Identification and Estimation of Treatment Effects from Network Data with Confounded Social Links

Licheng Liu*

This draft: November 25, 2024
First draft: August 30, 2023

Abstract

We introduce a generalized propensity score (GPS) based approach to the identification and estimation of treatment effects from observational network data, wherein formation of social link between a pair of units depends on individual level characteristics. Ignoring the tie formation process, its interaction with the treatment assignment mechanism, and interference induced by the social network may lead to biased estimation of treatment effects. We propose a unified framework that addresses these challenges by jointly modeling treatment assignment and network formation. Generalized propensity score can be estimated given probabilistic models for these two processes and functional form defining effective treatment. Average potential outcomes and treatment effects are estimated with inverse probability weighting estimators. We illustrate the proposed method in several Monte Carlo studies and an empirical analysis that investigates the effect of a new political communication technology on political participation in Uganda.

Keywords: Causal Inference, Social Network Data, Interference
Word Count: 7,254

*Postdoctoral Scholar, Department of Political Science, Stanford University. Email: lichengl@stanford.edu. I thank Robert Franzese, Danny Hidalgo, Yue Hou, In Song Kim, Jialu Li, Xun Pang, Lily Tsai, Yiqing Xu, Teppei Yamamoto, Han Zhang, members of the Kim Research Group at MIT, and participants at seminars at University of Michigan, University of Rochester, Hong Kong University of Science and Technology, and Annual Meeting of the Society for Political Methodology in 2023 and 2024 for their helpful comments. All remaining errors are my own.

1 Introduction

Causal inference with network data has emerged as a vibrant topic in both empirical social science studies and methodological research, as units under study are inter-connected at micro or macro level in many scenarios. In behavioral studies, individuals might establish social ties, like family members, friendships, based on shared characteristics like age or educational background (e.g., Sinclair et al., 2012). For cross-country analyses, connections between countries can be assessed by geographical proximity or bilateral metrics such as trade volume and joint membership in international organizations (e.g., Simmons and Elkins, 2004) . These examples illustrate the concern of violation of SUTVA in empirical studies when researchers want to investigate the casual effects of a treatment: when units under study are interconnected, a treatment assigned to one unit can influence potential outcomes of other units, known as *interference* or *spillover effects* in the literature of causal inference.

A popular approach to addressing interference is to fit regression models that incorporate the treatment variable, a measure of interference, like weighted average of neighbor units' treatment status, and a vector of unit level control variables (e.g., Arpino and Mattei, 2016). And treatment effects are derived from estimated coefficients. This approach may be plausible when connections between unit pairs are unconfounded[1], like geographical location. However, when units under study are socially connected, the formation of social ties between a pair of units depends on some unit-level features. If these features also affect the behavioral outcome, they confound the relationship between network formation and outcome (Goldsmith-Pinkham and Imbens, 2013). The problem becomes even more complex if some of these features also affect the adoption of treatment. In such scenarios, regression model based approaches may fail to account for the confoundedness of social network formation, causing bias in estimation of treatment effects.

To address this problem, we propose a generalized propensity score (GPS) based approach for identification and estimation of treatment effects. In network data, treatment effects are defined as differences in average potential outcomes under different levels of "effective treatment" (Manski, 2013), which is a function, termed "exposure mapping"

[1]It means that no variables affect both outcome and formation of connections between units.

in Aronow and Samii (2017), of the treatment assignment vector and the social network. The chosen form of this exposure mapping reflects how researchers perceive interference between units. In this paper, we assume it is flexible but correctly specified. Average potential outcomes can be identified with generalized propensity score for each level of effective treatment (Imbens, 2000). If we properly specify probabilistic models for both the treatment assignment mechanism and network formation process, we can then estimate generalized propensity score for each unit, given that the exposure mapping depends on the treatment assignment vector and the social network.

Generalized propensity scores can be estimated via analytical expression if the exposure mapping has a simple functional form. In more complex scenarios, like continuous effective treatment, simulations based on the probabilistic models can be utilized for their estimation (e.g., Aronow and Samii, 2017). Once generalized propensity scores are estimated, average potential outcomes as well as treatment effects can be estimated with the inverse-probability weighting estimators. Uncertainty estimates can be obtained by implementing the network HAC estimator (Kojevnikov et al., 2021; Leung and Loupos, 2022) that accounts for correlations among observations in the social network.

The confoundedness of social network formation complicates the identification and estimation of treatment effects. Even in experimental studies where treatments are randomized, confounded social links can pose challenges in identifying these effects. With observational data, the situation becomes even more complicated due to the potential interactions between treatment assignment and network formation. For example, treatment assignment mechanism and network formation can be simultaneously determined by exogenous covariates (Franzese and Hays, 2008; Han et al., 2021), treatment assignment may affect tie formation process (Comola and Prina, 2021), and social network may induce diffusion of treatment adoption (Leung and Loupos, 2022)[2]. Moreover, network formation and treatment adoption could even mutually reinforce each other.

When both treatment and network are confounded, various mechanisms could underlie the treatment effects. Consider an instance where we want to evaluate the impact of a new political communication technology (PCT) on political participation. If residents who

[2]It means that the adoption of treatment by one unit might also affect the likelihood that other units adopt the treatment.

hold official positions are typically more engaged in adopting PCT and participating in political activities than other residents, and if they are more likely to form social ties, then these village leaders tend to connect with more adopters than other residents. Even after addressing the confounding influence of local official position, neglecting the spillover effects could bias the estimated effect of PCT adoption on political participation. This stems from the fact that we are essentially estimating the combination of both direct and spillover effects. A positive spillover effect can lead to an overestimation of the direct effect, whereas a negative spillover effect can result in its underestimation. Therefore, it is essential to jointly model treatment assignment mechanism and network formation process to identify both direct and spillover effects.

In fact, joint modeling of treatment assignment mechanism and network formation process is central to the proposed method. Yet, estimating a combined probabilistic model for them can be challenging, particularly when we introduce additional constraints like row normalization on network entries. To circumvent this complexity, we propose to factorize these two processes so that researchers can model them separately or sequentially to simplify estimation under alternative assumptions.

This paper contributes to the burgeoning literature on causal inference with social network data (Ogburn et al., 2024). It extends the framework of Aronow and Samii (2017) to observational settings and relaxes the assumption of unconfounded network. The idea of modeling network formation process echoes the insights of Toulis et al. (2021) that studies the problem of network dynamics as treatment, while our approach distinctively evaluates the joint effect of a treatment assignment variable and a social network, which is static but random. Highlighting other relevant studies, Forastiere et al. (2021) and Sanchez-Becerra (2022) also propose propensity score based approach to casual inference with observational network data. Working under tighter assumptions, they further suggest parametric models for a direct estimation of generalized propensity scores. In contrast, the proposed method is flexible to accommodate multiple types of treatment variables, networks and their interactions. The trade-off is computational demand, which may limit its feasibility on large scale network datasets. Additionally, Jackson et al. (2022) introduces a peer-influenced propensity score and Leung and Loupos (2022) proposed a graph neural

network (GNN) based propensity score. Both methods can incorporate the diffusion of treatment adoption induced by the network, operating under the assumption that networks are unconfounded.

The rest of this paper is organized as follows. Section 2 outlines the contextual background of a motivating example. It investigates the effect of a new political communication technology on political participation in Uganda. In Section 3, we set up the potential outcome framework, define the causal estimands, and develop key identification assumptions. Section 4 introduces the generalized propensity score based approach. We illustrate details of estimation and inference of treatment effects based on generalized propensity score. Section 5 reports results of several Monte Carlo studies designed to investigate the finite sample properties of the proposed method. Section 6 provides a comparison of estimation results from our motivating example, contrasting the proposed method with an existing estimation strategy. The last section concludes.

2 A Motivating Example

Whether adoption of information and communication technology (ICT) promotes political engagement in developing countries has been a long-standing debate in comparative politics. In theory, ICT innovations reduce communication costs, facilitating political participation among marginalized groups have limited opportunities to communicate with politicians (Grossman et al., 2020). However, existing literature provides mixed evidence on the effect of ICT adoption on political participation. In this paper, we investigate the effect of the U-Bridge program, a new political communication technology, on political participation in 16 Ugandan villages, combing replication data from Ferrali et al. (2020) and Eubank et al. (2021).

The U-Bridge program was implemented in a district located in northwestern Uganda. It allows residents to contact district officials via text message, both freely and anonymously. It was implemented in a field experiment that encouraged usage in 131 randomly selected villages. Residents from treatment villages were invited to periodic community meetings about ways to communicate with local officials. The first round of meetings was held in late 2014. To investigate the pattern of adoption of U-Bridge, Ferrali et al. (2020) conducted

in-person surveys in 16 treatment villages, where the U-Bridge program was advertised, in 2016, two years after launch of the program. These surveys gathered multiple individual-level variables such as age, gender, attendance at meetings, U-Bridge adoption status, and various social ties between resident pairs.

Survey responses from the 16 villages are utilized in the empirical analysis. The treatment variable is a binary indicator for whether a resident adopted U-Bridge. While U-Bridge was advertised in these villages, whether to adopt it is determined by the residents. The outcome variable is a continuous summary index of political participation that aggregates political actions in the last 12 months. Other individual level covariates include age, gender, levels of income, binary indicators representing whether a resident has attained secondary education, whether they occupy a formal leadership role in the village, and whether they own a phone, and a behavioral proxy measure of care for the community [3]. The original survey data comprises information on 3,184 respondents across the 16 treatment villages. After excluding entries with missing values, the dataset used in this paper covers 3,018 respondents, with 135 of them having adopted U-Bridge.

There are various social relations among the residents. Ferrali et al. (2020) collect four types of social ties: family relationships, friendships, lender relationships, or problem-solving connections. These social ties may channel the spillover effect of U-Bridge adoption via inter-personal communication (Sinclair et al., 2012). In addition, the formation of social ties may depend on individual characteristics that also affect political participation, i.e., it is confounded. To construct social networks based on social ties, we assume that two residents within the same village are connected if they share any of the four types of social ties above. Consequently, connections between residents are undirected. We exclude social ties spanning different villages, which reduces the overall network connecting residents to 16 distinct components [4]. Figure 1 suggests that village leaders generally have more connections. In fact, village leaders tend to have nearly doubled connections compared to ordinary residents.

In this motivating example, we face several challenges in identifying and estimating the effect of U-Bridge on political participation. First, the adoption of U-Bridge is not

[3]Detailed variable descriptions are available in the appendix of Ferrali et al. (2020).

[4]The construction of social network is consistent with Ferrali et al. (2020) and Sanchez-Becerra (2022).

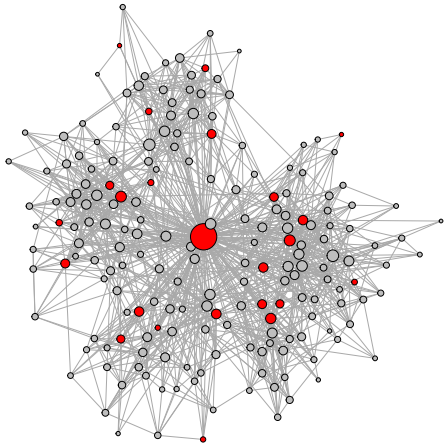


Figure 1: Network Visualization for a Village

Note: Size of each vertex is proportional to the square root of degree. Vertex in red represents respondent who occupies formal leadership role within village.

random, and it depends on resident level covariates like education. Second, the presence of social network induces interference of U-Bridge adoption on political participation, a violation of the Stable Unit Treatment Value Assumption (SUTVA). Lastly, the formation of social ties are nonrandom, and it depends on resident level covariates, like leadership, that may also affect political participation. Ignoring the social network and failing to address confoundedness of social network formation will lead to bias in treatment effect estimation. We propose a new potential outcome framework incorporating social network and a propensity score based estimation strategy to address these issues.

3 Setup

3.1 Notation

Suppose we have a cross-sectional dataset that includes N units. Let $\boldsymbol{W}$ denote an observed network[5], a $(N \times N)$ random adjacency matrix with entry W_{ij} specifying the connection between unit i and j[6]. In the terminology of network analysis, we use $G(\mathcal{V}, \mathcal{E})$ to denote the associated random graph for $\boldsymbol{W}$, where $\mathcal{V} = \{1, \ldots, N\}$ is the set of units [7] and $\mathcal{E}$ is the set of edges, i.e., $(i, j) \in \mathcal{E}$ if $W_{ij} > 0$. If G is undirected, $W_{ij} \equiv W_{ji}$ for all $i, j \in \mathcal{V}$. Otherwise, W_{ij} and W_{ji} may be different. $\boldsymbol{W}$ can be a weighted matrix, wherein the intensities of connections between units i, j and i, k are different if $W_{ij} \neq W_{ik}$. Throughout this paper we assume that $\boldsymbol{W}$ is unweighted. That is, $W_{ij} \in \{0, 1\} \quad \forall i, j$. In addition, we assume $W_{ii} \equiv 0$ for all $i \in \mathcal{V}$. Since $\boldsymbol{W}$ is random, let $\mathcal{W}$ denote the sample space of $\boldsymbol{W}$.

We use $D_i \in \{0, 1\}$ to denote a binary treatment assigned to unit i. Y_i is the observed outcome of interest. In vector form, we denote $\boldsymbol{D} = (D_1, \ldots, D_N)$ as a $(N \times 1)$ vector of treatments and $\boldsymbol{Y} = (Y_1, \ldots, Y_N)$ as the vector of observed outcomes for all units. Let $\mathcal{D} = \{0, 1\}^N$ be the sample space of $\boldsymbol{D}$.

We also observe a $(p \times 1)$ vector of pre-treatment covariates X_i for each unit. For network data, X_i can be decomposed into two parts: confounders that affect adoption of the treatment, and covariates contributing to homophily that induces network formation. We denote X_i^D as the first part and X_i^W as the second part. It is worth noting that these two subsets of covariates are not always mutually exclusive, meaning there may be covariates simultaneously impact both the adoption of treatment and network formation. In addition, X_i^W could also incorporate dyadic covariates, such as geographical distance or the bilateral trade volume between country pairs. In matrix form, let $\boldsymbol{X} = (X_1', \ldots, X_N')$ denote the matrix that aggregates all covariate vectors. Similarly, $\boldsymbol{X}^D$ and $\boldsymbol{X}^W$ are used

[5] We use bold uppercase letters to represent random vectors and matrices, bold lowercase letters their corresponding realizations.

[6] For notational convenience, we focus on the case of a single large network, while the results can be generated to datasets consisting of multiple networks or clusters (e.g. Hudgens and Halloran, 2008; Sanchez-Becerra, 2022) by regarding no cross-cluster connections, i.e., $W_{ij} \equiv 0$ if unit i and j belong to different clusters.

[7] They are also named "nodes" or "vertices" in network analysis.

to denote corresponding matrices for the aforementioned subsets of covariates.

3.2 Potential outcomes in social network data

In most social network data, not only the observed network $\boldsymbol{W}$ is random, but also its formation is confounded just as the assignment of treatment, meaning covariates affecting network formation also determine the potential outcome. In addition, network structure induces interference, as treatment assigned to one unit may affect potential outcome of other units. To address these two issues, we extend the potential outcome framework (Rubin, 1974) to model each unit's potential outcome as a function of the treatment assignment vector and network structure: $Y_i(\boldsymbol{d}, \boldsymbol{w})$ for $i \in \mathcal{V}$, $\boldsymbol{d} \in \mathcal{D}$ and $\boldsymbol{w} \in \mathcal{W}$. Under the assumption of no multiple versions of treatment (consistency, Rubin, 1986), the observed outcome $Y_i = Y_i(\boldsymbol{d}, \boldsymbol{w})$ if $\boldsymbol{D} = \boldsymbol{d}$ and $\boldsymbol{W} = \boldsymbol{w}$.

In this generalized potential outcome framework, even if the treatment assignment vector remains the same, as long as network changes, the potential outcome may be different. For example, network $\boldsymbol{w}$ on the left panel of Figure 2 is denser than network $\boldsymbol{w}'$ on the right panel, while the treatment status for each unit is the same. $Y_i(\boldsymbol{d}, \boldsymbol{w})$ may not equal $Y_i(\boldsymbol{d}, \boldsymbol{w}')$ as unit i has connections with more treated units when the network is denser. Therefore, we regard $\boldsymbol{D}$ and $\boldsymbol{W}$ as "joint" assignments of treatment and network.

Given the notations above, a non-parametric structural equation model (NP-SEM, Pearl, 2009) for network formation, treatment assignment and outcomes is represented as:

$$\begin{aligned} W_{ij} &= g_w(i, j, \boldsymbol{X}, \epsilon_{ij}^w), \\ D_i &= g_d(i, \boldsymbol{X}, \epsilon_i^d), \\ Y_i(\boldsymbol{d}, \boldsymbol{w}) &= g_y(i, \boldsymbol{d}, \boldsymbol{w}, \boldsymbol{X}, \epsilon_i^y), \end{aligned} \tag{1}$$

where ϵ_{ij}^w, ϵ_i^d and ϵ_i^y are random errors. The correlation between ϵ_{ij}^w and ϵ_i^d characterizes interactions between network formation and treatment assignment. When the network is undirected and link formation between pair of units depends only on pairwise covariates, the network formation model can be simplified as $W_{ij} = g_w(X_i, X_j, \epsilon_{ij}^w)$. In social network data, the covariate matrix $\boldsymbol{X}$ for all units may determine treatment adoption and behavioral outcome for each unit i. To see this, In NP-SEM (1), $\boldsymbol{X}$ can be written as $(X_i, \boldsymbol{X}_{-i})$, where

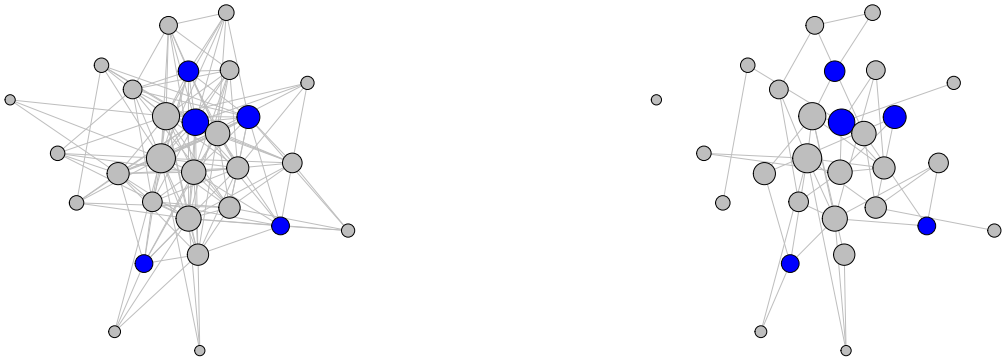


Figure 2: Potential Outcomes with Different Networks

Note: Blue nodes denote units under treatment, and grey nodes denote units under control.

$\boldsymbol{X}_{-i}$ is named *contextual variable* in the literature of network analysis (Jackson, 2008; Jackson et al., 2022). By incorporating such variables in functions g_d and g_y, researchers consider the influence of *social norm* on each unit's action on treatment adoption and behavioral outcome. If researchers have strong prior belief that such influence does not exist, the models can be further simplified as $D_i = g_d(X_i, \epsilon_i^d)$ and $Y_i(\boldsymbol{d}, \boldsymbol{w}) = g_y(\boldsymbol{d}, \boldsymbol{w}, X_i, \epsilon_i^y)$ by dropping the contextual variable.

Since the treatment vector and network structure jointly determine the potential outcomes, g_y for potential outcome is a function of high-dimensional inputs, making the identification and estimation of treatment effects a challenging task. Following existing literature, we assume that there exists a low-dimensional, and possibly vector-valued, function of treatment assignment vector and network that represents the potential outcomes.

Assumption 1. Exposure mapping. There exists a known function $e : \mathcal{N} \times \mathcal{D} \times \mathcal{W} \to \mathcal{T}$, such that

$$Y_i(\boldsymbol{d}, \boldsymbol{w}) = Y_i(t_i)$$

if $e(i, \boldsymbol{d}, \boldsymbol{w}) = t_i$ [8]. Under this assumption, $Y_i(\boldsymbol{d}, \boldsymbol{w}) = Y_i(\boldsymbol{d}', \boldsymbol{w}')$ if $e(i, \boldsymbol{d}, \boldsymbol{w}) = e(i, \boldsymbol{d}', \boldsymbol{w}')$.

[8]The potential outcome with exposure mapping is denoted as $Y_i(\boldsymbol{d}, \boldsymbol{w}) = \tilde{Y}_i(t_i)$ in Leung and Loupos (2022) to distinguish between original treatment assignment and exposure mapping. For notational con-

In the literature of casual inference under interference, $T_i = e(i, \boldsymbol{D}, \boldsymbol{W})$ is called "effective treatment" (Manski, 2013) and e is the "exposure mapping" function (Aronow and Samii, 2017).

The exposure mapping approach has been adopted in methodological research on interference (e.g., Forastiere et al., 2021) as well as empirical research on spillover effect (e.g., Arpino and Mattei, 2016). In the literature, scholars usually assume that the network structure is fixed and part of the exposure mapping function that defines interference structure. In this paper, we regard network as an input like the treatment assignment vector. Therefore, $\boldsymbol{W}$ and $\boldsymbol{D}$ jointly determine the effective treatment for each unit. Specifically, treatment assignment vector and network structure are common inputs for all units under study, while output of the exposure mapping, the effective treatment, is unit-specific.

Note that the specification of exposure mapping function needs substantial knowledge on the treatment. Here are some examples. When there is no interference, the effective treatment just equals individual treatment, and $T_i = e(i, \boldsymbol{D}, \boldsymbol{W}) = D_i$. When network structure induces interference, if the potential outcome for each unit is determined by its own treatment status as well as the number of connections with treated units, a vector-valued exposure mapping function

$$T_i = e(i, \boldsymbol{D}, \boldsymbol{W}) = (D_i, \sum_j W_{ij} D_j) = (D_i, Z_i) \tag{2}$$

is specified. While its form can be flexible, we assume the exposure mapping function is known and correctly specified by the researcher [9] . Under Assumption 1, we replace the outcome model in NP-SEM (1) with $Y_i(t) = g_y(i, t, \boldsymbol{X}, \epsilon_i^y)$.

3.3 Causal quantities of interest

When treatment is binary, there are only two potential outcomes for each unit, and causal estimands are well-defined. The average treatment effect is $ATE = \mathbb{E}(Y_i(1) - Y_i(0))$, differences in average potential outcomes under treatment and under control. And average treatment effect on the treated (control) are defined for units under treatment (control).

sistency, we keep using $Y_i(\cdot)$ for potential outcomes throughout this paper.

[9] Readers may refer to Sävje (2023) for the results of misspecification of exposure mapping function.

In network data, effective treatment represented by the exposure mapping function is often multi-valued. Since treatment effects are differences in average potential outcomes, we first define the average potential outcome at a given level of effective treatment as:

$$\mu(t) = \mathbb{E}(Y_i(t)). \tag{3}$$

Average potential outcome (3) is also called *average dose-response function* (ADRF), a concept in medical statistics and recently adopted in the casual inference literature to describe the relation between an outcome of interest and a continuous treatment. The effect of effective treatment t compared to t' is defined as:

$$\tau(t, t') = \mathbb{E}(Y_i(t) - Y_i(t')) = \mu(t) - \mu(t'). \tag{4}$$

While simple in its form, treatment effect (4) may have rich meanings given researchers' choices of exposure mapping function. In this paper, we focus on direct and spillover effects of a treatment assignment vector. To identify these treatment effects, we use the exposure mapping function (2) as an illustration, and then the average potential outcome is $\mu(d, z) = \mathbb{E}[Y_i(d, z)]$. Our second causal estimand is *conditional direct effect*:

$$\tau(z) = \mathbb{E}[Y_i(1, z) - Y_i(0, z)] = \mu(1, z) - \mu(0, z), \tag{5}$$

which measures the effect of individual treatment adoption d while holding the level of interference z constant. And the *marginal direct effect* is defined by averaging $\tau(z)$ over the distribution of z: $\tau = \int \tau(z) f_Z(z) dz$. Next, the *conditional spillover effect* is defined as:

$$\delta(z, z', d) = \mathbb{E}[Y_i(d, z) - Y_i(d, z')] = \mu(d, z) - \mu(d, z'), \tag{6}$$

which is the difference between average potential outcomes when we fix the value of individual treatment d and compare two alternative levels of interference z and z'. When z' is some benchmark value, e.g., $z' = 0$, we can simplify the notation by dropping z' in the equation and write $\delta(z, d) = \mu(d, z) - \mu(d, 0)$. Like the marginal direct effect, we define the *marginal spillover effect* as $\delta(d) = \int \delta(z, d) f_Z(z) dz$.

3.4 Unconfoundedness of treatment assignment and network formation

For causal inference with observational data, when SUTVA holds, identification results are based on the conditional unconfoundedness (ignorability) assumption: $\{Y_i(0), Y_i(1)\} \perp\!\!\!\perp D_i|X_i$. It states that, conditional on observed confounders X_i, potential outcomes are independent of treatment assignment. For network data, potential outcomes are determined by the effective treatment, a function of both treatment assignment vector and social network, and thus, the ignorability assumption must be restated. We introduce the following assumption to account for confounders that affect treatment assignment, network formation, and potential outcomes.

Assumption 2. Unconfoundedness of treatment assignment and network formation.

$$Y_i(t_i) \perp\!\!\!\perp \{\boldsymbol{D}, \boldsymbol{W}\}|\boldsymbol{X} \quad \forall i \in \mathcal{N}, \boldsymbol{d} \in \mathcal{D}, \boldsymbol{w}' \in \mathcal{W}, \tag{7}$$

where $t_i = e(i, \boldsymbol{d}', \boldsymbol{w}')$ is a certain level of effective treatment. Condition (7) states that, given observed covariates matrix $\boldsymbol{X}$, potential outcomes are independent of treatment assignment and social network formation. It differs from conventional assumption of unconfoundedness as we condition on the observed covariates for all units rather than individual covariates. This is because $\boldsymbol{X}$ determines assignment of treatment vector as well as formation of social network, and conditioning solely on X_i is not sufficient to address the confounding between $\boldsymbol{D}$, $\boldsymbol{W}$ and the potential outcomes. Since effective treatment $T_i = e(i, \boldsymbol{D}, \boldsymbol{W})$ is a function of $\boldsymbol{D}$ and $\boldsymbol{W}$, equivalently, we have $Y_i(t_i) \perp\!\!\!\perp T_i|\boldsymbol{X}$.

Remark 1. Literature on causal inference with network data has made several alternative assumptions. Condition (7) is similar to the high-dimensional condition of unconfoundedness [10] in Leung and Loupos (2022) that incorporates diffusion of treatment adoption. It generalizes the condition of unconfoundedness proposed by Forastiere et al. (2021), assuming individualized treatment adoption. In addition, Jackson et al. (2022) make a *Societal Conditional Unconfoundedness* assumption [11] to account for the equilibrial

[10] The original assumption made in Leung and Loupos (2022) focuses on treatment assignment: $\{Y_i(\cdot)\} \perp\!\!\!\perp \boldsymbol{D}|\boldsymbol{X}, \boldsymbol{W}$. Given that network formation is unconfounded, we replace the treatment assignment vector with the exposure mapping.

[11] Jackson et al. (2022) also assume that potential outcomes only depend on individual treatment status.

behavior in treatment adoption when units under study are strategic. These approaches regard the network as fixed and need to condition on the network structure. In a recent paper, Sanchez-Becerra (2022) derives the condition of unconfoundedenss incorporating the network formation process. However, unlike condition (7), it assumes as-if unconfoundedness in the formation of social ties.

4 A Generalized Propensity Score Based Approach

Propensity score is the conditional probability of receiving a (certain level of) treatment given observed confounders (Rosenbaum and Rubin, 1983). When treatment is multi-valued or continuous, the conditional probability is called generalized propensity score (GPS, Hirano and Imbens, 2004). In this section, we apply propensity score based method to the identification and estimation of treatment effects from social network data under condition (7). Unlike existing approaches (e.g., Sanchez-Becerra, 2022; Forastiere et al., 2021) that assume conditional independence of (or exchangeable) treatment assignment, condition (7) allows the effective treatment assigned to each unit to be interdependent, as inputs of the exposure mapping function, treatment assignment vector and network structure, are common to all units.

Formally, we define the generalized propensity score for effective treatment $T_i = t$ of unit i, given observed covariate matrix, as:

$$r(i, t; \boldsymbol{X}) = Pr(T_i = t|\boldsymbol{X}) = Pr(e(i, \boldsymbol{D}, \boldsymbol{W}) = t|\boldsymbol{X}). \tag{8}$$

GPS (8) is similar to conventional propensity score for a binary treatment indicator, which is probability that the treatment assigned to unit i. Propensity score for binary treatment indicator usually only depends on individual covariates, and differences in propensity scores across units come from heterogeneity of their values. For GPS (8) in network setting, the input $\boldsymbol{X}$ is common for all units, and we use index i to denote unit-level heterogeneity. We make the following assumption on the overlapping of generalized propensity score.

Assumption 3. Positivity of the generalized propensity score.

$$0 < r(i, t; \boldsymbol{X}) < 1, \quad \forall i \in \mathcal{N}, t \in \mathcal{T}. \tag{9}$$

Given the definition of generalized propensity score $r(i,t;\boldsymbol{X})$, under Assumptions 1 and 2, we have the following propositions.

Proposition 1. Balancing property of generalized propensity score.

$$Pr(T_i = t|\boldsymbol{X}, r(i,t;\boldsymbol{X})) = Pr(T_i = t|r(i,t;\boldsymbol{X})). \tag{10}$$

The balancing property implies conditional unconfoundedness given generalized propensity score:

$$\{Y_i(\cdot)\} \perp\!\!\!\perp T_i|r(i,t;\boldsymbol{X}). \tag{11}$$

Propositions 1 states that, conditioning on the generalized propensity score is sufficient to address confounding in treatment assignment and network formation. Under Assumption 3, the balancing property of generalized propensity score implies identification of average potential outcome given generalized propensity score.

Proposition 2. Identification of the average potential outcome $\mu(t)$.

$$\mathbb{E}\{\frac{Y_i \mathbb{1}\{T_i = e(i,\boldsymbol{D},\boldsymbol{W}) = t\}}{r(i,t;\boldsymbol{x})}\} = \mathbb{E}\{Y_i(t)\} = \mu(t). \tag{12}$$

4.1 Joint modeling of treatment assignment and network formation

For binary treatment indicators, propensity score is usually estimated by a treatment assignment model like Logit or Probit model. If we adopt a similar approach to estimate generalized propensity score (8), we need to fit a model with high-dimensional inputs $\boldsymbol{X}$. When network formation is unconfounded, Leung and Loupos (2022) propose a graph neural network (GNN) based estimator for such generalized propensity scores, conditioning on the network structure. However, in the proposed potential outcome framework, the social network itself determines the potential outcome, and conditioning on it will cause bias.

To tackle this problem, we propose an alternative approach to estimate GPS (8) by jointly modeling the assignment mechanism of treatment and the formation of network. Under condition (7), these two processes can be represented by a probabilistic model

$f_{\boldsymbol{D},\boldsymbol{W}|\boldsymbol{X}}(\boldsymbol{d},\boldsymbol{w})$ given covariate matrix $\boldsymbol{X}$, and then GPS (8) is re-formulated as:

$$\begin{aligned} r(i,t;\boldsymbol{X}) &= Pr(e(i,\boldsymbol{D},\boldsymbol{W})=t|\boldsymbol{X}) \\ &= \int_{\boldsymbol{d}\in\mathcal{D}}\int_{\boldsymbol{w}\in\mathcal{W}} \mathbb{1}\{e(i,\boldsymbol{d},\boldsymbol{w})=t\} f_{\boldsymbol{D},\boldsymbol{W}|\boldsymbol{X}}(\boldsymbol{d},\boldsymbol{w})\,\mathrm{d}\boldsymbol{d}\,\mathrm{d}\boldsymbol{w}. \end{aligned} \tag{13}$$

Therefore, if we properly model the joint distribution of treatment assignment vector and social network, we can estimate GPS (8) based on its alternative formulation (13) given the exposure mapping function.

While jointly modeling the distribution of $\boldsymbol{D}$ and $\boldsymbol{W}$ incorporates complex dynamics between them in observational data, in practice it is not straightforward to specify a probabilistic model for treatment assignment together with network formation, as these two processes are qualitatively different. To simplify modeling, we can factorize these two processes given substantive prior knowledge. For example, we can assume that these two processes are independent given the "common exposure" (Franzese and Hays, 2008). The upper left panel of Figure 3 is a diagram of such factorization. The joint distribution can be written as:

$$f_{\boldsymbol{D},\boldsymbol{W}|\boldsymbol{X}}(\boldsymbol{d},\boldsymbol{w}) = f_{\boldsymbol{D}|\boldsymbol{X}}(\boldsymbol{d}) f_{\boldsymbol{W}|\boldsymbol{X}}(\boldsymbol{w}) = f_{\boldsymbol{D}|\boldsymbol{X}^D}(\boldsymbol{d}) f_{\boldsymbol{W}|\boldsymbol{X}^W}(\boldsymbol{w}). \tag{14}$$

Since treatment assignment and network formation are two processes, it is possible that one precedes the other and intervenes the latter. For example, the network may cause the diffusion of treatment adoption (lower left panel of Figure 3). If treatment assignment precedes network formation instead, the network becomes a mediator (upper right panel of Figure 3). For these two scenarios, researchers can sequentially model these two processes. Finally, it is also possible that treatment assignment and network formation mutually reinforce each other, known as a co-evolution process (lower right panel of Figure 3). In this case, researchers need to specify a simultaneous equation model to account for the simultaneity[12]. Through out this paper, we use the case of common exposure to illustrate factorization.

Once we model the joint distribution of treatment assignment and network formation, given the form of exposure mapping function, we can estimate the generalized propensity

[12]One estimation strategy for such simultaneous equation models for co-evolution is the latent variable model proposed by Han et al. (2021) based on the latent space model (Hoff, 2021) for network formation.

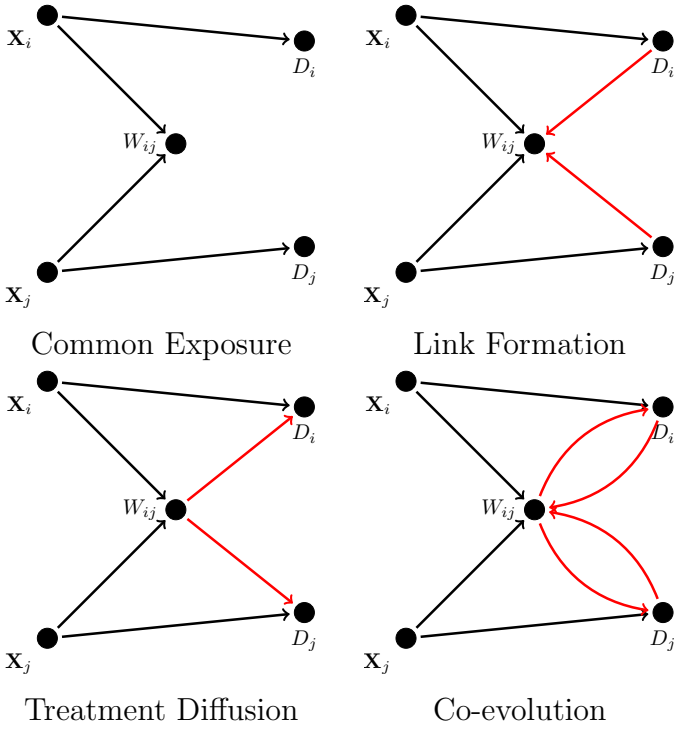


Figure 3: Factorization of Treatment Assignment and Network Formation

Note: Arrows in red represent interactions between treatment assignment and network formation.

score for effective treatment of each unit. Assuming that treatment assignment is individualized and tie formation is independent, factorization (14) becomes

$$f_{\boldsymbol{D}|\boldsymbol{X}^D}(\boldsymbol{d}) = \prod_{i=1}^{N} Pr(D_i = d_i | X_i^D),$$

$$f_{\boldsymbol{W}|\boldsymbol{X}^D}(\boldsymbol{w}) = \prod_{i=1}^{N} \prod_{\substack{j=1 \\ j \neq i}}^{N} Pr(W_{ij} = w_{ij} | X_i^W, X_j^W).$$

We consider the exposure mapping function (2), meaning the potential outcome is determined by the unit's own treatment status and the number of units under treatment that

it shares a common link. For unweighted network, GPS (13) becomes:

$$
\begin{aligned}
r(i,(d,z);\boldsymbol{X}) =& Pr\left[e(i,\boldsymbol{D},\boldsymbol{W})=(d,z)|\boldsymbol{X}\right] \\
=& Pr\left[D_i=d,\sum_{j=1}^{N}W_{ij}D_j=z|\boldsymbol{X}\right] \\
=& Pr(D_i=d|X_i^D) \\
&\times \sum_{\substack{d_j\in\{0,1\},w_{ij}\in\{0,1\}\\ j\neq i}} \mathbb{1}\{\sum_{\substack{j=1\\ j\neq i}}^{N} w_{ij}d_j=z\}\prod_{\substack{j=1\\ j\neq i}}^{N} Pr(D_j=d_j|X_j^D)Pr(W_{ij}=w_{ij}|X_i^W,X_j^W).
\end{aligned}
\tag{15}
$$

Essentially, the formulation of GPS (15) is a synthesis of probabilistic models of treatment assignment and network formation.

Like the example above, for simple functional forms of exposure mapping, special types of network like unweighted networks, and binary or general categorical treatment indicators, it is possible to get closed-form expression for the generalized propensity scores. For more complicated cases, like continuous treatment, weighted networks and complex functional form of exposure mapping incorporating influences from higher order neighbors, it is difficult to get closed-form expression for the generalized propensity score, and simulation-based methods (Aronow and Samii, 2017; Toulis et al., 2021) can be used for its estimation.

4.2 Estimation and inference

In the preceding part of this section, we introduce the generalized propensity score $r(i,t;\boldsymbol{X})$ for effective treatment and show its balancing property. We further show the identification of average potential outcomes and treatment effects defined as differences in average potential outcomes. Next, we discuss estimation and inference of treatment effects. We focus on conditional effects, and marginal effects are weighted averages of them.

Given the exposure mapping function, once the probabilistic model for treatment assignment and network formation is estimated, generalized propensity score, $\hat{r}(i,t;\boldsymbol{X})$, can be constructed based on it. Then we can estimate average potential outcome $\hat{\mu}(t)$ with the Horvitz-Thompson estimator:

$$
\hat{\mu}(t)^{HT}=\sum_{i=1}^{N}\mathbb{1}\{T_i=t\}\frac{Y_i}{\hat{r}(i,t;\boldsymbol{X})}. \tag{16}
$$

However, when $\hat{r}(i,t;\boldsymbol{X})$ is close to 0, weight for observation i, $1/\hat{r}(i,t;\boldsymbol{X})$, can be extremely large. Therefore, the Horvitz-Thompson estimator (16) is of high variance. Alternatively, we can estimate $\hat{\mu}(t)$ with the Hájek estimator, which is proved to improve efficiency at the cost of finite sample bias:

$$\hat{\mu}(t)^H = \frac{\sum_{i=1}^N \mathbb{1}\{T_i = t\}\frac{Y_i}{\hat{r}(i,t;\boldsymbol{X})}}{\sum_{i=1}^N \mathbb{1}\{T_i = t\}\frac{1}{\hat{r}(i,t;\boldsymbol{X})}}. \tag{17}$$

Compared to the Horvitz-Thompson estimator (16), the Hájek estimator (17) allows the denominator to vary according to sum of the weights. It improves estimation efficiency by shrinking the magnitude of the estimate when it is large, and increasing its magnitude when it is small.

To further improve estimation efficiency, we can also estimate $\mu(t)$ with the augmented inverse probability weighting (AIPW) estimator, with a properly-specified outcome model $\mu(i,T,\boldsymbol{X})$:

$$\hat{\mu}(t)^{AIPW} = \sum_{i=1}^N \mathbb{1}\{T_i = t\}\frac{Y_i - \hat{\mu}(i,t,\boldsymbol{X})}{\hat{r}(i,t;\boldsymbol{X})} + \hat{\mu}(i,t,\boldsymbol{X}). \tag{18}$$

The AIPW estimator (18) is also known as the doubly robust estimator, as the correct specification of either the generalized propensity score or the outcome model ensures consistency estimation (Glynn and Quinn, 2010). To deal with the high-dimensionality of $\boldsymbol{X}$ in (18), we can fit $\hat{\mu}(i,t,\boldsymbol{X})$ with machine learning models like random forests. Otherwise, we can fit a linear model by replacing $\boldsymbol{X}$ in the outcome model with individual covariates.

Taking the AIPW estimator (18) as an example, an estimator for average effect of effective treatment at level t relative to level t' is

$$\hat{\tau}(t,t') = \hat{\mu}(t)^{AIPW} - \hat{\mu}(t')^{AIPW}. \tag{19}$$

The AIPW estimator (19) for treatment effects estimation is used in the rest of the paper given its nice property of double robustness. For inference, we derive the asymptotic distribution of (19) for $\tau(t,t')$. Define

$$\begin{aligned}\tau_i(t,t') =& [\mathbb{1}\{T_i = t\}\frac{Y_i - \mu(i,t,\boldsymbol{X})}{r(i,t;\boldsymbol{X})} + \mu(i,t,\boldsymbol{X})] \\ &- [\mathbb{1}\{T_i = t'\}\frac{Y_i - \mu(i,t',\boldsymbol{X})}{r(i,t';\boldsymbol{X})} + \mu(i,t',\boldsymbol{X})]\end{aligned}$$

and $\varphi_i(t,t') = \tau_i(t,t') - \tau(t,t')$. Under a set of regularity conditions,

$$\sigma_{t,t'}^{-1}\sqrt{N}(\hat{\tau}(t,t') - \tau(t,t')) \xrightarrow{d} \mathcal{N}(0,1), \tag{20}$$

where $\sigma_{t,t'}^2 = \text{var}(\varphi_i(t,t')|\boldsymbol{X})$. A formal proof of the asymptotic distribution (20) can be found in Section A of the appendix. For variance estimation, Leung and Loupos (2022) show that the network HAC estimator (Kojevnikov et al., 2021) is a consistent estimator for $\sigma_{t,t'}^2$, which is written as:

$$\begin{aligned}\hat{\sigma}_{t,t'}^2 &= \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\left(\hat{\tau}_i(t,t') - \hat{\tau}(t,t')\right)\left(\hat{\tau}_j(t,t') - \hat{\tau}(t,t')\right)\mathbb{1}\{\ell_{\boldsymbol{W}}(i,j) \leq b\} \\ &= \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\hat{\varphi}_i(t,t')\hat{\varphi}_j(t,t')\mathbb{1}\{\ell_{\boldsymbol{W}}(i,j) \leq b\},\end{aligned} \tag{21}$$

where $\ell_{\boldsymbol{W}}(i,j)$ is the shortest path length from i to j on the observed network $\boldsymbol{W}$ and b is a pre-specified bandwidth. Note that the HAC estimator (21) incorporates clustered data as a special case, wherein $\ell_{\boldsymbol{W}}(i,j) = \infty$ if unit i and j belong to different clusters, and then (21) equals the clustered robust variance estimator.

Remark 2. In empirical studies, linear models are common practices for estimating direct and spillover effects (e.g., Cai et al., 2015). For example, researchers can fit the following regression model:

$$Y_i = \alpha + \tau D_i + \gamma\sum_j W_{ij}D_j + X_i'\beta + \varepsilon_i. \tag{22}$$

Based on this model, we can recover causal quantities of interest from the parameters τ and γ. While model (22) is easy to implement and interpret, when social network formation is confounded, individual covariates X_i in (22) may not be sufficient to adjust for confounding. In addition, the outcome model may be mis-specified, and thus, the estimated direct and spillover effects are biased. We illustrate this argument through several Monte Carlo studies.

5 Monte Carlo Studies

In this section, we conduct a series of Monte Carlo studies to investigate finite sample property of the proposed generalized propensity score based method for estimating treatment

effects from observational network data. We focus on the AIPW estimator (19) and consider bias, root mean squared error (RMSE) and coverage rate of 95% confidence interval as key performance metrics. Difference between average estimated standard error (SE) and sampling variation (SD) is also reported.

We use clustered data for simulation studies, where each simulated dataset consists of multiple independent clusters. Units within the same cluster may form link between each other, but there are no connections across clusters. Clustered data can be seen as a special case of a single network, where the probability of cross-cluster link formation is 0 and known to the researcher. For a given simulated dataset, We use $c(\cdot)$ to denote the cluster that each unit belongs to. Thus $c(i) = 1$ if unit i belongs to cluster 1. Suppose $\boldsymbol{W}$ is the observed network that specifies connections between units for this simulated dataset, then if $c(i) \neq c(j)$, $W_{ij} \equiv 0$, otherwise $W_{ij} \in \{0, 1\}$. For notational convenience, we assume that sizes of clusters are equal. We consider binary treatment indicator and unweighted network with no directions. The data generating process (DGP) for treatment assignment and network formation is as follows:

$$X_i \sim \begin{cases} N(1,1) & \text{if } i \text{ is odd }, \\ N(-1,1) & \text{if } i \text{ is even }, \end{cases}$$

$$W_{ij} = \begin{cases} 0 & \text{if} \quad c(i) \neq c(j), \\ \mathbb{1}\{0.1 - 0.25|X_i - X_j| + \nu_i \geq 0\}, \quad \nu_i \sim N(0,1) & \text{if} \quad c(i) = c(j), \end{cases} \tag{23}$$

$$D_i = \mathbb{1}\{0.1 + 0.5X_i + e_i \geq 0\}, \quad e_i \sim N(0,1),$$

where cluster $c(\cdot)$ is pre-specified. For each cluster, observed covariate X_i is drawn from independent but not identical normal distributions[13]. For simplicity, we assume that cluster size equals 4 and cluster is determined by the index of each unit, i.e., units 1, 2, 3, and 4 belong to cluster 1, units 5, 6, 7, and 8 belong to cluster 2, and so on. We use the exposure mapping function (2), and obviously $Z_i \in \{0, 1, 2, 3\}$. Therefore, the effective treatment T_i takes 8 different levels: $(0,0)$, $(0,1)$, $(0,2)$, $(0,3)$, $(1,0)$, $(1,1)$, $(1,2)$, $(1,3)$. Model for the potential outcome is:

$$Y_i(d,z) = 1 + X_i + \frac{\sum_{j \in c(i), j \neq i} X_j}{3} + d + 0.5z + dz + \varepsilon_i, \quad \varepsilon_i \sim N(0,1), \tag{24}$$

[13]In this case, the assumption of i.i.d. X_i in Sanchez-Becerra (2022) is violated.

where the term $\frac{\sum_{j\in c(i), j\neq i} X_j}{3}$ represents contextual effect.

Since there are 8 levels of effective treatment, $\binom{8}{2} = 28$ different treatment effects can be estimated. We focus on a direct treatment effect, $\tau((1,1),(0,1)) = \mu(1,1) - \mu(0,1)$, and a spillover treatment effect, $\tau((1,1),(1,0)) = \mu(1,1) - \mu(1,0)$, as quantities of interest. Outcome model (24) implies constant treatment effects, i.e., individual level treatment effects are constant across units and do not depend on covariate X_i. We also consider an alternative DGP for potential outcome with heterogeneous treatment effect, and the results are reported in section B of the appendix. Simulated datasets are generated based on the DGP above with varying sample sizes. Since size of cluster is fixed, we generate datasets with different numbers of clusters. Specifically, We consider the number of clusters equals $N_c = 50$, 100, 200, 500 and 1000. For each number of clusters, We conduct 2,000 simulations.

To estimate treatment effects, we first estimate models for treatment assignment and network formation, assuming they are correctly specified as in (23), and then we construct generalized propensity score using the closed-form expression (15). Next, we use the AIPW estimator (19) to estimate treatment effects. For uncertainty estimates, we calculate clustered robust standard error, which is equivalent to the network HAC variance estimator (21) in the simulated data setting, to construct 95% confidence interval. Results of Monte Carlo studies for the two treatment effects are summarized in Table 1 and Table 2.

Table 1: Finite Sample Properties: $\tau((1,1),(0,1))$

N_c	Bias	RMSE	Coverage Rate	SE	SE - SD
50	0.002	0.368	0.929	0.330	−0.039
100	−0.003	0.246	0.948	0.240	−0.006
200	0.000	0.180	0.948	0.173	−0.007
500	0.004	0.118	0.952	0.112	−0.006
1000	0.002	0.079	0.951	0.079	0.000

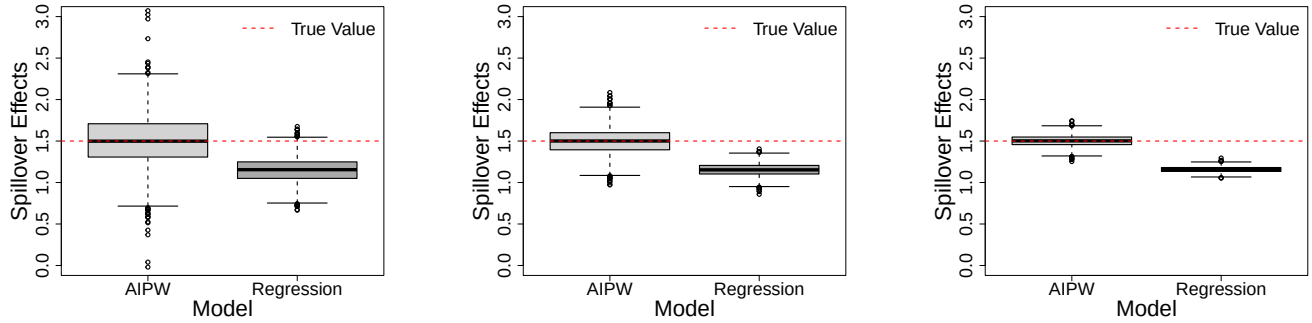
We find that, for both direct and spillover effects, the AIPW estimator produces small bias even when sample sizes are small. Meanwhile, root mean squared error also decreases as number of clusters increases, as both bias and sampling variation decrease. Next, We investigate properties of variance estimator. We find that, the coverage rate of 95% confidence interval is close to the nominal rate. In addition, difference between average estimated

Table 2: Finite Sample Properties: $\tau((1,1),(1,0))$

N_c	Bias	RMSE	Coverage Rate	SE	SE - SD
50	0.006	0.314	0.946	0.295	−0.019
100	−0.009	0.220	0.955	0.215	−0.004
200	−0.001	0.157	0.948	0.153	−0.004
500	−0.003	0.100	0.951	0.099	−0.001
1000	0.002	0.069	0.953	0.069	0.000

standard error and sampling variation of the proposed estimator decreases as the number of clusters increases, indicating that the variance estimator is consistent.

For model comparison, we also fit a regression model and derive the treatment effects from relevant coefficients. Robust standard errors clustered at group level are calculated. The adopted regression model is same as specification (22). Based on this model, estimates are $\hat{\tau}$ for the direct effect and $\hat{\gamma}$ for the spillover effect. Note that the outcome model (22) is mis-specified by ignoring the interaction term $D_i \cdot Z_i$ and contextual effect of the observed covariate. Estimation results based on regression model are reported in section B of the appendix. Sampling distributions of the AIPW estimator as well as regression based estimator are displayed in Figure 4, and bias and RMSE plots are shown in Figure 5. When the number of clusters is small, the proposed estimator has bias close to 0, and it decreases quickly as the number of clusters increases. In terms of RMSE, it also decreases quickly as the number of clusters increases. For the regression based estimator, it decreases only marginally because of large bias.



50 Clusters 200 Clusters 1000 Clusters

Figure 4: Sampling Distribution with Different Number of Clusters

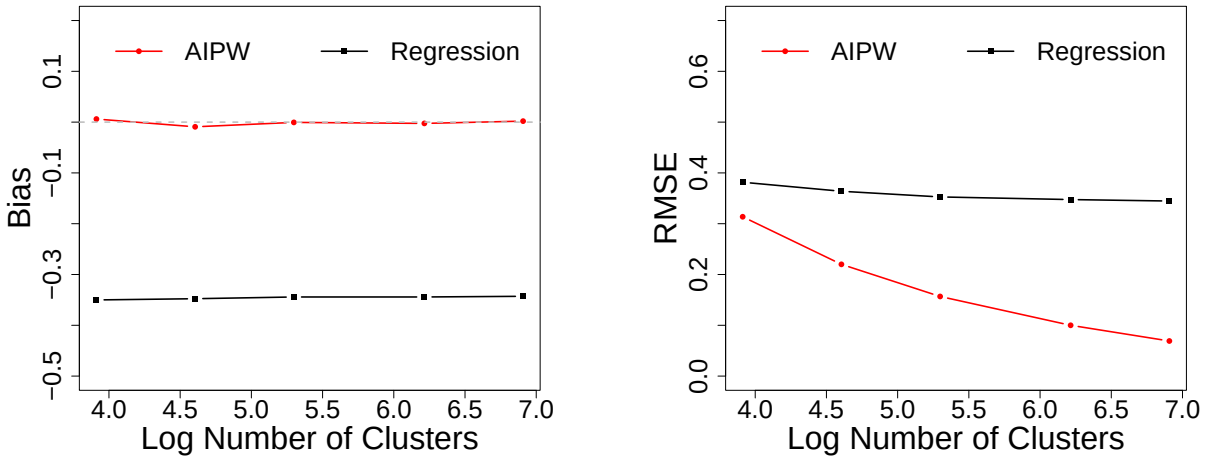


Figure 5: Finite Sample Properties: Bias and RMSE

Additional Monte Carlo studies indicate that properties of the proposed AIPW estimator are robust to heterogeneous treatment effect. Details of results are available in Section B of the appendix.

6 Empirical Analysis

To demonstrate its applicability in empirical social science research, we apply the proposed method to the motivating example that investigates the effect of U-Bridge on political participation in Uganda. Additionally, we compare the GPS based method with regression models, discussing the plausibility of their underlying assumptions in the empirical setting described in Section 2.

We start with several specifications of regression models. First, consider the scenario wherein we ignore the network structure, assuming that the Stable Unit Treatment Value Assumption (SUTVA) holds. Given that resident-level covariates might influence both the adoption of U-Bridge and political participation, we control for these confounders in the first regression model. Estimation results are presented in Column (1) of Table 3. We observe a positive and statistically significant coefficient of adoption, suggesting that adopting U-Bridge increases political participation.

Next, recognizing that residents are inter-connected through various forms of social ties within each village, we are interested in whether adoption of U-Bridge by one resident affects political participation of other residents. Such spillover effects are conceivable due to the transmission of information across social networks. Indeed, Sanchez-Becerra (2022) has identified a notable spillover effect from attending U-Bridge meetings on political participation. To identify spillover effects, we need to specify the exposure mapping. We consider effective treatment $T_i = e(i, \boldsymbol{D}, \boldsymbol{W}) = (D_i, \sum_j W_{ij} D_j) = (D_i, Z_i)$, same as exposure mapping (2). In this application, it means that political participation is determined by both a resident's own adoption of U-Bridge and number of her connections who adopt it. The term $\sum_j W_{ij} D_j$ is then incorporated as an additional treatment variable in the regression model, with its coefficient representing the spillover effect. Estimations are detailed in Column (2) of Table 3, which reveal a positive and substantial spillover effect. The coefficient of D_i, which represents the effect of direct treatment, diminishes, though remains positive and significant. Lastly, an interaction term between D_i and $\sum_j W_{ij} D_j$ is added to capture treatment effect heterogeneity. The results of these estimations are displayed in Columns (3) of Table 3. When incorporating the interaction term, both the coefficient of D_i and the coefficient of $\sum_j W_{ij} D_j$ see a marginal increase. The coefficient of the interaction term is negative, though statistically insignificant, indicating decreasing effect of D_i as $\sum_j W_{ij} D_j$ increases.

While regression models are flexible and easy to implement, the estimates may be biased due to model mis-specification. Moreover, it is hard to interpret the treatment effect when the social network itself is part of the exposure mapping. As a result, we employ the proposed method, a design-based alternative that directly models treatment assignment mechanism and network formation process, to estimate treatment effects. We use the third specification of regression model as outcome model in the proposed AIPW estimator, and then fit network formation and treatment assignment models to construct generalized propensity scores.

We fit network formation models separately for each village to account for village-level heterogeneity. We fit a Logit model, first line in Model (25), for the presence of social tie between a pair of residents. In the model specification, $c(i)$ represents the village that

Table 3: Regression Models for the Effect of U-Bridge on Political Participation

Outcome Variable	Political Participation Index (Y_i)		
	Model (1)	Model (2)	Model (3)
Adoption (D_i)	0.340***	0.261***	0.365***
	(0.700)	(0.069)	(0.096)
Interference ($\sum_j W_{ij} D_j$)		0.048***	0.053***
		(0.007)	(0.007)
Interaction ($D_i \cdot \sum_j W_{ij} D_j$)			-0.029
			(0.015)
Control Variable (X_i)	✓	✓	✓
Village Fixed Effects	✓	✓	✓
Observations	3,018	3,018	3,018

Notes: Robust standard errors clustered at village level are in the parentheses.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

resident i belongs to. Control variables include absolute differences in age, gender, levels of income, secondary education, and proxy measure of care for the community. These variables represent observed homophily. We include binary indicators for whether at least one of them own a phone and whether at least one of them occupy a leadership role within the village. In addition, we add village-specific intercept, $\gamma_{w,c(i)}$, to control for additional village-level variations, like trust and informal institutions, in formation of social ties. Estimation results are displayed in Section C of the appendix. While estimated coefficients are heterogeneous across villages, for most villages, coefficients for the homophily measures are negative, and corresponding 95% confidence intervals exclude 0. In addition, the effects of occupying formal leadership role and owning a phone are positive and statistically significant on social tie formation.

To model treatment assignment mechanism, we fit another Logit model, second line in Model (25), for adoption of U-Bridge that includes resident-level covariates, and village-level fixed effects ($\gamma_{d,c(i)}$) to control for unobserved village-level heterogeneity. Estimation results are summarized in Table 10 in Section C of the appendix. We find that coefficients for covariates like gender, education and leadership are statistically significant, while coefficients for income and age are insignificant. We then estimate generalized propensity score according to formulation (15) given estimated models for network formation and treatment assignment.

$$\begin{aligned} Pr(W_{ij}=1) &= \text{logit}^{-1}(|X_i - X_j|'\beta_{w1,c(i)} + X'_{ij}\beta_{w2,c(i)} + \gamma_{w,c(i)}), \\ Pr(D_i=1) &= \text{logit}^{-1}(X'_i\beta_{d1} + \gamma_{d,c(i)}). \end{aligned} \tag{25}$$

We evaluate treatment effect estimates based on these two different estimation strategies. Given that only a small proportion of respondents have adopted the U-Bridge program, certain levels of effective treatment contain limited observations. As a result, our analyses primarily focus on two direct treatment effects: $\tau_{d1} = \mu(1,1) - \mu(0,1)$ and $\tau_{d2} = \mu(1,2) - \mu(0,2)$, and two spillover effects: $\tau_{i1} = \mu(0,1) - \mu(0,0)$ and $\tau_{i2} = \mu(0,2) - \mu(0,0)$. For proposed AIPW estimator, we trim the top and bottom 2.5% estimated propensity score to avoid extreme weights and calculate robust standard error clustered at village level for inference. For regression models, we focus on the third specification used in the AIPW estimator for model comparison. We derive estimates of treatment effects using esti-

mated coefficients. Estimated treatment effects with 95% confidence intervals are displayed in Figure 6.

For direct effect $\hat{\tau}_{d1}$, both AIPW estimator and regression model yield positive and statistically significant estimates, and the magnitudes of estimated effects are similar. We may conclude that, conditional on connection to one resident that adopted U-Bridge, the adoption of U-Bridge increased residents' political participation. Next, we investigate whether the direct effect is constant or is it moderated by number of residents' connections. Regarding the direct effect $\hat{\tau}_{d2}$, point estimate of AIPW estimator decreases, though the 95% CI overlaps with that of $\hat{\tau}_{d1}$. Estimate based on regression model also decreases marginally, as the coefficient for the interaction term is negative.

For both spillover effects, estimates based on AIPW estimator and regression model are all positive and statistically significant. We also find that point estimates $\hat{\tau}_{s2}$ are larger than those of $\hat{\tau}_{s1}$. For AIPW estimator, difference between $\hat{\tau}_{s2}$ and $\hat{\tau}_{s1}$ is greater than the corresponding difference based on regression model. Given the results, we may conclude that the adoption of U-Bridge has positive spillover effect. Substantively, it means that if a resident is connected to other adopters, she may also be more active in political participation. In addition, the magnitude of such spillover effects increase as number of connections with adopters increases. While the patterns of estimates are similar, discrepancies between AIPW estimator and regression model may arise from neglecting the network formation process and misspecification of the outcome model. The AIPW estimator, however, is doubly robust and does not impose strong assumptions on the structure of spillover effects.

7 Conclusion

In this paper, we propose a generalized propensity score based approach to identification and estimation of treatment effects from observational social network data. In such data, treatment assigned to one unit may affect potential outcome of other units. Meanwhile, there may be some covariates that determines adoption of the treatment, formation of social ties and observed behavioral outcomes, making identification and estimation a challenging task. To incorporate the rich interactions between treatment assignment and network formation process, we propose to jointly model these two processes. Given a known func-

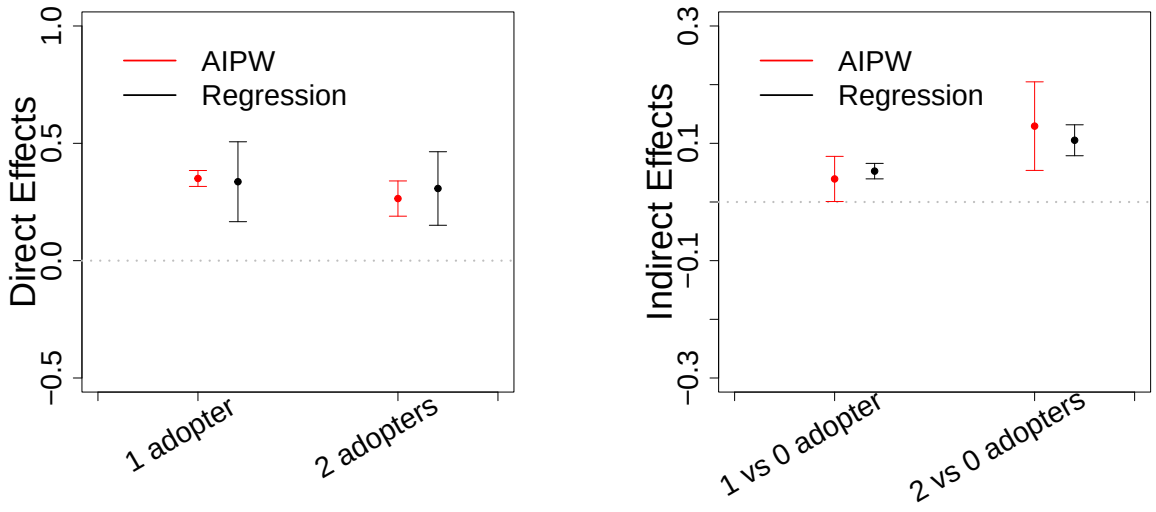


Figure 6: Estimated Effects of U-Bridge Adoption on Political Participation

Note: The left panel shows estimated direct effects and the right panel shows estimated spillover effects based on different approaches. Abbreviations "AIPW" represents the proposed method.

tional form of exposure mapping that determines the effective treatment level, generalized propensity score for each treatment level is estimated based on the probabilistic models for treatment assignment and network formation. An estimate of average potential outcome and treatment effect is obtained by implementing inverse probability weighting estimators.

We investigate its performance and finite sample properties in several Monte Carlo studies and illustrate its applicability in an empirical application on the effect of adoption of a new political communication technology on political participation. We compare the estimation results obtained from the proposed AIPW estimator and regression model. Their differences may arise from misspecification of regression model and ignoring network formation process. For practical insights, researchers should consider network formation process, treatment assignment mechanism, and potential interactions between these processes when estimating treatment effects from social network data. Only then can they select the most appropriate estimation strategy for their specific empirical setting.

While the proposed method can incorporate different types of treatment variables and networks and complex interactions between them in observational data, it has several lim-

itations. First, it still relies on strong assumptions about the parametric forms of treatment assignment mechanism, network formation process, and functional form of exposure mapping. Misspecification in any of these can lead to bias in estimation. Second, the method is computationally more intensive compared to some existing approaches that directly estimate generalized propensity scores through parametric models, limiting its practical application to relatively small datasets. Finally, the proposed method is designed for cross-sectional data where the network is random but static. Extending this framework to account for interactions and feedback among treatment assignment, network dynamics, and behavioral outcomes is essential for advancing causal inference in longitudinal social network data and represents an important avenue for future research.

References

Aronow, P. M. and C. Samii (2017). Estimating average causal effects under general interference, with application to a social network experiment. *Annals of Applied Statistics 11*(4), 1912–1947. 2, 3, 10, 17

Arpino, B. and A. Mattei (2016). Assessing the causal effects of financial aids to firms in tuscany allowing for interference. *The Annals of Applied Statistics*, 1170–1194. 1, 10

Cai, J., A. D. Janvry, and E. Sadoulet (2015). Social networks and the decision to insure. *American Economic Journal: Applied Economics 7*(2), 81–108. 19

Cai, Y. (2022). Linear regression with centrality measures. *arXiv preprint arXiv:2210.10024*.

Comola, M. and S. Prina (2021). Treatment effect accounting for network changes. *The Review of Economics and Statistics 103*(3), 597–604. 2

Eubank, N., G. Grossman, M. Platas, J. Rodden, et al. (2021). Viral voting: Social networks and political participation. *Quarterly Journal of Political Science 16*(3), 265–284. 4

Feng, P., X.-H. Zhou, Q.-M. Zou, M.-Y. Fan, and X.-S. Li (2012). Generalized propensity score for estimating the average treatment effect of multiple treatments. *Statistics in medicine 31*(7), 681–697.

Ferrali, R., G. Grossman, M. R. Platas, and J. Rodden (2020). It takes a village: Peer effects and externalities in technology adoption. *American Journal of Political Science 64*(3), 536–553. 4, 5

Forastiere, L., E. M. Airoldi, and F. Mealli (2021). Identification and estimation of treatment and interference effects in observational studies on networks. *Journal of the American Statistical Association 116*(534), 901–918. 3, 10, 12, 13

Forastiere, L., F. Mealli, A. Wu, and E. M. Airoldi (2022). Estimating causal effects under network interference with bayesian generalized propensity scores. *Journal of Machine Learning Research 23*(289), 1–61.

Frangakis, C. E. and D. B. Rubin (1999). Addressing complications of intention-to-treat analysis in the combined presence of all-or-none treatment-noncompliance and subsequent missing outcomes. *Biometrika 86*(2), 365–379.

Franzese, R. J. and J. C. Hays (2008). Contagion, common exposure, and selection: Empirical modeling of the theories and substance of interdependence in political science. *Available at SSRN 1323047*. 2, 15

Glynn, A. N. and K. M. Quinn (2010). An introduction to the augmented inverse propensity weighted estimator. *Political analysis 18*(1), 36–56. 18

Goldsmith-Pinkham, P. and G. W. Imbens (2013). Social networks and the identification of peer effects. *Journal of Business & Economic Statistics 31*(3), 253–264. 1

Griffith, A. (2022). Random assignment with non-random peers: A structural approach to counterfactual treatment assessment. *Review of Economics and Statistics*, 1–40.

Grossman, G., M. Humphreys, and G. Sacramone-Lutz (2020). Information technology and political engagement: Mixed evidence from uganda. *The Journal of Politics 82*(4), 1321–1336. 4

Han, X., C.-S. Hsieh, and S. I. Ko (2021). Spatial modeling approach for dynamic network formation and interactions. *Journal of Business & Economic Statistics 39*(1), 120–135. 2, 15

Hirano, K. and G. W. Imbens (2004). The propensity score with continuous treatments. *Applied Bayesian modeling and causal inference from incomplete-data perspectives 226164*, 73–84. 13

Hoff, P. (2021). Additive and multiplicative effects network models. *Statistical Science 36*(1), 34–50. 15

Horvitz, D. G. and D. J. Thompson (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American statistical Association 47*(260), 663–685.

Hudgens, M. G. and M. E. Halloran (2008). Toward causal inference with interference. *Journal of the American Statistical Association 103*(482), 832–842. 7

Imbens, G. W. (2000). The role of the propensity score in estimating dose-response functions. *Biometrika 87*(3), 706–710. 2

Jackson, M. (2008). Social and economic networks. 9

Jackson, M. O., Z. Lin, and N. N. Yu (2022). Adjusting for peer-influence in propensity scoring when estimating treatment effects. *Available at SSRN 3522256*. 3, 9, 12

Kinne, B. J. (2012). Multilateral trade and militarized conflict: Centrality, openness, and asymmetry in the global trade network. *The Journal of Politics 74*(1), 308–322.

Kojevnikov, D., V. Marmer, and K. Song (2021). Limit theorems for network dependent random variables. *Journal of Econometrics 222*(2), 882–908. 2, 19, 36

Leung, M. P. and P. Loupos (2022). Unconfoundedness with network interference. *arXiv preprint arXiv:2211.07823*. 2, 3, 9, 12, 14, 19, 36

Manski, C. F. (2013). Identification of treatment response with social interactions. *The Econometrics Journal 16*(1), S1–S23. 1, 10

McCaffrey, D. F., B. A. Griffin, D. Almirall, M. E. Slaughter, R. Ramchand, and L. F. Burgette (2013). A tutorial on propensity score estimation for multiple treatments using generalized boosted models. *Statistics in medicine 32*(19), 3388–3414.

Ogburn, E. L., O. Sofrygin, I. Diaz, and M. J. Van der Laan (2024). Causal inference for social network data. *Journal of the American Statistical Association 119*(545), 597–611. 3

Pearl, J. (2009). *Causality*. Cambridge university press. 8

Rosenbaum, P. R. and D. B. Rubin (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika 70*(1), 41–55. 13

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational Psychology 66*(5), 688. 8

Rubin, D. B. (1986). Which ifs have causal answers; comment on holland (1986). *Journal of the American Statististical Association 81*, 961–962. 8

Sanchez-Becerra, A. (2022). The network propensity score: Spillovers, homophily, and selection into treatment. *arXiv preprint arXiv:2209.14391*. 3, 5, 7, 13, 20, 24

Sävje, F. (2023). Causal inference with misspecified exposure mappings: separating definitions and assumptions. *Biometrika*, asad019. 10

Simmons, B. A. and Z. Elkins (2004). The globalization of liberalization: Policy diffusion in the international political economy. *American political science review 98*(1), 171–189. 1

Sinclair, B., M. McConnell, and D. P. Green (2012). Detecting spillover effects: Design and analysis of multilevel experiments. *American Journal of Political Science 56*(4), 1055–1069. 1, 5

Toulis, P., A. Volfovsky, and E. M. Airoldi (2021). Estimating causal effects when treatments are entangled by network dynamics. 3, 17

SUPPLEMENTARY MATERIAL

A Proofs

1. Proof for proposition 1. First, it is obvious that $\boldsymbol{X}$ contains all information about GPS $r(i,t;\boldsymbol{X})$. Therefore,

$$\begin{aligned} & Pr(T_i = t|\boldsymbol{X}, r(i,t;\boldsymbol{X})) \\ =&\mathbb{E}\left[\mathbb{1}\{e(i,\boldsymbol{D},\boldsymbol{W}) = t\}|\boldsymbol{X}, r(i,t;\boldsymbol{X})\right] \\ =&\mathbb{E}\left[\mathbb{1}\{e(i,\boldsymbol{D},\boldsymbol{W}) = t\}|\boldsymbol{X}\right] \\ =&Pr(T_i = t|\boldsymbol{X}) \\ =&r(i,t;\boldsymbol{X}). \end{aligned} \tag{26}$$

   By the law of iterated expectation, we have

$$\begin{aligned} & Pr(T_i = t|r(i,t;\boldsymbol{X})) \\ =&\mathbb{E}\left[\mathbb{1}\{e(i,\boldsymbol{D},\boldsymbol{W}) = t\}|r(i,t;\boldsymbol{X})\right] \\ =&\mathbb{E}\left[\mathbb{E}[\mathbb{1}\{e(i,\boldsymbol{D},\boldsymbol{W}) = t\}|\boldsymbol{X}]|r(i,t;\boldsymbol{X})\right] \\ =&\mathbb{E}\left[r(i,t;\boldsymbol{X})|r(i,t;\boldsymbol{X})\right] \\ =&r(i,t;\boldsymbol{X}). \end{aligned} \tag{27}$$

   Therefore, $Pr(T_i = t|\boldsymbol{X}, r(i,t;\boldsymbol{X})) = Pr(T_i = t|r(i,t;\boldsymbol{X}))$. Next, we prove the balancing property of GPS.

$$\begin{aligned} & Pr(T_i = t|Y_i(\cdot), r(i,t;\boldsymbol{X})) \\ =&\mathbb{E}\left[\mathbb{1}\{e(i,\boldsymbol{D},\boldsymbol{W}) = t\}|Y_i(\cdot), r(i,t;\boldsymbol{X})\right] \\ =&\mathbb{E}\left[\mathbb{E}[\mathbb{1}\{e(i,\boldsymbol{D},\boldsymbol{W}) = t\}|\boldsymbol{X}, Y_i(\cdot), r(i,t;\boldsymbol{X})]|Y_i(\cdot), r(i,t;\boldsymbol{X})\right] \\ =&\mathbb{E}\left[\mathbb{E}[\mathbb{1}\{e(i,\boldsymbol{D},\boldsymbol{W}) = t\}|\boldsymbol{X}, Y_i(\cdot)]|Y_i(\cdot), r(i,t;\boldsymbol{X})\right] \\ =&\mathbb{E}\left[\mathbb{E}[\mathbb{1}\{e(i,\boldsymbol{D},\boldsymbol{W}) = t\}|\boldsymbol{X}]|Y_i(\cdot), r(i,t;\boldsymbol{X})\right] \\ =&\mathbb{E}\left[r(i,t;\boldsymbol{X})|Y_i(\cdot), r(i,t;\boldsymbol{X})\right] \\ =&r(i,t;\boldsymbol{X}). \end{aligned} \tag{28}$$

We have the third line in Equation (28) by the law of iterated expectation, and then the fourth line because $\boldsymbol{X}$ contains all information about GPS $r(i,t;\boldsymbol{X})$. By condition (7), we have the fifth line.

2. Proof for proposition 2.

$$
\begin{aligned}
&\mathbb{E}\{\frac{Y_i\mathbb{1}\{T_i=g(i,\boldsymbol{D},\boldsymbol{W})=t\}}{r(i,t;\boldsymbol{X})}\}\\
=&\mathbb{E}\{\mathbb{E}\{\frac{Y_i\mathbb{1}\{T_i=g(i,\boldsymbol{D},\boldsymbol{W})=t\}}{r(i,t;\boldsymbol{X})}|\boldsymbol{X}\}\}\\
=&\mathbb{E}\{\mathbb{E}\{\frac{Y_i}{r(i,t;\boldsymbol{X})}|T_i=t,\boldsymbol{X}\}Pr(T_i=t|\boldsymbol{X})\}\\
=&\mathbb{E}\{\mathbb{E}\{\frac{Y_i}{r(i,t;\boldsymbol{X})}|T_i=t,\boldsymbol{X}\}r(i,t;\boldsymbol{X})\}\\
=&\mathbb{E}\{\mathbb{E}\{Y_i|T_i=t,\boldsymbol{X}\}\}\\
=&\mathbb{E}\{\mathbb{E}\{Y_i(t)|\boldsymbol{X}\}\}\\
=&\mathbb{E}\{Y_i(t)\}\\
=&\mu(t).
\end{aligned}
\tag{29}
$$

Therefore, average potential outcome $\mu(t)$ is identified given GPS $r(i,t;\boldsymbol{X})$.

3. Proof for the asymptotic distribution (20) of the AIPW estimator (19). Define

$$
\begin{aligned}
\tau_i(t,t') =&[\mathbb{1}\{T_i=t\}\frac{Y_i-\mu(i,t,\boldsymbol{X})}{r(i,t;\boldsymbol{X})}+\mu(i,t,\boldsymbol{X})]\\
&-[\mathbb{1}\{T_i=t'\}\frac{Y_i-\mu(i,t',\boldsymbol{X})}{r(i,t';\boldsymbol{X})}+\mu(i,t',\boldsymbol{X})]
\end{aligned}
$$

and $\varphi_i(t,t')=\tau_i(t,t')-\tau(t,t')$. First, we list the set of regularity conditions:

(a) Denote $\mathcal{C}_N$ the σ-algebra generated by $\boldsymbol{X}$. $\{\tau_i(t,t')\}_{i=1}^N$ is ψ-dependent given $\mathcal{C}_N$.

(b) Moment conditions.

i. There exists $M<\infty$ and $p>4$ such that for all $N\in\mathbb{N}, i\in\{1,\dots,N\}$, $\boldsymbol{d}\in\{0,1\}^N$, and $\boldsymbol{w}\in\mathcal{W}$, $\mathbb{E}\left[|Y_i(\boldsymbol{d},\boldsymbol{w})|^p\,|\boldsymbol{X}\right]<M$ a.s.

ii. There exists $[\underline{\pi},\bar{\pi}]\subset(0,1)$ such that for all $N\in\mathbb{N}, i\in\{1,\dots,N\}$, $t\in\mathcal{T}$, $\hat{r}(i,t;\boldsymbol{X})$ and $r(i,t;\boldsymbol{X})\in[\underline{\pi},\bar{\pi}]$ a.s.

(c) Convergence rates of estimated generalized propensity score and outcome model.

i. $N^{-1}\sum_{i=1}^{N}\left(\hat{r}(i,t;\boldsymbol{X})-r(i,t;\boldsymbol{X})\right)^2=o_p(1)$ and $N^{-1}\sum_{i=1}^{N}\left(\hat{\mu}(i,t,\boldsymbol{X})-\mu(i,t,\boldsymbol{X})\right)^2=o_p(1)$.

ii. $N^{-1}\sum_{i=1}^{N}\left(\hat{r}(i,t;\boldsymbol{X})-r(i,t;\boldsymbol{X})\right)^2 N^{-1}\sum_{i=1}^{N}\left(\hat{\mu}(i,t,\boldsymbol{X})-\mu(i,t,\boldsymbol{X})\right)^2=o_p\left(N^{-1}\right)$.

iii. $N^{-1}\sum_{i=1}^{N}\left(\hat{\mu}_t(i,\boldsymbol{X},\boldsymbol{A})-\mu_t(i,\boldsymbol{X},\boldsymbol{A})\right)\left(1-\mathbb{1}\{T_i=t\}r(i,t;\boldsymbol{X})^{-1}\right)=o_p\left(N^{-\frac{1}{2}}\right)$.

(d) Weak dependence. Error terms $\{(\epsilon_{ij}^w,\epsilon_i^d,\epsilon_i^y)\}_{i,j=1}^N$ in NP-SEM (1) are independently distributed conditional on $\boldsymbol{X}$ a.s.

Then we decompose $\sqrt{N}\left(\hat{\tau}\left(t,t'\right)-\tau\left(t,t'\right)\right)$ as:

$$\sqrt{N}\left(\hat{\tau}\left(t,t'\right)-\tau\left(t,t'\right)\right)=\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\varphi_i(t,t')-R_{1t}+R_{1t'}-R_{2t}+R_{2t'},\tag{30}$$

where

$$R_{1t}=\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{\mathbb{1}\{T_i=t\}\left(Y_i-\mu(i,t,\boldsymbol{X})\right)}{\hat{r}(i,t;\boldsymbol{X})r(i,t;\boldsymbol{X})}\left(\hat{r}(i,t;\boldsymbol{X})-r(i,t;\boldsymbol{X})\right),$$

$$R_{1t'}=\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{\mathbb{1}\{T_i=t'\}\left(Y_i-\mu(i,t',\boldsymbol{X})\right)}{\hat{r}(i,t';\boldsymbol{X})r(i,t';\boldsymbol{X})}\left(\hat{r}(i,t';\boldsymbol{X})-r(i,t';\boldsymbol{X})\right),$$

$$R_{2t}=\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\hat{\mu}(i,t,\boldsymbol{X})-\mu(i,t,\boldsymbol{X})\right)\left(1-\frac{\mathbb{1}\{T_i=t\}}{\hat{r}(i,t;\boldsymbol{X})}\right),$$

$$R_{2t'}=\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\hat{\mu}(i,t',\boldsymbol{X})-\mu(i,t',\boldsymbol{X})\right)\left(1-\frac{\mathbb{1}\{T_i=t'\}}{\hat{r}(i,t';\boldsymbol{X})}\right).$$

We follow Leung and Loupos (2022) to apply Theorem 3.2 in Kojevnikov et al. (2021) and obtain

$$\sigma_{t,t'}^{-1}\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\varphi_i(t,t')\xrightarrow{d}\mathcal{N}(0,1).$$

By Theorem 3 in Leung and Loupos (2022), $R_{1t}=o_p(1)$, $R_{1t'}=o_p(1)$, $R_{2t}=o_p(1)$, and $R_{2t'}=o_p(1)$. Thus,

$$\sigma_{t,t'}^{-1}\sqrt{N}(\hat{\tau}(t,t')-\tau(t,t'))\xrightarrow{d}\mathcal{N}(0,1).\tag{31}$$

B Additional Results for Monte Carlo Studies

Results for treatment effects estimation based on regression model for DGP (23) are summarized in Table 4 and Table 5.

Table 4: Regression: Finite Sample Properties: $\tau((1,1),(0,1))$

N_c	Bias	RMSE	Coverage Rate	SE	SE - SD
50	−0.401	0.449	0.488	0.201	−0.002
100	−0.388	0.413	0.228	0.142	0.000
200	−0.389	0.402	0.028	0.100	0.000
500	−0.389	0.394	0.000	0.063	0.000
1000	−0.390	0.393	0.000	0.045	0.000

Table 5: Regression: Finite Sample Properties: $\tau((1,1),(1,0))$

N_c	Bias	RMSE	Coverage Rate	SE	SE - SD
50	−0.350	0.382	0.370	0.151	−0.001
100	−0.348	0.364	0.105	0.107	−0.001
200	−0.344	0.353	0.007	0.076	−0.001
500	−0.344	0.348	0.000	0.048	0.000
1000	−0.343	0.345	0.000	0.034	0.000

To investigate whether the proposed method is robust to heterogeneous treatment effects (HTE), we consider an alternative data generating process for the potential outcome:

$$Y_i(d,z) = 1 + X_i + \frac{\sum_{j\in c(i), j\neq i} X_j}{3} + d + 0.5z + dz + X_i dz + \varepsilon_i, \quad \varepsilon_i \sim N(0,1). \tag{32}$$

Therefore, the individual level treatment effect $Y_i(d,z) - Y_i(\tilde{d},\tilde{z})$ depends on the covariate X_i. For the AIPW estimator, results for both treatment effects are summarized in Table 6 and Table 8. And for regression based estimator, results are summarized in Table 7 and Table 9. For model comparison, bias and RMSE plots for the spillover effect are displayed in Figure 7. We find that the results are quite similar to the case of constant treatment effect, though sampling variation for the AIPW estimator is larger when treatment effects are heterogeneous. Therefore, the proposed method performs well even when treatment effects are heterogeneous.

C Additional Results and Plots for the Empirical Application

This section includes additional results and plots for the empirical application. Figure 8 and Figure 9 display network structures for the rest 15 villages in the data sample. Residents

Table 6: Finite Sample Properties (HTE): $\tau((1,1),(0,1))$

N_c	Bias	RMSE	Coverage Rate	SE	SE - SD
50	0.023	0.504	0.938	0.444	−0.065
100	0.003	0.350	0.953	0.333	−0.022
200	0.001	0.257	0.949	0.246	−0.013
500	0.003	0.172	0.950	0.162	−0.011
1000	0.002	0.115	0.958	0.115	−0.001

Table 7: Regression: Finite Sample Properties (HTE): $\tau((1,1),(0,1))$

N_c	Bias	RMSE	Coverage Rate	SE	SE - SD
50	−0.432	0.489	0.527	0.228	−0.004
100	−0.422	0.450	0.254	0.161	0.002
200	−0.421	0.437	0.046	0.114	−0.001
500	−0.422	0.428	0.000	0.072	0.001
1000	−0.422	0.425	0.000	0.051	0.000

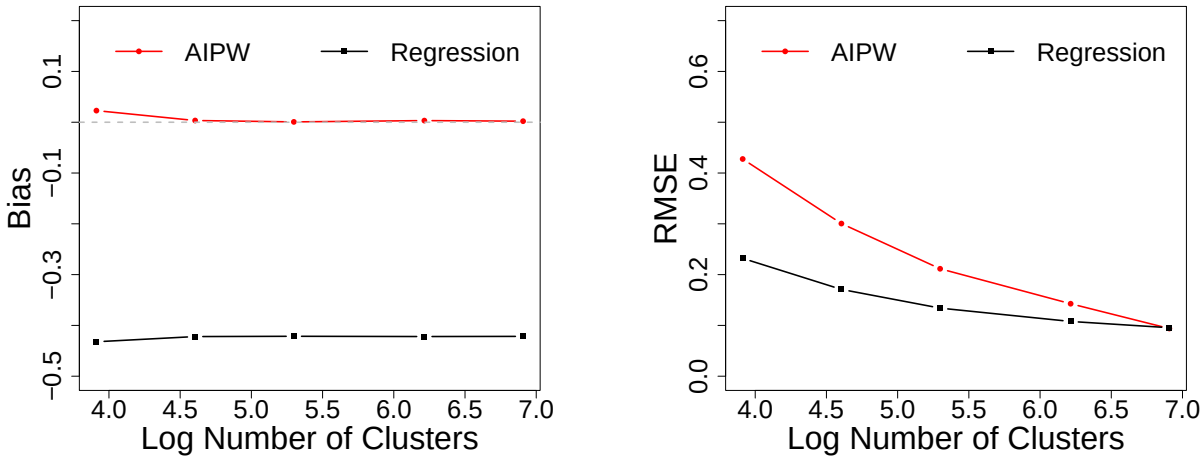


Figure 7: Finite Sample Properties (HTE): Bias and RMSE

who adopted the U-Bridge program are highlighted in blue and Residents who occupy a formal leader role within village are highlighted in red. Figure 10 and Figure 11 show estimated coefficients with 95% confidence intervals in the network formation models across the 16 villages under study. Table 10 summarizes estimation results for treatment assignment mechanism based on Logit model.

Table 8: Finite Sample Properties (HTE): $\tau((1,1),(1,0))$

N_c	Bias	RMSE	Coverage Rate	SE	SE - SD
50	0.033	0.428	0.931	0.385	−0.046
100	0.001	0.301	0.944	0.290	−0.016
200	0.004	0.212	0.945	0.212	−0.002
500	−0.002	0.143	0.952	0.140	−0.004
1000	0.002	0.094	0.964	0.099	0.004

Table 9: Regression: Finite Sample Properties (HTE): $\tau((1,1),(1,0))$

N_c	Bias	RMSE	Coverage Rate	SE	SE - SD
50	−0.091	0.232	0.879	0.204	−0.011
100	−0.091	0.171	0.882	0.146	−0.002
200	−0.085	0.134	0.854	0.104	−0.001
500	−0.086	0.108	0.740	0.067	0.000
1000	−0.083	0.095	0.566	0.047	0.001

Table 10: Estimation Results for Treatment Assignment Mechanism

Treatment Variable	Individual Adoption of U-Bridge Program (D_i)
Age	-0.008
	(0.007)
Female	-1.189***
	(0.233)
Income	0.018
	(0.087)
Education	1.960***
	(0.234)
Leader	0.771**
	(0.241)
Care Community	-1.164**
	(0.437)
Has Phone	1.304***
	(0.309)
Village Fixed Effects	✓

Notes: Robust standard errors clustered at village level are in the parentheses.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

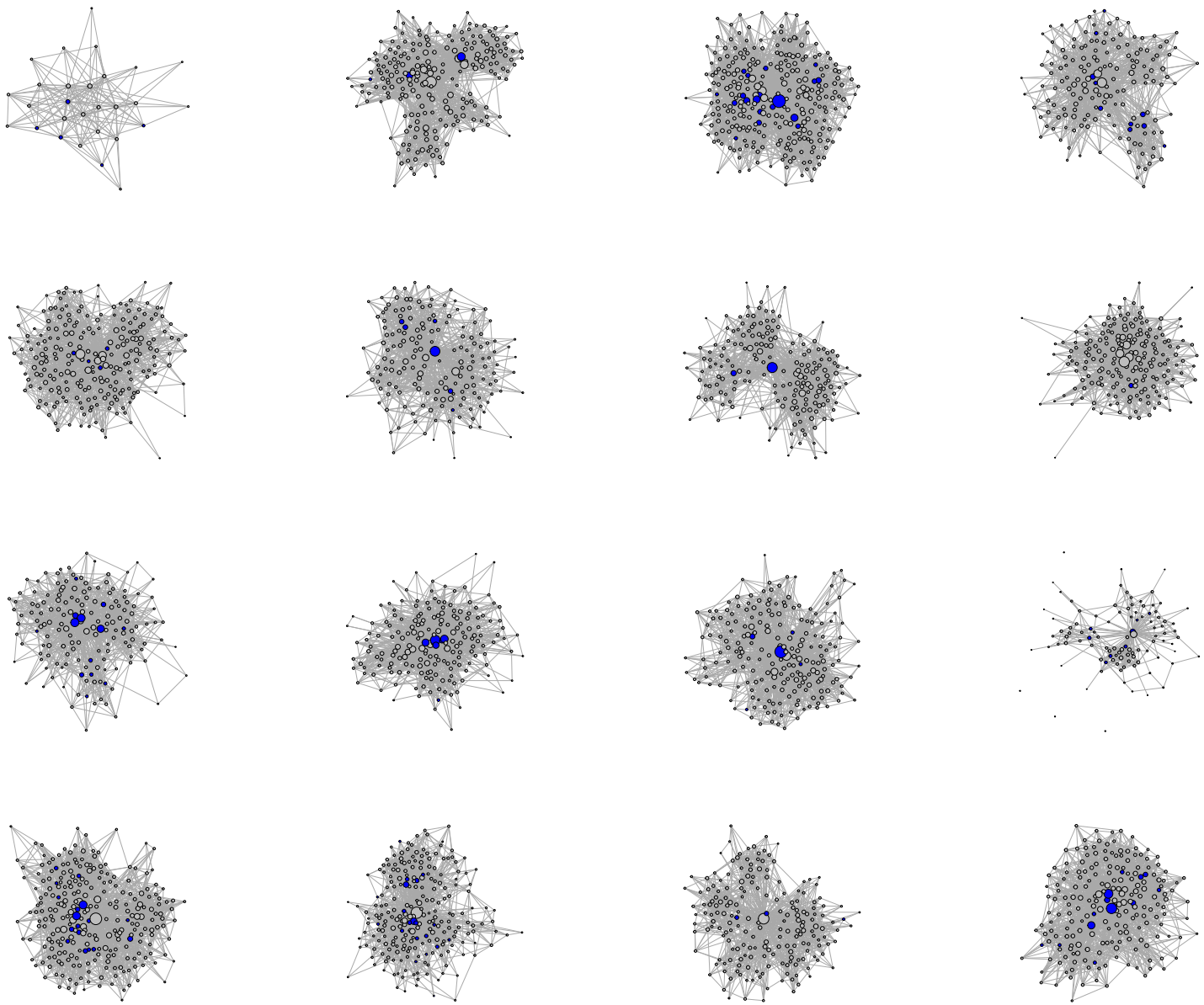


Figure 8: Additional Network Visualizations for Villages (Treatment)

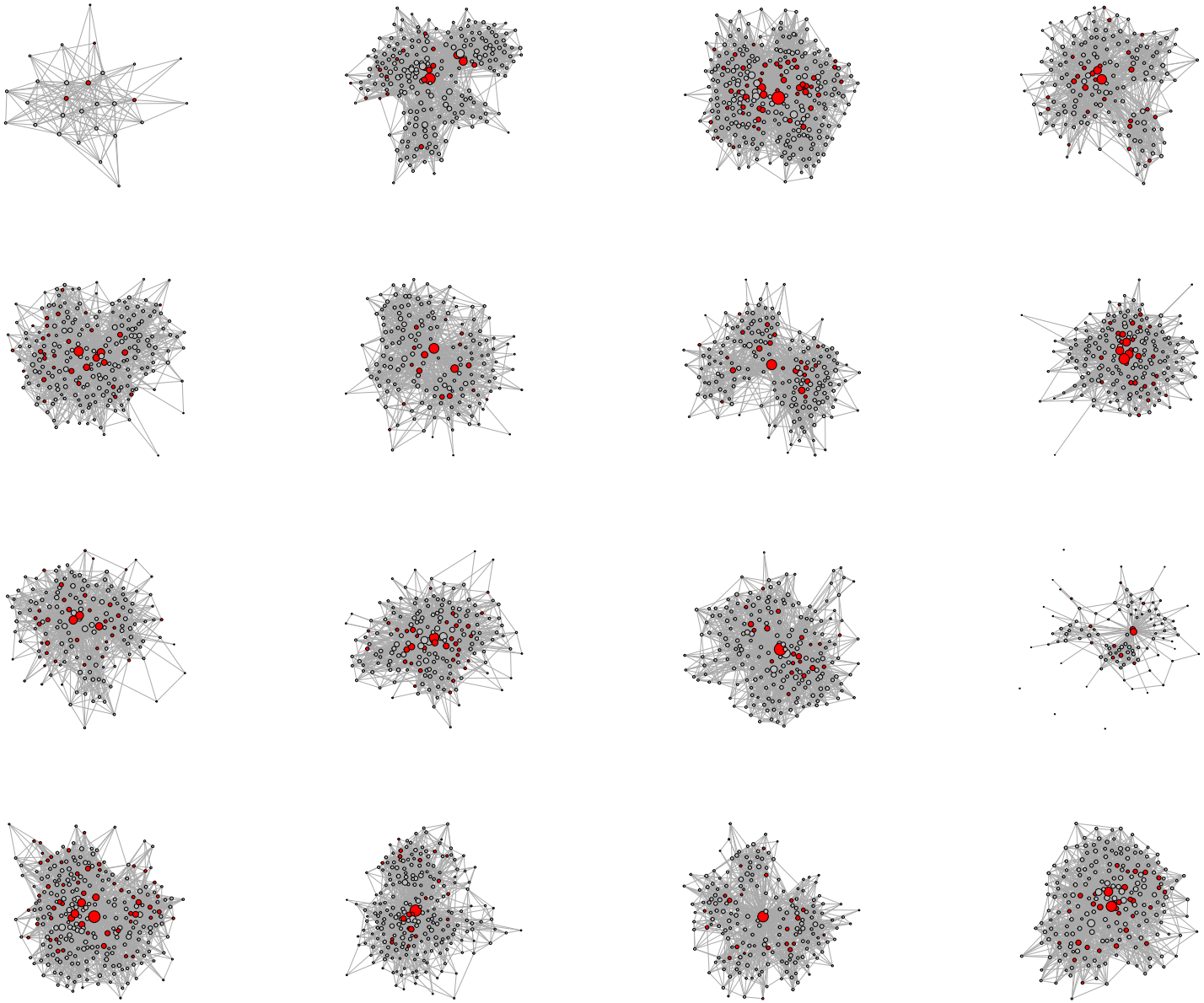


Figure 9: Additional Network Visualizations for Villages (Leadership)

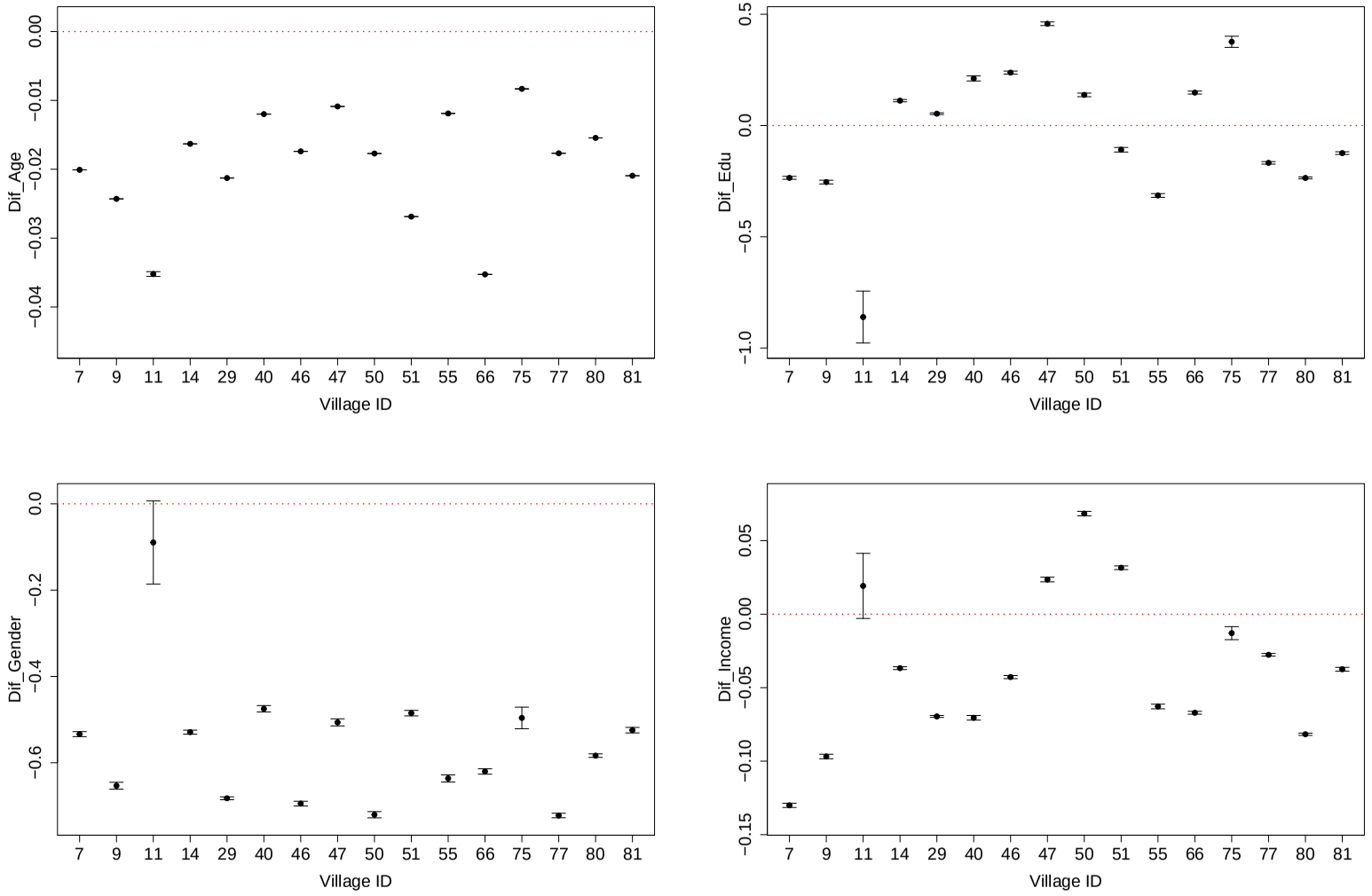


Figure 10: Heterogeneous Coefficients in Network Formation Models

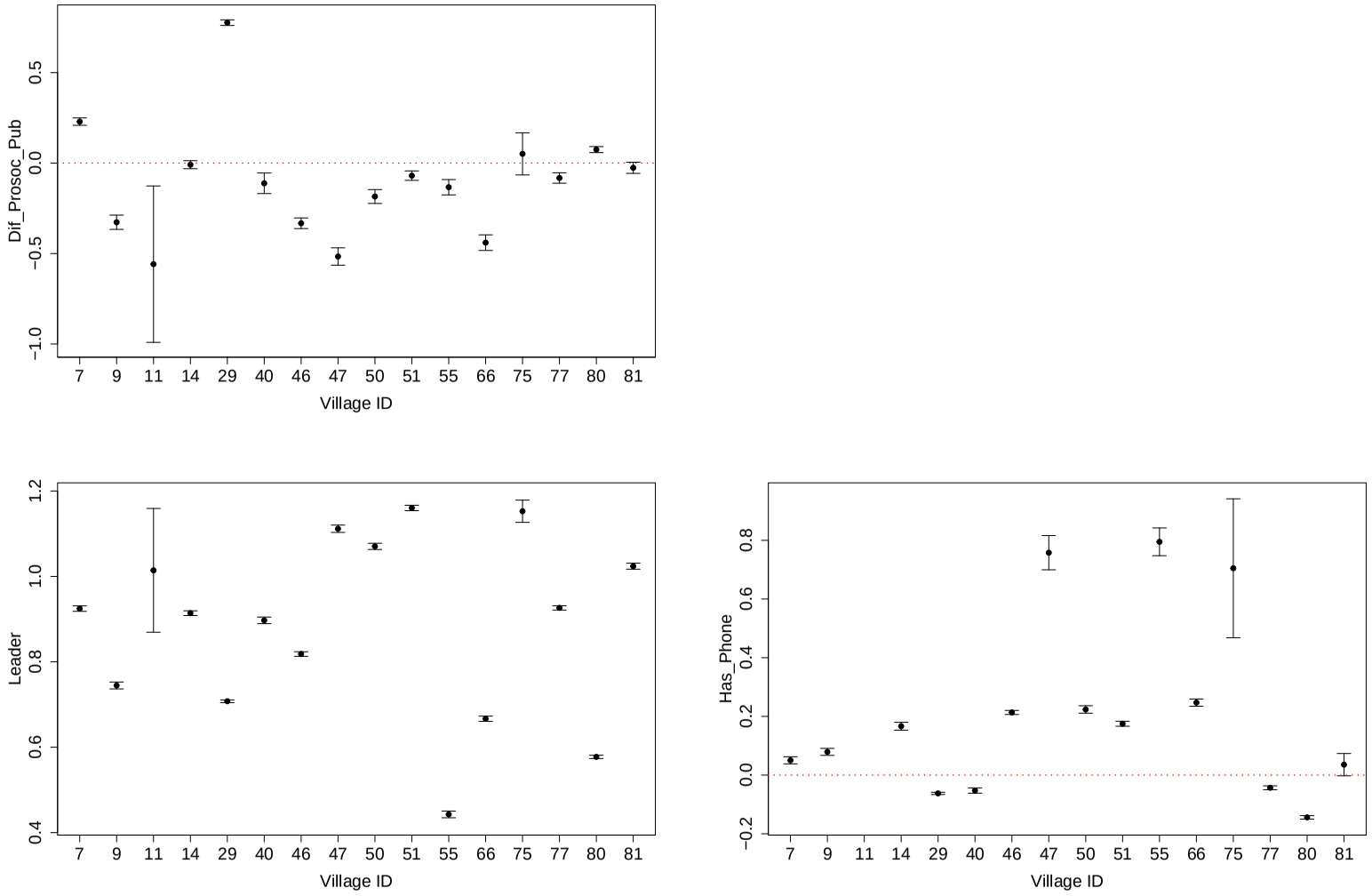


Figure 11: Heterogeneous Coefficients in Network Formation Models Cont'