0.9381
800	0.9492	0.9437	0.9506	0.9487	0.9506	0.9441	0.9487	0.9441

Notes: This table presents the simulation results for $B = 10,000$ 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.

Table 3: Bias When Assuming Exogenous Network (Design 1)

	Design 1			Design 2		
G	β	Exo.	Indep.	β	Exo.	Indep.
MAE						
50	0.4821	1.7681	2.7246	0.6544	0.9526	0.8959
100	0.3399	1.7658	2.7249	0.4603	0.9406	0.8959
200	0.2377	1.7655	2.7251	0.3242	0.9323	0.8959
400	0.1665	1.765	2.7257	0.2259	0.9267	0.8964
800	0.1181	1.7651	2.7258	0.1581	0.9232	0.8966
MSE						
50	1.8748	16.2161	15.2326	3.1601	5.1218	2.2042
100	0.9221	16.1391	15.2213	1.5362	5.0755	2.1751
200	0.4503	16.1128	15.2174	0.7585	5.0521	2.1659
400	0.2229	16.0941	15.2187	0.3688	5.0381	2.1566
800	0.1104	16.0925	15.2189	0.1789	5.0357	2.1551

Notes: This table presents the simulation results based on $B = 10,000$ replications. Column G represents the number of independent groups, with each group consisting of $N = 20$ individuals. The first panel reports the mean absolute errors across all replications, and the second panel shows mean squared errors. Columns labeled β present the results from the proposed estimation method. Columns labeled "Exo." display the results from regressing Y_i on $1, D_i, Q_i, R_i$ for design 1, and from regressing ΔY_i on $1, D_i, Q_{i1}, R_{i1} - S_{i0}$ for design 2. Columns labeled "Indep." present the results from regressing Y_i on $1, D_i$ for each design.

5 Empirical Illustration

In this section, I apply the proposed method to data from a randomized experiment conducted by Comola and Prina (2020). The experiment took place in villages surrounding Pokhara, Nepal, from 2009 to 2011, and involved providing households with access to savings accounts. The pre-treatment survey was conducted in February 2009, and the treatment was randomly assigned to half of the households in June 2010 through a public lottery.

As reported by Prina (2015), formal banking services in Nepal are 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 sample consists of 915 households across 19 villages, with detailed information on financial networks. The network 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 interactions occur within villages, resulting in a total of 56,308 dyads.

The outcome variable of interest in Comola and Prina (2021) is household meat consumption, which is considered a luxury good in these areas and serves as a proxy for wealth or conspicuous consumption. Given the potential influence of peer consumption, it is reasonable to expect social interactions to affect meat consumption. Indeed, the study found positive direct and indirect effects of savings account access on meat consumption. Table 4 shows the distribution of meat consumption at the post-treatment period.

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 calculate the direct and indirect effects as the derivatives of the reduced-form outcome equation, which is derived from the linear-in-means structure. These derivatives account for changes in links in response to the treatment, corresponding to the average treatment effects on row-normalized links. The authors estimate this effect by regressing first-differenced row-normalized links on indicator of *some treated*, i.e., $\max\{D_i, D_j\}$.

Estimating average treatment effect on link is similar to the method proposed in this paper. Table 5 presents the first-step regression results. The first column replicates the dyadic regression from Comola and Prina (2021), where the coefficient for the dyadic treatment is estimated to be 0.002, indicating that the average increase in row-normalized link is 0.002 percentage points in response to the treatment. The second column shows the results from a dyadic regression of row-normalized links on the dyadic regressors $\boldsymbol{W}_{ij} = (1, D_i, D_j, D_i D_j)$, again showing an average causal effect on row-normalized links of approximately 0.002. The interpretation becomes clearer in the third column, which reports estimates from the same dyadic regression using raw link (non-normalized). Here, the results indicate that when an individual is treated, the probability of forming a link increases by 0.39%p. This estimate is

Table 4: Distribution of outcome across villages

village	N	Mean	Std.Dev	Min	Med	Max
Bindabasini	36	1078.3	1064.8	0	1000	4000
Chorepatan	60	1708.2	2494.5	0	940	10800
Chorsangu	82	1034.7	1382.6	0	520	6800
GONESA	12	1070	983.6	0	780	3400
Hanuman Tole	74	913.4	1144.3	0	910	8000
Hemja	61	888.5	1168.3	0	520	4500
Kotre	64	1125.6	1403.1	0	560	7120
Kranti tole	119	917	1037.4	0	920	4160
Lower Goste	28	540	722.4	0	0	2000
Mahat Gaunda	47	988.5	1249.2	0	960	5000
Miyapatan	25	1100	938.2	0	1040	3360
Nagintole	48	913.3	1160.6	0	480	5000
Paropakar	51	1203.5	1060.9	0	1040	4080
Pragati Tole	26	1213.8	1591.6	0	920	7000
Rato Pahira	26	1455.4	1993.4	0	960	9200
Sarankot	74	1120	1352.6	0	920	5000
Tutunga	38	901.1	905.8	0	560	3920
Upper Goste	11	1118.2	1139.1	0	1200	3000
Yamdi	33	984.2	1112.8	0	520	4400
Total	915	1057.4	1345.2	0	700	10800

Notes: This table presents the descriptive statistics of the outcome at post-treatment period over 19 villages.

statistically significant, suggesting that the treatment has a causal impact on the formation of network links.

In Comola and Prina (2021), the direct and indirect effects are computed as derivatives of the reduced-form outcome with respect to the treatment vector, i.e., $\partial E[\mathbf{y}|\mathbf{D}]/\partial D_k$. Direct effects represent the average partial effect of one's own treatment, while indirect effects capture the average partial effect of others' treatments. The authors estimate the direct effect at 342.3 and the indirect effect at 260.9. However, causal interpretation of these estimate is not straightforward without further assumptions. In particular, it is valid only when potential outcome is additively separable with respect to others' treatment statuses. Additionally, both direct and indirect effects are mixed effects of treatment and network, which are difficult to disentangle in their method.

Table 6 presents the direct and indirect treatment and network effects as proposed in this

Table 5: Average Treatment Effects on Treated of Links

Var	ΔA^s	ΔA^s	ΔA
Constant	-0.0009	-0.0009	-0.0031
	(0.0012)	(0.0012)	(0.0024)
Some Treated	0.0021		
	(0.0016)		
D_i		0.0021	0.0039*
		(0.002)	(0.0023)
D_j		0.0023	0.0039*
		(0.0018)	(0.0023)
$D_i \times D_j$		-0.0025	-0.0034
		(0.003)	(0.0034)
Observations		56,308	

Notes: 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.

paper. Village fixed effects are included to account for variations in the outcome distribution across villages as shown in Table 4. The dependent variable in the first two columns (M1, M2) is Y_1, while in the last two columns (M3, M4), it is $\log(Y_1)$. Model (M2) additionally controls for a dummy variable $\mathbb{1}\{Y_1 = 0\}$, to account for individuals who do not consume meat. In model (M3), the dependent variable is set to zero for individuals with $Y_1 = 0$, while also controlling for the dummy variable $\mathbb{1}\{Y_1 = 0\}$. In model (M4), observations with $Y_1 = 0$ are dropped entirely. For models (M3) and (M4), the coefficients are adjusted by multiplying them by the mean of Y_1 (1,057.43) to enable comparison with the first two columns.

I find positive total direct and indirect effects. The direct treatment effects are significantly estimated, while the direct network effects are not significant. Furthermore, the direct network effects are negative, resulting in an insignificant total direct effect. In contrast, the indirect network effects are relatively small, but both the indirect treatment effect and the total indirect effect are significantly positive.

Although the direct network effects are not significant, their direction is opposite to that of the direct treatment effects. This suggests there may be an opposing effect arising from changes in the network structure. Specifically, opening a savings account appears to directly increase consumption, but it may also reduce consumption by altering the network. As shown in Table 6, since the treatment increases the probability of forming new links, this could be

Table 6: Decomposition of Treatment Effects

	M1		M2		M3		M3	
	Direct	Indirect	Direct	Indirect	Direct	Indirect	Direct	Indirect
Treatment	240.7**	185***	207.4**	215.4***	211.8***	138.4***	275.2***	195.6***
	(115.8)	(38.3)	(99.3)	(44.8)	(54.7)	(28.9)	(52.3)	(41.2)
Network	-169.9	1.6	-202.4	1.6	-135	0.8	-188.7	1.2
	(385.7)	(3.6)	(459.9)	(3.7)	(306.8)	(1.8)	(428.8)	(2.8)
Total	70.9	186.6***	4.9	217.1***	76.8	139.2***	86.5	196.8***
	(380.1)	(36.4)	(445)	(43)	(292.7)	(28.1)	(415.1)	(39.9)
Obs.	915		915		915		612	
R^2	0.40		0.58		0.98		0.99	

Notes: The dependent variable in the first two columns (M1, M2) is Y_1. In model (M2) a dummy variable $\mathbb{1}\{Y_1 = 0\}$ is used to control individuals who do not consume meat. In model (M3) the dependent variable is $\log(Y_1)$ for $Y_1 > 0$, set to zero for $Y_1 = 0$, and controls a dummy variable $\mathbb{1}\{Y_1 = 0\}$. In model (M4) the dependent variable is $\log(Y_1)$ and drop the observations with $Y_1 = 0$. The coefficients in columns (M3) and (M4) are adjusted by multiplying by the mean of Y_1 (1,057.43) to allow for comparison with the first two columns. Village fixed effects are included to account for variations in meat consumption across different villages. 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.

interpreted as a spillover effect, with savings behavior spreading through the network.

Overall, the results in this section demonstrate that the method proposed in this paper can effectively decompose causal effects into pure treatment effects and those driven by causal changes in the network. This approach is particularly useful when the treatment and network influence the outcome in opposite directions, as seen in the direct effects presented in Table 6.

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

method 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, this study assumes the availability of full network information, which is often unavailable. Future work would 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". *The Review of Economic Studies* 72.1, pp. 1–19.

Ananth, Abhishek (2021). "Optimal treatment assignment rules on networked populations". *working paper*.

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

Auerbach, Eric, Yong Cai, and Ahnaf Rafi (2024). "Regression Discontinuity Design with Spillovers". *arXiv preprint arXiv:2404.06471*.

Auerbach, Eric and Max Tabord-Meehan (2021). "The local approach to causal inference under network interference". *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". *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". *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". *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". *Journal of Econometrics* 150.1, pp. 41–55.

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

Cai, Jing, Alain De Janvry, and Elisabeth Sadoulet (2015). "Social networks and the decision to insure". *American Economic Journal: Applied Economics* 7.2, pp. 81–108.

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". *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". *Journal of Econometrics* 235.2, pp. 1589–1624.

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". *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". *Journal of the American Statistical Association* 116.534, pp. 901–918.

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

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

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

Hoshino, Tadao and Takahide Yanagi (2023). "Causal inference with noncompliance and unknown interference". *Journal of the American Statistical Association*, pp. 1–12.

Hudgens, Michael G and M Elizabeth Halloran (2008). "Toward causal inference with interference". *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". *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". *arXiv preprint arXiv:1609.04464*.

Kelejian, Harry H and Gianfranco Piras (2014). "Estimation of spatial models with endogenous weighting matrices, and an application to a demand model for cigarettes". *Regional Science and Urban Economics* 46, pp. 140–149.

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

König, Michael D, Xiaodong Liu, and Yves Zenou (2019). "R&D networks: Theory, empirics, and policy implications". *Review of Economics and Statistics* 101.3, pp. 476–491.

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

Lee, Lung-Fei, Xiaodong Liu, Eleonora Patacchini, and Yves Zenou (2021). "Who is the key player? A network analysis of juvenile delinquency". *Journal of Business & Economic Statistics* 39.3, pp. 849–857.

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

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

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

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

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

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

Sánchez-Becerra, Alejandro (2021). "Spillovers, homophily, and selection into treatment: the network propensity score". *Job market paper*.

Vazquez-Bare, Gonzalo (2023a). "Causal spillover effects using instrumental variables". *Journal of the American Statistical Association* 118.543, pp. 1911–1922.

— (2023b). "Identification and estimation of spillover effects in randomized experiments". *Journal of Econometrics* 237.1, p. 105237.

Viviano, Davide (2024). "Policy targeting under network interference". *Review of Economic Studies*, pp. 1–36.

Xu, Ruonan (2023). "Difference-in-differences with Interference: a finite population perspective". *arXiv preprint arXiv:2306.12003*.

Appendix

A Proofs

A.1 Proofs of Propositions in Section 2.4

Proof of Proposition 1. Recall that $Q_i = \sum_{j\neq i} A_{ij}D_j$ and $R_i = \sum_{j\neq i} A_{ij}(1-D_j)$. Assumption 1 implies:

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

By taking conditional expectation on the observed outcome (3), 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 Assumptions 2, and 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

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

Therefore, $\zeta_2 = \bar{A}(1,0) - \bar{A}(0,0) = H(1,0)$, and $\zeta_3 = \bar{A}(0,1) - \bar{A}(0,0) = H(0,1)$. Moreover, $\zeta_1 = E[A_{ij}(0,0)] = E[A_{ij}(0,0)|D_i = 0, D_j = 1] = m(0,1)$. 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 (A.1), 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 Assumptions 2, 6, and 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

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

Note that $E[\Delta\varepsilon_i(0)|D_i = 1] = E[\Delta\varepsilon_i(0)|D_i = 0]$ by Assumption 7, and it implies $E[\varepsilon_{i1}(0)|D_i = 1] = E[\varepsilon_{i0}(0)|D_i = 1] + E[\Delta\varepsilon_i|D_i = 0]$. Here, the first term is $E[\varepsilon_{i0}(0)|D_i = 1] = E[\varepsilon_{i0}|D_i = 1]$ by Assumption 6, and therefore $E[\varepsilon_{i1}(1) - \varepsilon_{i1}(0)|D_i = 1] = E[\varepsilon_{i1}|D_i = 1] - E[\varepsilon_{i0}|D_i = 1] - E[\Delta\varepsilon_i|D_i = 0]$. It follows that

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

by construction of the individual error term in Assumption 2. Therefore, (i) and (iii) follow by the standard identification result of least squares estimator of projection coefficients. Next, note that

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

where $\bar{A}_t(d,e) = E[A_{ijt}|D_i = d, D_j = e]$. The third equation is by Assumption 6-(ii) and Assumption 7-(ii). Therefore, because $\zeta_{2t} = \bar{A}_t(1,0) - \bar{A}_t(0,0)$, we have $\xi_2 = H(1,0)$ and $\xi_3 = H(0,1)$. Moreover,

$$\begin{aligned} m(0,1) &:= E[A_{ij1}(0,0)|D_i = 0, D_j = 1] \\ &= E[A_{ij1}(0,1)|D_i = 0, D_j = 1] - H(0,1) = \zeta_{31} + \zeta_{11} - \xi_3 = \zeta_{30} + \zeta_{11}. \end{aligned}$$

As a result, the decomposition in Section 2 can be recovered by using $\boldsymbol{\zeta} = (\boldsymbol{\zeta}_1, \boldsymbol{\zeta}_2)$ and $\boldsymbol{\beta}$. □

A.2 Proofs of Propositions in Section 3

The following lemma is a version of Lemma 4.3 in Newey and McFadden (1994), and is used to prove Propositions 3, and 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 $E\left[\sup_{\phi \in \mathcal{N}} \left\|\ell(\boldsymbol{V}_g, \phi)\right\|\right] < \infty$. 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 theorem, 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 $\hat{\zeta}$

Note that $\left\|\boldsymbol{W}_{ijg}\right\|^2 = 1 + D_{ig} + D_{jg} + D_{ig}D_{jg} \leq 4$, and $|\boldsymbol{W}'_{ijg}\boldsymbol{\zeta}| = |\zeta_1 + D_{ig}\zeta_2 + D_{jg}\zeta_3 + D_{ig}D_{jg}\zeta_4| \leq |\zeta_1 + \zeta_2 + \zeta_3 + \zeta_4| = \bar{A}(1,1) \in [0,1]$, because

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

where $\bar{A}(d,e) = E[A_{ij}|D_i = d, D_j = e]$. 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

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

as $G\to\infty$, where $\boldsymbol{\Sigma}_{\zeta}:=\mathrm{Var}\left(\frac{1}{N(N-1)}\sum_{(i,j):i\neq j}\boldsymbol{W}_{ijg}\omega_{ijg}\right)$. 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 $\boldsymbol{\Sigma}_{\zeta}$ of last limiting distribution exists because

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

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

$$\sqrt{G}(\hat{\boldsymbol{\zeta}}-\boldsymbol{\zeta}^{\star})=\frac{1}{\sqrt{G}}\sum_{g=1}^{G}\psi_{\zeta}(\mathcal{V}_g,\boldsymbol{\zeta}^{\star})+o_p(1)\xrightarrow{d}N(0,\boldsymbol{B}_{\boldsymbol{W}}^{-1}\boldsymbol{\Sigma}_{\zeta}\boldsymbol{B}_{\boldsymbol{W}}^{-1}),\tag{A.2}$$

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

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

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

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

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

$$\begin{aligned}
\left\|\nabla_\zeta \mathbf{Z}_{ig}(\boldsymbol{\zeta})\right\|^2 &= \operatorname{tr}\left(\sum_{j\neq i} \boldsymbol{W}_{ijg} D_{jg} \sum_{j\neq i} \boldsymbol{W}'_{ijg} D_{jg} + \sum_{j\neq i} \boldsymbol{W}_{ijg}(1-D_{jg}) \sum_{j\neq i} \boldsymbol{W}'_{ijg}(1-D_{jg})\right) \\
&= \sum_{j\neq i} \boldsymbol{W}'_{ijg} D_{jg} \sum_{j\neq i} \boldsymbol{W}_{ijg} D_{jg} + \sum_{j\neq i} \boldsymbol{W}'_{ijg}(1-D_{jg}) \sum_{j\neq i} \boldsymbol{W}_{ijg}(1-D_{jg}) \\
&\leq 2\left(\sum_{j\neq i} D_{jg}\right)^2 + 2\left(\sum_{j\neq i} D_{ig} D_{jg}\right)^2 + \left\{\sum_{j\neq i}(1-D_{jg})\right\}^2 + \left\{\sum_{j\neq i} D_{ig}(1-D_{jg})\right\}^2 \\
&\leq 6(N_g-1)^2 = B_2 < \infty. \qquad \text{(A.3)}
\end{aligned}$$

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

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

where

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

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

$$\begin{aligned}
\frac{1}{G}\sum_{g=1}^{G} h(\mathcal{V}_g, \bar{\boldsymbol{\zeta}}, \bar{\boldsymbol{\beta}}) &\xrightarrow{p} E\left[h(\mathcal{V}_g, \boldsymbol{\zeta}^\star, \boldsymbol{\beta}^\star)\right] = 0 \\
\frac{1}{G}\sum_{g=1}^{G} \nabla_\beta h(\mathcal{V}_g, \bar{\boldsymbol{\zeta}}, \bar{\boldsymbol{\beta}}) &\xrightarrow{p} E\left[\nabla_\beta h(\mathcal{V}_g, \boldsymbol{\zeta}^\star, \boldsymbol{\beta}^\star)\right] = -E\left[\mathbf{Z}_{ig}(\boldsymbol{\zeta}^\star) \mathbf{Z}_{ig}(\boldsymbol{\zeta}^\star)'\right] =: -\boldsymbol{B}_{\mathbf{Z}} \qquad \text{(A.4)} \\
\frac{1}{G}\sum_{g=1}^{G} \nabla_\zeta h(\mathcal{V}_g, \bar{\boldsymbol{\zeta}}, \bar{\boldsymbol{\beta}}) &\xrightarrow{p} E\left[\nabla_\zeta h(\mathcal{V}_g, \boldsymbol{\zeta}^\star, \boldsymbol{\beta}^\star)\right] = -E[\mathbf{Z}_{ig}(\boldsymbol{\zeta}^\star) \nabla_\zeta (\mathbf{Z}_{ig}(\boldsymbol{\zeta}^\star)'\boldsymbol{\beta}^\star)] =: -\boldsymbol{C}_\zeta,
\end{aligned}$$

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

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

$$0 = \frac{1}{G}\sum_{g=1}^{G} h(\mathcal{V}_g, \hat{\boldsymbol{\zeta}}, \hat{\boldsymbol{\beta}}), \tag{A.5}$$

First two convergences in (A.4) imply

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

Next, by applying mean value theorem on (A.5) and by (A.2) and (A.4), we have

$$\begin{aligned}0 &= \frac{1}{\sqrt{G}}\sum_{g=1}^{G} h(\mathcal{V}_g, \boldsymbol{\zeta}^\star, \boldsymbol{\beta}^\star) + \frac{1}{G}\sum_{g=1}^{G} \nabla_{\boldsymbol{\beta}} h(\mathcal{V}_g, \bar{\boldsymbol{\zeta}}, \bar{\boldsymbol{\beta}})\sqrt{G}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^\star) + \frac{1}{G}\sum_{g=1}^{G} \nabla_{\boldsymbol{\zeta}} h(\mathcal{V}_g, \bar{\boldsymbol{\zeta}}, \bar{\boldsymbol{\beta}})\sqrt{G}(\hat{\boldsymbol{\zeta}} - \boldsymbol{\zeta}^\star) \\ &= \frac{1}{\sqrt{G}}\sum_{g=1}^{G} h(\mathcal{V}_g, \boldsymbol{\zeta}^\star, \boldsymbol{\beta}^\star) - \boldsymbol{B}_{\mathbf{Z}}\sqrt{G}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^\star) - \boldsymbol{C}_{\boldsymbol{\zeta}}\frac{1}{\sqrt{G}}\sum_{g=1}^{G} \psi_{\boldsymbol{\zeta}}(\mathcal{V}_g, \boldsymbol{\zeta}^\star) + o_p(1).\end{aligned} \tag{A.6}$$

By rearranging (A.6), we have

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

where $\psi_{\boldsymbol{\beta}}(\mathcal{V}_g, \boldsymbol{\zeta}, \boldsymbol{\beta}) := \boldsymbol{B}_{\mathbf{Z}}^{-1}\left[h(\mathcal{V}_g, \boldsymbol{\zeta}, \boldsymbol{\beta}) - \boldsymbol{C}_{\boldsymbol{\zeta}}\psi_{\boldsymbol{\zeta}}(\mathcal{V}_g, \boldsymbol{\zeta})\right]$. Note that

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

provided that $E[Y_i^4] < \infty$, and we showed $E\left[\left\|\psi_{\boldsymbol{\zeta}}(\mathcal{V}_g, \boldsymbol{\zeta}^\star)\right\|^2\right] < \infty$ in Part A. Thus, by the Cauchy-Schwarz inequality and the Pythagorean theorem,

$$E\left[\left\|\psi_{\boldsymbol{\beta}}(\mathcal{V}_g, \boldsymbol{\zeta}^\star, \boldsymbol{\beta}^\star)\right\|^2\right] \leq \left\|\boldsymbol{B}_{\mathbf{Z}}^{-1}\right\|^2\left[E\left[\left\|h(\mathcal{V}_g, \boldsymbol{\zeta}^\star, \boldsymbol{\beta}^\star)\right\|^2\right] + \left\|\boldsymbol{C}_{\boldsymbol{\zeta}}\right\|^2 E\left[\left\|\psi_{\boldsymbol{\zeta}}(\mathcal{V}_g, \boldsymbol{\zeta}^\star)\right\|^2\right]\right] < \infty.$$

As a result, the asymptotic variance of $\sqrt{G}(\boldsymbol{\beta} - \boldsymbol{\beta}^\star)$ is the variance of $\psi_{\boldsymbol{\beta}}(\mathcal{V}_g, \boldsymbol{\zeta}^\star, \boldsymbol{\beta}^\star)$.

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

By applying the Delta-method, we have

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

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

$$\begin{pmatrix} \psi_{\boldsymbol{\beta},2}(\mathcal{V}_g, \boldsymbol{\zeta}, \boldsymbol{\beta}) \\ (N-1)\psi_{\boldsymbol{\beta},4}(\mathcal{V}_g, \boldsymbol{\zeta}, \boldsymbol{\beta})\zeta_2^\star + \beta_4^\star \psi_{\boldsymbol{\zeta},2}(\mathcal{V}_g, \boldsymbol{\zeta}) \\ \left(\psi_{\boldsymbol{\beta},3}(\mathcal{V}_g, \boldsymbol{\zeta}, \boldsymbol{\beta}) - \psi_{\boldsymbol{\beta},4}(\mathcal{V}_g, \boldsymbol{\zeta}, \boldsymbol{\beta})\right)\zeta_1^\star + (\beta_T^\star - \beta_U^\star)\psi_{\boldsymbol{\zeta},1}(\mathcal{V}_g, \boldsymbol{\zeta}) \\ \psi_{\boldsymbol{\beta},3}(\mathcal{V}_g, \boldsymbol{\zeta}, \boldsymbol{\beta})\zeta_3^\star + \beta_3^\star \psi_{\boldsymbol{\zeta},3}(\mathcal{V}_g, \boldsymbol{\zeta}) \end{pmatrix},$$

where $\psi_{\boldsymbol{b},k}(\cdot)$ denote k-th element in vector $\psi_{\boldsymbol{b}}(\cdot)$ for $\boldsymbol{b} \in \{\boldsymbol{\zeta}, \boldsymbol{\beta}, \boldsymbol{\pi}\}$.

Recall that $E\left[\left\|\psi_{\boldsymbol{\zeta}}(\mathcal{V}_g, \boldsymbol{\zeta}^\star)\right\|^2\right]$, and $E\left[\left\|\psi_{\boldsymbol{\beta}}(\mathcal{V}_g, \boldsymbol{\zeta}^\star, \boldsymbol{\beta}^\star)\right\|^2\right]$ are bounded, and therefore $E[\left\|\psi_{\boldsymbol{\zeta}}(\mathcal{V}_g, \boldsymbol{\zeta}^\star)\psi_{\boldsymbol{\beta}}(\mathcal{V}_g, \boldsymbol{\zeta}^\star, \boldsymbol{\beta}^\star)'\right\|]$ is also bounded. It follows that $E\left[\left\|\psi_{\boldsymbol{\pi}}(\mathcal{V}_g, \boldsymbol{\zeta}^\star, \boldsymbol{\beta}^\star)\right\|^2\right] < \infty$.

D. Consistency of plug-in standard errors

From Parts A and B, we have $E\left[\sup_{\boldsymbol{\zeta}}\left\|\psi_{\boldsymbol{\zeta}}(\mathcal{V}_g, \boldsymbol{\zeta})\right\|^2\right] < \infty$, and $E\left[\sup_{(\boldsymbol{\zeta},\boldsymbol{\beta})}\left\|\psi_{\boldsymbol{\beta}}(\mathcal{V}_g, \boldsymbol{\zeta}, \boldsymbol{\beta})\right\|^2\right] < \infty$. Therefore, by the Cauchy-Schwarz inequality,

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

This implies $E\left[\sup_{(\boldsymbol{\zeta},\boldsymbol{\beta})}\left\|\psi_{\boldsymbol{\pi}}(\mathcal{V}_g, \boldsymbol{\zeta}, \boldsymbol{\beta})\right\|^2\right] < \infty$, since $\left\|\psi_{\boldsymbol{\pi}}(\mathcal{V}_g, \boldsymbol{\zeta}, \boldsymbol{\beta})\psi_{\boldsymbol{\pi}}\right\|$ is computed by elements in $\psi_{\boldsymbol{\zeta}}(\mathcal{V}_g, \boldsymbol{\zeta})$, and $\psi_{\boldsymbol{\beta}}(\mathcal{V}_g, \mathcal{V}_g, \boldsymbol{\zeta}, \boldsymbol{\beta})$. Therefore, by consistency of $(\hat{\boldsymbol{\zeta}}, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\pi}})$, and applying Lemma 1, we have

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

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

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

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

Proof or Proposition 4. By applying the same argument of Part A in the proof of Proposition 3 to ζ_1, ζ_2, and by the fact that $\xi = \zeta_1 - \zeta_0$, we have $\hat{\xi} \xrightarrow{p} \xi^\star$, and

$$\begin{aligned}\sqrt{G}(\hat{\xi} - \xi^\star) &= \frac{1}{\sqrt{G}} \sum_{g=1}^{G} \psi_\xi(\mathcal{V}_g, \xi^\star) + o_p(1) \xrightarrow{d} N(0, \boldsymbol{B}_{\boldsymbol{W}}^{-1} \boldsymbol{\Sigma}_\xi \boldsymbol{B}_{\boldsymbol{W}}^{-1}), \\ \hat{V}_\zeta &= \frac{1}{G} \sum_{g=1}^{G} \psi_\xi(\mathcal{V}_g, \hat{\xi}) \psi_\xi(\mathcal{V}_g, \hat{\xi})' \xrightarrow{p} \boldsymbol{B}_{\boldsymbol{W}}^{-1} \boldsymbol{\Sigma}_\xi \boldsymbol{B}_{\boldsymbol{W}}^{-1},\end{aligned} \tag{A.7}$$

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

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

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

Also, $\nabla_{\zeta_1} \boldsymbol{X}_{ig}(\zeta) = (\boldsymbol{0}, \boldsymbol{0}, \sum_{j \neq i} \boldsymbol{W}_{ijg} D_{jg}, \sum_{j \neq i} \boldsymbol{W}_{ijg}(1 - D_{jg}))'$. Thus, by the same argument in (A.3), $\left\| \nabla_{\zeta_1} \boldsymbol{X}_{ig}(\zeta) \right\|^2 \leq 6(N_g - 1)^2 = B_2 < \infty$. Next, $\nabla_{\zeta_0} \boldsymbol{X}_{ig}(\zeta) = (\boldsymbol{0}, \boldsymbol{0}, \boldsymbol{0}, \sum_{j \neq i} \boldsymbol{W}_{ijg})'$. Thus,

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

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

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

where

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

for $t \in \{0, 1\}$, By the Cauchy-Schwarz inequality for the Frobenius inner product, and the definitions of the Frobenius norm and the L_2 norm. Also, since $\nabla_\beta \ell(\mathcal{V}_g, \zeta, \beta)$ and $\nabla_{\zeta_t} \ell(\mathcal{V}_g, \zeta, \beta)$

are continuous on ζ, β, we can apply Lemma 1 to conclude

$$\frac{1}{G}\sum_{g=1}^{G}\ell(\mathcal{V}_g,\bar{\zeta},\bar{\beta}) \xrightarrow{p} E[\ell(\mathcal{V}_g,\zeta^\star,\beta^\star)] = 0$$

$$\frac{1}{G}\sum_{g=1}^{G}\nabla_\beta\ell(\mathcal{V}_g,\bar{\zeta},\bar{\beta}) \xrightarrow{p} E\left[\nabla_\beta\ell(\mathcal{V}_g,\zeta^\star,\beta^\star)\right] = -E\left[\boldsymbol{X}_{ig}(\zeta^\star)\boldsymbol{X}_{ig}(\zeta^\star)'\right] =: -\boldsymbol{B}_{\boldsymbol{X}}$$

$$\frac{1}{G}\sum_{g=1}^{G}\nabla_{\zeta_t}\ell(\mathcal{V}_g,\bar{\zeta},\bar{\beta}) \xrightarrow{p} E\left[\nabla_{\zeta_t}\ell(\mathcal{V}_g,\zeta^\star,\beta^\star)\right] = -E[\boldsymbol{X}_{ig}(\zeta^\star)\nabla_{\zeta_t}(\boldsymbol{X}_{ig}(\zeta^\star)'\beta^\star)] =: -\boldsymbol{C}_{\zeta_t},$$

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

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

$$\hat{V}_\beta = \frac{1}{G}\sum_{g=1}^{G}\psi_\beta(\mathcal{V}_g,\hat{\zeta},\hat{\beta})\psi_\beta(\mathcal{V}_g,\hat{\zeta},\hat{\beta})' \xrightarrow{p} V_\beta,$$

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

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

$$\hat{V}_\pi = \frac{1}{G}\sum_{g=1}^{G}\psi_\pi(\mathcal{V}_g,\hat{\zeta},\hat{\beta})\psi_\pi(\mathcal{V}_g,\hat{\zeta},\hat{\beta})' \xrightarrow{p} V_\pi,$$

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

$$\begin{pmatrix} \psi_{\beta,2}(\mathcal{V}_g,\zeta,\beta) \\ (N-1)\psi_{\beta,4}(\mathcal{V}_g,\zeta,\beta)\xi_2^\star + \beta_4^\star\psi_{\xi,2}(\mathcal{V}_g,\xi) \\ \left(\psi_{\beta,3}(\mathcal{V}_g,\zeta,\beta) - \psi_{\beta,4}(\mathcal{V}_g,\zeta,\beta)\right)(\zeta_{30}^\star+\zeta_{11}^\star) + (\beta_3^\star-\beta_4^\star)\left(\psi_{\zeta_0,3}(\mathcal{V}_g,\zeta)+\psi_{\zeta_1,1}(\mathcal{V}_g,\zeta)\right) \\ \psi_{\beta,3}(\mathcal{V}_g,\zeta,\beta)\xi_3^\star + \beta_3^\star\psi_{\xi,3}(\mathcal{V}_g,\xi) \end{pmatrix},$$

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

□

B Tables

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

Table B.1: 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
Median												
50	0.1587	0.0254	0.0254	0.3694	0.0667	0.0689	0.0689	0.037	0	0.0205	0.0205	0.1718
100	0.1587	0.0254	0.0254	0.3698	0.0668	0.069	0.069	0.0373	0	0.0208	0.0208	0.1719
200	0.1587	0.0254	0.0254	0.3698	0.0668	0.0689	0.0689	0.0373	0	0.0206	0.0206	0.1719
400	0.1586	0.0254	0.0254	0.3697	0.0668	0.0688	0.0688	0.0375	0	0.0206	0.0206	0.172
800	0.1586	0.0254	0.0254	0.3698	0.0668	0.0688	0.0688	0.0375	0	0.0206	0.0206	0.1721
TRUE	0.1587	0.0254	0.0254	0.3698	0.0668	0.0689	0.0689	0.0374	0	0.0206	0.0206	0.1721
MSE												
50	0.6	0.9	0.9	3	0.3	0.5	0.5	2	0.5	1	1	3.2
100	0.3	0.4	0.4	1.4	0.1	0.3	0.3	1	0.3	0.5	0.5	1.7
200	0.1	0.2	0.2	0.7	0.1	0.1	0.1	0.5	0.1	0.2	0.2	0.8
400	0.1	0.1	0.1	0.4	0	0.1	0.1	0.2	0.1	0.1	0.1	0.4
800	0	0.1	0.1	0.2	0	0	0	0.1	0	0.1	0.1	0.2
Coverage Rate												
50	0.9412	0.9408	0.9408	0.9367	0.9461	0.9415	0.9415	0.944	0.9448	0.9464	0.9464	0.9467
100	0.9467	0.9473	0.9473	0.9474	0.9458	0.941	0.941	0.9421	0.9468	0.9416	0.9416	0.9459
200	0.9464	0.9472	0.9472	0.9467	0.9446	0.947	0.947	0.9506	0.9474	0.951	0.951	0.9497
400	0.9497	0.9472	0.9472	0.9498	0.9485	0.9491	0.9491	0.9534	0.9493	0.9439	0.9439	0.9513
800	0.9501	0.9489	0.9489	0.9495	0.9527	0.9503	0.9503	0.951	0.9504	0.9505	0.9505	0.949

Notes: This table presents the simulation results for $B = 10,000$ replications. Column G denotes the number of independent groups, with each group containing $N = 20$ individuals. The first panel shows the median across all replications, and the row labeled "TRUE" presents the true values for each decomposition. The second and third columns display the 1000×MSE and 95% coverage rates, respectively.

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

Table B.2: Simulation Result of Outcome Coefficients

	Design 1				**Design 2**			
G	β_0	β_I	β_T	β_U	$\Delta\beta_0$	β_I	β_T	β_U
Median								
50	1.988	1.0011	0.7996	0.6071	0.9963	1.0052	0.8007	0.6
100	2.0075	1.0026	0.7993	0.5969	0.998	1.0045	0.7974	0.5968
200	1.9923	0.9998	0.8007	0.6065	1.0022	0.9969	0.7986	0.6024
400	2.0022	0.9995	0.7997	0.5984	0.9946	1.0058	0.8007	0.5973
800	2.0012	1.0033	0.8	0.5985	1.0008	0.9992	0.7998	0.6034
TRUE	2	1	0.8	0.6	1	1	0.8	0.6
MSE								
50	1.0006	0.6166	0.0303	0.2273	0.5897	1.6011	0.1624	0.8069
100	0.4933	0.3012	0.0148	0.1128	0.2862	0.7761	0.0774	0.3965
200	0.2418	0.1463	0.0073	0.0549	0.1387	0.384	0.0394	0.1964
400	0.1193	0.073	0.0036	0.027	0.0678	0.1869	0.019	0.0951
800	0.0593	0.0358	0.0018	0.0135	0.0338	0.0897	0.009	0.0464
Coverage Rate								
50	0.9359	0.9363	0.9345	0.9346	0.9334	0.9436	0.9362	0.9415
100	0.9375	0.9394	0.9402	0.9378	0.9359	0.9429	0.9427	0.9402
200	0.9452	0.9433	0.9451	0.9442	0.9437	0.943	0.9446	0.9417
400	0.9459	0.9428	0.9406	0.9476	0.9462	0.9443	0.9463	0.9437
800	0.9487	0.9492	0.9521	0.95	0.9462	0.9506	0.9517	0.9508

Notes: This table presents the simulation results for $B = 10,000$ replications. Column G denotes the number of independent groups, with each group containing $N = 20$ individuals. The first panel shows the median 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.

This table shows the mean absolute error (MAE) computed by 10,000 replications for all parameters in Designs 1 and 2.

Table B.3: Overall MAE with Different Group Size

		Design 1			**Design 2**			
N	G	ζ	β	π	ζ_1	ξ	β	π
5	50	0.038	0.952	0.364	0.029	0.04	3.196	1.202
5	100	0.027	0.477	0.153	0.021	0.028	1.372	0.46
5	200	0.019	0.308	0.095	0.015	0.02	0.544	0.149
5	400	0.013	0.212	0.064	0.01	0.014	0.345	0.089
5	800	0.009	0.147	0.044	0.007	0.01	0.241	0.061
5	1600	0.007	0.105	0.031	0.005	0.007	0.168	0.042
10	50	0.018	0.508	0.206	0.014	0.018	0.818	0.346
10	100	0.013	0.345	0.137	0.01	0.013	0.523	0.207
10	200	0.009	0.243	0.095	0.007	0.009	0.355	0.137
10	400	0.006	0.169	0.066	0.005	0.007	0.249	0.095
10	800	0.004	0.12	0.047	0.003	0.005	0.175	0.067
10	1600	0.003	0.084	0.033	0.002	0.003	0.121	0.046
20	50	0.009	0.482	0.223	0.007	0.009	0.654	0.341
20	100	0.006	0.34	0.157	0.005	0.006	0.46	0.238
20	200	0.004	0.238	0.109	0.003	0.004	0.324	0.167
20	400	0.003	0.167	0.077	0.002	0.003	0.226	0.116
20	800	0.002	0.118	0.054	0.002	0.002	0.158	0.081
20	1600	0.002	0.082	0.037	0.001	0.002	0.113	0.058

Notes: This table presents the simulation results for $B = 10,000$ replications. Column G denotes the number of independent groups, and column N denotes the number of individuals in each group. The table shows the overall mean absolute error (MAE) across all replications, i.e., $\frac{1}{4B}\sum_{b=1}^{B}\sum_{k=1}^{4}|\hat{\theta}_{kb} - \theta_k^{\star}|$.

This table shows the mean squared error (MSE) computed by 10,000 replications for all parameters in Designs 1 and 2.

Table B.4: Overall MSE with Different Group Size

		Design 1			**Design 2**			
N	G	ζ	β	π	ζ_1	ξ	β	π
5	50	9.95	154.646	53.899	6.42	11.14	7235.119	1760.309
5	100	4.94	3.89	1.034	3.13	5.49	1051.371	191.949
5	200	2.47	0.784	0.112	1.58	2.76	6.278	0.926
5	400	1.23	0.353	0.046	0.79	1.37	1.124	0.108
5	800	0.61	0.167	0.021	0.4	0.68	0.53	0.047
5	1600	0.3	0.084	0.01	0.2	0.35	0.252	0.022
10	50	2.2	1.901	0.604	1.4	2.4	19.279	11.341
10	100	1.11	0.847	0.244	0.69	1.23	2.109	0.663
10	200	0.55	0.412	0.115	0.36	0.61	0.925	0.271
10	400	0.27	0.2	0.054	0.17	0.3	0.445	0.127
10	800	0.14	0.1	0.027	0.09	0.16	0.219	0.062
10	1600	0.07	0.049	0.013	0.04	0.07	0.107	0.03
20	50	0.53	1.875	0.71	0.33	0.57	3.16	1.819
20	100	0.26	0.922	0.344	0.17	0.29	1.536	0.868
20	200	0.13	0.45	0.166	0.08	0.14	0.759	0.425
20	400	0.06	0.223	0.083	0.04	0.07	0.369	0.206
20	800	0.03	0.11	0.041	0.02	0.04	0.179	0.099
20	1600	0.02	0.053	0.019	0.01	0.02	0.091	0.051

Notes: This table presents the simulation results for $B = 10,000$ replications. Column G denotes the number of independent groups, and column N denotes the number of individuals in each group. The table shows the overall mean squared error (MSE) across all replications, i.e., $\frac{1}{4B}\sum_{b=1}^{B}\sum_{k=1}^{4}(\hat{\theta}_{kb} - \theta_k^{\star})^2$. For $\boldsymbol{\zeta}, \boldsymbol{\zeta}_1, \boldsymbol{\xi}$, it displays $1000 \times MSE$.