Decomposition of Causal Effect with Interference Accounting for Network Change

Siwon Ryu

August 22, 2024

Abstract

Recent empirical studies emphasize the importance of indirect, or spillover, effects of a program or policy. These studies usually assume that the underlying network is fixed or unaffected by the program. However, there has been some empirical evidence suggesting the treatment can also have significant affect on network. This paper analyzes the identification and estimation of causal treatment effects while explicitly considering possible network changes resulting from a program. The main finding of this study is the decomposition of the causal effect into two distinct parts: the effect of treatment when the network remains unchanged and the effect when only the network structure is changed by the treatment. This contributes to the literature by enhancing our understanding of the mechanisms of a policy or program through the consideration of counterfactual scenarios in which the network is either changed or unchanged due to treatment. The proposed method is applicable in quasi-experimental situations using a Difference-in-Differences approach, as well as in randomized experiments. A straightforward multi-step estimation procedure for the causal effects and their decompositions is introduced, and its performance is evaluated through a Monte Carlo study, and an empirical illustration.

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

1 Introduction

Evaluating a program is an important topic in empirical economics, typically involving the estimation of causal effects of a program. Many methods are based on the potential outcome framework (e.g., Rubin (1974)), and the baseline setting is the Stable Unit Treatment Value Assumption (SUTVA) that excludes interference between units or individuals. However, SUTVA can be restrictive, as economic agents frequently interact with one another. Furthermore, 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. This situation presents primary challenges for researchers. It is not only difficult to define causal parameters compared to the situations under SUTVA, but also difficult to identify meaningful parameters as there are many number of potential outcomes. For example, if N units interact, the possible number of treatment assignments profile, and hence the number of possibly distinct potential outcome is 2^N. To address these challenges, studies commonly employ the constant treatment response (CTR) assumption noted by Manski (2013) or use the exposure mapping approach to significantly reduce dimensionality (e.g., Leung (2020), Leung (2022), Vazquez-Bare (2022a), Vazquez-Bare (2022b), Aronow and Samii (2017), Bramoullé, Djebbari, and Fortin (2009), Forastiere, Airoldi, and Mealli (2021), Auerbach and Tabord-Meehan (2021)).

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

While exposure mapping and double randomization make the problem easier to handle, they assume that the underlying network structure is not causally affected by

the treatment or remains fixed. Exposure itself usually depends on the underlying network structure. Therefore, considering possible network changes with the exposure mapping approach is not straightforward, as disentangling changes in exposure from changes in the network will be challenging without strong assumptions. Studies based on double randomization also usually assume a fixed network or independence of the network from the treatment. This assumption might hold in the short run, where there is not enough time for the network structure to change.

However, recent empirical studies suggest that the underlying network can be significantly affected by the treatment. Comola and Prina (2021) use an experimental data in Nepal and observe that providing savings accounts to households led to changes in network degrees. Specifically, the probability of being linked with other of households with at least one treated member shifted from 81% to 76%. Banerjee et al. (2024) examine data from Karnataka, India, and a field experiment in Hyderabad to investigate the effects of exposure to formal financial institutions on network density. Dupas, Keats, and Robinson (2019) use an experimental data from Kenya where households were provided with free savings accounts. Their results show that households became less reliant on distant family members and more supportive of neighbors and friends within their village. These studies emphasize the importance of considering network effects in program evaluations.

This paper introduces a method to identify and estimate the causal effect of a program, taking into account the possibility of network changes resulting from the treatment. Additionally, this study decomposes the treatment (or spillover) effect into two parts. The first part examines the impact of the treatment when the network remains unchanged, while the second part focuses on the effect of the treatment when it changes the network structure. Therefore, if the treatment does not cause changes in the network, the second part has no effect, and the first part will align with existing methods that estimate treatment and spillover effects.[1]

The proposed method is based on dyadic links and a linear-in-sums outcome response to avoid the use of an exposure map. Two experimental settings are considered: first, quasi-experiments with parallel trends. In this setting, the difference-in-differences

[1] Could you do a test?

estimand plays a key role in identifying and decomposing causal effects. It can also be applied to randomized experiments as a special case. A straightforward multi-stage estimation procedure is proposed, and its performance is evaluated through a Monte Carlo study. The method is illustrated using experimental data conducted by Comola and Prina (2020) in Nepal, which provided access to savings accounts for households. The empirical results show that while the overall direct effect is positive, it is decomposed into a positive treatment effect and a negative network effect. This decomposition and interpretation cannot be obtained from the computation of direct and indirect effects in Comola and Prina (2021).

The main contributions of this paper can be summarized in three key aspects. First, it proposes a novel method to analyze causal effects while accounting for potential network changes within the potential outcome framework. Second, it introduces a decomposition of causal effects, allowing for an in-depth examination of a program's mechanisms. Third, the methods presented are applicable to both randomized experiments and quasi-experimental datasets, particularly using a Difference-in-Differences approach. The proposed methods can also be seen as a generalization of existing approaches. If the network remains unchanged, the method simplifies to existing methods such as those by Vazquez-Bare (2022b), Leung (2020), or Aronow and Samii (2017), making the decomposition trivial. If there is no interference, the proposed method reduces to the standard potential outcome framework under SUTVA.

This study is closely related to the literature on the identification and estimation of causal effects of a program when units interact with each other. Studies use exposure map approach under randomized experiments (e.g., Leung (2020), Leung (2022), and Vazquez-Bare (2022b)). However, treatment is barely exogenous, as it can be endogenously given due to imperfect compliance even under randomization. In those cases, causal parameters are defined and identified under various settings (e.g., Vazquez-Bare (2022a), DiTraglia et al. (2023), Hoshino and Yanagi (2023), Kormos, Lieli, and Huber (2023), Kang and Imbens (2016), Blackwell (2017)). On the other hand, the endogeneity of treatments arising from selection is addressed by using quasi-experimental variation such as Difference-in-Differences approach (e.g., Xu (2023) and Butts (2021)). Since treatment assignments under interference can be considered as multiple treatments, this

paper also relates to the problems of multiple treatments in quasi-experimental settings (e.g., Frölich (2004), Fricke (2017)).

While the aforementioned studies assume a fixed or independent network, Comola and Prina (2021) explicitly address methods to estimate treatment effects while accounting for network changes. The authors estimate a two-period linear-in-means model (e.g., Manski (2013), Bramoullé, Djebbari, and Fortin (2009)) and define the direct and indirect treatment effects as the partial derivatives of the conditional mean of the reduced-form outcome with respect to the treatment vector. Their findings suggest that indirect effects could be underestimated if potential network changes resulting from the treatment are not considered. However, the computed direct and indirect effects are difficult to interpret as causal effects without further conditions. Although they consider possible network changes in the structural equation of outcomes, it is unclear how much of the effect is due to network changes. In contrast, the decomposition proposed in this paper explicitly derives each component of the effect, providing a clear understanding of the underlying mechanisms related to network changes. The difference between the decomposition introduced in this paper and the methods in Comola and Prina (2021) is illustrated using the same experimental data, highlighting the advantages of the proposed method.

This paper is organized as follows: Section 2 outlines the setting, defines the parameters of interest, and addresses their identification. The decomposition of causal effects is introduced in Section 2.2. Section Section 3 presents the details of the estimation procedure. Section 4 evaluates performance through Monte Carlo simulations. Section 5 provides an empirical illustration of the proposed method. Section 6 concludes.

2 Model and Identification

In this section, I provide an overview of the setting and define the parameters of interest. I begin by discussing response functions for potential links and potential outcomes. Then, I establish key causal parameters, specifically the direct and indirect effects. The direct effect focuses on how a unit's treatment status impacts its outcome, while the indirect effects consider the effects of changes in others' treatment statuses. Subsequently,

I propose a decomposition of these effects into two parts. The first part examines the impact when the underlying network remains unchanged, referred to as the *treatment effect*, as the conventional treatment effect in the literature. The second part of the decomposition is about the effect solely arising from changes in the network structure, referred to as the *network effect* in this study.

2.1 Response Functions for Links and Potential Outcomes

Suppose there are G independent groups with N units in each group. If we observe two periods, let $t \in \{0, 1\}$ denote the time periods, where some units are assigned to a treatment group after $t = 0$, i.e., the period $t = 0$ represent the pre-treatment period and $t = 1$ the post-treatment period. Also, let $D_{ig} \in \{0, 1\}$ be the indicator for a unit receiving treatment, and assume there is no imperfect compliance. Each unit interacts with others through an underlying network structure. Specifically, let $A_{ijtg} \in \{0, 1\}$ represent the link between individuals i and j in group g at time t, with no self-links, i.e., $A_{ijtg} = 1$ if two different units i and j are linked at time t in group g, and $A_{iitg} = 0$ for all i. The network can be both directed and undirected in this setting. For each unit i in group g, let $Y_{itg} \in \mathbb{R}$ be the outcome of interest at time t.

Potential outcomes and potential links can be defined as functions of the entire treatment vector in each group. Let $\boldsymbol{d} \in \{0, 1\}^N$ represent a vector of treatment assignments for N units. Corresponding to an assignment $\boldsymbol{d}$, let $A_{ijtg}(\boldsymbol{d})$ denote the *potential* link between units i and j, and $Y_{itg}(\boldsymbol{d})$ denote the *potential* outcome of unit i at time t. Because there are 2^N possible potential treatment statuses, defining the causal effect of interest becomes challenging, particularly if the number of units is large. To address this issue, I first assume for potential network to be formed from a dyadic model as stated in Assumption 1.

Assumption 1 (Dyadic Response on Potential Network Links)**.** *For each pair of units (i, j) in group g at time period t, for any $\boldsymbol{d}, \boldsymbol{d}' \in \{0, 1\}^N$, if $d_i = d'_i$ and $d_j = d'_j$, then $A_{ijtg}(\boldsymbol{d}) = A_{ijtg}(\boldsymbol{d}')$ with probability 1. Therefore, by abusing notation, denote the*

potential link as follows: for $\boldsymbol{d} = (d_1, ..., d_N) \in \{0,1\}^N$,

$$A_{ijtg}(\boldsymbol{d}) = \begin{cases} A_{ijtg}(d_i, d_j) & i \neq j, \\ 0 & i = j. \end{cases}$$

Assumption 1 imposes the restriction that each pair's potential link is determined by the pair's treatment status (d_i, d_j) only, but not affected by the other's treatment status. For example, Assumption 1 is satisfied for a dyadic link formation model (e.g., Graham (2017)):

$$A_{ij}(\boldsymbol{d}) = \mathbb{1}\{\psi_1 d_i + \psi_2 d_j + \delta_i + \delta_j + u_{ij} > 0\}. \tag{1}$$

The potential outcome is generally expressed as a function of the treatment vector $\boldsymbol{d} \in \{0,1\}^N$. The conventional approach to handling potential outcomes under interference is by using exposure maps or the constant treatment response assumption (e.g., Manski (2013)). For instance, if for any $\boldsymbol{d}, \boldsymbol{d}' \in \{0,1\}^N$, $\psi(\boldsymbol{d}) = \psi(\boldsymbol{d}')$ implies $Y(\boldsymbol{d}) = Y(\boldsymbol{d}')$ with probability 1, then ψ is called an exposure map.

The existence or specific functional form of an exposure mapping is largely unknown without restrictions on the response function. However, in some cases, it is possible to specify an appropriate exposure map for a potential outcome. For example, if the network is anonymous then only the number of treated and untreated neighbors will matter. Indeed, Leung (2020) demonstrated that assuming (i) local spillover (interference within a network distance of 1) and (ii) exchangeability is equivalent to having an exposure map based on one's own treatment (d_i), the number of treated neighbors $(q_i := \sum_j A_{ij} d_j)$, and the number of untreated neighbors $(r_i := \sum_j A_{ij}(1 - d_j))$. Therefore, the potential outcome can be written as $Y(d_i, q_i, r_i)$ instead of a function of the entire treatment vector.

To focus on the identification of causal effects and their decomposition, I simply assume that the response function of the potential outcome is linear in the exposure (d_i, q_i, r_i), as stated in Assumption 2.[2]

[2]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.

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

$$Y_{ig}(\boldsymbol{d}) = \alpha + \beta d_i + \varepsilon_{ig}(d_i) + \gamma_1 q_{ig}(\boldsymbol{d}) + \gamma_2 r_{ig}(\boldsymbol{d}),$$

where $q_{ig}(\boldsymbol{d}) = \sum_{j=1,j\neq i}^{N} A_{ijg}(d_i, d_j)d_j$ and $r_{ig}(\boldsymbol{d}) = \sum_{j=1,j\neq i}^{N} A_{ijg}(d_i, d_j)(1-d_j)$.

The parameter β represents the effect of one's own treatment d_i when all links and others' treatments are fixed. Thus, β captures the direct treatment effect. Next, since $q_{ig}(\boldsymbol{d})$ denotes the number of potentially treated neighbors, γ_1 captures the spillover (or exposure) effect from marginal changes in the number of treated neighbors. Similarly, γ_2 captures the effects from the untreated neighbors. These parameters can be interpreted as causal effects under various settings if the network link is unchanged by the treatments. However, because one's own treatment d_i also affects potential links $\{A_{ijg}(d_i, d_j)\}_{i,j}$, it has an additional effect on outcomes, driven by the changed links. I define this type of effect as the *network effect*. Changes in others' treatments d_j can have similar effects, which will be formally defined below.

Assumption 2 implies that the observed outcome can be expressed as a linear-in-sums model:

$$Y_{ig} = \alpha + \beta D_i + \gamma_1 \sum_{j=1}^{N} A_{ij} D_j + \gamma_2 \sum_{j=1}^{N} A_{ij}(1 - D_j) + \varepsilon_{i1g}.$$

If the network is unaffected by the treatment or is fixed, i.e., $A_{ij}(d_i, d_j) = A_{ij}$, then the response function is just a linear response model commonly used in the literature (e.g., Leung (2020), Forastiere, Airoldi, and Mealli (2021)). Furthermore, if $\gamma_2 = \gamma_3 = 0$, the model reduces to the standard potential outcome model under SUTVA. From this perspective, this model generalizes existing methods to account for cases where the network structure can be affected by the treatment.

When we observe outcomes and links over two periods $t \in \{0,1\}$ where $t = 0$ is pre-treatment, $t = 1$ is post-treatment periods, then the potential links at period $t = 0$ can be written as $A_{ij0g} = A_{ij0g}(0,0)$ then the potential outcome at period $t = 0$ is

written as $Y_{i0g} = Y_{i0g}(\mathbf{0}) = \alpha + \varepsilon_{i0g}(0) + \gamma_2 r_{i0g}(\mathbf{0})$.

2.2 Causal Parameters and Decomposition

In this subsection, I omit the group and time index subscripts (t, g) for simplicity to define parameters of interest and their decompositions. First, consider a case when there are 2 units in each group ($N = 2$). Assumption 2 implies that we can write the potential outcome as a function of own treatment, neighbor's treatment, and their potential link: $Y_i(\boldsymbol{d}) = h(d_i, d_j, A_{ij}(d_i, d_j))$. Then, the effect of own treatment (d_i) on i's outcome is expressed as

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

Also, the effect of neighbor j's treatment (d_j) on i's outcome is expressed as

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

The direct and indirect network effects represent the causal impact resulting from changes in links driven by treatment. The second terms capture the pure effects of treatment when links are fixed.

In a general case with N individuals, each indirect effect can be influenced by the treatment status of all neighbors. However, the potential outcome is affected by the treatment of others solely through the count of treated or untreated neighbors.[3] Therefore, I focus on the marginal effect of a neighbor's treatment, specifically the impact of *one additional* treated or untreated neighbor.

Let $\boldsymbol{e}_1, ..., \boldsymbol{e}_N$ be the standard Euclidean basis, i.e., for each i, $e_{ii} = 1$, and $e_{ij} = 0$ for all $j \neq i$. Then, the average direct effect of the treated (π^D) on the outcome is

[3]This is due to the exposure mapping used to construct the response function being implied by the exchangeability of networks.

defined as

$$\begin{aligned}\text{Average Direct Effect } (\pi^D) &:= E[Y_i(\boldsymbol{e}_i) - Y_i(\mathbf{0})|\boldsymbol{D} = \boldsymbol{e}_i] \\ &= \underbrace{\beta}_{:=\pi^{DT}} + \underbrace{\gamma_2 \sum_{j=1}^{N} E[A_{ij}(1,0) - A_{ij}(0,0)|D_i = 1, D_j = 0]}_{:=\pi^{DN}}.\end{aligned}$$

Assumption 2 provides the causal interpretation of the decomposition. π^{DT} is the average direct treatment effect that represents the effect of one's own treatment when the network is fixed. On the other hand, π^{DN} is the average indirect treatment effect that captures the effect of changes in links driven by one's own treatment. Similarly, the average indirect effect of the treated (π^I) on the outcome is defined for some j.

$$\begin{aligned}\text{Average Indirect Effect} &:= E[Y_i(\boldsymbol{e}_j) - Y_i(\mathbf{0})|\boldsymbol{D} = \boldsymbol{e}_j] \\ &= \underbrace{(\gamma_1 - \gamma_2)E[A_{ij}(0,1)|D_i = 0, D_j = 1]}_{:=\pi^{IT}} \\ &\quad + \underbrace{\gamma_2 E[A_{ij}(0,1) - A_{ij}(0,0)|D_i = 0, D_j = 1]}_{:=\pi^{IN}},\end{aligned}$$

where π^{IT} is the average indirect treatment effect and π^{IN} is the average indirect network effect that are interpreted similarly to the direct effects. From the linearity of the response function, any direct and indirect effects (i.e., any $\boldsymbol{d} \in \{0,1\}^N$) can be expressed by the sum of $(\pi^{DT}, \pi^{DN}, \pi^{IT}, \pi^{IN})$s. To identify these four effects separately, we need the values of $\beta, \gamma_1, \gamma_2$, and the conditional distribution of $A_{ij}(0,1) - A_{ij}(0,0)$, $A_{ij}(1,0) - A_{ij}(0,0)$, and $A_{ij}(0,1)$.

2.3 Identification

This section discusses the identification of causal effects and their decomposition. I first address the setting where outcomes 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. Then, I discuss the case where only the post-treatment period is observed, but the treatment is exogenous (i.e., randomized experiments).

Note that we have two types of data: (i) unit-level data and (ii) dyadic-level data. Thus, the identification involves using both types of data. First, by using dyadic-level data, which includes links and dyadic characteristics, the average treatment effects of links on the treated are identified using a dyadic difference-in-differences estimand. Second, I provide an interpretation of coefficients from a difference-in-differences regression of outcomes and show that they are related to the parameters in the outcome response function (i.e., $\beta, \gamma_1, \gamma_2$) as a linear transformation. Therefore, the parameters of the response function can be recovered by taking the inverse map. With this information, we have all the ingredients to recover the decomposition $\boldsymbol{\pi} = (\pi^{DT}, \pi^{DN}, \pi^{IT}, \pi^{IN})$.

2.3.1 Identifying Assumptions

To begin with, I list the identifying assumptions for the main identification results.

Assumption 3 (Local Exogeneity)**.** *Let $\boldsymbol{D}_g = (D_{1g}, ..., D_{Ng})$ be observed treatment vector of group g. For each time t and group g,*

1. $E[\varepsilon_{itg}(d)|\boldsymbol{D}_g] = E[\varepsilon_{itg}(d)|D_{ig}] = 0$*, for all individual i and $d \in \{0, 1\}$,*
2. $E[A_{ijtg}(d, e)|\boldsymbol{D}_g] = E[A_{ijtg}(d, e)|D_{ig}, D_{jg}]$*, for all pairs (i, j) and $(d, e) \in \{0, 1\}^2$.*

The first part of Assumption 3 states that the individual error term $\varepsilon_{itg}(d)$ is mean independent of the others' treatment statuses given the individual's own treatment. Therefore, when $\gamma_2 = \gamma_3 = 0$, the model reduces to a standard linear causal model. The second part of Assumption 3 strengthens Assumption 1 by assuming that potential links are not only formed by a dyadic model but are also mean independent of other pairs' treatment statuses given their own treatments. For example, in the dyadic model (1), it is implied by $(\delta_i, \delta_j, u_{ij})|\boldsymbol{D} \sim (\delta_i, \delta_j, u_{ij})|D_i, D_j$.

Assumption 4 (Distributions)**.** *(i) The individual level data $\{(Y_{i1g}, Y_{i0g}, D_{ig})\}_{i,g}$ are identically distributed over both individuals and groups, and independent across groups. And the dyadic data $\{(D_{ig}, D_{jg}, A_{ij1g}, A_{ij0g})\}_{(i,j),g}$ are identically distributed over all pairs (i, j) and groups g, and independent across groups; (ii) For all pairs (i, j) and g, $\Pr(D_{ig} = d, D_{jg} = e) \in (0, 1)$ for all $(d, e) \in \{0, 1\}^2$.*

The first part of Assumption 4 is about the distribution of the population, stating that we have independent groups, and the dependence between individuals or between dyads is unrestricted. The second part is a standard requirement that guarantees the existence of corresponding conditional distributions.

Assumption 5 (No Anticipation)**.** *For each group g,*

1. $E[\varepsilon_{i0g}(0)|D_{ig}=1]=E[\varepsilon_{i0g}(1)|D_{ig}=1]$*, for all individual i,*
2. $E[A_{ij0g}(d,e)|D_i=d,D_j=e]=E[A_{ij0g}(0,0)|D_i=d,D_j=e]$*, for all pairs (i,j), and $(d,e)\in\{0,1\}^2$.*

Assumption 5 states the required condition that there is no anticipation of the treatment. Because no one is treated at period $t=0$, it can be thought of as $\varepsilon_{i0g}=\varepsilon_{i0g}(0)$, and therefore the first part of Assumption 5 follows. By the same argument, the second part follows because $A_{ij0g}=A_{ij0g}(0,0)$.

Assumption 6 (Parallel Trend)**.** *Let Δ be the first-difference operator, i.e., $\Delta Z=Z_1-Z_0$. For each group g,*

1. $E[\Delta\varepsilon_{ig}(0)|D_{ig}=1]=E[\Delta\varepsilon_{ig}(0)|D_{ig}=0]$*, for each individual i,*
2. $E[\Delta A_{ijg}(0,0)|D_{ig}=d,D_{jg}=e]=E[\Delta A_{ijg}(0,0)|D_{ig}=0,D_{jg}=0]$*, for all pairs (i,j) and $(d,e)\in\{0,1\}^2$.*

Assumption 6 is the key identifying assumption for the difference-in-differences estimands. For instance, in the identification of $E[A_{ij1g}(d,e)-A_{ij1g}(0,0)|D_i=d,D_j=d]$, the first term is directly observed, while the second term remains counterfactual. The second part of Assumption 6 recovers this counterfactual term by exploiting the exogenous parallel trend.

Assumptions 5, 6 are additional restrictions on the response function in Assumption 2. In particular, by construction of outcome response function, Assumption 5 implies there is no anticipation about treatment on potential outcomes at the pre-treatment period:

$$E[Y_{i0g}(\boldsymbol{d})|\boldsymbol{D}_g=\boldsymbol{d}]=E[Y_{i0g}(\boldsymbol{0})|\boldsymbol{D}_g=\boldsymbol{d}],\quad\forall\boldsymbol{d}\in\{0,1\}^N,$$

and Assumption 6 implies parallel trend on potential outcomes:[4]

$$E[Y_{i1g}(\mathbf{0}) - Y_{i0g}(\mathbf{0})|\boldsymbol{D}_g = \boldsymbol{d}] = E[Y_{i1g}(\mathbf{0}) - Y_{i0g}(\mathbf{0})|\boldsymbol{D}_g = \mathbf{0}], \quad \forall \boldsymbol{d} \in \{0,1\}^N.$$

2.3.2 Dyadic Regression

Because each group is independent and identical, I suppress the group index for simplicity in this subsection. For each $(d, e) \in \{0,1\}^2$, the average effect on the link between two individuals (i, j) with $D_i = d, D_j = e$ is defined as follows:

$$H(d, e) := E[A_{ij1}(d, e) - A_{ij1}(0, 0)|D_i = d, D_j = e].$$

Identifying this conditional expectation is considered equivalent to identifying the average treatment effect on the treated for multiple treatments (quaternary treatment in this case). It can be established under dyadic parallel trend and no-anticipation assumptions, as investigated in studies such as Frölich (2004) and Fricke (2017). Proposition 1 summarizes the identification results for dyadic difference-in-differences estimand:

Proposition 1 (Difference-in-differences of links)**.** *For each $t \in \{0,1\}$, and $(d, e) \in \{0,1\}^2$, define $m_t(d, e) := E[A_{ijt}|D_i = d, D_j = e]$ as the probability of i and j being linked conditional on the treatment statuses, and $\Delta m(d, e) := m_1(d, e) - m_0(d, e)$ as the difference of that conditional probability over time. Under Assumptions 4, 5-2, 6-2, we have for each $(d, e) \in \{0,1\}^2$,*

$$H(d, e) = \Delta m(d, e) - \Delta m(0, 0), \tag{2}$$

and $H(0, 0) = 0$ by definition.

Note that if the network is undirected, then we have $A_{ijt} = A_{jit}$ with probability 1, and therefore it follows that $H(1, 0) = H(0, 1)$ with probability 1.

[4]The coefficient γ_2 is the effect of untreated neighbor, and it is assumed to be same for each period $t \in \{0,1\}$ on this purpose. If we allow time varying coefficient for γ_2, the parallel trend for outcome holds if and only if $\gamma_{20} = \gamma_{21}$, or $A_{ij0g}(d, e) = A_{ij0g}$ with probability one.

2.3.3 Outcome Regression

Consider the following conditional expectations:

$$E[A_{ij1g}|D_{ig}, D_{jg}] = \widetilde{\boldsymbol{W}}'_{ijg}\boldsymbol{\zeta}, \quad E[\Delta A_{ijg}|D_{ig}, D_{jg}] = \widetilde{\boldsymbol{W}}'_{ijg}\boldsymbol{\xi}, \tag{3}$$

where $\widetilde{\boldsymbol{W}}_{ijg} = (1, D_{ig}, D_{jg}, D_{ig}D_{jg})' \in \mathbb{R}^4$. Thus, Proposition 1 implies that the coefficients in (3) are given by:[5]

$$\boldsymbol{\zeta} = \begin{pmatrix} m_1(0,0) \\ m_1(1,0) - m_1(0,0) \\ m_1(0,1) - m_1(0,0) \\ m_1(1,1) - m_1(1,0) - m_1(0,1) + m_1(0,0) \end{pmatrix}, \boldsymbol{\xi} = \begin{pmatrix} \Delta m(0,0) \\ H(1,0) \\ H(0,1) \\ H(1,1) - H(1,0) - H(0,1) \end{pmatrix}. \tag{4}$$

In addition, define

$$\boldsymbol{\omega} := \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix} \times \boldsymbol{\zeta} = \begin{pmatrix} m_1(0,1) \\ m_1(1,1) - m_1(0,1) \end{pmatrix}. \tag{5}$$

Proposition 2 introduces an estimating equation for first-differentiated outcomes with a parameter vector $\boldsymbol{\delta}$. This coefficient vector $\boldsymbol{\delta}$ is expressed by a function of the parameters in the response function: $\boldsymbol{\theta} = (\Delta\alpha, \beta, \gamma_1, \gamma_2)'$, and the parameters related to the distributions of potential links: $\boldsymbol{\zeta}$ and $\boldsymbol{\xi}$ in (4).

Proposition 2 (Identification of parameters in outcome model)**.** *Let* $\Delta Y_{ig} := Y_{i1g} - Y_{i0g}$. *Then, under Assumptions 1-6,*

$$E[\Delta Y_{ig}|\boldsymbol{D}_g] = \tilde{\boldsymbol{D}}'_{ig}\boldsymbol{\delta}, \tag{6}$$

[5]For any discrete random variable $X \in \{x_0, x_1, ..., x_M\}$ and a random vector Y, we have

$$E[Y|X] = E[Y|X = x_0] + \sum_{j=1}^{M} \mathbb{1}\{X = x_j\}(E[Y|X = x_j] - E[Y|X = x_0]).$$

where $\boldsymbol{\delta} = (\delta_1, \delta_2, \delta_3, \delta_4)' \in \mathbb{R}^4$, $\tilde{\boldsymbol{D}}_{ig} = \left(1, D_{ig}, \sum_{j \neq i} D_{jg}, D_{ig} \sum_{j \neq i} D_{jg}\right)' \in \mathbb{R}^4$, *and*

$$
\begin{aligned}
\delta_1 &= \Delta\alpha + (N-1)\Delta m(0,0)\gamma_2, \\
\delta_2 &= \beta + (N-1)H(1,0)\gamma_2, \\
\delta_3 &= m_1(0,1)\gamma_1 + (H(0,1) - m_1(0,1))\gamma_2, \\
\delta_4 &= (m_1(1,1) - m_1(0,1))\gamma_1 + (H(1,1) - H(1,0) - H(0,1) - m_1(1,1) + m_1(0,1))\gamma_2.
\end{aligned}
$$

Further, if $E[\tilde{\boldsymbol{D}}_{ig}\tilde{\boldsymbol{D}}_{ig}']$ *is nonsingular, then* $\boldsymbol{\delta}$ *is identified by a least squares estimand:* $\boldsymbol{\delta} = E[\tilde{\boldsymbol{D}}_{ig}\tilde{\boldsymbol{D}}_{ig}']^{-1} E[\tilde{\boldsymbol{D}}_{ig}\Delta Y_{ig}]$.

For $E[\tilde{\boldsymbol{D}}_{ig}\tilde{\boldsymbol{D}}_{ig}']$ to be nonsingular, we need between-group variation. Let $\bar{D}_g = \sum j = 1^N D_{jg}$. Therefore, as $\sum_{j \neq i} D_{jg} = \bar{D}_g - D_{ig}$, D_{ig} and $\sum_{j \neq i} D_{jg}$ are not linearly independent if only one group is available.

2.3.4 Parameters in the Response Function and the Decomposition

$\boldsymbol{\delta}$ in (6) has no specific causal interpretations yet. However, it can be used to recover causal parameters. Define the following matrix

$$
\begin{aligned}
\boldsymbol{H} &= \begin{pmatrix}
1 & 0 & 0 & (N-1)\Delta m(0,0) \\
0 & 1 & 0 & (N-1)H(1,0) \\
0 & 0 & m_1(0,1) & H(0,1) - m_1(0,1) \\
0 & 0 & m_1(1,1) - m_1(0,1) & H(1,1) - H(1,0) - H(0,1) - m_1(1,1) + m_1(0,1)
\end{pmatrix}. \\
&= \begin{pmatrix}
1 & 0 & 0 & (N-1)\xi_1 \\
0 & 1 & 0 & (N-1)\xi_2 \\
0 & 0 & \omega_1 & \xi_3 - \omega_1 \\
0 & 0 & \omega_2 & \xi_4 - \omega_2
\end{pmatrix}
\end{aligned} \tag{7}
$$

Note that $\boldsymbol{H}$ is computed using $\boldsymbol{\omega}$ (or $\boldsymbol{\zeta}$) and $\boldsymbol{\xi}$ defined in (5) and (4), which are identified by Proposition 1. Furthermore, Proposition 2 implies that $\boldsymbol{\delta} = \boldsymbol{H}\boldsymbol{\theta}$, where $\boldsymbol{\theta} = (\Delta\alpha, \beta, \gamma_1, \gamma_2)$ represents the parameters in the response function, as in Assumption 2. Consequently, if $\boldsymbol{H}$ is invertible, $\boldsymbol{\theta}$ is recovered. Further, once $\boldsymbol{\theta}$ is identified, the

causal parameters $\boldsymbol{\pi} = (\pi^{DT}, \pi^{DN}, \pi^{IT}, \pi^{IN})'$ defined in Section 2.2 are also identified by using the identified parameters $(\boldsymbol{\theta}, \boldsymbol{\delta}, \boldsymbol{\zeta}, \boldsymbol{\xi})$. Proposition 3 summarizes this result of identification, representing the main contribution of this study.

Proposition 3 (Identification of Decompositions on Treatment Effects)**.** *Suppose* $\boldsymbol{\delta} = (\delta_1, \delta_1, \delta_2, \delta_3)'$ *in (6),* $\boldsymbol{\xi}$ *in (4), and* $\boldsymbol{\omega}$ *in (5) are identified. Then, the parameters* $\boldsymbol{\theta} := (\Delta\alpha, \beta, \gamma_1, \gamma_2)'$ *are identified by* $\boldsymbol{\theta} = \boldsymbol{H}^{-1}\boldsymbol{\delta}$*, where* $\boldsymbol{H}$ *is defined in (7), provided that* $\omega_1\xi_4 \neq \omega_2\xi_3$ *(i.e.,* $\boldsymbol{H}$ *is invertible). In addition, the parameters of decompositions defined in Section 2.2 are identified by*

$$\boldsymbol{\pi} = \begin{pmatrix} \pi^{DT} \\ \pi^{DN} \\ \pi^{IT} \\ \pi^{IN} \end{pmatrix} = \begin{pmatrix} \beta \\ \gamma_2(N-1)H(1,0) \\ (\gamma_1 - \gamma_2)m_1(0,1) \\ \gamma_2 H(0,1) \end{pmatrix} = \begin{pmatrix} \theta_2 \\ (N-1)\theta_4\xi_2 \\ (\theta_3 - \theta_4)\omega_1 \\ \theta_4\xi_3 \end{pmatrix}. \tag{8}$$

Remark (Identification Under Randomized Experiment)**.** Consider a situation when we only have a post-treatment data. Additionally suppose the treatment is exogenous, i.e., $\{(\varepsilon_{ig}(d), A_{ij1g}(d,e))\}$ are independent of $\boldsymbol{D}_g$. Recall that $H(d,e) = E[A_{ij1}(d,e) - A_{ij1}(0,0)|D_i = d, D_j = e]$. In this setting, the second term is identified as $E[A_{ij1}(0,0)|D_i = d, D_j = e] = E[A_{ij1}|D_i = 0, D_j = 0] = m_1(0,0)$. Therefore, $H(d,e) = m_1(d,e) - m_1(0,0)$ and we have $\boldsymbol{\zeta} = \boldsymbol{\xi}$. Similarly, the coefficient $\boldsymbol{\delta}_1$ of the outcome regression $E[Y_{i1g}|\boldsymbol{D}] = \tilde{\boldsymbol{D}}_{ig}'\boldsymbol{\delta}_1$ is written as $\boldsymbol{\delta}_1 = \boldsymbol{H}_1\boldsymbol{\theta}$, where

$$\boldsymbol{H}_1 = \begin{pmatrix} 1 & 0 & 0 & (N-1)\zeta_1 \\ 0 & 1 & 0 & (N-1)\zeta_2 \\ 0 & 0 & \omega_1 & \zeta_3 - \omega_1 \\ 0 & 0 & \omega_2 & \zeta_4 - \omega_2 \end{pmatrix}.$$

As a result, $\boldsymbol{\pi}$ is identified by the same transformation replacing $\boldsymbol{H}$ to $\boldsymbol{H}_1$, and $\boldsymbol{\xi}$ to $\boldsymbol{\zeta}$.

Remark (Identification Under Fixed Network)**.** If links are not affected by treatment, then $H(d,e) = 0$, indicating the absence of network effects. Therefore, we don't need to

estimate dyadic regression to estimate effects on links. The direct and indirect treatment effects are identified by the coefficients in the outcome regression as $\pi^{DT} = \beta = \delta_2$, and $\pi^{IT} = (\gamma_1 - \gamma_2)m_1(0,1) = \delta_3$.

Remark (Identification Without Interactions)**.** If $A_{ijtg} = 0$ for all i, j, t, g, then the only parameter of interest is β, which is identified by the canonical difference-in-differences estimand.

3 Estimation and Inference

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

3.1 Estimators

I propose a three-stage procedure to estimate the parameters of interest. The first-stage estimators estimate the distributions of potential links $(\boldsymbol{\omega}, \boldsymbol{\xi})$ and the coefficient of outcome regression $(\boldsymbol{\delta})$. Subsequently, the parameters in the response function $(\boldsymbol{\theta})$ in Assumption 2 are estimated in the second stage, by using the first-stage estimates. Finally, the decompositions of causal effects $(\boldsymbol{\pi})$ are estimated in the third stage, by using the estimates from the first and second stages. To this end, define the following

notations:

$$\tilde{\boldsymbol{W}}_{ijg} := (1, D_{ig}, D_{jg}, D_{ig}D_{jg})' \in \mathbb{R}^4, \qquad \tilde{\boldsymbol{W}}_g := (\tilde{\boldsymbol{W}}_{11g}, ..., \tilde{\boldsymbol{W}}_{NNg})' \in \mathbb{R}^{N(N-1)\times 4},$$

$$\tilde{\boldsymbol{D}}_{ig} := \left(1, D_{ig}, \sum_{j\neq i} D_{jg}, D_{ig}, \sum_{j\neq i} D_{jg}\right)' \in \mathbb{R}^4, \qquad \tilde{\boldsymbol{D}}_g := (\tilde{\boldsymbol{D}}_{1g}, ..., \tilde{\boldsymbol{D}}_{Ng})' \in \mathbb{R}^{N\times 4},$$

$$\boldsymbol{A}_{tg} := (A_{11tg}, ..., A_{NNg})' \in \mathbb{R}^{N(N-1)},$$

$$\boldsymbol{Y}_{tg} := (Y_{1tg}, ..., Y_{Ntg})' \in \mathbb{R}^N.$$

3.1.1 First-Stage Estimators for $\boldsymbol{\zeta}, \boldsymbol{\xi}$

The first-stage parameters, denoted as $\boldsymbol{\xi}_0, \boldsymbol{\omega}_0, \boldsymbol{\delta}_0$, are defined in (4), (5), and (6). They are estimated using the following least squares estimators:

$$\hat{\boldsymbol{\omega}} = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix} \left[\frac{1}{G}\sum_{g=1}^{G} \tilde{\boldsymbol{W}}'_g \tilde{\boldsymbol{W}}'_g\right]^{-1} \frac{1}{G}\sum_{g=1}^{G} \tilde{\boldsymbol{W}}'_g \boldsymbol{A}_{1g},$$

$$\hat{\boldsymbol{\xi}} = \left[\frac{1}{G}\sum_{g=1}^{G} \tilde{\boldsymbol{W}}'_g \tilde{\boldsymbol{W}}'_g\right]^{-1} \frac{1}{G}\sum_{g=1}^{G} \tilde{\boldsymbol{W}}'_g (\boldsymbol{A}_{1g} - \boldsymbol{A}_{0g}),$$

$$\hat{\boldsymbol{\delta}} = \left[\frac{1}{G}\sum_{g=1}^{G} \tilde{\boldsymbol{D}}'_g \tilde{\boldsymbol{D}}'_g\right]^{-1} \frac{1}{G}\sum_{g=1}^{G} \tilde{\boldsymbol{D}}'_g (\boldsymbol{Y}_{1g} - \boldsymbol{Y}_{0g}).$$

The estimator $\hat{\boldsymbol{\omega}}$ estimates the conditional expectations of links at $t = 1$. On the other hand, $\hat{\boldsymbol{\xi}}$ is the difference-in-differences estimator for links, using dyadic data to estimate the average treatment effect of the treated on network links. $\hat{\boldsymbol{\delta}}$ represents the coefficient of the conditional mean model of the outcome in (6).

3.1.2 Second-Stage Estimator for $\boldsymbol{\theta}$

Once these estimators, $\hat{\boldsymbol{\omega}}$, $\hat{\boldsymbol{\xi}}$, and $\hat{\boldsymbol{\delta}}$, are computed, the parameter $\boldsymbol{\theta}_0$ is estimated by $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{H}}^{-1}\hat{\boldsymbol{\delta}}$, where

$$\hat{\boldsymbol{H}} = \begin{pmatrix} 1 & 0 & 0 & \hat{\xi}_1 \\ 0 & 1 & 0 & \hat{\xi}_2 \\ 0 & 0 & \hat{\omega}_1 & \hat{\xi}_3 - \hat{\omega}_1 \\ 0 & 0 & \hat{\omega}_2 & \hat{\xi}_4 - \hat{\omega}_2 \end{pmatrix}.$$

3.1.3 Third-Stage Estimator for $\boldsymbol{\pi}$

Lastly, the decompositions $\boldsymbol{\pi}_0$ is estimated by the following plug-in estimator:

$$\hat{\boldsymbol{\pi}} = \begin{pmatrix} \hat{\delta}_2 \\ (N-1)\hat{\gamma}_2\hat{\xi}_2 \\ (\hat{\delta}_3 - \hat{\delta}_4)\hat{\omega}_1 \\ \hat{\delta}_4\hat{\xi}_3 \end{pmatrix}.$$

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

3.2 Inference

Since the proposed estimators are least squares estimators for projection coefficients, standard large sample theories can be applied. The random sample of independent groups plays a key role in the application of asymptotic theories. Additionally, the following assumptions are required as standard regularity conditions.

Assumption 7 (Regularity Conditions)**.** *(i)* $E[\tilde{\boldsymbol{W}}_g\tilde{\boldsymbol{W}}_g']$, $E[\tilde{\boldsymbol{D}}_g\tilde{\boldsymbol{D}}_g']$ *are nonsingular; (ii)* $E\left[|Y_{itg}|^4\right] < \infty$.

The first part of Assumption 7 is a regularity condition to ensure the uniqueness and thus identification of parameters as the projection coefficients. The second part is

employed to derive the asymptotic distribution, as appropriate bounded moments are required. Since indicator variables are bounded, only the boundedness of the outcome is stated. Let "$\Rightarrow \mathcal{N}$" denote convergence in distribution to some multivariate normal distribution. Proposition 4 summarizes the influence functions for parameters in each stage:

Proposition 4. *Let* $\boldsymbol{V}_g = (\tilde{\boldsymbol{W}}_g, \boldsymbol{A}_{1g}, \boldsymbol{A}_{0g}, \boldsymbol{Y}_{1g}, \boldsymbol{Y}_{0g}, \tilde{\boldsymbol{D}}_g)$, $\boldsymbol{M} = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix}$, $\boldsymbol{Q} = E[\tilde{\boldsymbol{W}}_g \tilde{\boldsymbol{W}}_g']$, *and* $\boldsymbol{R} = E[\tilde{\boldsymbol{D}}_g \tilde{\boldsymbol{D}}_g']$. *Under Assumptions 1-4, and 7, we have*

$$\sqrt{G}\begin{pmatrix} \hat{\boldsymbol{\omega}} - \boldsymbol{\omega}_0 \\ \hat{\boldsymbol{\xi}} - \boldsymbol{\xi}_0 \\ \hat{\boldsymbol{\delta}} - \boldsymbol{\delta}_0 \\ \hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 \\ \hat{\boldsymbol{\pi}} - \boldsymbol{\pi}_0 \end{pmatrix} = \frac{1}{\sqrt{G}} \sum_{g=1}^{G} \begin{pmatrix} \boldsymbol{M}\psi_{\boldsymbol{\zeta}}(\boldsymbol{V}_g) \\ \psi_{\boldsymbol{\xi}}(\boldsymbol{V}_g) \\ \psi_{\boldsymbol{\delta}}(\boldsymbol{V}_g) \\ \psi_{\boldsymbol{\theta}}(\boldsymbol{V}_g) \\ \psi_{\boldsymbol{\pi}}(\boldsymbol{V}_g) \end{pmatrix} + o_p(1) \Rightarrow \mathcal{N},$$

where $\psi_{\boldsymbol{\zeta}}(\boldsymbol{V}_g) = \boldsymbol{Q}^{-1}\tilde{\boldsymbol{W}}_g'(\boldsymbol{A}_{1g} - \tilde{\boldsymbol{W}}_g\boldsymbol{\zeta}_0)$, $\psi_{\boldsymbol{\xi}}(\boldsymbol{V}_g) = \boldsymbol{Q}^{-1}\tilde{\boldsymbol{W}}_g'(\Delta\boldsymbol{A}_g - \tilde{\boldsymbol{W}}_g\boldsymbol{\xi}_0)$, $\psi_{\boldsymbol{\delta}}(\boldsymbol{V}_g) = \boldsymbol{R}^{-1}\tilde{\boldsymbol{D}}_g'(\Delta\boldsymbol{Y}_g - \tilde{\boldsymbol{D}}_g\boldsymbol{\delta}_0)$, $\psi_{\boldsymbol{\theta}} = \boldsymbol{H}_0^{-1}(\psi_{\boldsymbol{\delta}} - \psi_{\boldsymbol{H}}\boldsymbol{\theta}_0)$, *and*

$$\psi_{\boldsymbol{H}} = \begin{pmatrix} 1 & 0 & 0 & \psi_{\xi_1} \\ 0 & 1 & 0 & \psi_{\xi_2} \\ 0 & 0 & \psi_{\omega_1} & \psi_{\xi_3} - \psi_{\omega_1} \\ 0 & 0 & \psi_{\omega_2} & \psi_{\xi_4} - \psi_{\omega_2} \end{pmatrix}, \psi_{\boldsymbol{\pi}} = \begin{pmatrix} \psi_{\theta_2} \\ (N-1)(\psi_{\theta_4} H(1,0) + \gamma_2\psi_{\xi_2}) \\ (\psi_{\theta_3} - \psi_{\theta_4})\, m_1(0,1) + (\gamma_1 - \gamma_2)\psi_{\omega_1} \\ \psi_{\theta_3} H(0,1) + \gamma_2\psi_{\xi_3} \end{pmatrix}.$$

Proposition 4 implies that $\hat{\boldsymbol{\pi}}$ has asymptotic normal distribution with zero mean and the asymptotic variance $\text{Avar}(\sqrt{G}(\hat{\boldsymbol{\pi}} - \boldsymbol{\pi}_0)) = E[\psi_{\boldsymbol{\pi}}(\boldsymbol{V}_g)\psi_{\boldsymbol{\pi}}(\boldsymbol{V}_g)']$. Let $\hat{\psi}_{\boldsymbol{\pi}}$ be the empirical influence function (plug-in estimator) for $\psi_{\boldsymbol{\pi}}$. Then, the clustered standard error is computed as the square root of the diagonal elements of the following matrix:

$$\frac{1}{G^2}\sum_{g=1}^{G} \hat{\psi}_{\boldsymbol{\pi}}(\boldsymbol{Z}_g)\hat{\psi}_{\boldsymbol{\pi}}(\boldsymbol{Z}_g)'. \tag{9}$$

Note that $\frac{1}{G}\sum_{g=1}^{G} \hat{\psi}_{\boldsymbol{\pi}}(\boldsymbol{Z}_g)\hat{\psi}_{\boldsymbol{\pi}}(\boldsymbol{Z}_g)' \xrightarrow{p} E[\psi_{\boldsymbol{\pi}}(\boldsymbol{Z}_g)\psi_{\boldsymbol{\pi}}(\boldsymbol{Z}_g)']$, and therefore the inference based on the clustered standard error is asymptotically valid. Other clustered standard

errors are defined similarly for the other parameters.

4 Monte Carlo Simulation

To investigate the finite and asymptotic characteristics of the estimators intoduced in Section 3, I conduct simulations using data generated based on the Assumptions studied in Section 2 with different sample sizes.

First, the treatment indicators D_{ig} are generated by a Bernoulli distribution with the probability $P_D = 0.5$. Next, the potential network links are generated according to the following binary choice model: For each $(i,j) \in \{1,...,N\}^2$, $t \in \{0,1\}$, $g \in \{1,...,G\}$, $A_{iitg}(d,e) = 0$ for all $(d,e) \in \{0,1\}^2$, and $A_{ijtg}(d,e) = \mathbb{1}\{f_t(d,e) \geq u_{ijtg}\}$ for $i\ nej$, where $f_t(d,e) = (1,d,e,de)\boldsymbol{a}_t$, $\boldsymbol{a}_0 = (0.1,0,0,0)'$, $\boldsymbol{a}_1 = (-0.3,0.5,0.1,0.6)'$, and $u_{ijtg} = v_{ijg} - g_t(D_{ig}, D_{jg})$, where $v_{ijg} \sim F_v$. Consequently, we have $E[A_{ijtg}(d,e)|D_{ig} = d', D_{jg} = e'] = F_v(f_t(d,e) + g_t(d',e'))$. Assumption 5 is satisfied as $f_0(d,e) = f_0(0,0)$ for all $(d,e) \in \{0,1\}^2$. The values for function g are set by $g_0(1,1) = 0.4$, $g_0(1,0) = 0.3$, $g_0(0,1) = 0.2$, $g_0(0,0) = 0.1$, $g_1(0,0) = 0.01$, and the remaining values are constructed for the potential links to satisfy Assumption 6.[6] Lastly, the outcome is generated from the response function defined in Assumption 2 :

$$Y_{i1g} = \alpha_1 + \beta D_{ig} + \gamma_1 \sum_{j=1}^{N} A_{ij1g} D_{jg} + \gamma_2 \sum_{j=1}^{N} A_{ij1g}(1 - D_{jg}) + u_{i1g},$$

$$Y_{i0g} = \alpha_2 + \gamma_2 \sum_{j=1}^{N} A_{ij0g} + u_{i0g},$$

where $u_{itg} \sim N(0,1) + s_b(D_{ig} - P_D)$ to consider both exogenous and endogenous treatment. The parameters of the response function are set by $\theta := (\alpha_0, \alpha_1, \beta, \gamma_1, \gamma_2) = (1, 1.2, 5, 0.6, 0.3)$.

The generated data consists of a dyadic level data $\{D_{ig}D_{jg}, A_{ij0g}, A_{ij1g}\}_{(i,j):i\neq j,g}$,

[6] across all $(d,e) \in \{0,1\}^2$, then the values of g_1 must also be same for Assumption 6, and hence u_{ijtg} independent of the treatment. Consequently, the difference-in-differences approach becomes unnecessary, as the differences in means are enough to identify the causal effect of treatment on potential links. To avoid this scenario, I initially specify the values of g_0 and $g_1(0,0)$ and then construct the remaining values as $g_1(d,e) = \Phi^{-1}(F_v(a_{10} + g_1(0,0)) - F_v(a_{00} + g_0(0,0)) + F_v(a_{00} + g_0(d,e)) - a_{10})$, which ensures that Assumption 6 is satisfied by construction.

and individual level data $\{D_{ig}, Y_{i0g}, Y_{i1g}\}_{i,g}$. The estimators defined in Section 3, and clustered standard errors for each coefficient are computed. The coverage rate is computed by the proportion of the cases that estimate is included in the 95% confidence interval, computed by the clusterd standard error, among all simulations. For instance, for a parameter $\phi \in \mathbb{R}^K$, let $\hat{\phi}_b$ be the estimator for a replication b, and $CI_b(\phi) = [\hat{\phi}_b \pm z_{\alpha/2} SE_b(\phi)]$, where $z_{\alpha/2} = \Phi^{-1}(1-\alpha/2)$, $\alpha = 0.05$. Then, MSE $= \frac{1}{B}\sum_{b=1}^{B}(\hat{\phi}_b - \phi_0)(\hat{\phi}_b - \phi_0)'$, and Coverage $= \frac{1}{B}\sum_{b=1}^{B} \mathbb{1}\{\phi_0 \in CI_b(\phi)\}$.

Table 1 and Table 2 display results for each parameters. The suggested asymptotic theory and the clustered standard errors work well, as all coverage rates are close to 95%, and the mean squared errors converges to zero. For detailed results on each parameter, see Table 5 and Table 6 in Appendix B.

Table 1: Root Mean Squared Errors

		Parameteres					
G	N	$\boldsymbol{\xi}$	$\boldsymbol{\zeta}$	$\boldsymbol{\omega}$	$\boldsymbol{\delta}$	$\boldsymbol{\theta}$	$\boldsymbol{\pi}$
100	15	0.014	0.019	0.011	0.387	2.781	2.55
500	15	0.006	0.009	0.005	0.162	1.096	1.004
1,000	15	0.005	0.006	0.003	0.121	0.82	0.751
5,000	15	0.002	0.003	0.001	0.055	0.377	0.344
10,000	15	0.001	0.002	0.001	0.037	0.247	0.226

notes. For each sample size N, G, the number of simulations is set to $B = 500$.

Table 2: Coverage Rates

		Parameteres					
G	N	$\boldsymbol{\xi}$	$\boldsymbol{\zeta}$	$\boldsymbol{\omega}$	$\boldsymbol{\delta}$	$\boldsymbol{\theta}$	$\boldsymbol{\pi}$
100	15	0.944	0.946	0.936	0.92	0.926	0.932
500	15	0.944	0.93	0.932	0.958	0.954	0.958
1,000	15	0.938	0.936	0.954	0.94	0.932	0.934
5,000	15	0.944	0.946	0.956	0.928	0.946	0.94
10,000	15	0.946	0.944	0.948	0.952	0.95	0.956

notes. For each sample size N, G, the number of simulations is set to $B = 500$.

5 Empirical Illustration

In this section, I apply the proposed estimation method to data from studies conducted by Comola and Prina (2020), Comola and Prina (2021). The dataset involves a randomized experiment carried out in villages surrounding Pokhara, Nepal, from 2009 to 2011. The treatment in this experiment was providing an access to a savings account to households. The pre-treatment survey took place in February 2009, and the treatment was randomly assigned to half of the households in June 2010 through a public lottery.

According to Prina (2015), formal banking services in Nepal are highly limited, with only 20% of households having a bank account. At the start of the experiment, only 17% of participants had savings accounts, with most keeping their cash at home. The experiment aimed to assess the impact of providing a savings account on economic behaviors, such as consumption. Specifically, the treatment involved offering 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 these actively using it.

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

Comola and Prina (2021) estimate a two-period version of the linear-in-means model using an IV estimation strategy similar to that of Bramoullé, Djebbari, and Fortin (2009). They compute the direct and indirect effects as derivatives of the reduced form outcome. The linear-in-means structure implies specific derivatives consisting of changes in links in response to the treatment. This relates to the average treatment effects on treated row-normalized links. However, the authors estimate a regression

of differenced row-normalized links on dyadic treatment, defined as *some* treated, i.e., $\max D_i, D_j$. In contrast, the estimation procedure proposed in this paper estimates the difference in links on D_i, D_j, and $D_i D_j$. Since the network is undirected, the coefficients of D_i and D_j are equivalent to that of $D_i + D_j$. Table 3 shows the results of these regressions. The first column presents the regression results from Comola and Prina (2021), where the coefficient of the dyadic treatment is estimated at 0.002. This means that the average change in row-normalized links (or the probability of being linked) increases by 0.002 percentage points in response to the dyadic treatment.

Table 3: Average Treatment Effects on Treated of Links

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

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

The second and third columns display the first-stage regression of the differenced link. The second column estimates the regression of differenced row-normalized links for comparison with the first column. Due to the undirected nature of the network, the coefficients for D_1 and D_2 are almost identical. Additionally, while the regression in the second column considers the interaction term, its coefficient estimating $H(1,1)$ is not significant, resulting in similar coefficients for D_1 and D_2 as in the first column.

The first-stage regression actually used in the proposed estimation procedure is in the third column. The coefficient estimates for D_1 and D_2 are about double those in the second column, with $H(1,0) = H(0,1)$ estimated at 0.004.

Table 4 illustrates how the causal effects are decomposed and estimated, by comparing with estimation result in Comola and Prina (2021). In Comola and Prina (2021), direct and indirect effects are computed as the derivative of the reduced form outcome with respect to the treatment vector, i.e., $\partial E[\boldsymbol{y}|\boldsymbol{D}]/\partial D_k$. Thus, the direct effects represent the average partial effect of one's own treatment, and the indirect effects represent the average partial effect of others' treatment. The linear-in-means structure implies a linear projection of $E[\boldsymbol{y}|\boldsymbol{D}]$. Therefore, these effects are challenging to interpret causally unless the potential outcome is additively separable with respect to the treatment statuses of all others. Additionally, both effects are mixed effects of treatment and network, i.e., $\pi^{\cdot T}$ and $\pi^{\cdot N}$. While Comola and Prina (2021) considers changes in the network driven by the intervention, their method cannot clearly decompose these two effects.

Table 4: Decomposition of Treatment Effects

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

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

The third column estimates the direct and indirect treatment and network effects $\boldsymbol{\pi}$. The direct treatment effect π^{DT} is estimated by 478.5, and the direct network effect is estimated by -243.2. Therefore, the total direct effect is 243.2, which is smaller than those estimated by Comola and Prina (2021). While the overall direct effect is positive, this result suggests that an opposite effect may exist due to changes in the network structure. Specifically, this can be interpreted as the offer of a savings account tending to increase consumption and enhance the financial network. However, increased financial links result in more savings and reducing the consumption. By contrast, the indirect effects are not significant.

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

Overall, the results presented in this section suggest that the method proposed in this paper can clearly decompose causal effects into pure treatment effects and those from causal changes in network links. In particular, this method will be useful when identifying treatment effects that have opposite directions, as illustrated in Table 3. However, the data requirements are substantial, as we need full information of the network links, and a sufficient number of independent groups.

6 Conclusion

This paper analyzes identification and estimation of the causal effects of programs, considering potential network changes resulting from the treatment. The method decomposes the treatment effect into two parts: the impact when the network remains

unchanged and the impact when the network structure changes. The approach is evaluated through a Monte Carlo study and illustrated with data from a study in Nepal conducted by Comola and Prina (2020). By introducing this method, the paper not only provides a novel way to estimate causal effects while considering network changes but also offers a decomposition that enhances understanding of the underlying mechanisms of the program.

The proposed method has two main limitations. First, the data requirements are significant, as full information about the network is typically costly to obtain. However, some inferences can still be made using only observed exposure and treatment assignments, although this is not addressed in the paper. For instance, if we can identify distributions of potential exposures, then the differences between those distributions will reflect the causal changes in the underlying networks. Second, the method relies on many restrictions, particularly on the response functions of links and outcomes. Potential links are assumed to be formed by a dyadic model, which could be extended to triadic or more complex interaction structures at the cost of higher dimensionality. The linear-in-sums response function for potential outcomes is also restrictive, though it can be relaxed by considering local interference up to a distance of 2 or 3. More crucially, the parametric response model imposes certain homogeneities; extending the outcome response to a general nonparametric or semi-parametric function would provide a more realistic model.

References

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

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

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

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

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

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

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

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

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

Chernozhukov, Victor, Juan Carlos Escanciano, Hidehiko Ichimura, Whitney K Newey, and James M Robins (2022). "Locally robust semiparametric estimation". In: *Econometrica* 90.4, pp. 1501–1535.

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

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

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

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

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

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

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

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

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

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

Ichimura, Hidehiko and Whitney K Newey (2022). "The influence function of semiparametric estimators". In: *Quantitative Economics* 13.1, pp. 29–61.

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

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

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

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

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

Leung, Michael P (2022). “Causal inference under approximate neighborhood interference”. In: *Econometrica* 90.1, pp. 267–293.

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

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

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

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

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

— (2022b). “Identification and estimation of spillover effects in randomized experiments”. In: *Journal of Econometrics*.

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

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

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

Appendix

A Proofs

Proof of Proposition 1. Let $(d, e) \in \{(0,1),(1,0),(1,1)\}$ be given. Then,

$$\begin{aligned}
&E[A_{ij1g}(0,0)|D_{ig}=d, D_{jg}=e] \\
&= E[A_{ij0g}(0,0)|D_{ig}=d, D_{jg}=e] + E[A_{ij1g}(0,0) - A_{ij0g}(0,0)|D_{ig}=0, D_{jg}=0] \\
&= E[A_{ij0g}(d,e)|D_{ig}=d, D_{jg}=e] + E[A_{ij1g} - A_{ij0g}|D_{ig}=0, D_{jg}=0] \\
&= E[A_{ij0g}|D_{ig}=d, D_{jg}=e] + E[A_{ij1g} - A_{ij0g}|D_{ig}=0, D_{jg}=0], \qquad (10)
\end{aligned}$$

where the first equation is by Assumption 6-2, and the second equation is by Assumption 5-2. It follows that

$$\begin{aligned}
H(d,e) &= E[A_{ij1g}(d,e) - A_{ij1g}(0,0)|D_{ig}=d, D_{jg}=e] \\
&= E[A_{ij1g}(d,e)|D_{ig}=d, D_{jg}=e] - E[A_{ij1g}(0,0)|D_{ig}=d, D_{jg}=e] \\
&= E[A_{ij1g}|D_{ig}=d, D_{jg}=e] - E[A_{ij0g}|D_{ig}=d, D_{jg}=e] \\
&\quad - E[A_{ij1g} - A_{ij0g}|D_{ig}=0, D_{jg}=0] \\
&= E[A_{ij1g} - A_{ij0g}|D_{ig}=d, D_{jg}=e] - E[A_{ij1g} - A_{ij0g}|D_{ig}=0, D_{jg}=0],
\end{aligned}$$

and $H_g(0,0) = 0$ by construction. Assumption Assumption 4 guarantees the existence of conditional expectations.

□

Lemma 1. *Let $\boldsymbol{D} = (D_1, ..., D_N) \in \{0,1\}^N$ be a random vector, and $Y_{it} : \{0,1\}^N \to \mathbb{R}$, $f_{ijt} : \{0,1\}^2 \to \mathbb{R}$, $h_{it} : \{0,1\} \to \mathbb{R}$ are functions. Assume*

- *(LR) $Y_{it}(\boldsymbol{d}) = h_{it}(d_i) + \sum_{j=1, j\neq i}^{n} f_{ijt}(d_i, d_j)$, for all $(d_i, d_j) \in \{0,1\}^2$.*
- *(LE-h) $E[h_{it}(d)|\boldsymbol{D}] = E[h_{it}(d)|D_i]$, for all $d \in \{0,1\}$.*
- *(LE-f) $E[f_{ijt}(d,e)|\boldsymbol{D}] = E[f_{ijt}(d,e)|D_i, D_j]$, for all $(d,e) \in \{0,1\}^2$.*

Then, $E[Y_{i1} - Y_{i0}|\boldsymbol{D}] = \tilde{\boldsymbol{D}}_i'\boldsymbol{\theta}$*, where*

$$\tilde{\boldsymbol{D}}_i = \begin{pmatrix} 1 \\ D_i \\ \sum_{j-1,j\neq i}^{n} D_j \\ D_i \sum_{j-1,j\neq i}^{n} D_j \end{pmatrix}, \quad \boldsymbol{\theta} = \begin{pmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ \theta_3 \end{pmatrix} = \begin{pmatrix} \alpha + \sum_{j=1,j\neq i}^{n} K_{ij}(0,0) \\ \beta + \sum_{j=1,j\neq i}^{n} (K_{ij}(1,0) - K_{ij}(0,0)) \\ K_{ij}(0,1) - K_{ij}(0,0) \\ K_{ij}(1,1) - K_{ij}(1,0) - K_{ij}(0,1) + K_{ij}(0,0) \end{pmatrix},$$

and $K_{ij}(d,e) := E[f_{ij1} - f_{ij0}|D_i = d, D_j = e]$ *for* $(d,e) \in \{0,1\}^2$*.*

Proof of Lemma 1. Let $\alpha := E[h_{i1} - h_{i0}|D_i = 0]$ and $\beta := E[h_{i1} - h_{i0}|D_i = 1] - E[h_{i1} - h_{i0}|D_i = 0]$. Then, Assumption (LE-h) implies

$$E[h_{i1} - h_{i0}|\boldsymbol{D}] = E[h_{i1} - h_{i0}|D_i] = \alpha + \beta D_i. \tag{11}$$

Next, let $K_{ij} = K_{ij}(D_i, D_j) = E[f_{ij1} - f_{ij0}|D_i, D_j]$. Then, (LE-f) implies

$$\begin{aligned} E[f_{ij1} - f_{ij0}|\boldsymbol{D}] = K_{ij} &= \sum_{(d,e)\in\{0,1\}^2} K_{ij}(d,e) \\ &= K_{ij}(0,0) + D_i(K_{ij}(1,0) - K_{ij}(0,0)) \\ &\quad + D_j(K_{ij}(0,1) - K_{ij}(0,0)) \\ &\quad + D_iD_j(K_{ij}(1,1) - K_{ij}(1,0) - K_{ij}(0,1) + K_{ij}(0,0)). \end{aligned} \tag{12}$$

By combining (LR), (11), and (12), we have

$$\begin{aligned} E[Y_{i1} - Y_{i0}|\boldsymbol{D}] &= E[h_{i1} - h_{i0}|\boldsymbol{D}] + \sum_{j=1,j\neq i}^{n} E[f_{ij1} - f_{ij0}|\boldsymbol{D}] \\ &= \alpha + \sum_{j=1,j\neq i}^{n} K_{ij}(0,0) \\ &\quad + D_i\left(\beta + \sum_{j=1,j\neq i}^{n} (K_{ij}(1,0) - K_{ij}(0,0))\right) + \sum_{j=1,j\neq i}^{n} D_j(K_{ij}(0,1) - K_{ij}(0,0)) \\ &\quad + D_i \sum_{j=1,j\neq i}^{n} D_j(K_{ij}(1,1) - K_{ij}(1,0) - K_{ij}(0,1) + K_{ij}(0,0)) = \tilde{\boldsymbol{D}}_i'\boldsymbol{\theta}. \end{aligned}$$

□

Proof of Proposition 2. Let

$$f_{ijtg}(d_i, d_j) = \begin{cases} \gamma_1 A_{ij1g}(d_i, d_j) d_j + \gamma_2 A_{ij1g}(d_i, d_j)(1 - d_j) & t = 1, \\ \gamma_3 A_{ij0g}(d_i, d_j) & t = 0. \end{cases}$$

$$h_{it}(d) = \begin{cases} \alpha_{1g} + \gamma_0 d_i + \varepsilon_{i1g} & t = 1, \\ \alpha_{0g} + \varepsilon_{i0g} & t = 0. \end{cases}$$

Then, Assumptions 2, 1, 3 imply Assumptions in the Lemma 1. Also,

$$f_{ij1g}(d, e) - f_{ij0g}(d, e) = A_{ij1g}(d, e)\left(\gamma_2 + (\gamma_1 - \gamma_2)e\right) - \gamma_3 A_{ij0g}(d, e).$$

It follows that

$$\begin{aligned} &E[f_{ij1g}(0,0) - f_{ij0g}(0,0)|D_i = d, D_j = e] \\ &= E[A_{ij1g}(0,0)|D_i = d, D_j = e]\gamma_2 - E[A_{ij0g}(0,0)|D_i = d, D_j = e]\gamma_3, \\ &= E[A_{ij1g}(0,0) - A_{ij0g}(0,0)|D_i = d, D_j = e]\gamma_2 + E[A_{ij0g}(0,0)|D_i = d, D_j = e](\gamma_2 - \gamma_3). \end{aligned}$$

by Assumption 6, the first term doesn't depend on d, e. Hence, for the potential outcome to satisfy Assumption 6, we need to assume $\gamma_2 = \gamma_3$ or $E[A_{ij0g}(0,0)|D_i = d, D_j = e] = E[A_{ij0g}(0,0)]$ for all $(d, e) \in \{0, 1\}^2$.

Case 1. Assume $\gamma_2 = \gamma_2$.

In this case, it follows that

$$f_{ij1g}(d, e) - f_{ij0g}(d, e) = (A_{ij1g}(d, e) - A_{ij0g}(d, e))\gamma_2 + A_{ij1g}(d, e)(\gamma_1 - \gamma_2)e,$$

and therefore

$$\begin{aligned} K_{ij}(d, e) &= \Delta m(d, e)\gamma_2 + m_1(d, e)(\gamma_1 - \gamma_2)e, \\ K_{ij}(d, e) - K_{ij}(0, 0) &= H(d, e)\gamma_2 + m_1(d, e)(\gamma_1 - \gamma_2)e. \end{aligned}$$

Thus we have the desired result

$$\boldsymbol{\theta} = \begin{pmatrix} \alpha + (N-1)\Delta m(0,0)\gamma_2 \\ \beta + (N-1)H(1,0)\gamma_2 \\ H(0,1)\gamma_2 + m_1(0,1)(\gamma_1 - \gamma_2) \\ (H(1,1) - H(1,0) - H(0,1))\gamma_2 + (m_1(1,1) - m_1(0,1))(\gamma_1 - \gamma_2) \end{pmatrix}$$
$$= \begin{pmatrix} 1 & 0 & 0 & (N-1)\Delta m(0,0) \\ 0 & 1 & 0 & (N-1)H(1,0) \\ 0 & 0 & m_1(0,1) & H(0,1) - m_1(0,1) \\ 0 & 0 & m_1(1,1) - m_1(0,1) & m_0(1,1) - H(1,0) - m_0(0,1) \end{pmatrix} \begin{pmatrix} \Delta\alpha \\ \beta \\ \gamma_1 \\ \gamma_2 \end{pmatrix}.$$

Case 2. Assume $E[A_{ij0g}(0,0)|D_i = d, D_j = e] = E[A_{ij0g}(0,0)]$ for all $(d,e) \in \{0,1\}^2$.

In this case,

$$\begin{aligned} K_{ij}(d,e) &= m_1(d,e)\gamma_2 + m_1(d,e)(\gamma_1 - \gamma_2)e - m_0\gamma_3 \\ &= \Delta m(d,e)\gamma_2 + m_1(d,e)(\gamma_1 - \gamma_2)e + (\gamma_2 - \gamma_3)m_0, \\ K_{ij}(d,e) - K_{ij}(0,0) &= H(d,e)\gamma_2 + m_1(d,e)(\gamma_1 - \gamma_2)e. \end{aligned}$$

Thus we have

$$\boldsymbol{\theta} = \begin{pmatrix} \alpha + (N-1)m_1(0,0)\gamma_2 - (N-1)m_0\gamma_3 \\ \beta + (N-1)H(1,0)\gamma_2 \\ H(0,1)\gamma_2 + m_1(0,1)(\gamma_1 - \gamma_2) \\ (H(1,1) - H(1,0) - H(0,1))\gamma_2 + (m_1(1,1) - m_1(0,1))(\gamma_1 - \gamma_2) \end{pmatrix}$$
$$= \begin{pmatrix} 1 & 0 & 0 & (N-1)m_1(0,0) \\ 0 & 1 & 0 & (N-1)H(1,0) \\ 0 & 0 & m_1(0,1) & H(0,1) - m_1(0,1) \\ 0 & 0 & m_1(1,1) - m_1(0,1) & m_0(1,1) - H(1,0) - m_0(0,1) \end{pmatrix} \begin{pmatrix} \Delta\alpha - (N-1)m_0\gamma_3 \\ \beta \\ \gamma_1 \\ \gamma_2 \end{pmatrix}.$$

As a result, $\Delta\alpha$ and γ_3 are not separately identified in this case. However, $\boldsymbol{\pi}$ is identified by the same transformation $\boldsymbol{H}$. □

Proof of Proposition 3. This follows directly by invertibility of $\boldsymbol{H}$. □

Proof of Proposition 4. Let $\boldsymbol{Q} = E[\tilde{\boldsymbol{W}}_g'\tilde{\boldsymbol{W}}_g]$, $\hat{\boldsymbol{Q}} = \frac{1}{G}\sum_{g=1}^{G}\tilde{\boldsymbol{W}}_g'\tilde{\boldsymbol{W}}_g$, $\boldsymbol{R} = E[\tilde{\boldsymbol{D}}_g'\tilde{\boldsymbol{D}}_g]$ and $\hat{\boldsymbol{R}} = \frac{1}{G}\sum_{g=1}^{G}\tilde{\boldsymbol{D}}_g'\tilde{\boldsymbol{D}}_g$. Then, $\hat{\boldsymbol{Q}}^{-1} = \boldsymbol{Q}^{-1} + o_p(1)$ and $\hat{\boldsymbol{R}}^{-1} = \boldsymbol{R}^{-1} + o_p(1)$ by a law of large numbers and the continuous mapping theorem, since Q, E are nonsingular and data is i.i.d. across groups. Hence,

$$\begin{aligned}
\sqrt{G}(\hat{\boldsymbol{\zeta}} - \boldsymbol{\zeta}_0) &= \frac{1}{\sqrt{G}}\sum_{g=1}^{G}\boldsymbol{Q}^{-1}\tilde{\boldsymbol{W}}_g'\varepsilon_g^{(\boldsymbol{\zeta})} + o_p(1)\frac{1}{\sqrt{G}}\sum_{g=1}^{G}\tilde{\boldsymbol{W}}_g'\varepsilon_g^{(\boldsymbol{\zeta})},\\
\sqrt{G}(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}_0) &= \frac{1}{\sqrt{G}}\sum_{g=1}^{G}\boldsymbol{Q}^{-1}\tilde{\boldsymbol{W}}_g'\varepsilon_g^{(\boldsymbol{\xi})} + o_p(1)\frac{1}{\sqrt{G}}\sum_{g=1}^{G}\tilde{\boldsymbol{W}}_g'\varepsilon_g^{(\boldsymbol{\xi})},\\
\sqrt{G}(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}_0) &= \frac{1}{\sqrt{G}}\sum_{g=1}^{G}\boldsymbol{R}^{-1}\tilde{\boldsymbol{D}}_g'\varepsilon_g^{(\boldsymbol{\delta})} + o_p(1)\frac{1}{\sqrt{G}}\sum_{g=1}^{G}\tilde{\boldsymbol{D}}_g'\varepsilon_g^{(\boldsymbol{\delta})},
\end{aligned}$$

where

$$\begin{aligned}
\varepsilon_g^{(\boldsymbol{\zeta})} &= \boldsymbol{A}_{1g} - \tilde{\boldsymbol{W}}_g\boldsymbol{\zeta}_0,\\
\varepsilon_g^{(\boldsymbol{\xi})} &= \Delta\boldsymbol{A}_g - \tilde{\boldsymbol{W}}_g\boldsymbol{\xi}_0,\\
\varepsilon_g^{(\boldsymbol{\delta})} &= \Delta\boldsymbol{Y}_g - \tilde{\boldsymbol{D}}_g\boldsymbol{\delta}_0.
\end{aligned}$$

Data are i.i.d. across groups, and by the condition $E\left[|Y_{itg}|^4\right] < \infty$, the second term in each equation is $O_p(1)$. Therefore the influence functions for $\boldsymbol{\zeta}, \boldsymbol{\xi}, \boldsymbol{\delta}$ are given by

$$\begin{aligned}
\psi_{\boldsymbol{\zeta}}(\boldsymbol{Z}_g) &= \boldsymbol{Q}^{-1}\tilde{\boldsymbol{W}}_g'\varepsilon_g^{(\boldsymbol{\zeta})},\\
\psi_{\boldsymbol{\xi}}(\boldsymbol{Z}_g) &= \boldsymbol{Q}^{-1}\tilde{\boldsymbol{W}}_g'\varepsilon_g^{(\boldsymbol{\xi})},\\
\psi_{\boldsymbol{\delta}}(\boldsymbol{Z}_g) &= R^{-1}\tilde{\boldsymbol{D}}_g'\varepsilon_g^{(\boldsymbol{\delta})}.
\end{aligned}$$

Because $\boldsymbol{\omega} = \boldsymbol{M}\boldsymbol{\zeta}$, we have $\psi_{\boldsymbol{\omega}} = \boldsymbol{M}\psi_{\boldsymbol{\zeta}}$ by delta method. Next, let $\boldsymbol{Z}_g \sim F$. When an estimand ψ is considered as a functional on the space of distributions that maps $F \mapsto \psi(F)$, then the influence function of that estimand is obtained by the Gáteaux derivative of the functional at the true distribution of data, on the direction to dirac measure to a specific observation (e.g., Ichimura and Newey (2022) Chernozhukov et al. (2022)), i.e., $\psi_\phi = \phi'(F)(\delta_{\boldsymbol{Z}_g} - F)$. Thus, the influence function for $\boldsymbol{H}$ is the matrix valued function in which each element is replaced by its influence function. Also,

note that we have $\boldsymbol{H}\boldsymbol{\theta} = \boldsymbol{\delta}$, and this holds for all underlying distribution of data by construction. Thus, it follows from the chain rule of Gáteaux derivative that

$$\psi_{\boldsymbol{H}}\boldsymbol{\theta}_0 + \boldsymbol{H}_0\psi_{\boldsymbol{\theta}} = \psi_{\boldsymbol{\delta}},$$

and hence if $\boldsymbol{H}_0$ is invertible, we have the influence function of $\boldsymbol{\theta}$ as

$$\psi_{\boldsymbol{\theta}} = \boldsymbol{H}_0^{-1}(\psi_{\boldsymbol{\delta}} - \psi_{\boldsymbol{H}}\boldsymbol{\theta}_0).$$

By the same argument, we have the influence function for $\boldsymbol{\pi}$ as in the Proposition 4. □

B Tables

Table 5: MSE, Variance, Bias for each parameters for different group size and $N = 30$

			1st stage			**2nd stage**		
Cat	N	G	θ_1	θ_2	θ_3	β	γ_1	γ_2
Mean	30	100	2.087	0.258	0.281	0.72	0.83	0.24
	30	500	2.147	0.261	0.276	1.024	0.797	0.196
	30	1000	2.144	0.26	0.277	1.014	0.799	0.198
	30	5000	2.15	0.261	0.276	1.034	0.796	0.195
True Values			0.214	0.26	0.277	1	0.8	0.2
MSE	30	100	0.126	<1e-5	0.001	2.898	0.031	0.057
	30	500	0.024	<1e-5	<1e-5	0.506	0.006	0.01
	30	1000	0.01	<1e-5	<1e-5	0.249	0.003	0.005
	30	5000	0.003	<1e-5	<1e-5	0.073	0.001	0.001
Variance	30	100	0.123	<1e-5	0.001	2.82	0.03	0.056
	30	500	0.024	<1e-5	<1e-5	0.506	0.006	0.01
	30	1000	0.01	<1e-5	<1e-5	0.249	0.003	0.005
	30	5000	0.003	<1e-5	<1e-5	0.072	0.001	0.001
Bias	30	100	-0.056	-0.002	0.004	-0.28	0.03	0.04
	30	500	0.003	<1e-5	<1e-5	0.024	-0.003	-0.004
	30	1000	0.001	<1e-5	<1e-5	0.014	-0.001	-0.002
	30	5000	0.007	<1e-5	<1e-5	0.034	-0.004	-0.005

notes. For each sample size N, G, the number of simulations is set to $B = 100$.

Table 6: MSE, Variance, Bias for each parameters for different group size and $N = 50$

			1st stage			**2nd stage**		
Cat	N	G	θ_1	θ_2	θ_3	β	γ_1	γ_2
Mean	50	100	2.927	0.26	0.277	0.952	0.803	0.204
	50	500	2.929	0.26	0.277	0.96	0.802	0.204
	50	1000	2.923	0.26	0.277	0.977	0.802	0.201
	50	5000	2.931	0.26	0.277	0.996	0.8	0.2
True Values			2.932	0.26	0.277	1	0.8	0.2
MSE	50	100	0.139	<1e-5	<1e-5	3.463	0.013	0.024
	50	500	0.029	<1e-5	<1e-5	0.624	0.002	0.004
	50	1000	0.014	<1e-5	<1e-5	0.304	0.001	0.002
	50	5000	0.004	<1e-5	<1e-5	0.091	<1e-5	0.001
Variance	50	100	0.139	<1e-5	<1e-5	3.461	0.013	0.024
	50	500	0.029	<1e-5	<1e-5	0.623	0.002	0.004
	50	1000	0.013	<1e-5	<1e-5	0.304	0.001	0.002
	50	5000	0.004	<1e-5	<1e-5	0.091	<1e-5	0.001
Bias	50	100	-0.005	<1e-5	<1e-5	-0.048	0.003	0.004
	50	500	-0.003	-0.001	<1e-5	-0.04	0.002	0.004
	50	1000	-0.009	<1e-5	<1e-5	-0.023	0.002	0.001
	50	5000	-0.002	<1e-5	<1e-5	-0.004	<1e-5	<1e-5

notes. For each sample size N, G, the number of simulations is set to $B = 100$.