# Spillovers, Homophily, and Selection into Treatment: The Network Propensity Score<sup>\*</sup>

Alejandro Sánchez-Becerra<sup>1</sup>

<sup>1</sup>Department of Economics, University of Pennsylvania

Job Market Paper Latest Version Here This version: March 19, 2021

#### Abstract

Propensity score matching is often used to estimate treatment effects when there is selection on observables; however, it fails to identify causal effects when one person's treatment affects another's outcome. This phenomenon is known as *spillovers*. I propose a novel *network propensity score matching* approach that identifies both the average treatment effects and the average spillover effects between individuals. My approach is grounded in an endogenous model of network formation with spillovers on the outcome. This methodology can be used to identify causal effects for individuals with similar observables, analogous to the propensity score. I then propose estimators that are consistent and asymptotically normal for settings with multiple networks. I apply my methodology to two empirical examples. First, I study the effects of an intervention on political participation in Uganda where I find evidence of spillovers on non-participants. Second, I evaluate a microfinance adoption intervention in India, and find large treatment effects but limited spillovers effects. In some extensions of the method, I show how to conduct robustness checks and how to interpret the network propensity score in stratified multi-stage experiments.

Keywords: Networks, Selection into Treatment, Causal Inference.

<sup>\*</sup>I would like to thank Xu Cheng, Francis J. DiTraglia, Petra Todd, Jere Behrman, Frank Schorfheide, Francis Diebold, Wayne Gao, Karun Adusumilli, Ben Golub, Mathew Jackson, Hyungsik Roger Moon, Bryan Graham, Suyong Song, Andrin Pelican, Luis Candelaria, Aureo de Paula, and Juan Camilo Castillo, for helpful comments and suggestions, as well as seminar participants at the Young Economists Symposium 2020, the Econometric Society World Congress 2020, the Warwick 2019 Ph.D. Conference and the UPenn Empirical Micro and Econometrics Workshops. Email: alesan@sas.upenn.edu

# 1 Introduction

Propensity score matching is a procedure for comparing the average outcomes of matched-pairs of individuals. It is used to estimate program treatment effects, typically from observational data on a cross-section of individuals whose outcome depends on their treatment status.<sup>1</sup> At its core, propensity score matching relies on a high level *unconfoundedness* (or *selection-onobservables*) condition, assuming that the key determinants of treatment take-up are observed, and a *support* condition, ensuring a match between comparable treated and control individuals.

Nevertheless, in many practical settings it is plausible that one person's treatment can affect the outcome of a friend in their social network. In these cases –known as spillovers– conventional unconfoundedness and support conditions are insufficient to identify program treatment effects. For example, job placement programs can displace non-participants from the labor market (Crépon et al., 2013), cash transfers can affect informal insurance networks (Meghir et al., 2020), and professional events can encourage the adoption of business practices (Fafchamps and Quinn, 2018). In the presence of spillovers, researchers model the outcome as a function of both an individual's treatment status and a sufficient statistic of friends' treatment decisions. To address identification, a recent literature proposes analogs of the unconfoundedness and support conditions (Forastiere et al., 2020; Liu et al., 2019; Sofrygin and van der Laan, 2017), intended for settings where the social network is observed by the researcher. In spite of its technical convenience, unconfoundedness with respect to a constructed statistic is harder to justify intuitively than unconfoundedness of the treatment, and recent work offers limited guidance on its validity for specific situations.

This paper establishes a set of primitive conditions on the treatment, the network formation process, and a flexible random coefficients model of the outcome, that rationalizes unconfoundedness and support conditions to identify the average partial effects (APE). The APE's in my model are summarized by a vector of program treatment effects that includes an explicit measure of spillovers. To build my argument, I exploit exchangeability: the key regressors of my outcome equation are constructed from a common vector of treatment indicators, weighted by another set of friendship indicators drawn from an exchangeable network process. I then establish two key findings. First, the researcher can satisfy the unconfoundedness condition by choosing individual determinants of treatment take-up and friendship choices. Second, there exists a three-dimensional individual statistic –that I call the network propensity score (NPS)–

<sup>&</sup>lt;sup>1</sup>This assumption is called the Stable Unit Treatment Value Assumption (SUTVA).

which can be used as a matching variable. Crucially, the validity of the support condition depends on heterogeneous take-up probabilities within a friend circle. From a theoretical perspective, the NPS can be expressed as an integrand of the take-up process, friend preferences over traits, and the measure of traits in the population. I use this characterization to establish when the NPS contains complementary information to the propensity score.

This paper builds on the intuitive idea that the determinants of network formation serve as valid controls for identifying program treatment effects. This approach has been leveraged by recent papers that assume exogenous treatment and homogeneous effects (Goldsmith-Pinkham and Imbens, 2013; Johnsson and Moon, 2019). My paper formalizes this idea in a setting with selection on observables and heterogeneous effects by proposing a model that nests the standard potential outcomes used in program evaluation (Fisher et al., 1960; Rubin, 1980). This generalization is important for two reasons. First, nesting the standard model allows for a setting with selection only and without spillovers, where propensity score methods should thrive. Second, it shows that the network determinants approach imposes testable restrictions on the matrix  $\mathbf{Q}_{xx}$ , defined as the (conditional) second moments of the endogenous regressors given covariates. I further show that when the researcher has incomplete set of network determinants, then  $\mathbf{Q}_{xx}$  can be expressed as a mixture of known matrices, and propose additional restrictions on the unobserved heterogeneity that guarantee unconfoundedness. I extend this idea to construct a covariate balancing test, as a robustness check.

I model program treatment effects using a random coefficients model, that allows for correlation with the treatment decisions. The model's average partial effects capture three distinct estimands of interest: Direct effects capture the impact of the program in isolation, spillover effects capture the impact's of other's treatment, and interaction effects capture the differential effect of spillovers on those that are already treated. I propose a two-step identification strategy based on inverse weighting in the style of Graham and Pinto (2018) and Wooldridge (2004). In the first step, I estimate the network propensity score parameters and use them to construct an individual-specific matrix  $\mathbf{Q}_{xx}$ . In the second stage, I estimate the average partial effects by inversely weighting each observation and then averaging the resulting estimands across individuals. I then propose an estimator that is consistent and asymptotically normal in a setting where the number of networks grows along the asymptotic sequence. I compute the standard errors by rewriting the first and second stages as a single GMM system and computing standard sandwich formulas.

I apply my methodology to two empirical examples. First, I consider an intervention de-

signed to increase political participation in Uganda (Eubank et al., 2019; Ferrali et al., 2020). Citizens voluntarily participated in quarterly information sessions about ways to engage with local district officials. I find evidence of spillovers because individuals with a higher number of friends participating in the sessions were more likely to be politically active, after controlling for covariates. The estimates of the spillover effects under my approach are statistically significant and about twice the size of comparable ordinary least squares (OLS) regressions with additive covariates. The network propensity score matching methodology is better equipped to handle heterogeneous spillover effects that can be correlated with the endogeneous regressors.

In the second example, I analyze the effects of an intervention to increase microfinance adoption (Banerjee et al., 2013). This example has been analyzed extensively by the econometrics literature (Candelaria, 2020; Chandrasekhar and Jackson, 2014) and has lead to many follow-up projects (Banerjee et al., 2017; Breza and Chandrasekhar, 2019; Chandrasekhar et al., 2018). The microfinance organization used a non-random selection rule based on occupation of household members (shopkeepers, teachers), who received in-depth information about the loans offered by the company. In practice, households with higher wealth and privileged castes were both more likely to receive treatment themselves and to be friends with others that received treatment as well. My network propensity score matching approach estimates large treatment effects but limited spillover effects. Nevertheless, these results do not necessarily rule out spillover effects through friends-of-friends in a diffusion model (Akbarpour et al., 2018; Banerjee et al., 2013). In this case, the researcher observes loan adoption decisions, but not whether they are informed at each moment in time. This might generate a possible attenuation bias in the spillovers from direct friend connections, because it underestimates the diffusion of information through the network.

Finally this paper considers applications of the network propensity score approach to stratified experiments. I analyze experiments that exogenously assign treatment probabilities across multiple networks (Baird et al., 2019; Crépon et al., 2013; Duflo and Saez, 2003; Vasquez-Bare, 2019) and experiments with treatment assignment variation within networks (Eckles et al., 2017; Ugander et al., 2011). I find that the network propensity score has a simple form in both cases under perfect compliance. I also consider settings with non-compliance and spillovers (DiTraglia et al., 2021; Imai et al., 2020; Vasquez-Bare, 2019). I discuss the applicability of the network propensity score to identify average spillover effects under non-compliance in sparse networks.

There have been three recent approaches in the literature that extend propensity score

methods for use with network data. The first approach uses relationship data and friend covariates to relax the selection on observables assumption. Jackson et al. (2020) assume that program participation is the result of a strategic game with friends (spillovers in treatment), but assume that there are no spillovers on outcomes. The second approach assumes selection on observables (without spillovers) but focuses on pairwise outcomes. Arpino et al. (2015), for example, compute the propensity score of adopting tariff agreements and use it to evaluate their effect on bilateral trade between countries. The third approach, gaining traction in biostatistics, incorporates spillovers by assuming anonymous interactions (Manski, 2013), which implies heterogeneous outcomes that depend on own treatment and the total number of treated friends. This approach is sometimes called *multi-treatment* matching because it assumes that individuals with different numbers of treated friends experience different intensities that satisfy unconfoundedness (Forastiere et al., 2018; Liu et al., 2019; Sofrygin and van der Laan, 2017). This literature focuses on predicting a propensity score for each intensity level, which is equivalent to modeling the distribution of treated friends. No restrictions are imposed on the network process but that generality comes at the cost of a very large vector of propensity scores that needs to be estimated.

My approach is closest to multi-treatment matching, in the sense that I assume anonymous interactions and focus on the distribution of treated friends. However, my paper shows that augmenting the model with a network formation process introduces *overidentifying restrictions* that reduce the dimensionality of the required propensity score. Moreover, the dyadic network model that I assume –where individuals become friends based on the similarity of their pairwise characteristics– is actually quite general. In two influential papers Aldous (1981) and Hoover (1979) showed that *any* network process whose distribution is ex-ante independent of the ordering of agents can be represented as a dyadic network with independent covariates and independent shocks. Furthermore, there is a growing theoretical literature that provides microfoundations of the dyadic model as a limiting network of a dynamic strategic game (Mele, 2017). The dyadic representation is extremely useful to analyze spillovers because it allows us to focus on individual confounders rather than complicated functions of the covariates of others. The network propensity score emerges quite naturally as a sufficient statistic that describes the distribution of treated friends after conditioning on personal information.

The key empirical challenge is whether the covariates of the Aldous-Hoover representation are actually observed or whether some of them may be latent. A recent literature (Gao, 2020; Graham, 2017; Johnsson and Moon, 2019) proposes empirical dyadic models with a single source of unobserved heterogeneity called the *popularity index* because it generates variation in the number of friends. Goldsmith-Pinkham and Imbens (2013) uses a bayesian model of network formation, spillover effects, and exogenous treatment with a similar popularity index. My model allows for plug-in estimates of the degree heterogeneity in the form of Johnsson and Moon (2019) when there are multiple large networks. Testing whether the popularity index is the only source of unobserved heterogeneity is an area of ongoing research. Pelican and Graham (2019) test the validity of the empirical dyadic network model with popularity indexes, against models with so-called strategic interactions (Chandrasekhar and Jackson, 2014; Leung, 2019a). Another recent literature (Auerbach, 2019; Zeleneev, 2020) focuses on a "graphon" metric to analyze more general forms of unobserved heterogeneity that allows for latent communities. Auerbach (2019), however, argues that this form of heterogeneity cannot be separately identified from spillovers in dense networks. Recent research typically assumes sequences of networks with increasing degree for consistent estimating models with this form of heterogeneity. However, the residual variance of the fraction of treated friends is inversely proportional to the degree, which means that identification in dense networks is at best weak without any further regularization. Handling these types of settings remains an open question.

There is a broader reduced form literature on networks and social interactions. Manski (1993) studies a linear model with group level averages of key variables as regressors. Manski's model is similar to mine in the sense that he also includes summary measures of friend treatment. However, he also includes an average of the outcome variable as a regressor and calls its associated coefficient the *endogenous peer effect*. There has been significant interest in estimating this coefficient (Bramoullé et al., 2009; Lee, 2007) although there has been skepticism about whether it is plausible to identify it in practice (Angrist, 2014). The models that I consider do not estimate the endogenous peer effects. There are many reasons to focus on the coefficient on friend treatment instead of the endogenous peer effect. On one hand, the coverage of the program can be adjusted by the policymaker, whereas changing the composition of peers or the network itself is not always feasible, except in tightly controlled environments. On the other hand, the average partial effects that I identify can be viewed as observational analogs of the estimands recovered by random saturation experiments (Baird et al., 2019; Crépon et al., 2013; Duflo and Saez, 2003; Vasquez-Bare, 2019), in which the researcher exogenously shifts the proportion of people treated by design. Overall, understanding the trade-offs in expanding the coverage of a program are essential to the policy maker's cost-benefit calculations.

Section 2 introduces the model and the identification results. Section 3 proposes feasible

estimator and presents the asymptotic results. Section 4 discusses the empirical examples. Section 5 presents a discussion about extensions. Section 6 concludes.

## 2 Model

I assume that there are  $g = \{1, \ldots, G\}$  disjoint groups that contain  $i = \{1, \ldots, N_g\}$  individuals each. We can interpret g as the identifier for a school, village or city. Treatment status is denoted by a binary variable  $D_{ig}$  that equals one if individual  $\{ig\}$  is treated and zero if she is not. A social network is denoted by a  $N_g \times N_g$  adjacency matrix  $A_g$  with binary entries. Each entry  $A_{ijg}$  equals one if individuals  $\{ig\}$  and  $\{jg\}$  are friends and zero otherwise, using the convention that  $A_{iig} = 0$ . To make the model tractable, I follow a recent literature that relies on summary measures of friends' treatment status (Aronow and Samii, 2017; Leung, 2019a; Manski, 2013). To this end, I define two additional measures: the total number of  $\{ig\}'s$  friends  $L_{ig} \equiv \sum_{j=1}^{N_g} A_{ijg}$  and the total number of  $\{ig\}'s$  treated friends by  $T_{ig} \equiv \sum_{j=1}^{N_g} A_{ijg}D_{jg}$ . The variables  $L_{ig}$  and  $T_{ig}$  are meant to capture peer influence in  $\{ig\}'s$  immediate friend circle.

I analyze a model where a scalar outcome  $Y_{ig}$  is determined by

$$Y_{ig} = \alpha_{ig} + D_{ig}\beta_{ig} + \varphi(T_{ig}, L_{ig})'\gamma_{ig} + D_{ig} \times \varphi(T_{ig}, L_{ig})'\delta_{ig}.$$
(1)

Here,  $\varphi : \mathbb{Z}^2_+ \to \mathbb{R}^k$  is a *known* function, which maps  $(T_{ig}, L_{ig})$  to a set of individual covariates. I define a vector of real-valued random coefficients  $\tau_{ig} \equiv (\alpha_{ig}, \beta_{ig}, \gamma'_{ig}, \delta'_{ig})' \in \mathbb{R}^{2+2k}$  that can be correlated with  $(D_{ig}, T_{ig}, L_{ig})$ . I am interested in identifying the average partial effects for a target population, defined as

$$\tau \equiv (\alpha, \beta, \gamma, \delta) = \mathbb{E}[\tau_{ig} \mid \mathcal{F}].$$
<sup>(2)</sup>

The average partial effects vector  $\tau$  integrates the coefficients in (1). The conditioning  $\mathcal{F}$  is important to emphasize that the average is computed for a specific subpopulation (men or women, old or young, etc.). When the conditioning set is empty, i.e  $\mathcal{F} = \emptyset$ , the average is computed for the entire population.

The potential outcomes model (Fisher et al., 1960; Rubin, 1980) that is routinely used in program evaluation is a special case of (1). In that case we set  $\gamma_{ig} = \delta_{ig} = 0$  and define individual-specific outcomes by treatment status as  $Y_{ig}(0) \equiv \alpha_{ig}$  and  $Y_{ig}(1) \equiv \alpha_{ig} + \beta_{ig}$ . The average treatment effect is defined as  $\beta = \mathbb{E}[\beta_{ig} | \mathcal{F}] = \mathbb{E}[Y_{ig}(1) - Y_{ig}(0) | \mathcal{F}]$ . Heterogeneity of  $\beta_{ig}$  is important to capture varying responses to treatment. Researchers are often interested in testing  $\beta = 0$ , the null hypothesis that the treatment has no effect on average. If  $\beta > 0$  then the treatment has a positive effect on the population of interest, and a negative effect if  $\beta < 0$ .

The more interesting case is when  $\gamma_{ig}$  and  $\delta_{ig}$  are not zero. For simplicity assume that  $\varphi(t,l) = t/l$  and l > 0, which implies that the model in (1) is a linear function of own treatment, the fraction of treated friends, and an interaction. We can define the potential outcomes as  $Y_{ig}(0,t,l) \equiv \alpha_{ig} + \gamma_{ig} \times (t/l)$  and  $Y_{ig}(1,t,l) \equiv \alpha_{ig} + \beta_{ig} + (\gamma_{ig} + \delta_{ig}) \times (t/l)$ . The direct average treatment effect is equal to  $\mathbb{E}[Y_{ig}(1,t,l) - Y_{ig}(0,t,l) \mid \mathcal{F}] = \beta + \delta \times (t/l)$ . In contrast to the Fisher-Rubin model, the magnitude of the treatment effect depends on how many friends are treated. For example, if  $\delta > 0$  then having more treated friends widens the gap between the treated and control. In addition to the ATE we can compute the spillover effect for control individuals  $\mathbb{E}[Y_{ig}(0,t,l) - Y_{ig}(0,0,l) \mid \mathcal{F}] = \gamma \times t/l$ . If  $\gamma > 0$  then control individuals have better outcomes when some of their friends are treated even if they are not participating in the treatment directly. Modeling heterogeneity of  $(\gamma_{ig}, \delta_{ig})$  is important to capture the fact that not everyone is equally susceptible to peer influence.

The choice of  $\varphi$  determines the shape of the potential outcomes function in terms of friend treatment status. Many empirical examples choose a linear specification with homogeneous coefficients where  $\varphi(t, l) = t$  or  $\varphi(t/l) = t/l$  following Manski (1993), although more general forms are also possible. It is worth noting that the choice of  $\varphi$  is not essential to the identification argument. Non-separable models are an alternative that can capture heterogeneous, non-linear relationships between an outcome an endogenous variable (Blundell and Powell, 2003; Florens et al., 2008; Imbens and Newey, 2009). In this case the equivalent of  $\tau$  is a function known as the average dose response or average structural function. In Section Appendix show that this function is identified by using analogous arguments to the random coefficients.

The more substantial restriction in (1) is that individuals are only affected by the average treatment status of their immediate friends rather than those of second order connections. This assumption is known in the literature as anonymous interactions (Aronow and Samii, 2017; Baird et al., 2019; DiTraglia et al., 2021; Leung, 2019a; Vasquez-Bare, 2019). This condition is typically violated in so-called endogenous peer effects models that include an additional term  $\rho \sum_{j=1}^{N_g} A_{ig} Y_{jg}$  on the right-hand side of (1). Bramoullé et al. (2009) show that in a setting with homogeneous coefficients, i.e.  $(\beta_{ig}, \gamma_{ig}, \delta_{ig}) = (\beta, \gamma, \delta)$ , as well as exogenous treatment and network, there is a reduced-form representation that depends on the entire treatment vector and the whole adjacency matrix. Some work has been done on accounting for endogenous network formation (Johnsson and Moon, 2019) but the general case with higher order connections, heterogeneous coefficients, endogenous treatment and network formation is still an ongoing area of research (see Bramoullé et al. (2020) for a recent review).

#### 2.1 Identification of $\tau$ : Main insights

Let  $X_{ig}$  be the vector of regressors (1) which is defined as

$$X_{ig} \equiv (1, D_{ig}, \varphi(T_{ig}, L_{ig}), D_{ig} \times \varphi(T_{ig}, L_{ig})).$$

The random variable  $X_{ig}$  has dimension 2 + 2k. This allows us to write down the model in (1) concisely as  $Y_{ig} = X'_{ig}\tau_{ig}$ .

The main barrier to identifying the average partial effect is that  $\tau_{ig}$  and  $X_{ig}$  might be correlated. To address this problem, I propose using individual covariates  $V_{ig}$  that capture the main determinants of treatment and network formation. I assume that  $V_{ig}$  satisfies the unconfoundedness condition  $\tau_{ig} \perp X_{ig} \mid V_{ig}$  and that  $\mathcal{F}$  is  $V_{ig}$  – measurable. For example,  $\mathcal{F}$  could include gender and  $V_{ig}$  could include a finer set of variables such as gender, age and wealth. I establish primitive assumptions on the network and treatment processes that justify these conditions in the next section. Under unconfoundedness,

$$\mathbb{E}[Y_{ig} \mid X_{ig} = x, V_{ig} = v] = x' \mathbb{E}[\tau_{ig} \mid V_{ig} = v] \equiv x' \tau(v)$$
(3)

Here,  $\tau(v)$  is a localized average of  $\tau_{ig}$  in terms of observables. Unconfoundedness allows us to separate the endogenous regressors from the random coefficients. Crucially, x and  $\tau(v)$  form a system of equations that can be used to solve for  $\tau(v)$ .

For example, consider a restricted case where  $\{ig\}$  does not have any friends and hence there are no spillovers. For convenience, we can express the regressors as  $X_{ig} = (1, D_{ig}, 0, 0)$ , setting the variables that involve peer treatment to zero. The system has two equations

$$\mathbb{E}[Y_{ig} \mid X_{ig} = (1, 0, 0, 0), V_{ig} = v] = \mathbb{E}[\alpha_{ig} \mid V_{ig} = v] \qquad \text{if } D_{ig} = 1$$

$$\mathbb{E}[Y_{ig} \mid X_{ig} = (1, 1, 0, 0), V_{ig} = v] = \mathbb{E}[\alpha_{ig} \mid V_{ig} = v] + \mathbb{E}[\beta_{ig} \mid V_{ig} = v] \qquad \text{if } D_{ig} = 0$$
(4)

We can solve for  $\mathbb{E}[\beta_{ig} \mid V_{ig} = v]$  by subtracting the first line from the second line of (4).

Intuitively, for each treated individual we have to find another person with similar characteristics in the control group, that proxies as a counter-factual. This matching process requires an support/overlap condition  $0 < \mathbb{P}(D_{ig} = 1 | V_{ig} = v) < 1$ , that ensures that the researcher can find such a match with high probability. The probability  $\mathbb{P}(D_{ig} = 1 | V_{ig} = v)$  is commonly known as the *propensity score*. If it is equal to either zero or one, then one of the outcomes in Equation (4) is not identified and we cannot solve  $\mathbb{E}[\beta_{ig} | V_{ig} = v]$ . If the overlap condition does hold over the support of  $V_{ig}$  then we can obtain the average treatment effect as  $\beta \equiv \mathbb{E}[\beta_{ig} | \mathcal{F}] = \mathbb{E}[\mathbb{E}[\beta_{ig} | V_{ig}] | \mathcal{F}]$  by applying the law of iterated expectations.

We now turn to the case where  $\{ig\}$  has one or more friends. To solve the system of equations involving  $\tau(v)$  we can pre-multiply (4) by  $X_{ig}$  and apply the law of iterated expectations once more. This means that  $\mathbb{E}[X_{ig}Y_{ig} | V_{ig} = v] = \mathbb{E}[X_{ig}X'_{ig} | V_{ig} = v]\tau(v)$ . We can solve for  $\tau(v)$  as

$$\tau(v) = \underbrace{\mathbb{E}[X_{ig}X'_{ig} \mid V_{ig} = v]}_{\equiv \mathbf{Q}_{xx}(v)}^{-1} \underbrace{\mathbb{E}[X_{ig}Y_{ig} \mid V_{ig} = v]}_{\equiv \mathbf{Q}_{xy}(v)}$$

Here, the weighting matrix  $\mathbf{Q}_{xx}(v)$  needs to be invertible over the support of  $V_{ig}$ . The estimand for  $\tau(v)$  resembles the form of a varying coefficients regression that conditions on  $V_{ig}$ . Invertibility depends on both the choice of basis functions  $\varphi$  and the distribution of  $(D_{ig}, L_{ig}, T_{ig})$  given  $V_{ig}$ . It ensures that the econometrician observes individuals with the same value of  $V_{ig}$  but different treatment status and different values of  $(T_{ig}, L_{ig})$ . If the weighting matrix is invertible uniformly in the support of  $V_{ig}$ , then we can identify  $\tau = \mathbb{E}[\tau(V_{ig}) | \mathcal{F}] = \mathbb{E}[\mathbf{Q}_{xx}(V_{ig})^{-1}\mathbf{Q}_{xy}(V_{ig}) | \mathcal{F}]$ . Graham and Pinto (2018) and Wooldridge (2004) show that in a generic random coefficients model with regressors  $X_{ig}$  we can write an expression for  $\tau$  that does not depend on  $\mathbf{Q}_{xy}(V_{ig})$ , by applying the law of iterated expectations.

**Theorem 1** (Average Partial Effects). Suppose that (i)  $Y_{ig} = X'_{ig}\tau_{ig}$ , (ii)  $X_{ig} \perp \tau_{ig} \mid V_{ig}$ , (iii)  $\mathcal{F}$  is  $V_{ig}$ -measurable and  $\mathbf{Q}_{xx}(v) = \mathbb{E}[X_{ig}X'_{ig} \mid V_{ig} = v]$  is invertible almost surely over the support of  $V_{ig} \mid \mathcal{F}$ . Then  $\tau$  defined in (2) is equal to  $\mathbb{E}[\mathbf{Q}_{xx}(V_{ig})^{-1}X_{ig}Y_{ig} \mid \mathcal{F}]$ .

Theorem 1 shows that the average partial effect can be identified by an inverse weighting strategy that only depends on  $\mathbf{Q}_{xx}(V_{ig})$ . In contrast to a generic random coefficients model, in the spillovers model  $X_{ig}$  is a function of own and friend treatment indicators, which constrains the form of  $\mathbf{Q}_{xx}(v)$ . Let  $\otimes$  denote the Kronecker product. The weighting matrix takes the form

$$\mathbf{Q}_{xx}(v) = \mathbb{E}\left[\begin{pmatrix} 1 & \varphi(T_{ig}, L_{ig})' \\ \varphi(T_{ig}, L_{ig}) & \varphi(T_{ig}, L_{ig})\varphi(T_{ig}, L_{ig})' \end{pmatrix} \otimes \begin{pmatrix} 1 & D_{ig} \\ D_{ig} & D_{ig} \end{pmatrix} | V_{ig} = v \right]$$

The overlap condition is still necessary for invertibility. If  $\mathbb{P}(D_{ig} = 1 | V_{ig} = v)$  is either zero or one, then some of the columns are colinear. However, the overlap condition is not sufficient because of the other entries that involve  $\varphi$ . The remaining entries of  $\mathbf{Q}_{xx}(v)$  can be interpreted as a generalized propensity score in the style of Hirano and Imbens (2004) to match the first and second conditional moments of  $X_{ig}$ . I show that imposing the network model introduces over-identifying restrictions that drastically reduce the number of entries that need to be computed, and provide guidance on the choice of  $V_{ig}$ .

#### 2.2 Endogenous Treatment and Network

I assume that the researcher has auxiliary covariates that explain  $\{ig\}'s$  participation in the treatment and choice of friends. Let  $C_{ig} \in \mathbb{R}^{d_c}$  be a vector of individual characteristics that are sampled at random from a super-population and let  $\Psi_g^* \in \mathbb{R}^{d_{\Psi}}$  be a vector of group characteristics. I next describe assumptions on the core structure that provide guidance on the choice of  $V_{ig}$ .

Assumption (Random Sampling).

- (i) (Across Groups)  $\{\tau_{ig}, D_{ig}, C_{ig}\}_{i=1}^{N_g}$ ,  $\Psi_g^*$  are *i.i.d.* across groups.
- (*ii*) (Within Groups)  $\{\tau_{ig}, D_{ig}, C_{ig}\}$  are *i.i.d.* within group given  $\Psi_g^*$ .

The first part of Random Sampling –stating that groups are i.i.d– is plausible when the groups are spatially, economically or socially separated. The second part states that the covariates within a group are conditionally independent within groups, which is a common assumption in the literature on network formation (Auerbach, 2019; Graham, 2017; Johnsson and Moon, 2019). This can also be interpreted as an exchangeability condition.

Assumption (Selection on Observables).  $\tau_{ig} \perp D_{ig} \mid C_{ig}, \Psi_g^*$ .

The Selection on Observables assumption states that the treatment status is independent of the treatment effects, after controlling for baseline characteristics. It puts the burden on researchers to identify relevant confounding variables (such as gender, income or age) that are motivated by either theory or practice. For example, the confounders can emerge from welldefined institutional rules that constrain the assignment of slots to treatment or the stratifying variables in experiments with perfect compliance. Selection on Observables is the same assumption discussed by Rosenbaum and Rubin (1983), which justifies propensity score analysis.

Assumption (Dyadic Network). Suppose that there exists an unobserved vector of pairspecific shocks  $\{U_{ijg}\}_{i,j=1}^{N_g} \in \mathbb{R}^{N_g}$  for  $g = 1, \ldots, G$  and an unknown link function  $\mathcal{L} : \mathbb{R}^{k_c} \times \mathbb{R}^{k_c} \times \mathbb{R}^{k_{\Psi}} \times \mathbb{R} \to \{0, 1\}$  such that

- (i) (Pairwise Links)  $A_{ijg} = \mathcal{L}(C_{ig}, C_{jg}, \Psi_g^*, U_{ijg}).$
- (*ii*) (Shocks)  $U_{ijg}$  are *i.i.d.* and mutually independent of  $\{\tau_{ig}, D_{ig}, C_{ig}\}_{i=1}^{N_g}$  given  $\Psi_q^*$ .

The Dyadic Network assumption states that friendships between pairs of individuals  $\{ig\}$ and  $\{jg\}$  depend on their observed characteristics  $(C_{ig}, C_{jg})$ , a group component  $\Psi_g^*$  and a pairspecific shock  $U_{ijg}$ . For example, let  $||c - c^*||$  be the Euclidean distance between two sets of covariates  $(c, c^*)$ . In a random geometric graph,  $\mathcal{L}(c, c, \Psi^*, u) = \mathbb{1}\{||c - c^*|| \leq u\}$ , which implies that individuals are more likely to be friends if their characteristics are similar. In economics, dyadic networks have been used to analyze risk sharing agreements, political alliances and business partnerships (Attanasio et al., 2012; Fafchamps and Gubert, 2007; Fafchamps and Quinn, 2018; Graham, 2017; Lai and Reiter, 2000). The function  $\mathcal{L}$  can be interpreted as a decision rule that encodes preferences over friends, as a random meeting process that brings two people together (Mele, 2017), or a combination of both.

Dyadic networks can also be motivated as reduced form objects by appealing to exchangeability. In two influential papers, Aldous (1981) and Hoover (1979) showed that any network whose distribution is invariant to the ordering of the sample (exchangeability) can be represented as a dyadic network, where some of the components of  $C_{ig}$  are possibly unobserved. From a practical point of view, the Dyadic Network assumption states that the relevant determinants are indeed observed by the researcher. Therefore it can be interpreted as a network analog of the Selection on Observables assumption.

#### 2.3 The Network Propensity Score

Define the propensity score and the friend propensity score, respectively as

$$p_{dig} \equiv \mathbb{P}(D_{ig} = 1 \mid C_{ig}, \Psi_g^*),$$
$$p_{fig} \equiv \mathbb{P}(D_{jg} = 1 \mid C_{ig}, \Psi_g^*, A_{ijg} = 1)$$

The scalar  $p_{dig}$  is the probability of treatment given individual characteristics, whereas  $p_{fig}$  is the probability that a potential friend is treated. The Random Sampling assumption ensures that every friend is ex-ante identical and hence the probability does not depend on the subscript  $\{jg\}$ . I call the three dimensional vector  $P_{ig} = (p_{dig}, p_{fig}, L_{ig})$  the network propensity score.

Before presenting the general results I focus on a special case where  $\tau$  has a closed form expression. The following result in Theorem 2 is a special case of Theorem 1, by setting  $V_{ig} = (C_{ig}, \Psi_g^*, L_{ig})$  and imposing a particular set of basis functions.

**Theorem 2** (Closed form  $\tau$ ). If  $\varphi(t, l) = t/l$ ,  $\mathcal{F} = \mathbb{1}\{L_{ig} > 0\}$ ,  $\mathbf{Q}_{xx}(V_{ig})$  is almost surely full rank and Random Sampling, Selection on Observables and Dyadic Network hold, then the average partial effects equal

(i) 
$$\alpha = \mathbb{E}\left[\left(1 - \frac{T_{ig} - L_{ig}p_{fig}}{1 - p_{fig}}\right)\left(\frac{(1 - D_{ig})Y_{ig}}{1 - p_{dig}}\right) \mid \mathcal{F}\right],$$

(*ii*) 
$$\beta = \mathbb{E}\left[\left(1 - \frac{T_{ig} - L_{ig}p_{fig}}{1 - p_{fig}}\right) \left(\frac{D_{ig}Y_{ig}}{p_{dig}} - \frac{(1 - D_{ig})Y_{ig}}{1 - p_{dig}}\right) \mid \mathcal{F}\right],$$

(*iii*) 
$$\gamma = \mathbb{E}\left[\left(\frac{T_{ig}-L_{ig}p_{fig}}{p_{fig}(1-p_{fig})}\right)\left(\frac{(1-D_{ig})Y_{ig}}{1-p_{dig}}\right) \mid \mathcal{F}\right],$$

$$(iv) \quad \delta = \mathbb{E}\left[\left(\frac{T_{ig} - L_{ig}p_{fig}}{p_{fig}(1 - p_{fig})}\right)\left(\frac{D_{ig}Y_{ig}}{p_{dig}} - \frac{(1 - D_{ig})Y_{ig}}{1 - p_{dig}}\right) \mid \mathcal{F}\right].$$

Theorem 2 shows that the average partial effects can be identified from  $(p_{dig}, p_{fig}, T_{ig}, L_{ig}, D_{ig}, Y_{ig})$  for the subsample of individuals with at least one friend. The network propensity score is not observed directly but it can be identified from the data.

The treatment effect  $\beta$ , in particular, looks very similar to its counterpart  $\beta^{ATE}$  in the absence of spillovers. Robins et al. (1994) and many others have shown that

$$\beta^{ATE} = \mathbb{E}\left[\frac{D_{ig}Y_{ig}}{p_{dig}} - \frac{(1-D_{ig})Y_{ig}}{1-p_{dig}} \mid \mathcal{F}\right].$$

By plugging in the outcome from (1), and applying the law of iterated expectations, it is possible to show that  $\beta^{ATE} = \beta + \mathbb{E}[p_{fig} \times \delta_{ig} | \mathcal{F}]$ . In the special case where the friend propensity score is independent of the spillover effect on the treated  $(\delta_{ig})$ , then this expression simplifies to  $\beta + \mathbb{E}[D_{jg} | \mathcal{F}] \times \delta$ . That means that the treatment effect that is recovered from propensity score matching can be interpreted for the average effect when  $\mathbb{E}[D_{jg} | \mathcal{F}]$  friends are treated. This quantity is not directly policy relevant because does not reflect the average outcomes when the program is implemented at a smaller or larger scale.

The example in Lemma 1 also highlights some of the relevant rank conditions for identification that hold for more general settings. As in standard propensity score matching the overlap condition  $0 < p_{dig} < 1$  needs to hold, otherwise the denominator is not well defined. There is a similar overlap condition for potential friends, where  $0 < p_{fig} < 1$ . This means that  $\{ig\}'s$ friend cannot all be part of the treatment or control with probability approaching one. Otherwise, there is no residual variation to identify the spillover effects. Lastly, the distribution of  $(T_{ig}, L_{ig})$  needs to have thin tails (not too many friends), otherwise expectation may not be well defined. This suggests a potential weak identification problem in dense network limits where  $L_{ig} \to \infty$ . This is not a problem for networks with a bounded number of friends.

The first step to prove the general result is to show that  $V_{ig} = (C_{ig}, \Psi_g^*)$  satisfies the key unconfoundedness condition of Theorem 1 and can hence be used as matching variable to compute the average causal effect  $\tau$ .

**Theorem 3** (Direct Confounders). Suppose that  $Y_{ig}$  is generated by (1). If Random Sampling, Selection on Observables and Dyadic Network hold, then  $(X_{ig}, L_{ig}) \perp \tau_{ig} \mid C_{ig}, \Psi_g^*$ .

Intuitively, Random Sampling and Dyadic Network imply that  $(C_{ig}, \Psi_g^*)$  controls for others' treatment whereas Selection on Observables ensures that it controls for own selection. From a practical standpoint, Theorem 3 suggests that the researcher should include all the covariates that she considers relevant for treatment participation and network formation in  $V_{ig}$ . The variables  $(C_{ig}, \Psi_g^*)$  control for  $\{ig\}'s$  friend preferences, and hence all the residual variation in  $X_{ig}$  is exogenous.

The second step is to prove that the network propensity score is a sufficient statistic for the distribution of the endogenous regressors.

Lemma 1 (Conditional Distribution). If Random Sampling and Dyadic Network, then

(i) 
$$D_{ig} \mid T_{ig}, L_{ig}, C_{ig}, \Psi_q^* \sim Bernoulli(p_{dig}),$$

(*ii*) 
$$T_{ig} \mid L_{ig}, C_{ig}, \Psi_q^* \sim Binomial(p_{fig}, L_{ig})$$
 and

Lemma 1 shows that the distribution of  $(D_{ig}, T_{ig})$  given  $(C_{ig}, \Psi_g^*)$  can be parametrized in terms of  $P_{ig}$ . Part (i) is an extension of the canonical result of Rosenbaum and Rubin (1983), whereas par (ii) is a new result. This factorization holds regardless of the primitive function  $(\mathcal{L})$  and shock distribution of network formation. The proof builds on the insight that  $T_{ig}$  is a sum of conditionally independent Bernoulli variables after conditioning on the key variables of network formation. Under model (1),  $X_{ig}$  is a deterministic function of  $(D_{ig}, T_{ig}, L_{ig})$  which means that  $P_{ig}$  also parametrizes the distribution of  $X_{ig} | C_{ig}, \Psi_g, L_{ig}$ . Lemma 1 also shows that  $P_{ig}$  is identified from the conditional means of  $(D_{ig}, T_{ig})$ .

The third and final step, is to show that the network propensity score can be used as a matching variable for causal comparisons.

**Theorem 4** (Balancing). If Random Sampling and Dyadic Network hold, then  $P_{ig}$  is a balancing score, in the sense that  $X_{ig} \perp (C_{ig}, \Psi_g^*) \mid P_{ig}$ . If Selection on Observables also holds, then  $X_{ig} \perp \tau_{ig} \mid P_{ig}$ .

Theorem 4 shows that  $P_{ig}$  is a suitable generalization of the propensity score to setting with spillovers and network formation by showing that inherits two key properties. First, it is a balancing score which means that two individuals with the same value of  $P_{ig}$  are guaranteed to have the same distribution of covariates  $(C_{ig}, \Psi_g^*)$ . This property is important for causal analyses because it ensures that any matching procedure based on  $P_{ig}$  will compare similar individuals. Second, it shows that  $P_{ig}$  satisfies the unconfoundedness property required to identify the average partial effect  $\tau$  in Theorem 1. The selection on observables ties the observed characteristics  $(C_{ig}, \Psi_g^*)$  to the random coefficients and is therefore crucial to prove the final step.

From an economic point of view, the network propensity score can be interpreted as a function of agents' underlying preferences. To this end, it is convenient to represent  $\{ig\}'s$  treatment indicator as  $D_{ig} = \mathcal{H}(C_{ig}, \Psi_g^*, \eta)$  where  $\mathcal{H}$  is a measurable function and  $\eta_{ig} \mid C_{ig}, \Psi_g^* \sim F(\eta \mid c, \Psi^*)$  is an unobserved participation shock. Since we can always define the participation shock as  $\eta = D_{ig} - \mathbb{P}(D_{ig} = 1 \mid C_{ig} = c, \Psi_g^* = \Psi^*)$ , this form does not entail any loss of generality. The function  $\mathcal{H}$  can also take the form of a threshold utility models or institutional assignment rules based on observables. The first component of the network propensity score is the propensity score conditional on  $(C_{ig}, \Psi_g^*)$ , which is defined as

$$\mathbb{P}(D_{ig} = 1 \mid C_{ig} = c, \Psi_g^* = \Psi^*) = \int \mathcal{H}(c, \Psi^*, \eta) \, dF(\eta \mid c, \Psi^*) \tag{5}$$

The propensity score depends on the preference function  $\mathcal{H}$  and the distribution of selection shocks. The integral averages out the individual heterogeneity  $\eta$ , holding the characteristics  $(c, \Psi^*)$  fixed.

The friend propensity score can be written in a similar way. Let  $F(\eta^*, c^*, u \mid \Psi^*)$  be the distribution of traits of a potential friend in each group  $(\eta^*, c^*)$  and the friendship shock (u) given  $\Psi_q^*$ . By Bayes' rule

$$\mathbb{P}(D_{jg} = 1 \mid C_{ig} = c, \Psi_g^* = \Psi^*, A_{ijg} = 1) \\
= \int \mathcal{L}(c, c^*, \Psi^*, u) \,\mathcal{H}(c^*, \Psi^*, \eta^*) \,\frac{dF(\eta^*, c^*, u \mid \Psi^*)}{\int \mathcal{L}(c, c^*, \Psi^*, u) \,dF(c^*, u^* \mid \Psi^*)}.$$
(6)

The friend propensity combines  $\{ig\}'s$  friendship preferences/meeting likelihood and  $\{jg\}'s$ preferences for participation in the program. In the extreme case that  $L = \mathbb{1}\{c = c^*\}$ , agents only befriend others with exactly the same characteristics and the friend propensity score is equal to the propensity score. At the other extreme, when  $L = \mathbb{1}\{u > 0\}$  the network is exogenous then (6) reduces to  $\int \mathcal{H}(c^*, \Psi^*, \eta) dF(\eta^*, c^* | \Psi^*)$ , which is a group-level constant. Conversely, when the treatment is exogenous, that is when  $\mathcal{H}(c^*, \Psi^*, \eta^*) = \eta$  and  $\eta$  is independent of the other characteristics, then the propensity score and the friend propensity score are constant. For intermediate cases the friend propensity score will not contain in the same information as the propensity score.

## 2.4 Mixture Representation of $Q_{xx}$

To compute the network propensity score,  $(C_{ig}, \Psi_g^*)$  is either fully observed or can be consistently estimated. Unobserved heterogeneity can be addressed in a variety of ways. For instance, by estimating group-specific network propensity score to capture variation in  $\Psi_g^*$ , by exploiting restrictions on the network structure (Johnsson and Moon, 2019) or constraints on compliance behavior in randomized experiments (DiTraglia et al., 2021). However, it is possible that all the relevant heterogeneity cannot be captured by the data available to the researcher. In this section I analyze the form of the weighting matrix when  $V_{ig}$  does not contain all the relevant determinants of selection and network formation.

To state the formal result we need some preliminary notation. Define the functions  $\tilde{\varphi}_1(p_f, l) = \mathbb{E}[\varphi(T_{ig}, L_{ig}) \mid p_{fig} = p_f, L_{ig} = l]$  and  $\tilde{\varphi}_2(p_f, l) = \mathbb{E}[\varphi(T_{ig}, L_{ig})\varphi(T_{ig}, L_{ig})' \mid p_{fig} = p_f, L_{ig} = l]$  which are the conditional first and second moments given the friend propensity score and the

total number of friends. Since Lemma 1 shows that  $(p_{fig}, L_{ig})$  parametrizes the distribution of  $(T_{ig}, L_{ig})$  given  $(C_{ig}, \Psi_g^*)$ , these are equivalent to conditioning on  $(C_{ig}, \Psi_g^*)$  directly by the decomposition axiom (Constantinou et al., 2017). Lemma 1 also implies that  $\tilde{\varphi}_1$  and  $\tilde{\varphi}_2$  are known functions that only change depending on the basis  $\varphi$ . In our running example, where  $\varphi(t, l) = t/l$ these function take a very simple form. In this case  $\tilde{\varphi}_1(p_f, l)$  and  $\tilde{\varphi}_2(p_f, l) = \frac{p_f(1-p_f)}{l} + p_f^2$ .

Lemma 2 shows that the matrix  $\mathbf{Q}_{xx}$  can be expressed as a mixture of known functions of the network propensity score.

**Lemma 2** (Mixture Representation). Suppose that Random Sampling and Dyadic Network hold, and that  $V_{ig}$  is measurable with respect to  $(C_{ig}, \Psi_q^*, L_{ig})$ , then

$$\mathbf{Q}_{xx}(v) = \int \begin{pmatrix} 1 & \widetilde{\varphi}_1(p_f, l)' \\ \widetilde{\varphi}_1(p_f, l, p_f, l) & \widetilde{\varphi}_2(p_f, l) \end{pmatrix} \otimes \begin{pmatrix} 1 & p_d \\ p_d & p_d \end{pmatrix} dF(p_d, p_f, l \mid V_{ig} = v).$$
(7)

In the special case where  $V_{ig} = (C_{ig}, \Psi_g^*)$  the distribution F is degenerate and we can drop the integral sign. Therefore, observing the key variables for selection and network formation imposes over-identifying restrictions on the weighting matrix. The integral is non-degenerate when some of these key variables are unobserved by the researcher. This assumption is testable by comparing the entries of  $\mathbf{Q}_{xx}$ . For example, in a parametric model F can be modeled as a latent distribution that nests the degenerate case and  $(p_{dig}, p_{fig})$  as link function such as probit or logit.

If F is non-degenerate, then  $V_{ig}$  is not guaranteed to satisfy the conditions of Theorem 1 from the Random Sampling, Selection on Observables and Dyadic Network assumptions alone. We need an additional exclusion restriction.

**Lemma 3.** If Random Sampling, Selection on Observables, Dyadic Network hold, and  $\tau_{ig} \perp P_{ig} \mid V_{ig}$ , where  $P_{ig} = (p_{dig}, p_{fig}, L_{ig})$ , then  $X_{ig} \perp \tau_{ig} \mid V_{ig}$ .

Lemma 3 provides a high-level condition that says that the variation in the network propensity is exogenous after conditioning on  $V_{ig}$ , that ensures the validity of  $V_{ig}$ . Since the network propensity score is itself a function of  $(C_{ig}, \Psi_g^*)$  this means that are exogenous shifter in individual behavior  $(C_{ig})$  or variation across groups  $(\Psi_g^*)$ . In the discussion I present some examples of experiments where this condition holds by construction.

## 2.5 Covariate Balancing ("Placebo") Test

The balancing property in Theorem 4 is testable. Parametric propensity score analyses typically conduct so-called covariate balancing tests. I propose an analogous "placebo" test, where the pretreatment covariates serve as an outcome variable. Let  $\widetilde{V}_{ig} \in \mathbb{R}$  be a variable in the covariate set  $V_{ig} = (C_{ig}, \Psi_g)$ . My test relies on the simple idea that  $\widetilde{V}_{ig}$  can be decomposed as

$$\widetilde{V}_{ig} = \underbrace{\widetilde{V}_{ig}}_{\widetilde{\alpha}_{ig}} + 0 \times D_{ig} + 0 \times \varphi(T_{ig}, L_{ig}) + 0 \times D_{ig} \times \varphi(T_{ig}, L_{ig})$$

Let  $\tilde{\tau}_{ig} = (\tilde{V}_{ig}, 0, 0, 0)$  is the vector of coefficients of the placebo outcome. It is easy to verify that  $X_{ig} \perp \tilde{\tau}_{ig} \mid V_{ig}$  since  $\tilde{\tau}_{ig}$  is a measuarable function of  $V_{ig}$ . Therefore by Theorem 1,  $\mathbb{E}[\mathbf{Q}_{xx}(V_{ig})^{-1}X_{ig}\tilde{V}_{ig}] = (\mathbb{E}[\tilde{V}_{ig}], 0, 0, 0)$ . Therefore, when  $\mathbf{Q}_{xx}(v)$  is properly specified the researcher can test the null hypothesis that the slope coefficients are zero. This test only uses information about the treatment, the network and the covariates but not the outcome.

In practice the test could be rejected in a parametric settings if the functional form is not flexible enough. However, it could also be rejected because a violation of the over-identifying restrictions imposed by the Random Sampling and Dyadic Network assumptions. The researcher may want to check whether there are omitted variables that might influence network formation or treatment.

## 3 Estimation

I outline a two-step procedure to estimate the causal effects for linear models as a sample analog of the estimand of  $\tau$ . In the first stage, I fit a parametric model for  $\mathbf{Q}_{xx}$  using data from the endogenous regressors  $X_{ig}$  and the control variable  $V_{ig}$ . In the second stage, I substitute the estimated weighting matrix  $\mathbf{Q}_{xx}$  to compute  $\tau$  by inverse weighting.

**Notation:** Let  $Z_{ig}$  denote a vector of individual variables, where  $Z_{ig} \equiv (X_{ig}, Y_{ig}, V_{ig})$  includes the endogenous regresors, the outcome and the observed control variables. I let  $\sum_{ig} f(Z_{ig})$  be the sum  $\sum_{g=1}^{G} \sum_{i=1}^{N_g} f(Z_{ig})$ , where  $f(\cdot)$  is an arbitrary function. I also let  $\bar{n} = \frac{1}{G} \sum_{g=1}^{G} N_g$  denote the average group size. By construction,  $\bar{n}G$  is equal to the total sample size. For convenience, let  $vec(\cdot)$  denote the vectorize operator, which stacks the columns of a matrix into a single vector. I also use ||x|| to denote the Euclidean norm of the vector x, defined as  $||x|| = \sqrt{\sum_{k=1}^{K} x_k^2}$ .

In the first stage, I consider a parametric class of functions to model the weighting matrix,  $\{\mathbf{Q}_{xx}(v, \boldsymbol{\theta}) : \boldsymbol{\theta} \in \boldsymbol{\Theta} \subseteq \mathbb{R}^{d_{\boldsymbol{\theta}}}\},$  that nest the true model. This means that there is a  $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}$  such that  $\mathbf{Q}_{xx}(v, \boldsymbol{\theta}_0) = \mathbb{E}[X_{ig}X'_{ig} \mid V_{ig} = v]$ . The matrix  $\mathbf{Q}_{xx}$  has to be symmetric and positive semi-definite. If Random Sampling, Selection on Observables and Dyadic Network hold, and  $V_{ig} = (C_{ig}, \Psi_g^*)$  the choice of parametric family can be disciplined by imposing over-identifying restrictions of the network formation model, so that  $\mathbf{Q}_{xx}(v, \boldsymbol{\theta})$  can be expressed as a function of the network propensity score. Alternatively we can use the mixture model representation of Lemma 2 to inform the choice of  $\mathbf{Q}_{xx}$  for other choices of  $V_{ig}$ . The control variable  $V_{ig}$  is valid as long as the conditions of Lemma 3 hold.

I define the vectorized residuals,

$$r(Z_{ig}, \boldsymbol{\theta}) \equiv \operatorname{vec}(X_{ig}X'_{ig} - \mathbf{Q}_{xx}(V_{ig}, \boldsymbol{\theta})).$$

The residuals capture how well the control variables fit  $X_{ig}$ . The sample criterion function computes the average of square residuals as

$$\widehat{\mathcal{R}}(\boldsymbol{\theta}) \equiv \frac{1}{\bar{n}G} \sum_{ig} \|r(Z_{ig}, \boldsymbol{\theta})\|^2.$$
(8)

The sample criterion  $\hat{\mathcal{R}}(\boldsymbol{\theta})$  is an approximation to  $\mathcal{R}(\boldsymbol{\theta}) = \mathbb{E}[\|r(Z_{ig}, \boldsymbol{\theta})\|^2]$ . The least squares criterion is appropriate for three reasons. First, the population criterion  $\mathcal{R}(\boldsymbol{\theta})$  is minimized at  $\boldsymbol{\theta}_0$  because the conditional mean of  $X_{ig}X'_{ig}$  given  $V_{ig}$  is the optimal prediction. This provides a rationale for minimizing  $\hat{\mathcal{R}}(\boldsymbol{\theta})$ . Second, joint-likelihood approaches are either impractical or infeasible without strong assumptions. The variables  $T_{ig}$  and  $L_{ig}$  that enter  $X_{ig}$  are constructed based on the treatment status of friends, which introduces a mechanical dependence. For example, when  $\{ig\}$  and  $\{jg\}$  have all their friends in common,  $T_{ig}$  and  $T_{jg}$  are functions of the same information. It is therefore difficult to write down a likelihood without specifying the full network formation model. Third quasi-likelihood approaches, such as those in Tchetgen et al. (2017) and Sofrygin and van der Laan (2017) are valid under certain assumptions, but are more sensitive to the specification of the model. My approach is more robust than quasi-likelihood methods because it targets the conditional mean directly, which is the main object required for identification. We can construct a feasible estimator by minimizing the sample criterion,

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \ \hat{\mathcal{R}}(\boldsymbol{\theta}). \tag{9}$$

The estimated parameter  $\hat{\boldsymbol{\theta}}$  can be plugged-in to compute a feasible weighting matrix  $\mathbf{Q}_{xx}(V_{ig}, \hat{\boldsymbol{\theta}})$ . I propose the following sample analog of the inverse-weighting estimand of  $\tau$ .

$$\hat{\boldsymbol{\tau}} \equiv \frac{1}{\bar{n}G} \sum_{ig} \mathbf{Q}_{xx} (V_{ig}, \hat{\boldsymbol{\theta}})^{-1} X_{ig} Y_{ig}$$

The vector  $\hat{\tau}$  is a feasible estimator of the average partial effects defined in (1). The estimator is subject to two sources of uncertainty. First, the sample average is an approximation to  $\mathbb{E}[\mathbf{Q}_{xx}(V_{ig})^{-1}X_{ig}Y_{ig}]$ . Second, the inverse weighting method is subject to first-stage uncertainty in the estimation of  $\hat{\boldsymbol{\theta}}$ . Under standard regularity conditions that I list in the Appendix,  $\hat{\boldsymbol{\theta}}$ and  $\hat{\boldsymbol{\tau}}$  are consistent but the standard errors need to be adjusted. This is analogous to the first stage uncertainty in propensity score methods, that can be corrected analytically or by bootstrap procedures (Abadie and Imbens, 2016).

To adjust the standard errors it is useful to view the first and second stages as a single system of equations. As before, let  $z \equiv (x, y, v)$ . I write down the first-order conditions in terms of the jacobian of the square residuals  $\psi_q(z, \theta) = \frac{\partial}{\partial \theta'} ||r(v, \theta)||^2$  and the second stage influence function  $\psi_{IW}(z, \theta) = \mathbf{Q}_{xx}(v, \theta)$ . I stack the first and second stage equations in a single influence function  $\psi \equiv [\psi_q, \psi'_{IW}]'$ . The estimated parameters solve

$$\frac{1}{\bar{n}G}\sum_{ig}\psi(Z_{ig},\hat{\boldsymbol{\tau}},\hat{\boldsymbol{\theta}}) = \mathbf{0}$$
(10)

To this end, I define the within-group average  $\overline{\psi}_g(\mathbf{Z}_g, \boldsymbol{\theta}) \equiv \frac{1}{N_{gt}} \sum_{i=1}^{N_g} \psi(Z_{ig}, \boldsymbol{\theta})$ , where  $\mathbf{Z}_g \equiv \{Z_{ig}\}_{i=1}^{N_{gt}}$  is a matrix of individual covariates for each group. This allows me to decompose (10) into group averages as  $\frac{1}{\bar{n}G} \sum_{ig} \psi(Z_{ig}, \hat{\boldsymbol{\tau}}, \hat{\boldsymbol{\theta}}) = \frac{1}{G} \sum_{g=1}^{G} \left(\frac{N_g}{\bar{n}}\right) \overline{\psi}_g(\mathbf{Z}_g, \boldsymbol{\theta})$ . The fraction  $(N_g/\bar{n})$  denotes the relative size of each group.

For inference, I compute heteroskedasticity-robust standard errors, clustered a the group level. Let  $\hat{\Omega}$  be an estimate of the second moments of the influence function (10) and let  $\hat{H}$  be a sample analog of the expected jacobian, defined as

$$\hat{H} \equiv \frac{1}{G} \sum_{g=1}^{G} \left( \frac{N_g}{\bar{n}} \right) \frac{\partial}{\partial (\boldsymbol{\theta}, \boldsymbol{\beta})'} \overline{\psi}_g(\boldsymbol{Z}_g, \hat{\boldsymbol{\tau}}, \hat{\boldsymbol{\theta}})$$
(11)

$$\widehat{\Omega} \equiv \frac{1}{G} \sum_{g=1}^{G} \left( \frac{N_g}{\bar{n}} \right)^2 \overline{\psi}_g(\boldsymbol{Z}_g, \widehat{\boldsymbol{\tau}}, \widehat{\boldsymbol{\theta}}) \overline{\psi}_g(\boldsymbol{Z}_g, \widehat{\boldsymbol{\tau}}, \widehat{\boldsymbol{\theta}})'$$
(12)

Then the covariance of the estimators is computed by the sandwich estimator  $\hat{\Sigma} \equiv \frac{1}{G}\hat{H}^{-1}\hat{\Omega}\hat{H'}^{-1}$ and the standard errors can be recovered from the square root of the diagonal of  $\hat{\Sigma}$ . Since the estimator  $\hat{\tau} \in \mathbb{R}^{d_{\tau}}$  only enters the second stage linearly,

$$\hat{H} = \frac{1}{G} \sum_{g=1}^{G} \begin{pmatrix} \underline{N}_{g} \\ \overline{n} \end{pmatrix} \begin{pmatrix} \frac{\partial}{\partial \boldsymbol{\theta}'} \overline{\psi}_{q,g}(\boldsymbol{Z}_{g}, \hat{\boldsymbol{\tau}}, \hat{\boldsymbol{\theta}}) & \boldsymbol{0}'_{d_{\tau}} \\ \frac{\partial}{\partial \boldsymbol{\theta}'} \overline{\psi}_{IW,g}(\boldsymbol{Z}_{g}, \hat{\boldsymbol{\tau}}, \hat{\boldsymbol{\theta}}) & I_{d_{\tau}} \end{pmatrix}$$

Here,  $\overline{\psi}_{q,g}$  and  $\overline{\psi}_{IW,g}$  decompose the within-group average influence functions into the first and second stages, respectively. Both  $\hat{H}$  and its inverse are lower triangular, which means that the limiting covariance matrix of  $\tau$  depends on the upper-left block of  $\hat{\Omega}$  (which captures the first-stage uncertainty).

#### 3.1 Large Sample Theory

For the remainder of this section I propose inference procedures for a setting with many groups  $G \to \infty$  and allow for the possibility that  $N_g$  is either fixed or growing with G. This is intended to approximate the situation faced by empirical researchers who randomly collect data from distinct geographic units, with few individuals (classrooms) or many individuals (villages, cities), which matches the data that I use in the empirical example. Formally, I assume that there is a sequence of probability distributions that is indexed by t, with  $G_t$  groups of unequal size  $N_{gt}$ , and let  $N_t \equiv \mathbb{E}[N_{gt}]$  denote the expected group size. There is a triangular array of covariates for individual  $\{ig\}$  for the point t in the sequence, which I denote by  $Z_{igt} = (X_{igt}, Y_{igt}, V_{igt})$ . The variables  $(L_{igt}, T_{igt})$  are the number of treated friends and number of friends, respectively. Similarly, for each t, I compute estimators  $(\hat{\theta}_t, \hat{\tau}_t)$ . The estimator  $\hat{\tau}_t$ , in particular is compared to the population quantity  $\boldsymbol{\tau}_{0t} = \mathbb{E}[\tau_{igt} \mid \mathcal{F}_t]$ . Centering the estimator around the mean of the triangular array is important to derive the right rate of convergence. For simplicity, I define  $\rho_{gt}$  as the relative group size. Let  $0 < \rho < \bar{\rho} < 1$  be an

arbitrary constant that I use throughout the derivation.

## Assumption (Bounded Group Ratios). $\rho_{gt} \equiv (N_{gt}/N_t) \in [\rho, \overline{\rho}] \subset (0, 1)$ almost surely.

Bounded Group Ratios implies that all groups are approximately the same size, within a range. It implies that the ratio of the largest to the smallest group is bounded by  $\overline{\rho}/\underline{\rho}$ . This assumption is automatically satisfied when  $N_{gt}$  is bounded. However, if  $N_t \to \infty$  as  $t \to \infty$ , then the assumption implies that the smallest group size is growing, because  $\inf_g N_{gt} \ge \underline{\rho}N_t \to \infty$  as  $N_t \to \infty$ . Bester and Hansen (2016) propose a weaker assumption for large unbalanced panels, where the bounds hold in the limit experiment rather than for each point along the sequence, which leads to qualitatively similar conclusions.

My asymptotic results allow for some or all of the regressors in  $V_{igt}$  to be estimated. For example, Johnsson and Moon (2019) show the estimator  $L_{ig}/(N_g - 1)$  converges uniformly to a measure of unobserved degree heterogeneity in dense networks, at rate  $\sqrt{(\log N_t)/N_t}$  in supnorm. In related work in DiTraglia et al. (2021) we find that in randomized experiments with non-compliance, the key dimensions of heterogeneity in spillover models is unobserved but can be consistently estimated in large groups, with a  $\sqrt{(\log G_t)/N_t}$  uniform rate of convergence. Finally, researchers may also want to estimate group-level averages of the covariates that are consistent in large groups. I define  $V_{igt}^0$  as the true, but unobserved value of the regressors. My asymptotic results simply require that  $\max_{g=1,...,G_t} \max_{i=1,...,N_{gt}} ||V_{igt} - V_{igt}^0|| = O_p(\lambda_t)$  and that  $\sqrt{G_t}\lambda_t = o(1)$ . In the two examples above, this means that the expected size of each group needs to be large relative to the number of groups. This is plausible in situations where data is collected on large villages or other geographical units. If the key confounders are observed without error then  $V_{igt} = V_{igt}^0$ . Otherwise the condition holds trivially and  $N_t$  does not need to grow with G at any particular rate.

I list additional Regularity Conditions in the Appendix, where I impose conditions on the moments of  $(X_{igt}, Y_{igt})$  and smoothness conditions on the function  $\mathbf{Q}_{xx}(\cdot, \boldsymbol{\theta})$ . In particular, I provide conditions that ensure that the weighting matrix is almost surely invertible, by imposing a lower bound on the eigenvalues of the matrix. When  $V_{igt} = (C_{igt}, \Psi_{gt}^*)$  and Random Sampling, Selection on Observables and Dyadic Network hold, this is equivalent to saying that  $L_{igt}$  is bounded, and that the remaining components of the network propensity score are bounded in a compact subset of the unit interval, i.e.  $p_d(C_{igt}, \Psi_{gt}^*), p_f(C_{igt}, \Psi_{gt}^*) \in [\underline{\rho}, \overline{\rho}] \subset (0, 1)$ . That avoids boundary cases, where there is not enough residual variation in the regressors after conditioning on the controls. Finally, I define two more objects,  $H_{0t} \equiv \mathbb{E}[(N_{gt}/N_t)\overline{\psi}_g(\mathbf{Z}_{gt}, \hat{\boldsymbol{\tau}}, \hat{\boldsymbol{\theta}})]$  and  $\Omega_{0t} \equiv$ 

 $\mathbb{E}[(N_{gt}/N_t)^2 \overline{\psi}_g(\mathbf{Z}_{gt}, \hat{\boldsymbol{\tau}}, \hat{\boldsymbol{\theta}}) \overline{\psi}_g(\mathbf{Z}_{gt}, \hat{\boldsymbol{\tau}}, \hat{\boldsymbol{\theta}})'] \text{ that are used to compute the covariance matrix } \Sigma_t \equiv H_{0t}^{-1} \Omega_{0t} H_{0t}^{-1}.$ 

**Theorem 5** (Limiting Distribution Estimators). Suppose that  $V_{igt}$  satisfied the conditions of Theorem 1 and define the covariance matrix  $\Sigma_t \equiv H_{0t}^{-1}\Omega_{0t}H_{0t}^{-1}$ . If Bounded Group Ratios and Regularity Conditions hold, then as  $t \to \infty$ , (i)  $\hat{\theta}_t \to^p \theta_{0t}$ ,  $\hat{\tau}_t \to^p \tau_{0t}$  and (ii)

$$\sqrt{G_t} \Sigma_t^{-1/2} \begin{pmatrix} \widehat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}_{0t} \\ \widehat{\boldsymbol{\tau}}_t - \boldsymbol{\tau}_{0t} \end{pmatrix} \to^d \mathcal{N}(\mathbf{0}, I)$$

Theorem 5 shows that the estimators are consistent and converge to a normal distribution. The estimator is centered around the value of  $(\boldsymbol{\theta}_{0t}, \boldsymbol{\beta}_{0t})$  that solves the population criterion, at each point of the sequence. This allows for estimators that are consistent, even if the networks itself does not converge to any particular structure. Theorem 5 can be viewed as an approximation to the finite sample behavior. Researchers can construct test statistics by substituting  $\Sigma_t$  with a sample analog  $\hat{\Sigma}_t$  to confidence intervals.

My results are agnostic about the dependence structure across groups, but it may be possible to improve the  $\sqrt{G_t}$  to  $\sqrt{G_t N_t}$  under stronger conditions. For example, Kojevnikov et al. (2020) develop a central limit theorem for network dependence and provide specific regularity conditions for a single Dyadic Network. This requires the network to be sparse  $L_{igt}$  small relative to  $N_{gt}$  so that individuals far apart in the network are approximately independent. In practice, this does not change the estimation procedure but rather the way in which we construct confidence intervals. Kojevnikov et al. (2020) propose a Network-HAC estimator and Kojevnikov (2019) proposes a bootstrap procedure. Leung (2019b) proposes similar limiting theory for spillover effects when the treatment is exogenously assigned, and Chandrasekhar and Jackson (2014) propose alternative limit theorems under network dependence.

## 4 Empirical Examples

#### 4.1 Political Participation in Uganda

I evaluate the role of an intervention on political participation in Uganda (Eubank et al., 2019; Ferrali et al., 2020). U-Bridge is a novel political communications technology that allows citizens to contact district officials via text-messages. In a pilot program, individuals in 16 villages were invited to participate in quarterly meetings, at a central location, where they received information about national service delivery standards and ways to communicate with local officials. The Governance, Accountability, Participation, and Performance (GAPP) program collected survey data on 82% of adults in the 16 villages as well as social network data. Ferrali et al. (2020) evaluated the adoption patterns of U-Bridge a couple years later. Eubank et al. (2019) study the role of social network structure on voting patterns. For my analysis, I evaluate the impact of attendance to UBridge meetings on political participation using the network propensity score matching methodology. Spillovers are likely to occur in this context because non-participants can receive information about ways to engage in politics from their friends, which can increase their own political activity.

The data collected by the researchers contains four types of social networks: Family ties, friendships, lenders and problem solvers. In my analysis,  $\{ig\}$  is an identifier for an adult in the pilot villages. The indicator  $A_{ijg}$  equals one if  $\{ig\}$  and  $\{jg\}$  have a connection along any of the four dimensions and zero otherwise. Under this definition, individuals have 10 connections on average. The indicator  $D_{ig}$  equals one if  $\{ig\}$  attended the Ubridge meetings, which is around 8.6% of the sample. The outcome is a continuous variable  $Y_{ig}$  that denotes a political participation index constructed by Ferrali et al. (2020). Table 3 presents summary statistics comparing the treatment and control group. The average adult in the sample is around 40 years old. Men are more likely to attend the session than women. Individuals that a leader position and/or completed their secondary education are more likely to attend as well.

I estimate the following linear model with random coefficients.

$$Y_{ig} = \alpha_{ig} + \beta_{ig} D_{ig} + \gamma_{ig} \left(\frac{T_{ig}}{L_{ig}}\right) + \delta_{ig} \left(\frac{T_{ig}}{L_{ig}}\right)$$
(13)

Heterogeneity of  $\beta_{ig}$  means that agents engage in varying levels of political activity after attending the meeting. In this case, we expect  $\beta_{ig}$  to be close to zero because individuals that are already politically engaged are the ones opting to go to the meetings. Conversely,  $\gamma_{ig}$  is the effect of peers on non-participant adults. If  $\gamma_{ig} > 0$ , then individuals with a larger fraction of treated friends are more politically active. The coefficient  $\gamma_{ig} + \delta_{ig}$  captures the spillovers for participants. In this case we expect  $\delta_{ig} < 0$  because the marginal effect of attending friends is lower because they are already receiving the information first hand.

There is a potential identification in this example because individuals select connections with similar preferences We expect  $(\gamma_{ig}, \delta_{ig})$  to be correlated with  $(T_{ig}/L_{ig})$ . To address this problem I leverage additional covariates collected by the researcher to tease out the causal effects. The network propensity score matching methodology is the appropriate tool to identify the average partial effect  $\tau$  because it allows to incorporate additional covariates while allowing for heterogeneous causal effects  $\tau_{ig} = (\alpha_{ig}, \beta_{ig}, \gamma_{ig}, \delta_{ig})$ .

#### 4.2 Feasible Network Propensity Score and Causal Effects

The propensity score in this case describes the probability of attending an Ubridge meeting given covariates  $C_{ig} = (C_{ig1}, \ldots, C_{igK})$ . These include an indicator for holding a leadership position in the village, gender, an indicator for secondary education, a self-reported relative income measure, distance to the meeting place, number of friends and age. Ferrali et al. (2020) also incorporated a public goods where participants were asked to donate part of their remuneration to the village that were match researchers. The donation amount is meant to capture pro-sociability attitudes.

I assume that the group-level variation  $\Psi_g^*$  has an observed and an unobserved component. For the observed component, I include a vector of group-level averages of the key variables in  $C_{ig}$ , which I denote by  $\Psi_g$ . I assume that  $\Psi_g^*$  has a bivariate structure with mean  $(\Psi_g' \theta_{d\Psi}, \Psi_g' \theta_{d\Psi})'$ , where  $(\theta_{d\Psi}, \theta_{f\Psi})$  is a vector of parameters to be estimated. The error term of  $\Psi_g^*$  follows a normally distributed random-effects structure with covariance matrix  $\Sigma \equiv (\sigma_{11}^2, \sigma_{12}, \sigma_{12}, \sigma_{22}^2)$ , that is assumed to be independent of the observed covariates and the random coefficients  $\tau_{ig}$ . The coefficient  $\sigma_{12}$  captures the correlation between the two unobserved components of  $\Psi_g^*$ . Formally,

$$\Psi_g^* = \begin{pmatrix} \Psi_{gd}^* \\ \Psi_{gf}^* \end{pmatrix} \sim \mathcal{N}(\mu_g, \Sigma), \qquad \mu_g = \begin{pmatrix} \Psi_g' \theta_{d\Psi} \\ \Psi_g' \theta_{f\Psi} \end{pmatrix}, \qquad \Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}.$$

I assume that the own propensity score takes the form of a logit link function with an associated vector of parameters  $\boldsymbol{\theta}_d = (\theta_{d0}, \theta_{d1}, \dots, \theta_{dK})$  as follows

$$p_d(C_{ig}, \Psi_g^*; \boldsymbol{\theta}_d) = \frac{\exp(\theta_{d0} + \sum_{k=1}^K C_{igk}\theta_{dk} + \Psi_{gd}^*)}{1 + \exp(\theta_{d0} + \sum_{k=1}^K C_{igk}\theta_{dk} + \Psi_{gd}^*)}$$

I similarly construct the friend propensity score using a logit link function. I use the same observables variables as the friend friend propensity score with different coefficients  $\theta_f$  =

|                             | Network Propensity |            | OLS with covariates |            |
|-----------------------------|--------------------|------------|---------------------|------------|
|                             | Co efficients      | Std. Error | Co efficients       | Std. Error |
| Direct Effect $(\beta)$     | 0.270              | (0.165)    | -0.010              | (0.060)    |
| Spillover Effect $(\gamma)$ | $0.348^{***}$      | (0.116)    | $0.156^{**}$        | (0.068)    |
| Interaction $(\delta)$      | -0.199             | (0.862)    | $0.563^{***}$       | (0.165)    |
| N                           | 2831               |            | 2831                |            |
| Villages                    | 16                 |            | 16                  |            |

Table 1: (Average Partial Effects Political Participation in Uganda) \* Significant at 10%. \*\* Significant at 5%. \*\*\* Significant at 1%. The second and third columns show the coefficients and standard errors of the inverse-weighted estimator, respectively. The fourth and fifth columns are the coefficients of am additive ordinary least squares (OLS) regression that regresses  $Y_{ig}$  on a constant,  $D_{ig}$ ,  $(T_{ig}/L_{ig})$ ,  $D_{ig} \times (T_{ig}/L_{ig})$  and the observed controls used in the inverse-weighting procedure.

 $(\theta_{f0}, \theta_{f1}, \ldots, \theta_{df})$  as follows

$$p_f(C_{ig}, \Psi_g^*; \boldsymbol{\theta}_f) = \frac{\exp(\theta_{f0} + \sum_{k=1}^K C_{igk}\theta_{fk} + \Psi_{gf}^*)}{1 + \exp(\theta_{f0} + \sum_{k=1}^K C_{igk}\theta_{fk} + \Psi_{gf}^*)}$$

The full vector of parameters to be estimated is

$$\boldsymbol{ heta} \equiv (\boldsymbol{ heta}_{d\Psi}, \boldsymbol{ heta}_{f\Psi}, \sigma_1^2, \sigma_{12}, \sigma_2^2, \boldsymbol{ heta}_d, \boldsymbol{ heta}_f)$$

Let  $F(\Psi_g^*; \boldsymbol{\theta})$  is the distribution of unobserved heterogeneity, which corresponds to that of a normal distribution with parameters  $(\mu_g, \Sigma)$ . I construct a weighting matrix that satisfies the mixture model representation of Lemma 2, where  $V_{ig} = (C_{ig}, \Psi_g, L_{ig})$ . To simplify notation I define the auxiliary matrix

$$\Lambda(C_{ig}, \Psi_g^*, L_{ig}; \boldsymbol{\theta}) = \begin{pmatrix} 1 & p_f(C_{ig}, \Psi_g^*; \boldsymbol{\theta}_f) \\ p_f(C_{ig}, \Psi_g^*; \boldsymbol{\theta}) & \frac{p_f(C_{ig}, \Psi_g^*; \boldsymbol{\theta}_f)(1 - p_f(C_{ig}, \Psi_g^*; \boldsymbol{\theta}_f))}{L_{ig}} + p_f(C_{ig}, \Psi_g^*; \boldsymbol{\theta}_f)^2 \end{pmatrix}$$

The feasible weighting matrix is equal to

$$\mathbf{Q}_{xx}(V_{ig};\boldsymbol{\theta}) = \int \Lambda(C_{ig}, \Psi_g^*, L_{ig}; \boldsymbol{\theta}) \otimes \begin{pmatrix} 1 & p_d(C_{ig}, \Psi_g^*; \boldsymbol{\theta}) \\ p_d(C_{ig}, \Psi_g^*; \boldsymbol{\theta}) & p_d(C_{ig}, \Psi_g^*; \boldsymbol{\theta}) \end{pmatrix} dF(\Psi_g^*; \boldsymbol{\theta}).$$
(14)

where I evaluate the integral numerically via quadrature methods and estimate the parameter  $\boldsymbol{\theta}$  by minimizing the sample criterion function in (8).

Table 4 reports the estimated parameters. Columns (2) shows the coefficients of the propensity score. None of the variables in  $C_{ig}$  appears to be statistically significant. Column (3) reports



(a) Histogram of the number of friends

(b) Histogram of same-caste friends

Figure 1: The figure shows the estimated  $p_{dig}$  and  $p_{fig}$  for the network graph of one Ugandan village. Individuals are represented as nodes, and the links between them represent the relationships reported in the baseline survey. Treated individuals are represented with larger nodes. In figure (a) a darker shade of blue indicates a higher estimated probability of treatment, whereas a darker shade of yellow indicates a low probability. Analogously, in figure (b) a darker shade of blue indicates a larger probability of friend treatment.

the coefficients of the friend propensity score, which are far more interesting. The evidence suggests that individuals that hold a leadership position and have completed a higher education or more likely to have a treated friend. This is likely due because leaders tend to come in contact with a greater variety of individuals. Similarly individuals in villages where individuals perceive themselves as wealthier are more likely to see engagement with the U-Bridge sessions. Figure 1 plots the propensity score and friend propensity score, integrating out the heterogeneity  $\Psi_g^*$ . Each score contains complementary information about the selection patterns. Finally to test the fit of the model I run a covariate test / placebo test by replacing the outcome variable in (13) with each of the controls used in the analysis. None of the placebo coefficients are statistically significant for 16 out of the 20 variables. There are slight imbalances on one the relative income indicators, the distance to meeting and the average sociability.

Table 1 reports estimates of the average partial effects. Column (2) shows the coefficients under the network propensity approach. The direct effect  $\beta$  is positive but not statistically significant at the 10% level. The spillover effect  $\delta$  increases the participation index by 0.348 points, which is significant at the 1% level. This effect is quantitatively large relative to the standard deviation of the political participation index, which is around 0.567 points. This finding appears to suggest that the intervention had a large spillovers on non-participants, who increased their political activity. The interaction coefficient  $\delta$  is negative but not statistically significant at the 1% level. The results are consistent with the idea that the intervention had limited effects direct treatment effects, but promoted spillover effects on participants' social connections. Column (3) shows benchmark coefficients from an OLS regression with additive covariates. On one hand, the OLS coefficient of  $\beta$  is also not statistically significant at the 10% level. On the other hand, the OLS coefficient of  $\gamma$  is statistically significant but roughly half the size of the network propensity estimate. Finally, the coefficient of  $\delta$  is positive and statistically significant. The discrepancies in the results for  $\gamma$  and  $\delta$  can be explained by interactive spillover effects  $\gamma_{ig}$  and  $\delta_{ig}$  that are not captured by the additive OLS model.

#### 4.3 Microfinance Adoption in India

In this section I re-evaluate a program that encouraged the adoption of microfinance in rural areas of Southern India, by inviting select households to participate in an information about the program (Banerjee et al., 2013). Participant households were more likely to take out a loan. Spillovers are likely to occur in this context due to information transmission between participants and non-participants, and peer pressure to adopt.

The outcome is a binary variable  $Y_{ig}$  that is equal to one if household  $\{ig\}$  took out a loan when researcher followed-up a few months later. I estimate the following linear probability model with random coefficients.

$$Y_{ig} = \alpha_{ig} + \beta_{ig} D_{ig} + \gamma_{ig} \left(\frac{T_{ig}}{L_{ig}}\right) + \delta_{ig} \left(\frac{T_{ig}}{L_{ig}}\right)$$
(15)

Heterogeneity of  $\beta_{ig}$  in the microfinance example means that some households are more likely to take-out a loan after the information session than others. Conversely, heterogeneity of  $\gamma_{ig}$  and  $\delta_{ig}$  means that not every household is equally likely to get in debt after receiving information from their friends. The coefficient  $\delta_{ig}$  is the difference in spillovers effects between participant and non-participant households.

Identification of the average partial effect  $\tau \equiv (\alpha, \beta, \gamma, \delta)$  is particularly challenging in this setting, however, because the treatment was not randomly assigned. The microfinance organization followed a fixed targeting strategy in each village, that selected shopkeepers, teachers

and related occupations. However, Table 6 shows that treated households were wealthier; they were more likely to have stone or concrete houses as opposed to tile or thatch, have private electricity, more bedrooms, and own a latrine. For instance, the treated were 13.45% more likely to have access to some form of sanitation, with either a private or public latrine. These differences are statistically significant at the 5% level, using clustered standard errors by village. There were also significant differences by caste, a hereditary social category that still defines many social boundaries, with household of so-called "general caste" more likely to be treated as opposed to minorities.

To measure social network links, Banerjee et al. (2013) collected twelve different definitions of the network at baseline, including favor exchange, commensality and community activities. I choose a conservative definition of the network, such that  $A_{ijg}$  is equal to one if respondents reported a link along any of the dimensions. Figure 2a plots the resulting degree distribution, which shows that the treated had a higher number of friends. Households have around ten friends on average, which is around 5% of the average village size. Figure 2b shows that households reported that most of their friends were in the same broad caste category. As a matter of fact a significant portion of the households reported that *all* of their friends were in the same category. The histogram shows that the treated had more diversified friendships, in the sense that they had fewer friends of the same caste.

To estimate the network propensity score I use the same specification as in the example for Uganda. The second and third columns of Table 7 show the coefficients of the own propensity score and the corresponding standard errors. The structural parameters confirm the descriptive evidence. The number of rooms in the house, as well as the access to sanitation and electricity are statistically significant at the 5% level. Individuals of general caste and more connections, are more likely to be part of the program, even after accounting for asset measures. The observed group covariates are not statistically significant at the 10% level. Conversely, the fourth and fifth columns show estimated coefficients of the friend propensity score and their standard errors. Only the sociability index and the general caste indicator are statistically significant. This suggests that caste plays a crucial role on the interplay between homophily and selection. Treated individuals of general caste are more likely to be friend other treated individuals in their same caste category. The results also show that the unobserved heterogeneity parameters are not statistically significant at the 10% level.

Table 2 computes the treatment effects using my proposed inverse-weighting (IW) procedure and an ordinary least squares (OLS) regression that includes the covariates as additive controls.



(a) Histogram of the number of friends

(b) Histogram of same-caste friends

**Figure 2:** Figure (a) shows a histogram with the number of friends of each households, broken down by the treated and control households. Leaders tend to have a higher number of friends. Figure (b) shows a histogram with the fraction of same caste-friends. The general survey which contains information on five broad categories "General", "Minority", "OBC", "Scheduled Caste" and "Scheduled Tribe". I computed the fraction of treated friends for each household in the same caste category.

The IW results show that participants in the information session (leaders) are 8.5% more likely to take-out a microfinance loan after controlling baselin characteristics, and is significant at the 1% level. The value of the direct effect is 1% higher than the effect estimated by OLS. The OLS regression only controls for additive heterogeneity, but it does not account for the possibility of heterogeneous slopes/treatment effects. The fact that the IW and OLS produce similar results even though the leaders are highly selected suggests that the determinants of treatment are exogenous. The spillover effect is not significant in either case. That means that local variation in treated friends does not affect the outcome.

|                             | Network Propensity |            | OLS with covariates |            |
|-----------------------------|--------------------|------------|---------------------|------------|
|                             | Co efficients      | Std. Error | Co efficients       | Std. Error |
| Direct Effect $(\beta)$     | $0.096^{**}$       | (0.046)    | 0.077***            | (0.029)    |
| Spillover Effect $(\gamma)$ | 0.092              | (0.091)    | 0.026               | (0.036)    |
| Interaction $(\delta)$      | -0.102             | (0.292)    | 0.000               | (0.121)    |
| Village Controls            | Yes                |            | Yes                 |            |
| Ν                           | 7480               |            | 7480                |            |
| Villages                    | 43                 |            | 43                  |            |

**Table 2:** (Average Partial Effects Microfinance in India) \* Significant at 10%. \*\* Significant at 5%. \*\*\* Significant at 1%. The table shows the coefficients of the causal effects. The second and third columns show the coefficients and standard errors of the inverse-weighted estimator, respectively. The fourth and fifth columns are the coefficients of a ordinary least squares (OLS) regression that regresses  $Y_{ig}$  on a constant,  $D_{ig}$ ,  $(T_{ig}/L_{ig}), D_{ig} \times (T_{ig}/L_{ig})$  and the observed controls used in the inverse-weighting procedure. This sample merges the census-level data with a detailed survey for a random subsample of households, to fill in missing caste data. The sample excludes households without friends, households with more than 30 friends, and those that have missing caste or electricity data, which is 0.77% of the overall sample. The standard errors are clustered at the village level.

# 5 Discussion

#### 5.1 Effects by subpopulation

In many cases social programs deliberately target individuals based on baseline characteristics, and the policy maker may not be interested in the effects for the overall population. The identification problem is that individuals are only observed in a single treatment status, which means that the researcher has to find appropriates comparison individuals in the control group that approximate the behavior of the treated under a different exposure. To this end, let us define average partial effect on the treated (APT) and untreated (APU)

$$\tau_{APT} \equiv \mathbb{E}[\tau_{ig} \mid D_{ig} = 1, \mathcal{F}]$$
$$\tau_{APU} \equiv \mathbb{E}[\tau_{ig} \mid D_{ig} = 0, \mathcal{F}]$$

Theorem 6 presents identification results for  $\tau_{APT}$  and  $\tau_{APU}$ ,

**Theorem 6** (Identification Subpopulations). Suppose that (i)  $Y_{ig} = X'_{ig}\tau_{ig}$ , (ii)  $(X_{ig}, D_{ig}) \perp \tau_{ig}$  $V_{ig}$ , (iii)  $\mathcal{F}$  is  $V_{ig}$ -measurable and  $\mathbf{Q}_{xx}(v) = \mathbb{E}[X_{ig}X'_{ig} \mid V_{ig} = v]$  is invertible almost surely over the support of  $V_{ig} \mid \mathcal{F}$ , then

$$\boldsymbol{\tau}_{APT} = \frac{1}{\mathbb{E}[D_{ig}]} \times \mathbb{E}\left[p_d(V_{ig}) \times \mathbf{Q}_{xx}(V_{ig})^{-1} X_{ig} Y_{ig} \mid \mathcal{F}\right]$$
$$\boldsymbol{\tau}_{APU} = \frac{1}{1 - \mathbb{E}[D_{ig}]} \times \mathbb{E}\left[(1 - p_d(V_{ig})) \times \mathbf{Q}_{xx}(V_{ig})^{-1} X_{ig} Y_{ig} \mid \mathcal{F}\right]$$

The main intuition is fairly similar to Theorem 1, in the sense that the inverse weighting ensures equal comparisons across with different strata of  $V_{ig}$  whereas the own propensity  $p_d(V_{ig})$ weights each strata by the relative number of treated individuals. Notice that the unconditional average partial effects and the (APT, APU) are mutually constrained by the law of iterated expectations  $\boldsymbol{\tau} = \mathbb{E}[D_{ig}]\boldsymbol{\tau}_{APT} + (1 - \mathbb{E}[D_{ig}])\boldsymbol{\tau}_{APU}$ .

Table 9 computes the average partial effects by subpopulation for the political participation example in Uganda. The coefficients  $(\beta_{APT}, \delta_{APT})$  and  $(\beta_{APU}, \delta_{APU})$  have similar magnitudes, standard errors and significance. There are however, there are large differences in the magnitudes and significance levels of the spillovers for the control group (in fact  $\gamma_{APT} > \gamma_{APU}$ ). This suggested that individuals with a higher likelihood of participating in the session are more likely to change the behavior if one of their friends is treated. Analogously, in Table 10, I compute ( $\beta_{APT}, \gamma_{APT}, \delta_{APT}$  and ( $\beta_{APU}, \gamma_{APU}, \delta_{APU}$ ). The coefficients are similar in magnitude, with comparable standard errors, which suggests that both groups of individuals are fairly similar. In both tables, I compute the standard errors by replacing the definition of  $\psi_{IW}$  using a sample analog of the moment conditions in Theorem 6.

#### 5.2 Network Propensity Score and Experiments

One of the most effective ways to identify spillovers is to use a random saturation design. This a two-stage design rising in popularity in the empirical literature (Bursztyn et al., 2019; Crépon et al., 2013; Giné and Mansuri, 2018) and studied in several recent econometrics papers (Baird et al., 2019; DiTraglia et al., 2021). I establish a tight connection between the network propensity and identification in experiments. I show the applicability of my methods to study non-compliance in sparse networks.

In the first stage each group is randomly a saturation, a real number  $S_g \in [0, 1]$ . In the second stage individuals within each group are randomly assigned to treatment with probability  $S_g$ . This design is an extension of Bernoulli designs that treat individuals with a fixed probability, such as  $S_g = 0.5$ , and cluster design that assign complete groups to treatment or control, where  $S_g \in \{0, 1\}$ . The more interesting case combines corner an interior saturations. For example, Crépon et al. (2013) chooses  $S_g \in \{0, 0.25, 0.5, 0.75, 1\}$ , which generates more experimental variation. To simplify my analysis I focus on the case where the experimenter uses Bernoulli draws to offer treatment in the second stage.

The experimental setting relaxes the assumptions considerably. To discuss the identification of  $\tau$  in this experimental context it is useful to assume that  $C_{ig}$  includes both baseline individual characteristics (observed and unobserved). Similarly, I assume that  $\Psi_g^*$  includes group characteristics (observed or unobserved) heterogeneity and the exogenous saturations  $S_g$ . Under this definition it is easy to see that Selection on Observables is automatically satisfied because the treatment is exogenous. It is also easy to satisfy the Random Sampling and Dyadic Network assumptions. We can invoke the Aldous (1981) and Hoover (1979) theorems that state that any exchangeable network can be represented as a dyadic network with randomly sampled (and possibly unobserved)  $C_{ig}$ . The purpose of this exercise is to show that in certain experiments there is a simple set of conditioning statistics suffices to identify the treatment effects, even if there is rich unobserved heterogeneity determining the treatment and network choices.

**Example 1 (Perfect Compliance):** The random assignment of saturations and offers means that the propensity score is equal to the group saturation when there is perfect compliance. That means that individuals participate in the program when they are offered and are part of the control when they are not offered. In that case

$$p_{dig} = \underbrace{\mathbb{E}[D_{ig} \mid C_{ig}, \Psi_g^*]}_{\text{Definition}} = \underbrace{\mathbb{E}[D_{ig} \mid C_{ig}, \Psi_g^*, S_g]}_{\text{Redundancy}} = \underbrace{\mathbb{E}[D_{ig} \mid S_g]}_{\text{Second Stage}} = \underbrace{S_g}_{\text{First Stage}}.$$
 (16)

Equation (16) breaks down the process to show that the propensity is equal to the group saturation. The first equality defines  $p_d$ . The second equality uses the fact that  $S_g$  is a group characteristic that contains redundant information. The last two equality uses the property of the design, that the treatment probability only depends on a saturation which is independent of other characteristics.

I perform a similar break down for the friend propensity score.

$$p_{fig} = \underbrace{\mathbb{E}[D_{jg} \mid G_{ijg} = 1, C_{ig}, \Psi_g^*]}_{\text{Definition}} = \underbrace{\mathbb{E}[D_{jg} \mid G_{ijg} = 1, C_{ig}, \Psi_g^*, S_g]}_{\text{Redundancy}} = \underbrace{\mathbb{E}[D_{jg} \mid S_g]}_{\text{Second Stage}} = \underbrace{S_g}_{\text{First Stage}}$$

Finally, the number of friends  $L_{ig}$  is not randomly determined by the experimental design and can still be a source of homophily bias that the researcher needs to account for. In networks where everyone is connected ( $L_{ig} = N_g - 1$ ) this is equivalent to condition on the size of the group, such as classroom size.

The saturation  $S_g$  is independent of the random coefficients  $\tau_{ig}$  and the baseline information. Formally  $\tau_{ig} \perp S_g \mid L_{ig}$  and hence we can apply Lemma 3 to show that  $X_{ig} \perp \tau_{ig} \mid L_{ig}$ . That means that matching individuals with similar numbers of friends suffices to identify the average partial effects  $\tau$  using Theorem 1.

Example 2: (One-sided compliance) In practice researchers randomly extend offers but subjects may not be compelled to accept them. Under one-sided compliance treatment status is defined by  $D_{ig} = \tilde{C}_{ig}Z_{ig}$  where  $\tilde{C}_{ig}$  is a binary indicator for whether  $\{ig\}$  is a "complier" and  $Z_{ig}$ is their offer. Compliers with  $\tilde{C}_{ig} = 1$  may perceive larger returns from the program and always participate if offered, where never-takers  $\tilde{C}_{ig} = 0$  do not consider the program worthwhile. In their empirical example from (Crépon et al., 2013),  $D_{ig}$  is a job placement program. The peer effects are potential displacement effect for non-participants that were disadvantaged in a tight labor market. To fit this example within my framework I assume that  $\tilde{C}_{ig}$  is a component of the individual covariates  $C_{ig}$ .

Non-compliance introduces additional complications because the treatment is no longer randomly assigned. To analyze this problem it is useful to first compute an infeasible propensity score that conditions on the latent complier indicator. If  $\tilde{C}_{ig}$  were known

$$p_{dig} = \mathbb{E}[\widetilde{C}_{ig}Z_{ig} \mid G_{ijg} = 1, C_{ig}, \Psi_g^*] = \widetilde{C}_{ig}S_g$$

The propensity score for never-takers is always zero, whereas the propensity score for compliers depends on the saturation. The friend propensity equals

$$p_{fig} = \mathbb{E}[\widetilde{C}_{jg}Z_{jg} \mid G_{ijg} = 1, C_{ig}, \Psi_g^*] = \mathbb{E}[\widetilde{C}_{jg} \mid G_{ijg} = 1, C_{ig}, \Psi_g^*] \times S_g$$

The first equality applies the definition of the friend propensity and substitutes the expression for  $D_{jg}$  under one-sided compliance. Theorem 3 implies that the key dimensions of endogeneity are captured by the vector  $V_{ig} = (\tilde{C}_{ig}, \mathbb{E}[\tilde{C}_{jg} \mid G_{ijg} = 1, C_{ig}, \Psi_g^*], L_{ig})$  since  $S_g$  is exogenous. The second component of  $V_{ig}$  can be interpreted as the probability that a potential complier is treated. This agrees with related work in DiTraglia et al. (2021), where we show which causal effects are identified and show that  $(S_g)$  for the spillover effects because of first-stage heterogeneity. We propose a procedure that subsets  $Z_{ig} = 1$  to recover complier status from  $D_{ig}$  and consistently estimates the probability of a friend complier using  $T_{ig}/S_{ig}$  to construct a valid IV. The procedure relies on complete networks where  $L_{ig} = N_g - 1$  and  $N_g \to \infty$  in the asymptotic experiment.

Identification of the causal effects in networks where  $L_{ig}$  is bounded remains an open question. Vazquez-Bare (2020) and Imai et al. (2020) develop identification results for complete networks where  $L_{ig} = N_g - 1$  such as syblings, partners or classrooms without friendship information. However, there are no equivalent results for the case where  $L_{ig} \neq N_g - 1$  which captures the majority of observed networks. For these situations, estimating the friend propensity score from observed covariates by predicting the conditional mean of  $T_{ig}/S_g$  can be a practical secondbest alternative to account for network endogeneity. Non-compliance in randomized saturations designs introduces over-identifying restrictions on the matrix  $\mathbf{Q}_{xx}$  that fit within the framework discussed in this paper. Similar analyses can be used for two-sided compliance.

Example 3: (Graph-clustering) Random saturation (RS) designs are infeasible in the type of single-connected networks that are prevalent in online social media platforms like Facebook, Twitter and Linkedin. (Ugander et al., 2011) and (Eckles et al., 2017) propose a variant that uses within-network variation. Consider a three stage design. In the first stage, the researcher runs a graph-clustering algorithm to split the sample into distinct communities  $\kappa \in \{1, \ldots, \mathcal{K}\}$ . In the second stage each community is assigned a saturation  $S_{\kappa} \in [0, 1]$ . In the third stage each individual in  $\kappa$  is assigned to treatment with probability  $S_{\kappa}$ . A graphclustering experiment is identical to a random saturation design when the algorithm partitions the network into disjoint groups, but will produce very different results otherwise.

To analyze this design within my framework I assume that each individual belongs to a community  $\kappa_{ig} \in \{1, \ldots, \mathcal{K}\}$  and that the vector of saturations is a group level random variable  $\{S_1, \ldots, S_{\mathcal{K}}\}$  are included in  $\Psi_g^*$ . The own propensity score is equal to the saturation in  $\{ig\}'s$ community

$$p_{dig} = \mathbb{E}[D_{ig} \mid C_{ig}, \Psi_g^*] = \mathbb{E}[D_{ig} \mid \kappa_{ig}, C_{ig}, \Psi_g^*] = \mathbb{E}[D_{ig} \mid \kappa_{ig}] = S_{\kappa_{ig}}$$

Unsurprisingly, the propensity score is equal to saturation assigned to  $\{ig\}$ 's community.

The friend propensity score is more complicated. In a graph-clustering experiment we can

define multiple network measures as

$$p_{fig} = \mathbb{E}[D_{jg} \mid G_{ijg} = 1, C_{ig}, \Psi_g^*] = \mathbb{E}[[D_{jg} \mid G_{ijg} = 1, C_{ig}, \Psi_g^*, \kappa_{jg}] \mid G_{ijg} = 1, C_{ig}, \Psi_g^*]$$
  
=  $\mathbb{E}[S_{\kappa_{jg}} \mid G_{ijg} = 1, C_{ig}, \Psi_g^*]$   
=  $\sum_{\kappa=1}^{\kappa} \mathbb{E}[\mathbb{1}\{\kappa_{jg} = \kappa\} \mid G_{ijg} = 1, C_{ig}, \Psi_g^*] \times S_{\kappa}$ 

The inner expectation is the probability that  $\{ig\}'s$  potential friend belong to community  $\kappa$ . In this case, since the saturations are exogenously assigned, we can apply Lemma 3 once more to show that the key dimension of endogeneity is the probability of friendships between communities and the number of friends  $L_{ig}$ . This is a consequence of imperfect partitioning which makes inference depending on the clustering algorithm used.

# 6 Conclusion

This paper proposed a novel strategy for identifying average treatment effects and average spillover effects in settings with endogenous network formation and selection on observables. My approach provides a simple and tractable way of estimating these types of average effects. This method will be useful to program evaluators in a wide variety of non-experimental settings, where there is network data and treatment that is not randomly assigned by an experiment. Spillovers are important because they allows us to understand the broader implications of an intervention. Causal estimates of spillover effects allow policy makers to perform more accurate cost-benefit calculations. Spillovers are also important for welfare analysis. Interventions can sometimes have harmful unintended consequences on non-participants that should be taken into account at the time of generating new policies. Conversely, interventions can also have positive consequences on non-participants that need to be properly understood before adapting programs to new contexts.
|                            | Treated | Control | Difference    | Std. Error |
|----------------------------|---------|---------|---------------|------------|
| Political Participation    | 0.370   | -0.035  | 0.406***      | (0.043)    |
| Leader                     | 0.288   | 0.132   | $0.156^{***}$ | (0.027)    |
| Prosociability Index       | 0.198   | 0.196   | 0.001         | (0.011)    |
| Female                     | 0.275   | 0.606   | -0.331***     | (0.044)    |
| Secondary Education        | 0.458   | 0.207   | $0.251^{***}$ | (0.046)    |
| Relative income: Low       | 0.296   | 0.278   | 0.018         | (0.043)    |
| Relative income: Avg       | 0.108   | 0.103   | 0.005         | (0.015)    |
| Relative income: High      | 0.375   | 0.324   | 0.051         | (0.034)    |
| Relative income: Very High | 0.025   | 0.022   | 0.002         | (0.010)    |
| Distance to meeting        | 1.702   | 1.788   | - 0.086       | (0.163)    |
| Number of Friends          | 11.775  | 9.413   | $2.362^{***}$ | (0.326)    |
| Age                        | 40.504  | 37.090  | $3.415^{***}$ | (0.101)    |
| N                          | 250     | 2591    |               |            |
| Villages                   | 16      | 16      |               |            |

Table 3: (Summary statistics political participation in Uganda) Differences between leader households selected by the microfinance organization and non-leader households. All the variables are measured at baseline. This sample merges the census-level data with a detailed survey for a random subsample of households, to fill in missing caste data. The sample excludes households without friends, households with more than 30 friends, and those that have missing caste or electricity data, which is 0.77% of the overall sample. The standard errors are clustered by village.

|                                   | Own Prope    | ensity Score | Friend Prop   | ensity Score |
|-----------------------------------|--------------|--------------|---------------|--------------|
|                                   | Co efficient | Std. Error.  | Co efficient  | Std. Error   |
| Leader                            | 0.662        | (1.154)      | 0.116***      | (0.037)      |
| Sociability Index                 | -1.122       | (1.428)      | -0.157        | (0.128)      |
| Female                            | -1.383       | (1.712)      | -0.185        | (0.055)      |
| Has secondary education           | 0.876        | (1.503)      | $0.179^{***}$ | (0.050)      |
| Relative income: Somewhat worse   | 0.411        | (0.431)      | 0.013         | (0.047)      |
| Relative income: About the same   | 0.15         | (0.367)      | $0.079^{*}$   | (0.046)      |
| Relative income: Somewhat better  | 0.296        | (0.325)      | 0.022         | (0.057)      |
| Relative income: Much better      | 0.137        | (0.758)      | -0.111        | (0.125)      |
| Distance to meeting               | -0.191       | (0.430)      | -0.084        | (0.042)      |
| Number of friends                 | 0.187        | (0.263)      | -0.03         | (0.010)      |
| Age                               | 0.17         | (0.210)      | 0.027         | (0.020)      |
| Share of leaders in village       | 0.167        | (10.215)     | -2.406        | (2.433)      |
| Average sociability index         | -8.028       | (11.269)     | -8.034        | (3.334)      |
| Share of women in village         | -2.332       | (12.777)     | -9.321        | (4.281)      |
| Share of high-school educated     | 1.428        | (4.434)      | 0.322         | (1.748)      |
| Share reporting "Somewhat worse"  | 0.913        | (15.164)     | -2.484        | (4.092)      |
| Share reporting "About the same"  | 16.559       | (17.025)     | $9.536^{***}$ | (2.345)      |
| Share reporting "Somewhat better" | 4.024        | (17.623)     | 1.501         | (4.555)      |
| Average distance to meeting       | 0.233        | (0.701)      | -0.07         | (0.122)      |
| Average age                       | -1.192       | (2.598)      | -1.167        | (0.544)      |
| $\log(\sigma_1)$                  | 1.697        | (3.071)      |               |              |
| $\sigma_{12}$                     |              |              | 0.178         | (0.190)      |
| $\log(\sigma_2)$                  |              |              | -3.324        | (0.323)      |
| Constant                          | -2.168       | (16.170)     | $9.934^{***}$ | (3.328)      |
| Number of Observations            | 2,831        |              | 2,831         |              |
| Number of Villages                | 16           |              | 16            |              |

Table 4: (Network Propensity Score Parameters Uganda) \* Significant at 10%. \*\* Significant at 5%. \*\*\* Significant at 1%. Columns (2) and (4) show the estimated coefficients for propensity score and friend propensity scores, respectively. Columns (3) and (5) show the corresponding standard errors, that are clustered by village. The relative income asks how an individual's perceives her household income relative the typical household. The baseline category is "Much worse than the typical household". I dropped the "Share reporting: Much Better" variable because there was very little variation (only 2% of the sample marked this category). The bottom half of the table reports village-level averages and shares of the key variables. I omit the share for the "Much better" category because there are two few individuals. The bottom rows displays the parameters of the covariance matrix of the unobserved heterogeneity parameters. The sample for the table excludes households without friends and missing data on distance to meeting, gender, age and income.

|                               |        | $\widetilde{\beta}$ |              | $\widetilde{\gamma}$ |         | $\widetilde{\delta}$ |
|-------------------------------|--------|---------------------|--------------|----------------------|---------|----------------------|
|                               | Coeff. | Std. Error.         | Coeff.       | Std. Error           | Co eff. | Std. Error           |
| Leader                        | 0.039  | (0.217)             | 0.097        | (0.081)              | 0.097   | (3.970)              |
| Pro-Sociability Index         | 0.627  | (1.961)             | -0.091       | (0.107)              | -1.637  | (0.359)              |
| Female                        | 0.064  | (0.111)             | 0.083        | (0.150)              | -0.242  | (2.802)              |
| Has secondary education       | 0.399  | (1.454)             | 0.079        | (0.102)              | -0.946  | (3.161)              |
| Income: Somewhat worse        | 0.456  | (1.463)             | 0.123        | (0.097)              | -1.323  | (14.137)             |
| Income: About the same        | 1.509  | (6.494)             | 0.571        | (0.638)              | -3.92   | (58.105)             |
| Income: Somewhat better       | 7.395  | (26.897)            | $4.249^{**}$ | (1.917)              | -20.639 | (18.984)             |
| Income: Much better           | 2.877  | (9.529)             | 0.796        | (0.647)              | -7.198  | (0.957)              |
| Distance to meeting           | 0.129  | (0.462)             | $0.048^{**}$ | (0.023)              | -0.327  | (3.375)              |
| Number of friends             | 0.453  | (1.687)             | 0.081        | (0.068)              | -1.108  | (1.486)              |
| Age                           | 0.235  | (0.749)             | 0.036        | (0.027)              | -0.567  | (3.031)              |
| Share of leaders in village   | 0.401  | (1.507)             | 0.087        | (0.074)              | -1.013  | (2.164)              |
| Average sociability index     | Index  | (1.080)             | $0.067^{**}$ | (0.033)              | -0.736  | (12.502)             |
| Share of women in village     | 1.504  | (5.791)             | 0.562        | (0.745)              | -3.377  | (22.216)             |
| Share of high-school educated | 3.032  | (11.044)            | 0.597        | (0.422)              | -7.452  | (53.987)             |
| Share of "Somewhat worse"     | 7.243  | (26.817)            | $1.966^{*}$  | (1.157)              | -18.381 | (0.343)              |
| Share of "About the same"     | 0.007  | (0.080)             | 0.021        | (0.027)              | -0.012  | (0.977)              |
| Share of "Somewhat better"    | 0.031  | (0.245)             | 0.063        | (0.072)              | -0.08   | (7.409)              |
| Average distance to meeting   | -0.098 | (1.251)             | 0.708        | (0.794)              | 0.776   | (11.434)             |
| Average age                   | 0.23   | (2.839)             | 0.668        | (0.907)              | -0.566  | 0.000                |

**Table 5:** (Covariate Balancing Participation Uganda) \* Significant at 10%. \*\* Significant at 5%. \*\*\* Significant at 1%. This table shows the coefficients of inverse-weighted estimators, where each of the baseline characteristics is treated as a (placebo) outcome variable. If the weighting matrix is correctly specified  $\tilde{\beta} = \tilde{\gamma} = \tilde{\delta} = 0$ . The relative income asks how an individual's perceives her household income relative the typical household. The baseline category is "Much worse than the typical household". I dropped the "Share reporting: Much Better" variable because there was very little variation (only 2% of the sample marked this category). The bottom half of the table reports village-level averages and shares of the key variables. I omit the share for the "Much better" category because there are two few individuals. The sample for the table excludes households without friends and missing data on distance to meeting, gender, age and income.

|                           | Non Leaders | Leaders   | Difference  | Std. Error  |
|---------------------------|-------------|-----------|-------------|-------------|
|                           | (N = 6,551) | (N = 929) | (N = 7,480) | (N = 7,480) |
| Roof Type                 |             |           |             |             |
| Thatch                    | 2~%         | 1 %       | -1.12 %     | 0.43~%      |
| Tile                      | 38~%        | 31~%      | -6.32~%     | 2.42~%      |
| Stone                     | 26~%        | 30~%      | 4.2~%       | 2.19~%      |
| Sheet                     | 21~%        | 20~%      | -0.6 %      | 1.52~%      |
| RCC (Reinforced Concrete) | 10~%        | $15 \ \%$ | 4.69~%      | 1.2~%       |
| Other                     | 4 %         | 3~%       | -0.85~%     | 0.78~%      |
| No. Rooms                 |             |           |             |             |
| Mean                      | 0.77        | 1.06      | 0.29        | 0.06        |
| Sd                        | 1.1         | 1.39      |             |             |
| Electricity               |             |           |             |             |
| Yes, Private              | 61~%        | 72~%      | 10.94~%     | 1.98~%      |
| Yes, Government           | 32~%        | 24~%      | -8.19 %     | 1.9~%       |
| No                        | 7 %         | 4 %       | -2.75 %     | 0.68~%      |
| Latrine                   |             |           |             |             |
| Owned                     | 25~%        | 39~%      | 13.5~%      | $1.7 \ \%$  |
| Common                    | 1 %         | 1 %       | -0.06 %     | 0.25~%      |
| None                      | 74~%        | 61~%      | -13.45 %    | 1.78~%      |
| Residence                 |             |           |             |             |
| Owned                     | 90~%        | 93~%      | 2.66~%      | 1.05~%      |
| Owned but shared          | 1 %         | 1 %       | 0.34~%      | 0.35~%      |
| Rented                    | 6~%         | 3~%       | -2.65~%     | 0.76~%      |
| Leased                    | 0 %         | 0 %       | 0.08~%      | 0.16~%      |
| Government                | 4 %         | 3~%       | -0.42~%     | 0.65~%      |
| Caste                     |             |           |             |             |
| General                   | $11 \ \%$   | 20~%      | 8.31~%      | 1.64~%      |
| Minority                  | 3~%         | 3~%       | -0.68 %     | 0.69~%      |
| OBC                       | 51~%        | 51~%      | 0.21~%      | 1.65~%      |
| Scheduled Caste           | 29~%        | 22~%      | -6.69~%     | 1.57~%      |
| Scheduled Tribe           | $5 \ \%$    | 4 %       | -1.14 %     | 0.79~%      |
| Religion                  |             |           |             |             |
| Hinduism                  | 95%         | 95~%      | 0.09~%      | 0.87~%      |
| Islam                     | $5 \ \%$    | $5 \ \%$  | -0.1 %      | 0.91~%      |
| Christianity              | 0.09~%      | 0.11~%    | 0.02~%      | 0.12~%      |
| Number of Connections     |             |           |             |             |
| Mean                      | 9.91        | 12.5      | 2.59        | 0.25        |
| Standard Deviation        | 6.64        | 7.31      |             |             |

Table 6: (Summary statistics microfinance in India) Differences between leader households selected by the microfinance organization and non-leader households. All the variables are measured at baseline. This sample merges the census-level data with a detailed survey for a random subsample of households, to fill in missing caste data. The sample excludes households without friends, households with more than 30 friends, and those that have missing caste or electricity data, which is 0.77% of the overall sample. The standard errors are clustered by village.

|  | Own Propensity Score |             | Friend Prop   | ensity Score |
|--|----------------------|-------------|---------------|--------------|
|  | Co efficient         | Std. Error. | Co efficient  | Std. Error   |
| Tile Roof                              | -0.074               | (0.128)     | -0.08         | (0.060)      |
| Stone Roof                             | 0.09                 | (0.123)     | 0.053         | (0.061)      |
| Sheet Roof                             | 0.013                | (0.134)     | -0.059        | (0.061)      |
| No. Rooms                              | $0.124^{***}$        | (0.037)     | 0.007         | (0.013)      |
| Access to Electricity                  | $0.226^{*}$          | (0.124)     | 0             | (0.041)      |
| Access to Latrine                      | $0.321^{**}$         | (0.143)     | $0.106^{**}$  | (0.052)      |
| General Caste (base OBC)               | $0.602^{***}$        | (0.194)     | $0.266^{***}$ | (0.094)      |
| Scheduled Caste (base OBC)             | -0.087               | (0.106)     | -0.139        | (0.076)      |
| Scheduled Tribe (base OBC)             | -0.099               | (0.234)     | 0.046         | (0.097)      |
| Share of general caste in village      | -0.19                | (0.856)     | 0.253         | (0.402)      |
| Share of scheduled caste in village    | 0.032                | (0.354)     | -0.208        | (0.241)      |
| Share of scheduled tribe in village    | 0.233                | (2.040)     | 0.609         | (1.255)      |
| Share of latrine access in village     | 0.597                | (0.852)     | 0.384         | (0.567)      |
| Share of electricity access in village | -1.107               | (0.793)     | -0.613        | (0.479)      |
| Total Friends / Village Size           | $9.371^{***}$        | (2.155)     | $2.233^{***}$ | (0.858)      |
| $\log(\sigma_1)$                       | -0.447               | (2.087)     |               |              |
| $\sigma_{12}$                          |                      |             | 0.255         | (0.339)      |
| $\log(\sigma_2)$                       |                      |             | -2.071        | (15.712)     |
| Constant                               | -2.663               | (0.677)     | -1.565        | (0.391)      |
| Number of Observations                 | 7,480                |             | 7,480         |              |
| Number of Villages                     | 43                   |             | 43            |              |

Table 7: (Network Propensity Score Microfinance India) \* Significant at 10%. \*\* Significant at 5%. \*\*\* Significant at 1%. Columns (2) and (4) show the estimated coefficients for the propensity score and friend propensity scores, respectively. Columns (3) and (5) show the corresponding standard errors, that are clustered by village. All the variables are measured at baseline. The bottom rows displays the parameters of the covariance matrix of the unobserved heterogeneity parameters. This sample merges the census-level data with a detailed survey for a random subsample of households, to fill in missing caste data. The sample excludes households without friends, households with more than 30 friends, and those that have missing caste or electricity data, which is 0.77% of the overall sample.

|  |           | $\widetilde{o}$ |           | ~          |           | ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~ |
|--|-----------|-----------------|-----------|------------|-----------|---|
|  |           | $\beta$         |           | $\gamma$   |           | 0                                       |
|  | Co eff.   | Std. Error.     | Co eff.   | Std. Error | Co eff.   | Std. Error                              |
| Tile Roof                              | 0.029     | (0.080)         | 0.092     | (0.098)    | 0.092     | (0.504)                                 |
| Stone Roof                             | 0.025     | (0.095)         | 0.05      | (0.085)    | -0.136    | (0.381)                                 |
| Sheet Roof                             | 0.033     | (0.067)         | 0.085     | (0.122)    | -0.147    | (4.082)                                 |
| No. Rooms                              | 0.214     | (0.708)         | 0.551     | (0.466)    | -1.436    | (1.062)                                 |
| Access to Electricity                  | 0.07      | (0.190)         | 0.186     | (0.168)    | -0.433    | (0.646)                                 |
| Access to Latrine                      | 0.057     | (0.127)         | 0.078     | (0.086)    | -0.334    | (0.075)                                 |
| General Caste (base OBC)               | -0.027    | (0.016)         | -0.05     | (0.128)    | 0.13      | (0.710)                                 |
| Scheduled Caste (base OBC)             | 0.069     | (0.114)         | 0.161     | (0.198)    | -0.551    | (0.090)                                 |
| Scheduled Tribe (base OBC)             | 0.005     | (0.016)         | 0         | (0.025)    | 0.009     | (0.148)                                 |
| Share of general caste in village      | 0.002     | (0.021)         | -0.014    | (0.050)    | 0.035     | (0.421)                                 |
| Share of scheduled caste in village    | 0.013     | (0.069)         | 0.091     | (0.099)    | -0.144    | (0.102)                                 |
| Share of scheduled tribe in village    | 0.008     | (0.017)         | 0.01      | (0.010)    | -0.036    | (0.438)                                 |
| Share of latrine access in village     | 0.02      | (0.078)         | 0.052     | (0.056)    | -0.178    | (1.060)                                 |
| Share of electricity access in village | 0.051     | (0.178)         | 0.144     | (0.133)    | -0.391    | (0.118)                                 |
| Total Friends / Village Size           | 0.009     | (0.021)         | 0.016     | (0.014)    | -0.05     | 0.000                                   |
| Number of Observations                 | $7,\!480$ |                 | $7,\!480$ |            | $7,\!480$ |   |
| Number of Villages                     | 43        |                 | 43        |            | 43        |   |

**Table 8:** (Covariate Balancing Microfinance India) \* Significant at 10%. \*\* Significant at 5%. \*\*\* Significant at 1%. This table shows the coefficients of inverse-weighted estimators, where each of the baseline characteristics is treated as a (placebo) outcome variable. If the weighting matrix is correctly specified  $\tilde{\beta} = \tilde{\gamma} = \tilde{\delta} = 0$ . This sample merges the census-level data with a detailed survey for a random subsample of households, to fill in missing caste data. The sample excludes households without friends, households with more than 30 friends, and those that have missing caste or electricity data, which is 0.77% of the overall sample.

|                             | Political Pa        | articipation     |
|-----------------------------|---------------------|------------------|
|                             | Average Effect      | Average Effect   |
|                             | on Treated          | $on \ Untreated$ |
| Direct Effect $(\beta)$     | 0.269               | 0.283            |
|                             | (0.165)             | (0.166)          |
| Spillover Effect $(\gamma)$ | 0.410**             | 0.222            |
|                             | (0.116)             | (0.170)          |
| Interaction $(\delta)$      | -0.185              | -0.303           |
|                             | (0.862)             | (1.138)          |
| Ν                           | 2831                | 2831             |
| Villages                    | 16                  | 16               |
| * p < 0.1, *                | * p < 0.05. $*** p$ | < 0.01           |

**Table 9:** (Average partial effects by subpopulation Uganda) The left column shows the estimated coefficients of the average partial effects on the treated  $\mathbb{E}[(\beta_{ig}, \gamma_{ig}, \delta_{ig}) | D_{ig} = 1]$ . The right column shows the estimated coefficients of  $\mathbb{E}[(\beta_{ig}, \gamma_{ig}, \delta_{ig}) | D_{ig} = 0]$ .

|                             | Political Pa       | articipation     |
|-----------------------------|--------------------|------------------|
|                             | Average Effect     | Average Effect   |
|                             | on Treated         | $on \ Untreated$ |
| Direct Effect $(\beta)$     | 0.089              | 0.102**          |
|                             | (0.165)            | (0.052)          |
| Spillover Effect $(\gamma)$ | 0.090              | 0.093            |
|                             | (0.099)            | (0.090)          |
| Interaction $(\delta)$      | -0.094             | -0.108           |
|                             | (0.304)            | (0.300)          |
| Ν                           | 7480               | 7480             |
| Villages                    | 43                 | 43               |
| * p < 0.1, *                | ** p < 0.05. *** p | < 0.01           |

Table 10: (Average partial effects by subpopulation India) The left column shows the estimated coefficients of the average partial effects on the treated  $\mathbb{E}[(\beta_{ig}, \gamma_{ig}, \delta_{ig}) \mid D_{ig} = 1]$ . The right column shows the estimated coefficients of  $\mathbb{E}[(\beta_{ig}, \gamma_{ig}, \delta_{ig}) \mid D_{ig} = 0]$ .

# References

- Abadie, A., Imbens, G. W., 2016. Matching on the estimated propensity score. Econometrica 84 (2), 781–807.
- Akbarpour, M., Malladi, S., Saberi, A., 2018. Just a few seeds more: Value of targeting for diffusion in networks. Tech. rep., Working paper.
- Aldous, D. J., 1981. Representations for partially exchangeable arrays of random variables. Journal of Multivariate Analysis 11 (4), 581–598.
- Angrist, J. D., 2014. The perils of peer effects. Labour Economics 30, 98–108.
- Aronow, P., Samii, C., 2017. Estimating average causal effects under general interference, with application to a social network experiment. Annals of Applied Statistics.
- Arpino, B., De Benedictis, L., Mattei, A., 2015. Implementing propensity score matching with network data: The effect of gatt on bilateral trade.
- Attanasio, O., Barr, A., Cardenas, J. C., Genicot, G., Meghir, C., 2012. Risk pooling, risk preferences, and social networks. American Economic Journal: Applied Economics 4 (2), 134–67.
- Auerbach, E., 2019. Identification and estimation of a partially linear regression model using network data. arXiv preprint arXiv:1903.09679.
- Baird, S., Bohren, J. A., McIntosh, C., Ozler, B., 2019. Designing experiments to measure spillover effects.
- Banerjee, A., Chandrasekhar, A. G., Duflo, E., Jackson, M. O., 2013. The diffusion of microfinance. Science 341 (6144), 1236498.
- Banerjee, A. V., Chandrasekhar, A. G., Duflo, E., Jackson, M. O., 2017. Gossip: Identifying central individuals in a social network. Working Paper.
- Bester, C. A., Hansen, C. B., 2016. Grouped effects estimators in fixed effects models. Journal of Econometrics 190 (1), 197–208.
- Billingsley, P., 1995. Probability and measure. John Wiley & Sons.
- Blundell, R., Powell, J. L., 2003. Endogeneity in nonparametric and semiparametric regression models. Econometric society monographs 36, 312–357.

- Bramoullé, Y., Djebbari, H., Fortin, B., 2009. Identification of peer effects through social networks. Journal of econometrics 150 (1), 41–55.
- Bramoullé, Y., Djebbari, H., Fortin, B., 2020. Peer effects in networks: A survey.
- Breza, E., Chandrasekhar, A. G., 2019. Social networks, reputation, and commitment: evidence from a savings monitors experiment. Econometrica 87 (1), 175–216.
- Bursztyn, L., Cantoni, D., Yang, D., Yuchtman, N., Zhang, J., 2019. Persistent political engagement: Social interactions and the dynamics of protest movements. Working Paper.
- Candelaria, L. E., 2020. A semiparametric network formation model with unobserved linear heterogeneity. arXiv preprint arXiv:2007.05403.
- Chandrasekhar, A. G., Jackson, M. O., 2014. Tractable and consistent random graph models. Tech. rep., National Bureau of Economic Research.
- Chandrasekhar, A. G., Kinnan, C., Larreguy, H., 2018. Social networks as contract enforcement: Evidence from a lab experiment in the field. American Economic Journal: Applied Economics 10 (4), 43–78.
- Constantinou, P., Dawid, A. P., et al., 2017. Extended conditional independence and applications in causal inference. The Annals of Statistics 45 (6), 2618–2653.
- Crépon, B., Duflo, E., Gurgand, M., Rathelot, R., Zamora, P., 2013. Do labor market policies have displacement effects? evidence from a clustered randomized experiment. The quarterly journal of economics 128 (2), 531–580.
- DiTraglia, F. J., Garcia-Jimeno, C., O'Keefe-O'Donnovan, R., Sanchez-Becerra, A., 2021. Identifying causal effects in experiments with social interactions and non-compliance. Working Paper.
- Duflo, E., Saez, E., 2003. The role of information and social interactions in retirement plan decisions: Evidence from a randomized experiment. The Quarterly journal of economics 118 (3), 815–842.
- Eckles, D., Karrer, B., Ugander, J., 2017. Design and analysis of experiments in networks: Reducing bias from interference. Journal of Causal Inference 5 (1).
- Eubank, N., Grossman, G., Platas, M., Rodden, J., 2019. Viral voting: Social networks and political participation. Tech. rep.

- Fafchamps, M., Gubert, F., 2007. Risk sharing and network formation. American Economic Review 97 (2), 75–79.
- Fafchamps, M., Quinn, S., 2018. Networks and manufacturing firms in africa: Results from a randomized field experiment. The World Bank Economic Review 32 (3), 656–675.
- Ferrali, R., Grossman, G., Platas, M. R., Rodden, J., 2020. It takes a village: Peer effects and externalities in technology adoption. American Journal of Political Science 64 (3), 536–553.
- Fisher, R. A., et al., 1960. The design of experiments. The design of experiments. (7th Ed).
- Florens, J.-P., Heckman, J. J., Meghir, C., Vytlacil, E., 2008. Identification of treatment effects using control functions in models with continuous, endogenous treatment and heterogeneous effects. Econometrica 76 (5), 1191–1206.
- Forastiere, L., Airoldi, E. M., Mealli, F., 2020. Identification and estimation of treatment and interference effects in observational studies on networks. Journal of the American Statistical Association, 1–18.
- Forastiere, L., Mealli, F., Wu, A., Airoldi, E., 2018. Estimating causal effects under interference using bayesian generalized propensity scores. arXiv preprint arXiv:1807.11038.
- Gao, W. Y., 2020. Nonparametric identification in index models of link formation. Journal of Econometrics 215 (2), 399–413.
- Giné, X., Mansuri, G., 2018. Together we will: experimental evidence on female voting behavior in pakistan. American Economic Journal: Applied Economics 10 (1), 207–35.
- Goldsmith-Pinkham, P., Imbens, G. W., 2013. Social networks and the identification of peer effects. Journal of Business & Economic Statistics 31 (3), 253–264.
- Graham, B. S., 2017. An econometric model of network formation with degree heterogeneity. Econometrica 85 (4), 1033–1063.
- Graham, B. S., Pinto, C. C. d. X., 2018. Semiparametrically efficient estimation of the average linear regression function. Tech. rep., National Bureau of Economic Research.
- Hirano, K., Imbens, G. W., 2004. The propensity score with continuous treatments. Applied Bayesian modeling and causal inference from incomplete-data perspectives 226164, 73–84.
- Hoover, D. N., 1979. Relations on probability spaces and arrays of random variables. Preprint, Institute for Advanced Study, Princeton, NJ 2.

- Imai, K., Jiang, Z., Malani, A., 2020. Causal inference with interference and noncompliance in two-stage randomized experiments. Journal of the American Statistical Association, 1–13.
- Imbens, G. W., Newey, W. K., 2009. Identification and estimation of triangular simultaneous equations models without additivity. Econometrica 77 (5), 1481–1512.
- Jackson, M. O., Yu, N. N., Lin, Z., 2020. Adjusting for peer-influence in propensity scoring when estimating treatment effects. Available at SSRN.
- Johnsson, I., Moon, H. R., 2019. Estimation of peer effects in endogenous social networks: Control function approach. Tech. rep., arXiv. org.
- Kojevnikov, D., 2019. The bootstrap for network dependent processes. University of British Columbia working paper.
- Kojevnikov, D., Marmer, V., Song, K., 2020. Limit theorems for network dependent random variables. Journal of Econometrics.
- Lai, B., Reiter, D., 2000. Democracy, political similarity, and international alliances, 1816-1992. Journal of Conflict Resolution 44 (2), 203–227.
- Lee, L.-f., 2007. Identification and estimation of econometric models with group interactions, contextual factors and fixed effects. Journal of Econometrics 140 (2), 333–374.
- Leung, M. P., 2019a. Treatment and spillover effects under network interference. Review of Economics and Statistics, 1–42.
- Leung, M. P., 2019b. A weak law for moments of pairwise stable networks. Journal of Econometrics 210 (2), 310–326.
- Liu, L., Hudgens, M. G., Saul, B., Clemens, J. D., Ali, M., Emch, M. E., 2019. Doubly robust estimation in observational studies with partial interference. Stat 8 (1), e214.
- Manski, C. F., 1993. Identification of endogenous social effects: The reflection problem. The review of economic studies 60 (3), 531–542.
- Manski, C. F., 2013. Identification of treatment response with social interactions. The Econometrics Journal 16 (1).
- Meghir, C., Mobarak, A. M., Mommaerts, C. D., Morten, M., 2020. Migration and informal insurance. Tech. rep., Working Paper.
- Mele, A., 2017. A structural model of dense network formation. Econometrica 85 (3), 825–850.

- Newey, W. K., 1991. Uniform convergence in probability and stochastic equicontinuity. Econometrica: Journal of the Econometric Society, 1161–1167.
- Pelican, A., Graham, B., 2019. Testing for strategic interaction in social and economic network formation. University of California at Berkeley Working Paper.
- Robins, J. M., Rotnitzky, A., Zhao, L. P., 1994. Estimation of regression coefficients when some regressors are not always observed. Journal of the American statistical Association 89 (427), 846–866.
- Rosenbaum, P. R., Rubin, D. B., 1983. The central role of the propensity score in observational studies for causal effects. Biometrika 70 (1), 41–55.
- Rubin, D. B., 1980. Randomization analysis of experimental data: The fisher randomization test comment. Journal of the American Statistical Association 75 (371), 591–593.
- Sofrygin, O., van der Laan, M. J., 2017. Semi-parametric estimation and inference for the mean outcome of the single time-point intervention in a causally connected population. Journal of causal inference 5 (1).
- Tchetgen, E. J. T., Fulcher, I., Shpitser, I., 2017. Auto-g-computation of causal effects on a network. arXiv preprint arXiv:1709.01577.
- Ugander, J., Karrer, B., Backstrom, L., Marlow, C., 2011. The anatomy of the facebook social graph. arXiv preprint arXiv:1111.4503.
- Vasquez-Bare, G., 2019. Identification and estimation of spillover effects in randomized experiments. In: Working Paper.
- Vazquez-Bare, G., 2020. Causal spillover effects using instrumental variables. arXiv preprint arXiv:2003.06023.
- Wooldridge, J. M., 2004. Estimating average partial effects under conditional moment independence assumptions. Tech. rep., cemmap working paper.
- Zeleneev, A., 2020. Identification and estimation of network models with nonparametric unobserved heterogeneity.

# A Appendix

## A.1 Non-Separable Models

In this section I relax the random coefficients assumption in (1) by assuming that  $Y_{ig} = m(X_{ig}, \tau_{ig})$  as in Leung (2019a), where  $X_{ig} = (D_{ig}, T_{ig}, L_{ig})$ , *m* is an unknown function and  $\tau_{ig}$  is a vector of unobserved heterogeneity of arbitrary dimension. The researcher is interested in identifying the average structural function, defined as

$$M(x) = \int m(x,\varepsilon) dF(\varepsilon)$$

The function M(x) identified the average effect if everyone was subject to the same exposure.

The proof of Theorem 3 does not make any explicit use of the functional form of the outcome. If the assumptions of the theorem hold, then  $X_{ig} \perp \tau_{ig} \mid (C_{ig}, \Psi_q^*)$  and

$$\mathbb{E}[Y_{ig} \mid X_{ig} = x, C_{ig} = c, \Psi_g^* = \Psi^*, L_{ig} = l] = \int m(x, \varepsilon) dF(\varepsilon \mid x, c, \Psi^*, l)$$
$$= \int m(x, \varepsilon) dF(\varepsilon \mid c, \Psi^*, l)$$

This first stage is analogous to matching individuals with similar characteristics and similar levels of exposure. The conditional mean is only identified over the conditional support of  $(C_{ig}, \Psi_g^*)$  given  $X_{ig}$ . When the conditional support of  $(C_{ig}, \Psi_g^*)$  given  $X_{ig}$  equals the unconditional support we say that the system has *full support*. This condition is similar to a rank condition. In that case the average structural function can be identified by integrating the conditional mean using standard arguments as in Imbens and Newey (2009).

$$\mathbb{E}[\mathbb{E}[Y_{ig} \mid X_{ig} = x, C_{ig} = c, \Psi_g^* = \Psi^*, L_{ig} = l]] = \int \int m(x, \varepsilon) dF(\varepsilon \mid c, \Psi^*, l) dF(c, \Psi^*, l) = M(x)$$

Consequently, the average structural function is identified. Imbens and Newey (2009) show how to extend this idea to identify quantile effects in addition to average outcomes. We can also use the same set of arguments to prove identification of the average structural function for the network propensity score using the result of Theorem 4.

# A.2 Spurious Peer Effects

Consider the following example where an ordinary least squares (OLS) regression recovers spurious peer effects. Suppose that  $Y_{ig} = \alpha_{ig}$ , with  $\mathbb{E}[\varepsilon_{ig}] = 0$  and that Random Sampling, Selection on Observables and Dyadic Network are satisfied. Let  $V_{ig} \equiv (C_{ig}, \Psi_q^*, L_{ig})$  denote the confounders and  $X_{ig} = (1, \bar{D}_{-ig})$ , where  $\bar{D}_{-ig}$  is the fraction of treated friends, defined as  $T_{ig}/L_{ig}$ . In this case there are no treatment effects, direct or indirect, but the outcome are correlated with the confounders. The researcher runs the following regression over the subset of individuals with at least one friend,  $\mathcal{F} = \mathbb{1}\{L_{ig} > 0\}$ ,

$$Y_{ig} = \beta_0 + \beta_1 \bar{D}_{-ig} + \varepsilon_{ig}^*.$$

The true value of the intercept is  $\beta_0 = 0$  and the slope is  $\beta_1 = 0$ . The population OLS coefficient is defined as

$$\beta_1^{OLS} = \frac{Cov(D_{-ig}, Y_{ig}) \mid \mathcal{F}}{Var(\bar{D}_{-ig} \mid \mathcal{F})}$$

Plugging in  $Y_{ig} = \alpha_{ig}$  and using the law of total covariance,

$$\beta_1^{OLS} = \frac{\mathbb{E}[\overbrace{Cov(\bar{D}_{-ig},\alpha_{ig} \mid V_{ig},\mathcal{F})}^{(a)} \mid \mathcal{F}] + Cov(\overbrace{\mathbb{E}[\bar{D}_{-ig} \mid V_{ig}]}^{(b)},\overbrace{\mathbb{E}[\alpha_{ig} \mid V_{ig},\mathcal{F}]}^{(c)} \mid \mathcal{F})}^{(c)}}_{Var(\bar{D}_{-ig} \mid \mathcal{F})}.$$
 (A.1)

Theorem 4 ensures that  $\overline{D}_{-ig} \perp \alpha_{ig} \mid V_{ig}$ , which means that (a) is equal to zero. The term (b) equals  $p_{fig}$ , the friend propensity. To simplify notation define  $\alpha(V_{ig}) = \mathbb{E}[\alpha_{ig} \mid V_{ig}, \mathcal{F}] = \mathbb{E}[\alpha_{ig} \mid V_{ig}]$ . Consequently,

$$\beta_1^{OLS} = \frac{Cov\left(p_f(V_{ig}), \alpha(V_{ig}) \mid \mathcal{F}\right)}{Var(\bar{D}_{-ig} \mid \mathcal{F})}$$
(A.2)

The OLS coefficient is biased when  $\alpha(V_{ig})$  are correlated with  $p_{fig}$ . For example, suppose that  $V_{ig}$  is a poverty index and that  $p_{fig}$  is positively correlated with  $V_{ig}$ . That means that vulnerable individuals are more likely to have a higher fraction of friends who are targeted by the program. Similarly, suppose that  $Y_{ig}$  is a measure of food insecurity and that  $\alpha(V_{ig})$  is increasing in  $V_{ig}$ . Then  $\beta_1^{OLS} > 0$  because  $V_{ig}$  drives both the homophily/selection patterns and the baseline outcomes. Alternatively, when the network and treatment assignment are exogenous,  $p_{fig}$  is a constant and the OLS estimator is unbiased because the covariance in the numerator of (A.2) equals zero.

## A.3 Regularity Conditions

In this section I present conditions that are required to derive the asymptotic distribution of the estimator. In order to do so I assume that there is a sequence of distributions indexed by t. I denote the realization of variables of agent  $\{ig\}$  at point t in the sequence by including the subscript  $\{igt\}$ . I assume that one or more of the regressors need to be estimated. Let  $V_{igt} = (V_{1igt}^0, V_{2igt})$  be the observed regressor and let  $V_{igt}^0 = (V_{1igt}^0, V_{2igt}^0)$ . The first vector of regressors is observed without error, but the second estimator is estimated at rate  $\max_{g=1,\ldots,G_t} \max_{i=1,\ldots,N_{gt}} \|V_{2igt} - V_{2igt}^0\| = O(\lambda_t)$ . As in the main text, I assume that  $Z_{igt} = (X_{igt}, Y_{igt}, V_{igt})$  is a vector of data.

I next outline the key regularity conditions for convergence. First, for the estimator to be consistent the weighting matrix needs to by almost surely full rank in a neighborhood of  $\boldsymbol{\theta}$  around the true parameter. A positive semi-definite matrix  $\mathbf{Q}_{xx}$  is full rank if and only if its smallest eigenvalue is positive. Consequently, I quantify the *almost sure* requirement by imposing a lower bound on the eigenvalues of the estimated matrix. Let  $\lambda_{min}(v_1, v_2, \boldsymbol{\theta})$  denote the smallest eigenvalue of  $\mathbf{Q}_{xx}((v_1, v_2), \boldsymbol{\theta})$  and let  $B(\boldsymbol{\theta}_{0t}, \delta)$  be a ball or radius  $\delta > 0$  around  $\boldsymbol{\theta}_{0t}$  and suppose that  $V_{2igt}$  belongs to a compact set  $\mathcal{V}_2$  with probability approaching one. Let  $\underline{\lambda}(V_{1igt}^0, \boldsymbol{\theta}_{0t}, \delta) \equiv \inf_{\boldsymbol{\theta} \in B(\boldsymbol{\theta}_{0t}, \delta)} \inf_{v_2 \in \mathcal{V}_2} \lambda_{min}(V_{1igt}^0, v_2, \boldsymbol{\theta})$  be a lower bound on the eigenvalues of  $\mathbf{Q}_{xx}$ . I assume infimum holds over all values of  $v_2$  to ensure that the matrix is full rank, even if the regressors are noisily estimated.

Second, the weighting matrix also needs to be sufficiently smooth in order to reduce the impact of measurement error from estimating  $V_{2igt}$  and  $\boldsymbol{\theta}$ . I define its Sobolev-norm as

$$Q_{xx}^{\partial}(v_1, v_2, \theta) \equiv \sup_{0 \leqslant \alpha_1 + \alpha_2 \leqslant 3, \ \alpha_1, \alpha_2 \leqslant 2} \left\| \frac{\partial^{\alpha_1 + \alpha_2} \mathbf{Q}_{xx}(v_1, v_2, \theta)}{\partial v_2^{\alpha_1} \boldsymbol{\theta}^{\alpha_2}} \right\|$$
(A.3)

Equation (A.3) indicates the derivatives of the weighting matrix up to order three need to be bounded. In settings without a generated regressor problem, i.e.  $V_{2igt} = V_{2igt}^0$ , we typically only require smoothness conditions over  $\boldsymbol{\theta}$ . In this case, however, bounding the derivatives with respect to  $v_2$  as well, allows us to control the generated regressor error. In particular, I require that certain moments of the Sobolev norm need to be bounded.

In addition, the following regularity conditions have to be satisfied.

Assumption (Regularity Conditions). (i) There exists a  $\theta_{0t} \in int(\Theta)$  such that  $\forall \delta > 0$ ,  $\inf_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_{0t}\|>\delta} \mathcal{R}_t(\boldsymbol{\theta}) > \mathcal{R}_t(\boldsymbol{\theta}_{0t})$ , (ii)  $\mathbf{Q}_{xx}(V_{igt};\boldsymbol{\theta})$  is three-times continuously differentiable almost surely and  $\mathbb{E}[\sup_{\boldsymbol{\theta}\in\Theta}\sup_{v_2\in\mathcal{V}_2} (Q_{xx}^{\partial}(V_{1igt}^0, v_2, \theta))^4] < \infty$ , (iii)  $\mathbb{E}[\|X_{igt}\|^4]$ ,  $\mathbb{E}[\|Y_{igt}\|^2] < \infty$ , (iv)  $\underline{\lambda}(V_{1igt}^0, \boldsymbol{\theta}_{0t}, \delta) > \underline{\lambda} > 0$  almost surely for some  $\nu > 0$ . (v)  $H_{0t} \equiv \mathbb{E}\left[\frac{\partial}{\partial(\boldsymbol{\theta},\boldsymbol{\beta})'}\psi(Z_{igt}, \boldsymbol{\theta}_{0t})\right]$  is full rank, (vi)  $\Omega \equiv \mathbb{E}\left[\rho_{gt}\overline{\psi}_g(\boldsymbol{Z}_{gt}, \boldsymbol{\theta}_{0t})\overline{\psi}_g(\boldsymbol{Z}_{gt}, \boldsymbol{\theta}_{0t})'\right]$  is positive-definite, (vii)  $\max_{ig}\|V_{2igt} - V_{2igt}^0\| = O_p(\tau_t)$ , and (viii)  $\tau_t\sqrt{G_t} = o(1)$  and  $(G_t, N_t) \to \infty$  as  $t \to \infty$ .

Condition (i) is an identification condition that says that the true weighting matrix is the unique minimizer of the residuals. This is satisfied as long as the parametric family nests the conditional mean and the true criterion has a unique minimum. Condition (ii) imposes bounds on the moments of the Sobolev norm that hold uniformly over  $(\boldsymbol{\theta}, v_2)$ . Condition (iii) are bounds on the moments of the endogenous variable  $X_{igt}$  and  $Y_{igt}$ . Condition (iv) is a full rank condition for the average causal effect. Condition (v) is a rank condition on the system of equation that is similar to non-colinearity. Condition (vi) says the group-level covariance matrix

is non-degenerate and finite. Condition (vii) states the rate of convergence of the generated regressors. Condition (viii) states that the rate needs to be more accurate than the rate of growth of the groups  $G_t$ .

# **B** Proofs

# **B.1** Main Proofs

**Proof of Theorem 1 (Average Partial Effects).** By (ii)  $X_{ig} \perp \tau_{ig} \mid V_{ig}$ . By the decomposition property in Lemma B.1,  $X_{ig}X'_{ig} \perp \tau_{ig} \mid V_{ig}$  and by (i)  $Y_{ig} = X'_{ig}\tau_{ig}$ , which means that

$$\mathbf{Q}_{xy}(v) \equiv \mathbb{E}[X_{ig}Y_{ig} \mid V_{ig} = v] = \mathbb{E}[X_{ig}X'_{ig}\tau_{ig} \mid V_{ig} = v]$$
$$= \mathbb{E}[X_{ig}X'_{ig} \mid V_{ig} = v]\mathbb{E}[\tau_{ig} \mid V_{ig} = v]$$
$$= \mathbf{Q}_{xx}(v)\tau(v).$$

If  $\mathbf{Q}_{xx}(v)$  is almost surely full rank then  $\tau(v) = \mathbf{Q}_{xx}(v)^{-1}\mathbf{Q}_{xy}(v)$  almost surely. Since  $\mathcal{F}$  is coarser than  $V_{ig}$ ,  $\mathbb{E}[\tau_{ig} \mid V_{ig}, \mathcal{F}] = \mathbb{E}[\tau_{ig} \mid V_{ig}]$  and

$$\int \mathbf{Q}_{xx}(v)^{-1} \mathbf{Q}_{xy}(v) \ dF(v \mid \mathcal{F}) = \int \tau(v) \ dF(v \mid \mathcal{F}) = \tau$$

Finally, by the law of iterated expectations

$$\mathbb{E}[\mathbf{Q}_{xx}(V_{ig})^{-1}X_{ig}Y_{ig} \mid \mathcal{F}] = \mathbb{E}[\mathbb{E}[\mathbf{Q}_{xx}(V_{ig})^{-1}X_{ig}Y_{ig} \mid V_{ig}, \mathcal{F}] \mid \mathcal{F}] = \mathbb{E}[\mathbf{Q}_{xx}(V_{ig})^{-1}\mathbf{Q}_{xy}(V_{ig}) \mid \mathcal{F}] = \tau$$

**Proof of Theorem 2 (Closed form**  $\tau$ ). I make use of the mixture representation of  $\mathbf{Q}_{xx}$  derived in Lemma 2, assuming Random Sampling, Selection on Observables and Dyadic Network. If  $V_{ig} = (C_{ig}, \Psi_g^*, L_{ig})$ , then the conditional distribution of the network propensity score is degenerate and hence

$$\mathbf{Q}_{xx}(v) = \begin{pmatrix} 1 & \widetilde{\varphi}_1(p_f, l) \\ \widetilde{\varphi}_1(p_f, l, p_f, l) & \widetilde{\varphi}_2(p_f, l) \end{pmatrix} \otimes \begin{pmatrix} 1 & p_d \\ p_d & p_d \end{pmatrix}.$$

When  $\varphi(t, l) = t/l$ , then  $\tilde{\varphi}_1(p_f, l) = p_f$  and  $\tilde{\varphi}_1(p_f, l) = p_f(1-p_f)/l + p_f^2$  by using the moments in Lemma 1. The inverse of kronecker product of matrices is equal to the inverse of the kronecker products, which means that

$$\mathbf{Q}_{xx}(v)^{-1} = \begin{pmatrix} 1 & p_f \\ p_f & \frac{p_f(1-p_f)}{l} + p_f^2 \end{pmatrix}^{-1} \otimes \begin{pmatrix} 1 & p_d \\ p_d & p_d \end{pmatrix}^{-1} \\ = \left(\frac{1}{p_d(1-p_d)}\right) \left(\frac{l}{p_f(1-p_f)}\right) \begin{pmatrix} \frac{p_f(1-p_f)}{l} + p_f^2 & -p_f \\ -p_f & 1 \end{pmatrix} \otimes \begin{pmatrix} p_d & -p_d \\ -p_d & 1 \end{pmatrix}$$

We can write the regressors in kronecker product form as  $X'_{ig} = (1, T_{ig}/L_{ig}) \otimes (1, D_{ig})$ . Hence  $\mathbf{Q}_{xx}(V_{ig})^{-1}X_{ig}Y_{ig}$  multiplies two kronecker products. I use the property that for conformable matrices  $(M_1, M_2, M_3, M_4)$ ,  $(M_1 \otimes M_2)(M_3 \otimes M_4) = (M_1M_2) \otimes (M_3M_4)$ . After some algebraic manipulations we can show that

$$\mathbf{Q}_{xx}(V_{ig})^{-1}X_{ig}Y_{ig} = \begin{pmatrix} 1 + \frac{p_f L_{ig} - T_{ig}}{1 - p_f} \\ \frac{-p_f L_{ig} + T_{ig}}{p_f(1 - p_f)} \end{pmatrix} \otimes \begin{pmatrix} \frac{(1 - D_{ig})Y_{ig}}{1 - p_d} \\ \frac{D_{ig}Y_{ig}}{p_d} - \frac{(1 - D_{ig})Y_{ig}}{1 - p_d} \end{pmatrix}.$$

By Theorem 3,  $V_{ig}$  satisfies  $\tau_{ig} \perp X_{ig} \mid V_{ig}$ . Assuming the inverse of  $\mathbf{Q}_{xx}(V_{ig})$  is well defined then we can apply Theorem 1 to show that  $\tau = \mathbb{E}[\mathbf{Q}_{xx}(V_{ig})^{-1}X_{ig}Y_{ig} \mid \mathcal{F}]$ . We can obtain the individual coefficients  $(\alpha, \beta, \gamma, \delta)$  by expanding the kronecker product inside the expectation.

**Proof of Theorem 3 (Direct Confounders)**. I represent  $\{ig\}'s$  treatment indicator as  $D_{ig} = \mathcal{H}(C_{ig}, \Psi_g^*, \eta)$  where  $\mathcal{H}$  is a measurable function and  $\eta_{ig} \mid C_{ig}, \Psi_g^* \sim F(\eta \mid c, \Psi^*)$  is an unobserved participation shock. Since we can always define the participation shock as  $\eta = D_{ig} - \mathbb{P}(D_{ig} = 1 \mid C_{ig} = c, \Psi_g^* = \Psi^*)$ , this form does not entail any loss of generality.

Let  $\zeta_{ig} \equiv (\tau_{ig}, \eta_{ig}, C_{ig})$ . By Random Sampling and Dyadic Network,

$$\zeta_{ig} \perp \{U_{ijg}\}_{j \neq i}^{N_g}, \{\zeta_{jg}\}_{j \neq i}^{N_g} \mid \Psi_g^*$$
(B.1)

By (B.1), as well as the weak union and decomposition properties in Lemma B.1,

$$\zeta_{ig} \perp \{U_{ijg}\}_{j \neq i}^{N_g}, \{\zeta_{jg}\}_{j \neq i}^{N_g} \mid \eta_{ig}, C_{ig}, \Psi_g^*$$

$$\implies \tau_{ig} \perp \{U_{ijg}\}_{j \neq i}^{N_g}, \{\eta_{jg}, C_{jg}\}_{j \neq i}^{N_g} \mid \eta_{ig}, C_{jg}, \Psi_g^*$$

The second line subsets the relevant variables on either side of the independence relation. The participation decisions are functions of personal covariates and selection shocks. Similarly, the friendship vector  $\{ig\}$  only depends on the list of preference shocks (U) and covariates (C). Since  $L_{ig} = \sum_{j=1, j\neq i}^{N_g} A_{ijg}$  and  $X'_{ig} = (1, D_{ig}) \otimes \left(1, \varphi \left(\sum_{j=1, j\neq i}^{N_g} A_{ijg} D_{jg}, \sum_{j=1, j\neq i}^{N_g} A_{ijg}\right)\right)$ , that means that  $(L_{ig}, X_{ig})$  are both measurable with respect to  $\{U_{ijg}\}_{j\neq i}^{N_g}, \{\zeta_{jg}\}_{j=1}^{N_g}$ . Then by the decomposition property,

$$\tau_{ig} \perp \!\!\!\perp (X_{ig}, L_{ig}) \mid \eta_{ig}, C_{ig}, \Psi_g^*. \tag{B.2}$$

By Selection on Observables, the outcome heterogeneity is conditionally independent of the selection unobservables,  $\tau_{ig} \perp \eta_{ig} \mid C_{ig}, \Psi_g^*$ . By the contraction and decomposition properties,

$$\tau_{ig} \perp (X_{ig}, L_{ig}, \eta_{ig}) \mid C_{ig}, \Psi_g^* \implies \tau_{ig} \perp (X_{ig}, L_{ig}) \mid C_{ig}, \Psi_g^*$$
(B.3)

**Proof of Lemma 1 (Conditional Distribution).** Let  $\tilde{C}_{ig} \equiv (C_{ig}, \Psi_g^*)$  and let  $A_{ig} = \{A_{ijg}\}_{j=1, j\neq i}^{N_g}$ . If Random Sampling and Dyadic Network holds, then we can apply Lemma B.3 (Egocentric Likelihood Factorization) to show that

$$\mathbb{P}(D_g, A_{ig} \mid \widetilde{C}_{ig}) = \mathbb{P}(D_{ig} \mid \widetilde{C}_{ig}) \prod_{j \neq i}^{N_g} \mathbb{P}(D_{jg}, A_{ijg} \mid \widetilde{C}_{ig})$$
(B.4)

By Bayes' rule,  $\mathbb{P}(D_{jg}, A_{ijg} | \tilde{C}_{ig}) = \mathbb{P}(D_{jg} | A_{ijg}, \tilde{C}_{ig})\mathbb{P}(A_{ijg} | \tilde{C}_{ig})$  and substituting into (B.4)

$$\mathbb{P}(D_g, A_{ig} \mid \widetilde{C}_{ig})$$

$$= \mathbb{P}(D_{ig} \mid \widetilde{C}_{ig}) \prod_{j \neq i}^{N_g} \mathbb{P}(A_{ijg} \mid \widetilde{C}_{ig}) \prod_{j:A_{ijg}=1}^{N_g} \mathbb{P}(D_{jg} \mid A_{ijg} = 1, \widetilde{C}_{ig}) \prod_{j:A_{ijg}=0}^{N_g} \mathbb{P}(D_{jg} \mid A_{ijg} = 0, \widetilde{C}_{ig})$$

This proves that  $D_{ig} \perp \{D_{jg}, A_{ijg}\}_{j \neq i}^{N_g} \mid \widetilde{C}_{ig}$ . Let  $L_{ig} \equiv \sum_{j \neq i} A_{ijg}$  be the total friends,  $T_{ig} \equiv \sum_{j \neq i} D_{jg} A_{ijg}$  the total number of treated friends and  $M_{ig} \equiv \sum_{j \neq i} D_{jg} (1 - A_{ijg})$  be the total number of treated non-friends. Consequently, by the decomposition property in Lemma B.1,

$$D_{ig} \perp (L_{ig}, T_{ig}, M_{ig}) \mid \widetilde{C}_{ig} \implies D_{ig} \perp (L_{ig}, T_{ig}) \mid \widetilde{C}_{ig}$$

Furthermore, the likelihood can be factorized in terms of four sets of Bernoulli random variables, with a distinct event probability and  $(1, N_g, L_{ig}, N_g - L_{ig})$  trials, respectively.

Let  $p_f(\tilde{C}_{ig})$  and  $p_m(z)$  denote the participation probability of friends and non-friends. Then

$$\mathbb{P}(A_{ijg} \mid \tilde{C}_{ig}) = p_{\ell}(\tilde{C}_{ig})^{A_{ijg}}(1 - p_{\ell}(\tilde{C}_{ig}))^{1 - A_{ijg}}$$
$$\mathbb{P}(D_{jg} \mid A_{ijg} = 1, \tilde{C}_{ig}) = p_{f}(\tilde{C}_{ig})^{D_{jg}}(1 - p_{f}(\tilde{C}_{ig}))^{1 - D_{jg}}$$
$$\mathbb{P}(D_{jg} \mid A_{ijg} = 0, \tilde{C}_{ig}) = p_{m}(\tilde{C}_{ig})^{D_{jg}}(1 - p_{m}(\tilde{C}_{ig}))^{1 - D_{jg}}$$
(B.5)

The product of the probabilities is

$$\prod_{j:A_{ijg}=1}^{N_g} \mathbb{P}(A_{ijg} \mid \tilde{C}_{ig}) = p_{\ell}(\tilde{C}_{ig})^{L_{ig}} (1 - p_{\ell}(\tilde{C}_{ig}))^{N_g - L_{ig}}$$

$$\prod_{j:A_{ijg}=1}^{N_g} \mathbb{P}(D_{jg} \mid A_{ijg} = 1, \tilde{C}_{ig}) = p_f(\tilde{C}_{ig})^{T_{ig}} (1 - p_f(\tilde{C}_{ig}))^{L_{ig} - T_{ig}}$$

$$\prod_{j:A_{ijg}=0}^{N_g} \mathbb{P}(D_{jg} \mid A_{ijg} = 0, \tilde{C}_{ig}) = p_m(\tilde{C}_{ig})^{M_{ig}} (1 - p_m(\tilde{C}_{ig}))^{N_g - L_{ig} - M_{ig}}$$
(B.6)

Let  $\mathcal{B}_{(d,l,t,m)}$  be the set of permutations of treatment and link formation decisions that produce  $B_{ig} = (d, l, t, m)$ , where  $B_{ig} \equiv (D_{ig}, L_{ig}, T_{ig}, M_{ig})$ . Then  $\mathbb{P}_{B_{ig}|\tilde{C}_{ig}}(d, l, t, m)$  is equal to  $\sum_{(D_g, A_{ijg}) \in \mathcal{B}(d, l, t, m)} \mathbb{P}(D_g, A_{ijg})$ . The resulting distribution has the form

$$L_{ig} \mid \widetilde{C}_{ig} \sim \operatorname{Binom}(p_{\ell}(\widetilde{C}_{ig}), N_g)$$
$$T_{ig} \mid L_{ig}, \widetilde{C}_{ig} \sim \operatorname{Binom}(p_f(\widetilde{C}_{ig}), L_{ig})$$
$$D_{ig} \mid T_{ig}, L_{ig}, \widetilde{C}_{ig} \sim \operatorname{Bernoulli}(p_d(\widetilde{C}_{ig}))$$
$$M_{ig} \mid D_{ig}, T_{ig}, L_{ig}, \widetilde{C}_{ig} \sim \operatorname{Binom}(p_m(\widetilde{C}_{ig}, N_g - L_{ig}))$$

To complete the statement of the lemma, we only report the distribution of  $(D_{ig}, T_{ig}) | \tilde{C}_{ig}, L_{ig}$ , which does not depend on  $M_{ig}$ . The resulting distribution does not involve  $p_m(\tilde{C}_{ig})$ .

**Proof Theorem 4 (Balancing).** If Random Sampling and Dyadic Network hold, then we can apply Lemma 1 to show that  $D_{ig} \perp (T_{ig}, L_{ig}) \mid C_{ig}$  and

$$D_{ig} \mid T_{ig}, L_{ig}, C_{ig}, \Psi_g^* \sim \text{Bernoulli}(p_{dig})$$
  
 $T_{ig} \mid L_{ig}, C_{ig}, \Psi_g^* \sim \text{Binomial}(p_{fig}, L_{ig})$ 

The distribution of  $(D_{ig}, T_{ig}, L_{ig})$  is parametrized by  $P_{ig} \equiv (p_{dig}, p_{fig}, L_{ig})$ , which means that  $(D_{ig}, T_{ig}, L_{ig}) \mid C_{ig}, \Psi_g^*, P_{ig} \sim (D_{ig}, T_{ig}, L_{ig}) \mid P_{ig}$ . Consequently, the network propensity score and the group size summarizes all the pretreatment information and

$$(D_{ig}, T_{ig}, L_{ig}) \perp C_{ig}, \Psi_g^* \mid P_{ig}$$

By construction  $X_{ig}$  is a measurable function of  $(D_{ig}, L_{ig}, T_{ig})$ . By applying the decomposition property in Lemma B.1,

$$X_{ig} \perp C_{ig}, \Psi_g^* \mid P_{ig}. \tag{B.7}$$

This shows that  $P_{ig}$  is a balancing score.

If Random Sampling, Selection on Observables and Dyadic Network hold, then Theorem 3 states that  $\tau_{ig} \perp (X_{ig}, L_{ig}) \mid C_{ig}, \Psi_g^*$  which implies  $\tau_{ig} \perp |C_{ig}, \Psi_g^*, L_{ig}$ . By combining the redundancy and weak union properties, it follows that  $\tau_{ig} \perp X_{ig} \mid C_{ig}, \Psi_g^*, P_{ig}$ . Consequently, by (B.7) and the contraction property,  $(\tau_{ig}, C_{ig}, \Psi_g^*) \perp X_{ig} \mid P_{ig}$ . We can simplify the final expression by the decomposition property,

$$\tau_{ig} \perp X_{ig} \mid P_{ig}$$

**Proof of Lemma 2 (Mixture Representation).** By construction we can write the covariates as  $X'_{ig} = (1, \varphi(T_{ig}, L_{ig})) \otimes (1, D_{ig})$ . Therefore we can write  $X_{ig}X'_{ig}$  as

$$X_{ig}X'_{ig} = \begin{pmatrix} 1 & \varphi(T_{ig}, L_{ig})' \\ \varphi(T_{ig}, L_{ig}) & \varphi(T_{ig}, L_{ig})\varphi(T_{ig}, L_{ig})' \end{pmatrix} \otimes \begin{pmatrix} 1 & D_{ig} \\ D_{ig} & D_{ig} \end{pmatrix}$$

Define the functions

$$\widetilde{\varphi}_1(p_f, l) = \mathbb{E}[\varphi(T_{ig}, L_{ig}) \mid p_{fig} = p_f, L_{ig} = l]$$
  
$$\widetilde{\varphi}_2(p_f, l) = \mathbb{E}[\varphi(T_{ig}, L_{ig})\varphi(T_{ig}, L_{ig})' \mid p_{fig} = p_f, L_{ig} = l].$$

Under Lemma 1,  $D_{ig}$  is conditionally independent of  $(T_{ig}, L_{ig})$  given  $(C_{ig}, \Psi_g^*, L_{ig})$ , and the distributions are parametrized by the components of the network propensity score. Therefore we can decompose the conditional moments of  $X_{ig}X'_{ig}$  as

$$\mathbb{E}[X_{ig}X'_{ig} \mid C_{ig} = c, \Psi_g^* = \Psi, L_{ig} = l] = \begin{pmatrix} 1 & \widetilde{\varphi}_1(p_f, l)' \\ \widetilde{\varphi}_1(p_f, l) & \widetilde{\varphi}_2(p_f, l)' \end{pmatrix} \otimes \begin{pmatrix} 1 & p_d \\ p_d & p_d \end{pmatrix}$$

Since  $V_{ig}$  is measurable with respect to  $(C_{ig}, \Psi_g^*, L_{ig})$  we can apply the law of iterated expectations to obtain

$$\mathbf{Q}_{xx}(v) = \int \begin{pmatrix} 1 & \widetilde{\varphi}_1(p_f, l) \\ \widetilde{\varphi}_1(p_f, l, p_f, l) & \widetilde{\varphi}_2(p_f, l) \end{pmatrix} \otimes \begin{pmatrix} 1 & p_d \\ p_d & p_d \end{pmatrix} dF(p_d, p_f, l \mid V_{ig} = v).$$
(B.8)

**Proof of Lemma 3 ().** Let  $\widetilde{C}_{ig} \equiv (C_{ig}, \Psi_g^*)$  and  $X_{ig}^* \equiv (X_{ig}, L_{ig})$ . If Random Sampling, Selection on Observables and Dyadic Network hold, then we can apply Theorem 4 to show that  $(X_{ig}, L_{ig}) \perp \widetilde{C}_{ig} \mid p_d(\widetilde{C}_{ig}), p_f(\widetilde{C}_{ig}), L_{ig}$ .

Under Random Sampling, Selection on Observables and Selection on Observables we can apply Theorem 3 to show that  $(X_{ig}, L_{ig}) \perp \tau_{ig} | \widetilde{C}_{ig}$ . By the weak union property,  $(X_{ig}, L_{ig}) \perp \tau_{ig} | \widetilde{C}_{ig}, p_d(\widetilde{C}_{ig}), p_f(\widetilde{C}_{ig}), L_{ig}$ . Applying the contraction axiom,

$$(X_{ig}, L_{ig}) \perp (\tau_{ig}, \widetilde{C}_{ig}) \mid p_d(\widetilde{C}_{ig}), p_f(\widetilde{C}_{ig}), L_{ig}$$

Since  $V_{ig}$  is  $(\tilde{C}_{ig}, L_{ig})$ -measurable, we can apply the weak union property, as

$$(X_{ig}, L_{ig}) \perp (\tau_{ig}, \widetilde{C}_{ig}) \mid p_d(\widetilde{C}_{ig}), p_f(\widetilde{C}_{ig}), L_{ig}, V_{ig}$$

' By decomposition  $X_{ig} \perp \tau_{ig} \mid p_d(\widetilde{C}_{ig}), p_f(\widetilde{C}_{ig}), L_{ig}, V_{ig}$ . Since by assumption of the theorem

 $p_d(\widetilde{C}_{ig}), p_f(\widetilde{C}_{ig}), L_{ig} \perp \tau_{ig} \mid V_{ig}$ , we can apply the contraction axiom again to show that

$$(X_{ig}, p_d(\widetilde{C}_{ig}), p_f(\widetilde{C}_{ig}), L_{ig}) \perp \tau_{ig} \mid V_{ig}.$$

Finally, by the decomposition property,  $X_{ig} \perp \tau_{ig} \mid V_{ig}$ .

| Г |  |  |
|---|--|--|
|   |  |  |
|   |  |  |

### **B.2 Proof Asymptotics**

**Proof of Theorem 5 (Limiting Distribution Estimators)**. Define the square residual function as  $\mathcal{R}(z, v_2, \theta) \equiv r((x, y, (v_1, v_2)), \theta)^2$ , so that the estimated and population criterion functions can be written as

$$\hat{\mathcal{R}}_t(\boldsymbol{\theta}) \equiv \frac{1}{G_t \bar{n}} \sum_{ig} \mathcal{R}(Z_{igt}, V_{2igt}, \boldsymbol{\theta}),$$
$$\mathcal{R}_t(\boldsymbol{\theta}) \equiv \mathbb{E}[\mathcal{R}(Z_{igt}, V_{2igt}^0, \boldsymbol{\theta})].$$

Our first task is to prove uniform convergence of the criterion function by verifying the conditions of Lemma B.8. First, by Assumption (vii)  $\max_{ig} \|V_{2igt} - V_{2igt}^0\| = O_p(\lambda_t)$ . By assumption (viii),  $\sqrt{G_t}\lambda_t = o(1)$  which means that the maximum discrepancy is  $\tau_t = o(1)$ , as required.

Second we verify the uniform bounds on the moments. Assumptions (ii) and (iii) in Regularity Conditions imply that  $\mathcal{R}_t(Z_{igt}, V_{2igt}^0, \boldsymbol{\theta})$  has bounded moments. Conversely, let  $\mathcal{R}_{igt}^V$  and  $S_{igt}$  be uniform bounds on the derivatives  $\frac{\partial \mathcal{R}}{\partial v_2}$  and the score  $\psi_q \equiv \frac{\partial \mathcal{R}}{\partial \theta}$  as defined in (B.17) and (B.18). These bounds hold uniformly over  $\boldsymbol{\tau}$  because the average effect parameter does not enter  $\mathcal{R}$ . The bound on the expectation of the Sobolev-norm in Regularity Conditions part (ii) and Lemma (B.6) imply that  $\mathbb{E}[\mathcal{R}_{igt}^V] < \infty$  and  $\mathbb{E}[S_{igt}] < \infty$ . Consequently,  $\mathcal{R}$  satisfies the requirements of Lemma B.8, and hence

$$\sup_{\boldsymbol{\theta}\in\Theta} \|\hat{\mathcal{R}}_t(\boldsymbol{\theta}) - \mathcal{R}_t(\boldsymbol{\theta})\| \to^p 0$$
(B.9)

Our next task is to show that  $\hat{\theta}_t$  is consistent. By Regularity Conditions part (i) for any  $\delta > 0$  there exists a  $\nu > 0$  such that

$$\mathbb{P}\left(\left\|\widehat{\boldsymbol{\theta}}_{t}-\boldsymbol{\theta}_{0t}\right\| > \delta\right) \leq \mathbb{P}(\mathcal{R}_{t}(\widehat{\boldsymbol{\theta}}_{t})-\mathcal{R}_{t}(\boldsymbol{\theta}_{0t}) \ge \nu) \\
= \mathbb{P}(\mathcal{R}_{t}(\widehat{\boldsymbol{\theta}}_{t})-\widehat{\mathcal{R}}_{t}(\widehat{\boldsymbol{\theta}}_{t})+\widehat{\mathcal{R}}_{t}(\widehat{\boldsymbol{\theta}}_{t})-\mathcal{R}_{t}(\boldsymbol{\theta}_{0t}) \ge \nu) \quad \text{Adding/subtracting } \mathcal{R}_{t}(\widehat{\boldsymbol{\theta}}_{t}) \\
\leq \mathbb{P}(\mathcal{R}_{t}(\widehat{\boldsymbol{\theta}}_{t})-\widehat{\mathcal{R}}_{t}(\widehat{\boldsymbol{\theta}}_{t})+\widehat{\mathcal{R}}_{t}(\boldsymbol{\theta}_{0t})-\mathcal{R}(\boldsymbol{\theta}_{0t}) \ge \nu) \quad \text{Since } \widehat{R}_{t}(\widehat{\boldsymbol{\theta}}_{t}) \leq \widehat{R}_{t}(\boldsymbol{\theta}_{0t}) \\
\leq \mathbb{P}\left(2\sup_{\boldsymbol{\theta}\in\Theta}\left\|\widehat{\mathcal{R}}_{t}(\boldsymbol{\theta})-\mathcal{R}_{t}(\boldsymbol{\theta})\right\| \ge \nu\right) \quad \text{Uniform Bound} \\
\rightarrow^{p} 0 \quad \text{By (B.33)}$$

Consequently  $\hat{\boldsymbol{\theta}}_t \rightarrow^p \boldsymbol{\theta}_{0t}$ .

We now turn to the task of proving asymptotic normality. In a slight abuse of notation, I use  $\psi(z, v_2, \boldsymbol{\tau}, \boldsymbol{\theta})$  to denote the influence function  $\psi((x, y, (v_1, v_2)), \boldsymbol{\tau}, \boldsymbol{\theta})$ 

$$o_p(\tau) = \frac{1}{G_t N_t} \sum_{ig} \psi(Z_{igt}, V_{2igt}, \hat{\boldsymbol{\tau}}_t, \hat{\boldsymbol{\theta}}_t)$$

By a first-order expansion

$$0 = \frac{1}{G_t N} \sum_{ig} \psi(Z_{igt}, V_{2igt}^0, \boldsymbol{\tau}_{0t}, \boldsymbol{\theta}_{0t}) + \frac{1}{G_t N} \sum_{ig} \frac{\partial}{\partial v_2} \psi(Z_{igt}, \widetilde{V}_{2igt}, \widetilde{\boldsymbol{\tau}}, \widetilde{\boldsymbol{\theta}}_t) \Delta_{igt} + \frac{1}{NG_t} \sum_{ig} \begin{pmatrix} \frac{\partial}{\partial \boldsymbol{\tau}'} \psi(Z_{igt}, \widetilde{V}_{2igt}, \widetilde{\boldsymbol{\tau}}, \widetilde{\boldsymbol{\theta}}_t) \\ \frac{\partial}{\partial \boldsymbol{\theta}'} \psi(Z_{igt}, \widetilde{V}_{2igt}, \widetilde{\boldsymbol{\tau}}, \widetilde{\boldsymbol{\theta}}_t) \end{pmatrix} \begin{pmatrix} \hat{\boldsymbol{\tau}}_t - \boldsymbol{\tau}_{0t} \\ \hat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}_{0t} \end{pmatrix}$$
(B.10)

Our next task is to show that the second term is  $O_p(\lambda_t \sqrt{G_t})$ . To this end it is useful to decompose that influence function into two sets of equations  $\psi = [\psi_q, \psi_{IW}]'$ , for the weighting matrix and the average effects, respectively. Let  $B(\boldsymbol{\theta}_{0t}, \nu)$  denote a ball of radius  $\nu$  around the true parameter. By assumption (iv) the smallest eigenvalue of  $Q_{xx}$  is bounded by a fixed constant for  $\boldsymbol{\theta} \in B(\boldsymbol{\theta}_{0t}, \nu)$ . Since  $\hat{\boldsymbol{\theta}}_t$  and  $\tilde{\boldsymbol{\theta}}_t$  are both consistent, the estimator is contained in the ball with probability approaching one as  $(G_t, N_t) \to \infty$ .

Define  $S_{igt}^V$  and  $\psi_{IW,ig}^\partial$  as uniform upper bounds for the partial derivatives of s and  $\psi_{IW}$  as defined in (B.19) and (B.21). Furthermore, let  $\Delta_{max} \equiv \max_{ig} \|V_{2igt} - V_{2igt}^0\|$  be the maximum discrepancy between the generated and true regressors. By the triangle inequality.

$$\begin{aligned} \left\| \frac{1}{G_t \bar{n}_t} \sum_{ig} \frac{\partial}{\partial v_2} \psi(Z_{igt}, \tilde{V}_{2igt}, \tilde{\tau}, \tilde{\theta}_t) \Delta_{igt} \right\| \\ &\leqslant \frac{1}{G_t \bar{n}_t} \sum_{ig} \left\| \frac{\partial}{\partial v_2} \psi(Z_{igt}, \tilde{V}_{2igt}, \tilde{\tau}, \tilde{\theta}_t) \right\| \cdot \Delta_{max} \\ &\leqslant \frac{1}{G_t \bar{n}_t} \sum_{ig} \left( \left\| \frac{\partial s(Z_{igt}, \tilde{V}_{2igt}, \tilde{\tau}, \tilde{\theta}_t)}{\partial v_2} \right\| + \left\| \frac{\partial \psi_{IW}(Z_{igt}, \tilde{V}_{2igt}, \tilde{\tau}, \tilde{\theta}_t)}{\partial v_2} \right\| \right) \cdot \Delta_{max} \quad \text{Component Bounds} \\ &\leqslant \left( \frac{1}{G_t \bar{n}_t} \sum_{ig} S_{igt}^V + \psi_{IW,ig}^\partial \right) \cdot \Delta_{max} + o_p(1) \quad \text{Since } \tilde{\theta}_t \in B(\theta_{0t}, \nu) \text{ w.p.a.1} \end{aligned}$$
(B.11)

The discrepancy  $\Delta_{max}$  is  $O_p(\lambda_t)$  by Assumption (vii). Conversely, the bounds on the expectation of the Sobolev-norm in Regularity Conditions part (ii) and the moments in (iii) can be used to show that  $\mathbb{E}[S_{igt}^V]$ ,  $\mathbb{E}[\psi_{IW,ig}^{\partial}] < \infty$  and  $\frac{1}{G_t \bar{n}_t} \sum_{ig} (S_{igt}^V + \psi_{IW,ig}^{\partial}) = O_p(1)$ , by Lemmas B.6 and B.7, respectively. By combining the two findings we conclude that the right-hand side of (B.11) is  $O_p(\lambda_t)$ .

The partial derivative with respect to  $\boldsymbol{\tau}$  in (B.10) has a simple form

$$\frac{\partial}{\partial \boldsymbol{\tau}'} \psi(Z_{igt}, \widetilde{V}_{2igt}, \widetilde{\boldsymbol{\tau}}, \widetilde{\boldsymbol{\theta}}_t) = \begin{pmatrix} \mathbf{0} \\ -I \end{pmatrix} \equiv H_{0,\tau}$$

The first set of rows is zero because the equations to compute to the weighting matrix and the second rows is the identity because  $\tau$  enters linearly in  $\psi_{IW}$ . In this case the derivative is constant and crucially, does not depend on the estimated parameters.

Since the components that contain  $\boldsymbol{\theta}$  and  $\boldsymbol{\tau}$  are additively separable, the partial derivative with respect to  $\boldsymbol{\theta}$  in (B.10) does not depend on  $\boldsymbol{\tau}$ . We write this concisely as

$$\frac{\partial}{\partial \boldsymbol{\theta}'}\psi(Z_{igt}, \widetilde{V}_{2igt}, \widetilde{\boldsymbol{\tau}}, \widetilde{\boldsymbol{\theta}}_t) = \frac{\partial}{\partial \boldsymbol{\theta}'}\psi(Z_{igt}, \widetilde{V}_{2igt}, \widetilde{\boldsymbol{\theta}}_t) = \begin{pmatrix} \frac{\partial}{\partial \boldsymbol{\theta}'}\psi_q(Z_{igt}, \widetilde{V}_{2igt}, \widetilde{\boldsymbol{\theta}})\\ \frac{\partial}{\partial \boldsymbol{\theta}'}\psi_{IW}(Z_{igt}, \widetilde{V}_{2igt}, \widetilde{\boldsymbol{\theta}}) \end{pmatrix}$$
(B.12)

Our next task is to impose integrable bounds on (B.12) in order to apply the uniform consistency result in B.8. On one hand, our bounds on the expectations in assumptions (ii) and (iii) allow us to apply the first part of Lemma B.6. The lemma shows that  $\frac{\partial \psi_q}{\partial \theta}$ ,  $\frac{\partial^2 \psi_q}{\partial \theta \partial \theta'}$ ,  $\frac{\partial^2 \psi_q}{\partial v_2 \partial \theta'}$  are uniformly bounded over  $(V_2, \theta) \in \mathcal{V}_2 \times \Theta$  by integrable random variables. On the other hand, assumption (ii), (iii) and (iv) allow us to apply the second part of the lemma, which implies  $\frac{\partial^2 \psi_{IW}}{\partial \theta \partial \theta'}$ ,  $\frac{\partial^2 \psi_{IW}}{\partial v_2 \partial \theta'}$  are uniformly bounded over  $(V_2, \theta) \in \mathcal{V}_2 \times B(\theta_{0t}, \nu)$  by an integrable random variable. Consequently, we can apply Lemma B.8 to show that

$$\sup_{\boldsymbol{\theta}\in B(\boldsymbol{\theta}_{0t},\nu)} \left\| \frac{1}{G_t \bar{n}_t} \sum_{ig} \frac{\partial}{\partial \boldsymbol{\theta}'} \psi(Z_{igt}, \widetilde{V}_{2igt}, \widetilde{\boldsymbol{\theta}}_t) - \mathbb{E} \left[ \frac{\partial}{\partial \boldsymbol{\theta}'} \psi(Z_{igt}, V_{2igt}^0, \boldsymbol{\theta}) \right] \right\| \to^p 0 \tag{B.13}$$

Since  $\hat{\boldsymbol{\theta}}_t$  is consistent  $\|\tilde{\boldsymbol{\theta}}_t - \boldsymbol{\theta}_{0t}\| \leq \|\hat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}_{0t}\| = o_p(1)$ . Therefore, by the uniform consistency result in (B.13),

$$\frac{1}{G_t \bar{n}_t} \sum_{ig} \frac{\partial}{\partial \theta'} \psi(Z_{igt}, \widetilde{V}_{2igt}, \widetilde{\theta}_t) \to^p \mathbb{E} \left[ \frac{\partial}{\partial \theta'} \psi(Z_{igt}, V_{2igt}^0, \theta_{0t}) \right] \equiv H_{0,\theta}$$

By assumption (iv),  $H_0 = [H_{0,\tau}, H_{0,\theta}]$  is full rank. Therefore, solving for the parameter in (B.10) and multiplying by  $\sqrt{G_t}$ ,

$$\sqrt{G_t} \begin{pmatrix} \hat{\boldsymbol{\tau}}_t - \boldsymbol{\tau}_{0t} \\ \hat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}_{0t} \end{pmatrix} = -\left(H_0 + o_p(1)\right)^{-1} \left[ \sqrt{G_t} \left( \frac{1}{NG_t} \sum_{ig} \psi(Z_{igt}, V_{2igt}^0, \boldsymbol{\theta}_{0t}) \right) + O_p\left(\lambda_t \sqrt{G_t}\right) \right]$$

Let  $\mathbb{E}_*$  and  $\mathbb{E}$  denote the sampling (equal-weighted-group) measure and the population measure respectively. By construction,  $\mathbb{E}_*[\rho_{gt}\psi(Z_{igt},\theta)] = \mathbb{E}[\psi(Z_{igt},\theta)]$ , where  $\rho_{gt} \equiv N_{gt}/N_t$  is the relative size of each group. By Lemma B.7,  $\frac{\bar{n}_t}{N_t} \rightarrow^p 1$  as  $(G_t, N_t) \rightarrow \infty$ . Conversely, define the withingroup average  $\overline{\psi}_g(\mathbf{Z}_g, \boldsymbol{\theta}_{0t}) \equiv \frac{1}{N_{gt}} \sum_{i=1}^{N_{gt}} \psi(Z_{igt}, V_{2igt}^0, \boldsymbol{\theta}_{0t})$ , where  $\mathbf{Z}_g \equiv \{(X_{igt}, Y_{igt}, (V_{igt}^0))\}_{i=1}^{N_{gt}}$  is a matrix of individual covariates. By some algebraic manipulations

$$\frac{\sqrt{G_t}}{\bar{n}_t G_t} \sum_{ig} \psi(Z_{igt}, V_{2igt}^0, \boldsymbol{\theta}_{0t}) = \left(\frac{N_t}{\bar{n}_t}\right) \left(\frac{1}{\sqrt{G_t}} \sum_{g=1}^{G_t} \rho_{gt} \overline{\psi}_g(\boldsymbol{Z}_g, \boldsymbol{\theta}_{0t})\right)$$

Our final task is to apply the central limit theorem. First, we check that the influence function is mean zero. By distributing the expectation

$$\mathbb{E}_*[\rho_{gt}\overline{\psi}_g(\boldsymbol{Z}_g,\boldsymbol{\tau}_{0t},\boldsymbol{\theta}_{0t})] = \mathbb{E}_*[\rho_{gt}\psi(Z_{igt},V_{2igt}^0,\boldsymbol{\tau}_{0t},\boldsymbol{\theta}_{0t})] = \mathbb{E}[\psi(Z_{igt},V_{2igt}^0,\boldsymbol{\tau}_{0t},\boldsymbol{\theta}_{0t})]$$

Recall that  $\psi = [s, \psi_{IW}]$ . The mean of s is equal to zero at the true value when the weighting matrix is properly specified. Similarly,  $\psi_{IW}$  is equal to zero by Theorem 1.

Finally, by assumption (v),  $\mathbb{E}_*[\rho_{gt}^2 \overline{\psi}_g \overline{\psi}_g'] = \mathbb{E}[\rho_{gt} \overline{\psi}_g \overline{\psi}_g'] \equiv \Omega_{0t}$  is a positive-definite matrix. By the Lindenber-Feller central limit theorem, as  $(G_t, N_t) \to \infty$ ,

$$\Omega_{0t}^{-1}\left(\frac{1}{\sqrt{G_t}}\sum_{g=1}^{G_t}\rho_{gt}\overline{\psi}_g(\boldsymbol{Z}_g,\boldsymbol{\theta}_{0t})\right) \to^d \mathcal{N}(\boldsymbol{0},I).$$

Combining the results we prove that the estimator converges to a normal distribution plus a bias term,

$$\sqrt{G_t} \Sigma_t^{-1/2} \begin{pmatrix} \widehat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}_{0t} \\ \widehat{\boldsymbol{\tau}}_t - \boldsymbol{\tau}_{0t} \end{pmatrix} \to^d \mathcal{N}(\mathbf{0}, I) + O_p \left( \lambda_t \sqrt{G_t} \right).$$

where  $\Sigma_t = H_{0t}^{-1} \Omega_{0t} H_{0t}^{-1}$ . Under assumption (viii) the second term is  $o_p(1)$ ,

#### **B.3** Proofs Extensions and Experiments

**Proof of Theorem 6 (Identification Subpopulations).** For the APT, our first objective is to rewrite the inner term of the expectation in terms of the localized effect  $\tau(v)$ , instead of  $(\mathbf{Q}_{xx}, X_{ig}, Y_{ig})$ . To this end we compute the conditional expectation given a particular value of the control variable  $V_{ig}$ . In this case  $p_d(V_{ig})$  is a constant given  $V_{ig}$ , so we can directly apply part (i) of Lemma 1,

$$\mathbb{E}[p_d(V_{ig}) \times \mathbf{Q}_{xx}(V_{ig})^{-1} X_{ig} Y_{ig} \mid V_{ig} = v] = p_d(V_{ig}) \times \mathbf{Q}_{xx}(v)^{-1} \mathbb{E}[X_{ig} Y_{ig} \mid V_{ig}]$$
$$= p_d(V_{ig}) \times \mathbf{Q}_{xx}(v)^{-1} \mathbf{Q}_{xy}(V_{ig})$$
$$= p_d(v) \times \boldsymbol{\tau}(v)$$
(B.14)

The second task is to express (B.14) in terms of the primitives  $(D_{ig}, \tau_{ig})$ . By definition  $p_d(v) = \mathbb{P}(D_{ig} = 1 \mid V_{ig} = v)$ . Since  $V_{ig}$  is a control variable for  $D_{ig}$ , it follows that

$$\boldsymbol{\tau} \equiv \mathbb{E}[\tau_{ig} \mid V_{ig} = v] = \mathbb{E}[\tau_{ig} \mid V_{ig} = v, D_{ig} = 1]$$

Then by the law of iterated expectations  $p_d(v) \times \boldsymbol{\tau}(v)$  equals

$$\mathbb{P}(D_{ig} = 1 \mid V_{ig} = v) \times \mathbb{E}[\tau_{ig} \mid V_{ig} = v, D_{ig} = 1] = \mathbb{E}[D_{ig}\tau_{ig} \mid V_{ig} = v]$$
(B.15)

Therefore (B.15) produces a simplified expression for the conditional expectation in (B.14). Applying the law of iterated expectations and substituting the expression in (B.15),

$$\mathbb{E}\left[p_d(V_{ig}) \times \mathbf{Q}_{xx}(V_{ig})^{-1} X_{ig} Y_{ig}\right] = \mathbb{E}\left[\mathbb{E}\left[D_{ig} \tau_{ig} \mid V_{ig}\right]\right] = \mathbb{E}\left[D_{ig} \tau_{ig}\right]$$

By Bayes's rule and the fact that  $D_{ig}$  is binary,

$$\frac{\mathbb{E}[D_{ig}\tau_{ig}]}{\mathbb{P}(D_{ig}=1)} = \mathbb{E}[\tau_{ig} \mid D_{ig}=1] = \boldsymbol{\tau}_{APT}$$
(B.16)

The unconditional effect, the APT and APU effects are mutually constrained by the law of iterated expectations, which implies that  $\boldsymbol{\tau} = \mathbb{P}(D_{ig} = 1)\boldsymbol{\tau}_{APT} + \mathbb{P}(D_{ig} = 0)\boldsymbol{\tau}_{APU}$ . Lemma 1 implies that  $\boldsymbol{\tau} = \mathbb{E}[\mathbf{Q}_{xx}(V_{ig})^{-1}X_{ig}Y_{ig}]$ . Therefore, we can solve for the APU effect by substituting the expressions for  $(\boldsymbol{\tau}, \boldsymbol{\tau}_{APT})$  and solving for  $\boldsymbol{\tau}_{APU}$ ,

$$\boldsymbol{\tau}_{APU} = \frac{1}{1 - \mathbb{E}[D_{ig}]} \times \mathbb{E}\left[ (1 - p_d(V_{ig})) \times \mathbf{Q}_{xx}(V_{ig})^{-1} X_{ig} Y_{ig} \right]$$

### **B.4** Supporting Lemmas

**Lemma B.1** (Properties of Conditional Independence). Let X, Y, Z, W be random vectors defined on a common probability space, and let h be a measurable function. Then:

- (i) (Symmetry):  $X \perp Y \mid Z \implies Y \perp X \mid Z$ .
- (ii) (Redundancy):  $X \perp Y | Y$ .
- (iii) (Decomposition):  $X \perp Y \mid Z$  and  $W = h(Y) \implies X \perp W \mid Z$ .
- (iv) (Weak Union):  $X \perp Y | Z$  and  $W = h(Y) \implies X \perp Y | (W, Z)$ .

**Proof.** Constantinou et al. (2017)

**Lemma B.2** (Combining Events). Let  $E, E^*, U, U^*, \Psi$  be random variables on a common probability space. Suppose that (i)  $E \perp E^* \mid \Psi$ , (ii)  $(E, E^*) \perp U^* \mid \Psi$  and (iii)  $U \perp (U^*, E, E^*) \mid \Psi$ . Then

$$(E,U) \perp (E^*,U^*) \mid \Psi$$

**Lemma B.3** (Egocentric Likelihood Factorization). Suppose that  $D_{ig}$  is  $(C_{ig}, \Psi_g^*, \eta_{ig})$ -measurable and  $A_{ijg}$  is  $(C_{ig}, C_{jg}, \Psi_g^*, U_{ijg})$ -measurable. If Random Sampling and Dyadic Network hold, then for  $V_{ig} \equiv (C_{ig}, \Psi_g^*)$ 

$$\mathbb{P}(D_g, A_{ig} \mid V_{ig}) = \mathbb{P}(D_{ig} \mid V_{ig}) \prod_{j \neq i}^{N_g} \mathbb{P}(D_{jg}, A_{ijg} \mid V_{ig})$$

**Lemma B.4** (Bounds Quotients). Let a, b be non-zero scalars and suppose that  $||b|| \ge \underline{b} > 0$ . Then

$$||a^{-1} - b^{-2}|| \le \frac{\underline{b}^{-2}||b - a||}{1 - \underline{b}^{-1}||b - a||}.$$

**Lemma B.5** (Derivative of Inverse Matrix). Let  $v \in \mathbb{R}$  and suppose that Q(v) is differentiable and full rank in an open set around  $v_0$ . Then  $\frac{\partial}{\partial v}Q^{-1}(v_0) = -Q^{-1}(v_0)\frac{\partial Q(v_0)}{\partial v}Q^{-1}(v_0)$ .

**Lemma B.6** (Uniform Bounds Criterion Derivatives). Let  $\lambda_{min}(v_1, v_2, \boldsymbol{\theta})$  denote the smallest eigenvalue of  $\mathbf{Q}_{xx}((v_1, v_2), \boldsymbol{\theta})$  and let  $B(\boldsymbol{\theta}_0, \delta)$  be a ball or radius  $\delta > 0$  around  $\boldsymbol{\theta}_0$ . Let

 $\underline{\lambda}(V_{1igt}^{0},\boldsymbol{\theta}_{0t},\delta) \equiv \inf_{\boldsymbol{\theta}\in B(\boldsymbol{\theta}_{0t},\delta)} \inf_{v_{2}\in\mathcal{V}_{2}} \lambda_{min}(V_{1igt}^{0},v_{2},\boldsymbol{\theta}) \text{ be a lower bound on the eigenvalues of } \mathbf{Q}_{xx}$ for parameters in that set. Furthermore, define

$$\mathcal{R}_{ig}^{A} \equiv \sup_{\theta \in \Theta} \sup_{v_{2} \in \mathcal{V}_{2}} \left\| \frac{\partial}{\partial v_{2}} \mathcal{R}(Z_{ig}, v_{2}, \theta) \right\|$$
(B.17)

$$S_{ig} \equiv \sup_{\theta \in \Theta} \sup_{v_2 \in \mathcal{V}_2} \| s(Z_{ig}, v_2, \theta) \|$$
(B.18)

$$S_{ig}^{A} \equiv \sup_{\theta \in \Theta} \sup_{v_{2} \in \mathcal{V}_{2}} \left\| \frac{\partial}{\partial v_{2}} s(Z_{ig}, v_{2}, \theta) \right\|$$
(B.19)

$$S_{ig}^{A\theta} \equiv \sup_{\theta \in \Theta} \sup_{v_2 \in \mathcal{V}_2} \left\| \frac{\partial^2}{\partial v_2 \partial \theta} s(Z_{ig}, v_2, \theta) \right\|$$
(B.20)

$$\psi_{IW,ig}^{\partial} \equiv \sup_{\theta \in B(\theta_0,\nu)} \sup_{v_2 \in \mathcal{V}_2} \sup_{0 \leq \alpha_1 + \alpha_2 \leq 2} \left\| \frac{\partial^{\alpha_1 + \alpha_2}}{\partial v_2^{\alpha_1} \partial \theta^{\alpha_2}} \psi_{IW}(Z_{ig}, v_2, \boldsymbol{\tau}, \boldsymbol{\theta}) \right\|$$
(B.21)

then the following statements hold

- (i) If  $\mathbb{E}[||X_{ig}||^4] < \infty$  and  $\mathbb{E}[\sup_{\theta \in \Theta} \sup_{v_2 \in \mathcal{V}_2} (Q_{xx}^{\partial}(V_{1ig}, v_2, \theta))^2]$  is bounded, then  $\mathbb{E}[\mathcal{R}_{ig}^A] \mathbb{E}[S_{ig}]$ ,  $\mathbb{E}[S_{ig}^A]$  and  $\mathbb{E}[S_{ig}^{A\theta}]$  are bounded.
- (ii) Suppose that in addition  $\mathbb{E}[||Y_{ig}||^2] < \infty$ ,  $\mathbb{E}[\sup_{\theta \in \Theta} \sup_{v_2 \in \mathcal{V}_2} (Q_{xx}^{\partial}(V_{1ig}, v_2, \theta))^4] < \infty$  and  $\underline{\lambda} > 0$  almost surely. Then  $\mathbb{E}[\psi_{IW, ig}^{\partial}]$  is also bounded.

**Lemma B.7** (Stochastically Bounded Averages). Let  $X_{igt}$  be a sequence of random variable such that  $\mathbb{E}[||X_{igt}||] < \infty$ ,  $\bar{X}_t = \frac{1}{GN} \sum_{ig} X_{igt}$  the sample average and  $N_t = \mathbb{E}[N_{gt}]$  be the expected group size, and  $(G_t, N_t) \to \infty$  as  $t \to \infty$ . Suppose that the groups are randomly sampled with equal weight from a superpopulation and that Bounded Group Ratios holds, then  $\mathbb{E}_*[\frac{\bar{n}_t}{N_t}\bar{X}_t] =$  $\mathbb{E}[X_{igt}]$  and  $\bar{X}_t = O_p(1)$  as  $(G, N_*) \to \infty$ , where  $\mathbb{E}_*$  is the sampling (equal-group-weight) measure and  $\mathbb{E}$  is the population measure. Furthermore, if Random Sampling holds then  $\bar{X}_t \to^p$  $\mathbb{E}_t[X_{ig}]$  and  $\frac{\bar{n}}{N_t} \to^p 1$ .

**Lemma B.8** (Uniform Consistency with Generated Regressors). Let f be a measurable function of  $(z, v_2, \tau, \theta)$  that is continuously differentially with respect to  $(v_2, \tau, \theta)$ . Suppose that

$$(i) \quad \max_{igt} \|V_{2igt} - V_{2igt}^0\| \to_p 0$$

(*ii*) 
$$\mathbb{E}\left[\sup_{(\boldsymbol{\tau},\boldsymbol{\theta},v_2)\in\mathcal{T}\times\Theta\times\mathcal{V}_2}\|f(Z_{igt},v_2,\boldsymbol{\tau},\boldsymbol{\theta})\|\right] < \infty.$$

(*iii*) 
$$\mathbb{E}\left[\sup_{(\boldsymbol{\tau},\boldsymbol{\theta},v_2)\in\mathcal{T}\times\Theta\times\mathcal{V}_2}\left\|\frac{\partial}{\partial(v_2,\boldsymbol{\tau},\boldsymbol{\theta})}f(Z_{igt},v_2,\boldsymbol{\tau},\boldsymbol{\theta})\right\|\right] < \infty.$$

If Random Sampling holds, and  $(G_t, N_t) \to \infty$  as  $t \to \infty$ , then

$$\sup_{\boldsymbol{\tau}\in\mathcal{T}}\sup_{\boldsymbol{\theta}\in\Theta}\left\|\frac{1}{G_t\bar{n}_t}\sum_{ig}f(Z_{igt},V_{2igt},\boldsymbol{\tau},\boldsymbol{\theta})-\mathbb{E}[f(Z_{igt},V_{2igt}^0,\boldsymbol{\tau},\boldsymbol{\theta})]\right\|\to^p 0$$

## **B.5** Proof Supporting Lemmas

**Proof of Lemma B.2 (Combining Events).** By property (ii), weak union and decomposition

$$(ii) \implies (E, E^*) \perp U^* \mid E^*, \Psi \implies E \perp U^* \mid E^*, \Psi$$
(B.22)

By property (i), (B.22) and contraction,  $E \perp (E^*, U^*) \mid \Psi$ .

Similarly, by property (iii), weak union and decomposition.

$$(iii) \implies U \perp (U^*, E, E^*) \mid E, \Psi \implies U \perp (E^*, U^*) \mid E, \Psi$$
(B.23)

Combining the two results via the contraction property,  $(E, U) \perp (E^*, U^*) \mid \Psi$ .

Proof of Lemma B.3 (Egocentric Likelihood Factorization). Let  $V_{ig} \equiv (C_{ig}, \Psi_g^*)$ . By Bayes' rule:

$$\mathbb{P}(D_g, A_{ig} \mid V_{ig}) = \prod_{j=1}^n \mathbb{P}(D_{jg}, A_{ijg} \mid \{D_{kg}, A_{ikg}\}_{k=1}^{j-1}, V_{ig})$$
(B.24)

We can factor the joint probability in any order, so I set i = 1 without loss of generality. By definition  $G_{iig} = 0$  (no self-loops in the network), so  $\mathbb{P}(G_{iig} = 0 \mid V_{ig}) = 1$  and we can denote the probability as  $\mathbb{P}(D_{ig} \mid A_{ig}, V_{ig}) = \mathbb{P}(D_{ig} \mid V_{ig})$  without loss of generality.

For j > 1, define the random variables  $E \equiv (\eta_{jg}, C_{jg})$  and  $E^* \equiv \{(\eta_{kg}, C_{kg})\}_{k=1}^j$ , which denote the personal covariates of j and a vector of the covariates of agents 1 through (j-1), respectively. Similarly, let  $U \equiv U_{ijg}$  and  $U^* \equiv \{U_{ikg}\}_{k=1}^{(j-1)}$ , denote the respective link formation shocks. Random Sampling.(i) allows us to ignore covariates across groups. Random Sampling.(ii) states that the covariates of different agents are conditionally independent, which implies  $E \perp E^* \mid \Psi_g$ . Dyadic Network says that the personal covariates are conditionally independent of the link shocks, which implies that  $(E, E^*) \perp U^* \mid \Psi_g$ . Furthermore, the links are also mutually conditionally independent of each other, which means that  $U \perp (U^*, E, E^*) \mid \Psi_g$ .

Consequently  $(E, E^*, U, U^*, \Psi_g)$  meet the conditions of Lemma B.2 and

$$(\eta_{jg}, C_{jg}, U_{ijg}) \perp \{\eta_{kg}, C_{kg}\}_{k=1}^{j-1}, \{U_{ijg}\}_{k=2}^{j-1} \mid \Psi_g$$
(B.25)

The right-hand side of (B.25) contains enough information to compute  $D_{kg} = h(C_{kg}, \Psi_g^*, \eta_{kg})$ and  $A_{ijg} = L(C_{ig}, C_{jg}, \Psi_g^*, U_{ijg})$ . Therefore, we can use the decomposition property to show that

$$(U_{ijg}, \eta_{jg}, C_{jg}) \perp \{D_{jg}, A_{ijg}\}_{k=1}^{j-1}, C_{ig} \mid \Psi_g^*$$
 (B.26)

By combining (B.26), weak union and decomposition,

$$(U_{ijg}, \eta_{jg}, C_{jg}) \perp \{D_{jg}, A_{ikg}\}_{k=1}^{j-1} \mid C_{ig}, \Psi_g^*$$
(B.27)

Finally  $D_{jg}$  is  $(C_{jg}, \Psi_g^*, \eta_{jg})$ -measurable and  $A_{ijg}$  is  $(C_{ig}, C_{jg}, \Psi_g^*, U_{ijg})$ -measurable. We can use the redundancy property to incorporate the variables in the conditioning set and then apply the decomposition property to show that

$$(D_{jg}, A_{ikg}) \perp \{D_{kg}, A_{ikg}\}_{k=1}^{j-1} \mid C_{ig}, \Psi_g^*$$

By applying this argument recursively, we can show that potential link and participation decisions are conditionally independent. By (B.24)

$$\mathbb{P}(D_g, A_{ig} \mid V_{ig}) = \mathbb{P}(D_{ig} \mid V_{ig}) \prod_{j \neq i}^{N_g} \mathbb{P}(D_{jg}, A_{ijg} \mid V_{ig})$$

**Proof of Lemma B.4 (Bounds Quotients).** By finding a common denominator,  $a^{-1}-b^{-1} = b^{-1}(b-a)a^{-1}$ . By the triangle inequality

$$\begin{aligned} |a^{-1} - b^{-1}|| &= ||b^{-1}|| ||b - a|| ||a^{-1}|| \\ &\leq ||b^{-1}|| ||b - a|| (||b^{-1}|| + ||a^{-1} - b^{-1}||) \\ &\leq \underline{b}^{-1} ||b - a|| (\underline{b}^{-1} + ||a^{-1} - b^{-1}||). \end{aligned}$$

Solving for  $||a^{-1} - b^{-1}||$ ,

$$\|a^{-1} - b^{-1}\| \leq \frac{\underline{b}^{-2} \|b - a\|}{1 - \underline{b}^{-1} \|b - a\|}.$$

**Proof of Lemma B.5 (Derivative of Inverse Matrix).** Let  $M(v) \equiv Q^{-1}(v)$  and define  $F(v) \equiv Q(v)M(v) - I$ . By construction  $F(v) \equiv \mathbf{0}$  uniformly for v in open set around  $v_0$ . Let  $F_{i\ell}$  denote the entry in the  $i^{th}$  and the  $\ell^{th}$  column of F, which can be decomposed as

$$F_{i\ell}(v) = \sum_{k\ell} Q_{ij} M_{k\ell} - a_{i\ell} = 0$$

where  $a_{i\ell}$  are the entries of the identity matrix I. We can differentiate each component by the

scalar v. By the product rule,

$$\frac{\partial F_{i\ell}(v)}{\partial v} = \sum_{k\ell} \frac{\partial Q_{ij}}{\partial v} M_{k\ell} + Q_{ij} \frac{\partial M_{k\ell}}{\partial v} = 0$$

Define  $\frac{\partial F}{\partial v}$  denote the matrices with entries  $\frac{\partial F}{\partial v}$ . Define  $\frac{\partial Q}{\partial v}$ ,  $\frac{\partial M}{\partial v}$  analogously. Then

$$\frac{\partial F(v_0)}{\partial v} = \frac{\partial Q(v_0)}{\partial v} M(v_0) + Q(v_0) \frac{\partial M(v_0)}{\partial v} = 0$$

Solving the equation,  $\frac{\partial M(v_0)}{\partial v} = -Q(v_0)^{-1} \frac{\partial Q(v_0)}{\partial v} M(v_0)$  and substituting the definition of M,

$$\frac{\partial}{\partial v}Q^{-1}(v_0) = -Q^{-1}(v_0)\frac{\partial Q(v_0)}{\partial v}Q^{-1}(v_0).$$

**Proof of Lemma B.6 (Uniform Bounds Criterion Derivatives).** The first task is to express the derivatives of  $\mathcal{R}(\cdot)$  and  $s(\cdot)$  in terms of  $X_{ig}$  and the weighting matrix  $\mathbf{Q}_{xx}(\cdot)$ . It will be convenient to work with the vectorized version of the weighting matrix, which I denote by  $\mathbf{q}(Z_{ig}, v_2, \boldsymbol{\theta})$ . Similarly, define  $\mathbf{x}_{ig} = vec(X_{ig}X'_{ig})$ . In matrix form the criterion can be expressed as

$$\mathcal{R}(Z_{ig}, v_2, \boldsymbol{\theta}) = (\mathbf{x} - \mathbf{q})'(\mathbf{x} - \mathbf{q})$$

Since the score function is defined as the jacobian of  $\mathcal{R}$ , then  $s(Z_{ig}, v_2, \theta) = -2(\mathbf{x} - \mathbf{q})' \frac{\partial \mathbf{q}}{\partial \theta'}$ . We can compute the following derivatives by applying the chain rule. Let  $(\mathbf{x}_k, \mathbf{q}_k)$  denote the  $k^{th}$  rows of  $(\mathbf{x}, \mathbf{q})$ , respectively. Then

$$\frac{\partial}{\partial v_2} \mathcal{R}(Z_{ig}, v_2, \boldsymbol{\theta}) = -2(\mathbf{x} - \mathbf{q})' \frac{\partial \mathbf{q}}{\partial v_2}$$

$$\frac{\partial}{\partial \boldsymbol{\theta}} s(Z_{ig}, v_2, \boldsymbol{\theta}) = 2 \frac{\partial \mathbf{q}}{\partial \boldsymbol{\theta}} \frac{\partial \mathbf{q}}{\partial \boldsymbol{\theta}'} - 2 \sum_k (\mathbf{x}_k - \mathbf{q}_k) \frac{\partial^2 \mathbf{q}_k}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}$$

$$\frac{\partial}{\partial v_2} s(Z_{ig}, v_2, \boldsymbol{\theta}) = 2 \times \frac{\partial \mathbf{q}'}{\partial v_2} \frac{\partial \mathbf{q}}{\partial \boldsymbol{\theta}'} - 2 \sum_k (\mathbf{x}_k - \mathbf{q}_k) \frac{\partial^2 \mathbf{q}_k}{\partial v_2 \partial \boldsymbol{\theta}'}$$

$$\frac{\partial}{\partial^2 A \partial \boldsymbol{\theta}} s(Z_{ig}, v_2, \boldsymbol{\theta}) = 2 \times \left( \frac{\partial^2 \mathbf{q}'}{\partial v_2^2} \frac{\partial \mathbf{q}}{\partial \boldsymbol{\theta}'} + \frac{\partial \mathbf{q}'}{\partial v_2} \frac{\partial^2 \mathbf{q}}{\partial v_2 \partial \boldsymbol{\theta}'} \right) - 2 \sum_k \left[ -\frac{\partial \mathbf{q}_k}{\partial v_2} \frac{\partial^2 \mathbf{q}_k}{\partial v_2 \partial \boldsymbol{\theta}'} + (\mathbf{x}_k - \mathbf{q}_k) \frac{\partial^2 \mathbf{q}_k}{\partial v_2^2 \partial \boldsymbol{\theta}'} \right]$$

Let  $\mathbf{Q}_{xx}^{\partial}(Z_{ig}, v_2, \boldsymbol{\theta})$  denote the Sobolev norm, as defined in (A.3), which is a bound on the derivatives of order  $\{0, 1, 2, 3\}$ . Similarly, let  $\|\mathbf{x}\|$  denote the Euclidean norm of  $\mathbf{x}$ . It is useful to use the fact that  $\sum_k \|\mathbf{x}_k\| \leq \kappa \|\mathbf{x}\|$ , for some universal constant  $\kappa$  that only depends on the dimension. We denote this inequality as  $\sum_k \|\mathbf{x}_k\| \leq \|\mathbf{x}\|$ .

By the triangle inequality,

$$\begin{aligned} \left\| \frac{\partial}{\partial v_2} \mathcal{R}(Z_{ig}, v_2, \boldsymbol{\theta}) \right\| &\leq 2( \|\mathbf{x}_{ig}\| + \|\mathbf{q}\|) \left\| \frac{\partial \mathbf{q}}{\partial v_2} \right\| \leq 2 \|\mathbf{x}_{ig}\| \mathbf{Q}_{xx}^{\partial}(Z_{ig}, v_2, \boldsymbol{\theta}) + 2 \mathbf{Q}_{xx}^{\partial}(Z_{ig}, v_2, \boldsymbol{\theta})^2 \\ &\left\| \frac{\partial}{\partial \boldsymbol{\theta}} s(Z_{ig}, v_2, \boldsymbol{\theta}) \right\| \leq 2 \left\| \frac{\partial \mathbf{q}}{\partial \boldsymbol{\theta}} \right\| \left\| \frac{\partial \mathbf{q}}{\partial \boldsymbol{\theta}'} \right\| + 2 \sum_k (\|\mathbf{x}_k\| - \|\mathbf{q}_k\|) \left\| \frac{\partial^2 \mathbf{q}_k}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right\| \\ &\lesssim 4 \mathbf{Q}_{xx}^{\partial}(Z_{ig}, v_2, \boldsymbol{\theta})^2 + 2 \|\mathbf{x}_{ig}\| \mathbf{Q}_{xx}^{\partial}(Z_{ig}, v_2, \boldsymbol{\theta}) \\ &\left\| \frac{\partial}{\partial v_2} s(Z_{ig}, v_2, \boldsymbol{\theta}) \right\| \leq 2 \left\| \frac{\partial \mathbf{q}}{\partial v_2} \right\| \left\| \frac{\partial \mathbf{q}}{\partial \boldsymbol{\theta}'} \right\| + 2 \sum_k (\|\mathbf{x}_k\| - \|\mathbf{q}_k\|) \left\| \frac{\partial^2 \mathbf{q}_k}{\partial v_2 \partial \boldsymbol{\theta}'} \right\| \\ &\lesssim 4 \mathbf{Q}_{xx}^{\partial}(Z_{ig}, v_2, \boldsymbol{\theta})^2 + 2 \|\mathbf{x}_{ig}\| \mathbf{Q}_{xx}^{\partial}(Z_{ig}, v_2, \boldsymbol{\theta}) \end{aligned}$$
(B.28)

At each step we bound the derivatives by the Sobolev-norm and Euclidean norms, respectively. By using a similar procedure we can show that

$$\left\| \frac{\partial}{\partial^2 A \partial \boldsymbol{\theta}} s(Z_{ig}, v_2, \boldsymbol{\theta}) \right\| \lesssim 8 \mathbf{Q}_{xx}^{\partial} (V_{1ig}^0, v_2, \boldsymbol{\theta})^2 + 2 \| \mathbf{x}_{ig} \| \mathbf{Q}_{xx}^{\partial} (V_{1ig}^0, v_2, \boldsymbol{\theta})$$
(B.29)

Our next task is to derive a uniform bounds for the expectations of the derivatives. All of the derivatives in (B.28) and (B.29) are bounded uniformly by combinations of  $\|\mathbf{x}_{ig}\|$  and  $\mathbf{Q}_{xx}^{\delta}(\cdot)$ . By assumption  $\mathbb{E}[\sup_{\theta \in \Theta} \sup_{v_2 \in \mathcal{V}_2} (Q_{xx}^{\delta}(V_{1ig}^0, v_2, \theta))^2] < \infty$ , which allows us to bound some of the terms directly. To bound the rest of the terms we use the Cauchy-Schwartz inequality,

$$\mathbb{E}\left[\sup_{\theta\in\Theta}\sup_{v_{2}\in\mathcal{V}_{2}}\mathbf{Q}_{xx}^{\partial}(V_{1ig}^{0},v_{2},\boldsymbol{\theta})\right] \leqslant \sqrt{\mathbb{E}\left[\sup_{\theta\in\Theta}\sup_{v_{2}\in\mathcal{V}_{2}}\left(Q_{xx}^{\partial}(V_{1ig}^{0},v_{2},\theta)\right)^{2}\right]} < \infty.$$

$$\mathbb{E}\left[\sup_{\theta\in\Theta}\sup_{v_{2}\in\mathcal{V}_{2}}\|\mathbf{x}_{ig}\|\mathbf{Q}_{xx}^{\partial}(V_{1ig}^{0},v_{2},\boldsymbol{\theta})\right] \leqslant \sqrt{\mathbb{E}\left[\|\mathbf{x}_{ig}\|^{2}\right] \times \mathbb{E}\left[\sup_{\theta\in\Theta}\sup_{v_{2}\in\mathcal{V}_{2}}\left(Q_{xx}^{\partial}(V_{1ig}^{0},v_{2},\boldsymbol{\theta})\right)^{2}\right]} < \infty.$$

Recall that  $\mathbf{x}_{ig} = vec(X_{ig}X'_{ig})$  (the product of  $X_{ig}$ ) which means that  $\mathbb{E}[\|\mathbf{x}_{ig}\|^2] \leq \mathbb{E}[\|X_{ig}\|^4]$ , which is finite by assumption.

Define  $(\mathcal{R}_{ig}^V, S_{ig}, S_{ig}^V, S_{ig}^{V\theta})$  as in the statement of the Lemma. By (B.28) and (B.29)

$$\mathcal{R}_{ig}^{V}, S_{ig}, S_{ig}^{V}, S_{ig}^{V\theta} \lesssim 8 \sup_{\theta \in \Theta} \sup_{v_2 \in \mathcal{V}_2} \mathbf{Q}_{xx}^{\vartheta} (V_{1ig}^0, v_2, \boldsymbol{\theta})^2 + 2 \sup_{\theta \in \Theta} \sup_{v_2 \in \mathcal{V}_2} (1 + \|\mathbf{x}_{ig}\|) \mathbf{Q}_{xx}^{\vartheta} (V_{1ig}^0, v_2, \boldsymbol{\theta}).$$

The expectations of the right-hand side is bounded and therefore  $\mathbb{E}[\mathcal{R}_{ig}^V], \mathbb{E}[S_{ig}], \mathbb{E}[S_{ig}^V], \mathbb{E}[S_{ig}^V] < \infty$ .

Now we turn our attention to the derivatives of the influence function

$$\psi_{IW}(Z_{igt}, v_2, \boldsymbol{\beta}, \boldsymbol{\theta}) = \mathbf{Q}_{xx}(Z_{igt}, v_2, \boldsymbol{\theta})^{-1} X_{ig} Y_{ig}$$

By Lemma B.5 we can compute the derivatives of the inverse.

$$\frac{\partial \psi}{\partial v_2} = -\mathbf{Q}_{xx}^{-1} \frac{\partial \mathbf{Q}_{xx}}{\partial v_2} \mathbf{Q}_{xx}^{-1} X_{ig} Y_{ig} \qquad \frac{\partial \psi}{\partial \theta_m} = -\mathbf{Q}_{xx}^{-1} \frac{\partial \mathbf{Q}_{xx}}{\partial v_2} \mathbf{Q}_{xx}^{-1} X_{ig} Y_{ig}$$

Similarly, by applying the product rule and grouping terms

$$\frac{\partial^2 \psi}{\partial v_2 \partial \theta} = \begin{bmatrix} 2 \ \mathbf{Q}_{xx}^{-1} \frac{\partial \mathbf{Q}_{xx}}{\partial v_2} \mathbf{Q}_{xx}^{-1} \frac{\partial \mathbf{Q}_{xx}}{\partial \theta_j} \mathbf{Q}_{xx}^{-1} + \mathbf{Q}_{xx}^{-1} \frac{\partial^2 \mathbf{Q}_{xx}}{\partial v_2 \partial \theta_j} \mathbf{Q}_{xx}^{-1} \end{bmatrix} X_{ig} Y_{ig}$$
$$\frac{\partial^2 \psi}{\partial \theta_m \partial \theta_j} = \begin{bmatrix} 2 \ \mathbf{Q}_{xx}^{-1} \frac{\partial \mathbf{Q}_{xx}}{\partial \theta_m} \mathbf{Q}_{xx}^{-1} \frac{\partial \mathbf{Q}_{xx}}{\partial \theta_j} \mathbf{Q}_{xx}^{-1} + \mathbf{Q}_{xx}^{-1} \frac{\partial^2 \mathbf{Q}_{xx}}{\partial \theta_m \partial \theta_j} \mathbf{Q}_{xx}^{-1} \end{bmatrix} X_{ig} Y_{ig}$$

By assumption, the smallest eigenvalue of  $\mathbf{Q}_{xx}$  is bounded below by  $\underline{\lambda} > \theta$  for  $\boldsymbol{\theta} \in B(\boldsymbol{\theta}_0, \delta)$ . For parameter values in this set,  $\|\mathbf{Q}_{xx}^{-1}\| \leq \underline{\lambda}^{-1}$  by Lemma X and

$$\left\| \frac{\partial \psi}{\partial v_2} \right\| \leq \left\| \mathbf{Q}_{xx}^{-1} \right\| \left\| \frac{\partial \mathbf{Q}_{xx}}{\partial v_2} \right\| \left\| \mathbf{Q}_{xx}^{-1} \right\| \left\| X \right\| \left\| Y \right\| \leq \underline{\lambda}^{-2} \mathbf{Q}_{xx}^{\partial} \left\| X_{ig} Y_{ig} \right\|$$

$$\left\| \frac{\partial \psi}{\partial \theta_m} \right\| \leq \left\| \mathbf{Q}_{xx}^{-1} \right\| \left\| \frac{\partial \mathbf{Q}_{xx}}{\partial \theta_m} \right\| \left\| \mathbf{Q}_{xx}^{-1} \right\| \left\| X \right\| \left\| Y \right\| \leq \underline{\lambda}^{-2} \mathbf{Q}_{xx}^{\partial} \left\| X_{ig} Y_{ig} \right\|$$

By bounding the respective terms, we can also show that

$$\left\|\frac{\partial^{2}\psi}{\partial v_{2}\partial\theta_{m}}\right\| \leq \left(2\underline{\lambda}^{-3}(\mathbf{Q}_{xx}^{\partial})^{2} + \underline{\lambda}^{-2}\mathbf{Q}_{xx}^{\partial}\right)\|X_{ig}Y_{ig}\|$$
$$\left\|\frac{\partial^{2}\psi}{\partial\theta_{m}\partial\theta_{j}}\right\| = \left(2\underline{\lambda}^{-3}(\mathbf{Q}_{xx}^{\partial})^{2} + \underline{\lambda}^{-2}\mathbf{Q}_{xx}^{\partial}\right)\|X_{ig}Y_{ig}\|$$

By the Cauchy Schwartz inequality,  $\mathbb{E}[\|X_{ig}Y_{ig}\|] < \sqrt{\mathbb{E}[\|X_{ig}\|^2]\mathbb{E}[\|Y_{ig}\|^2]}$ , which is bounded by assumption.

By applying the Cauchy-Schwartz inequality a second time,

$$\mathbb{E}\left[\sup_{\theta\in\Theta}\sup_{v_{2}\in\mathcal{V}_{2}}\mathbf{Q}_{xx}^{\partial}(V_{1ig}^{0},v_{2},\boldsymbol{\theta})\|X_{ig}Y_{ig}\|\right] \leqslant \sqrt{\mathbb{E}\left[\sup_{\theta\in\Theta}\sup_{v_{2}\in\mathcal{V}_{2}}\left(Q_{xx}^{\partial}(V_{1ig}^{0},v_{2},\theta)\right)^{2}\right]\mathbb{E}\left[\|X_{ig}Y_{ig}\|\right]} < \infty.$$
$$\mathbb{E}\left[\sup_{\theta\in\Theta}\sup_{v_{2}\in\mathcal{V}_{2}}\mathbf{Q}_{xx}^{\partial}(V_{1ig}^{0},v_{2},\boldsymbol{\theta})^{2}\|X_{ig}Y_{ig}\|\right] \leqslant \sqrt{\mathbb{E}\left[\sup_{\theta\in\Theta}\sup_{v_{2}\in\mathcal{V}_{2}}\left(Q_{xx}^{\partial}(V_{1ig}^{0},v_{2},\theta)\right)^{4}\right]\mathbb{E}\left[\|X_{ig}Y_{ig}\|\right]} < \infty.$$

The fourth moment of the Sobolev norm is bounded by assumption. Consequently,  $\mathbb{E}[\psi_{IW,ig}^{\partial}]$  as defined in (B.21) is bounded.

**Proof of Lemma B.7 (Stochastically Bounded Averages).** We start by writing  $\bar{X}_t$  in terms of within-group averages  $\bar{X}_g$ .

$$\bar{X}_{t} = \frac{1}{G} \sum_{g=1}^{G} \frac{N_{gt}}{\bar{n}_{t}} \left( \frac{1}{N_{gt}} \sum_{i=1}^{N_{gt}} X_{igt} \right) \equiv \frac{1}{G} \sum_{g=1}^{G} \frac{N_{gt}}{\bar{n}_{t}} (\bar{X}_{gt})$$

Determining the properties of the average is slightly complicated by the fact that  $N_{gt}$  is a random variable, which means that  $\bar{X}_{gt}$  is an average with a random number of terms.

Let  $\mathbb{E}_*$  be a measure where groups are given equal weight regardless of their size, which satisfies two properties: (i)  $\mathbb{E}_*[\rho_{gt}X_{igt}] = \mathbb{E}[X_{igt}]$  and (ii)  $\mathbb{E}_*[X_{igt} | N_{gt} = n] = \mathbb{E}[X_{igt} | N_{gt} = n]$ , where  $\rho_{gt} \equiv N_{gt}/N_t$  and  $N_t = \mathbb{E}_*[N_{gt}]$ . Property (i) links the equal-weighted measure to the population measure by including importance weights, whereas property (ii) states the two measures are identical after conditioning on group size.

Our first task is to show that  $\frac{N}{N_t}\bar{X}$  is unbiased. By substituting the definition of  $\rho_{gt}$ ,

$$\frac{N}{N_t}\bar{X}_t = \frac{1}{G}\sum_{g=1}^G \rho_{gt}\bar{X}_{gt} \tag{B.30}$$

Conditional on group size,  $\bar{X}_{gt}$  is an average of a fixed number of terms and hence  $\mathbf{E}[\bar{X}_{gt} | N_{gt} = n] = \mathbb{E}[X_{igt} | N_{gt} = n]$ . Therefore by the law of iterated expectations and distributing the expectation over each group

$$\mathbb{E}_*\left[\frac{N}{N_t}\bar{X}_t\right] = \mathbb{E}_*\left[\rho_{gt}\bar{X}_{gt}\right] = \mathbb{E}_*\left[\rho_{gt}\mathbb{E}_*\left[X_{igt} \mid N_{gt}\right]\right] = \mathbb{E}_*\left[\rho_{gt}X_{igt}\right] = \mathbb{E}\left[X_{igt}\right]$$

Our next task is to show that  $\bar{X}$  is bounded in probability. By the triangle inequality and the law of iterated expectations.

$$\mathbb{E}_*\left[\|\bar{X}\|\right] \leqslant \frac{1}{G} \sum_{g=1}^G \mathbb{E}_*\left[\frac{N_{gt}}{N} \mathbb{E}_*\left[\|\bar{X}_{gt}\| \mid \mathbf{N}\right]\right] \leqslant \frac{1}{G} \sum_{g=1}^G \mathbb{E}_*\left[\frac{N_{gt}}{N} \mathbb{E}_*\left[\|X_{igt}\| \mid \mathbf{N}\right]\right] = \mathbb{E}\left[\frac{N_{gt}}{N} \|X_{igt}\|\right]$$

Assumption Bounded Group Ratios states that the  $N_{gt} \in [\underline{\rho}, \overline{\rho}] \subset (0, 1)$  which means that the sample size average  $\overline{n}_t \in [\rho, \overline{\rho}]$ . Consequently,

$$\frac{N_{gt}}{\bar{n}_t} = \frac{N_{gt}}{N_t} \times \frac{N_t}{\bar{n}} = \rho_{gt} \times \frac{N_t}{\bar{n}} \leqslant \rho_{gt} \times (1/\underline{\rho})$$

Hence  $\mathbb{E}_{*}[\|\bar{X}_{t}\|] \leq (1/\underline{\rho})\mathbb{E}_{*}[\rho_{gt}\|X_{igt}\|] = 1/\underline{\rho}\mathbb{E}[\|X_{igt}\|]$ , which is bounded. Then by Markov's inequality, for fixed  $\delta > 0$ ,

$$\mathbb{P}(\|\bar{X}_t\| > \delta) \leqslant \frac{\mathbb{E}[\|X_t\|]}{\delta}$$

Therefore  $\bar{X}_t = O_p(1)$ .

Finally, under Random Sampling the observations in each group are independent. Since the (B.30) is an average of i.i.d variables with finite moments, then we can apply the strong law of large numbers in (Billingsley, 1995, p.282), to show that  $\frac{N}{N_t} \bar{X}_t \rightarrow^p \mathbb{E}[X_{igt}]$ . As a special case,  $\frac{N}{N_t} \rightarrow^p 1$ . By combining the two results, we find that  $\bar{X}_t \rightarrow^p \mathbb{E}[X_{igt}]$ .

Proof of Lemma B.8 (Uniform Consistency with Generated Regressors). I start by proving point-wise convergence of the criterion function. For simplicity define  $\Delta_{igt} \equiv ||V_{2igt} - V_{2igt}^0||$ . By a first-order Taylor expansion

$$\hat{f}(\boldsymbol{\tau},\boldsymbol{\theta}) \equiv \frac{1}{G_t \bar{n}_t} \sum_{ig} f(Z_{igt}, V_{2igt}, \theta) = \frac{1}{G_t \bar{n}_t} \sum_{ig} f(Z_{igt}, V_{2igt}^0, \theta) + \frac{1}{G_t \bar{n}_t} \sum_{ig} \frac{\partial}{\partial v_2} f(Z_{igt} \tilde{V}_{2igt}, \theta) \Delta_{igt}$$

I apply the triangle inequality to bound the second term. By the triangle inequality

$$\left\| \frac{1}{G_t \bar{n}_t} \sum_{ig} \frac{\partial}{\partial v_2} f(Z_{igt} \tilde{V}_{2igt}, \boldsymbol{\tau}, \boldsymbol{\theta}) \Delta_{igt} \right\| \leq \left( \frac{1}{G_t \bar{n}_t} \sum_{ig} \left\| \frac{\partial}{\partial v_2} f(Z_{igt} \tilde{V}_{2igt}, \boldsymbol{\theta}) \right\| \right) \max_{ig} \Delta_{igt}$$
$$\leq \left( \frac{1}{G_t \bar{n}_t} \sum_{ig} \sup_{(\boldsymbol{\tau}, \boldsymbol{\theta}, v_2) \in \mathcal{T} \times \Theta \times \mathcal{V}_2} \left\| \frac{\partial}{\partial v_2} f(Z_{igt}, v_2, \boldsymbol{\tau}, \boldsymbol{\theta}) \right\| \right) \max_{ig} \Delta_{igt}$$
$$\equiv \left( \frac{1}{G_t \bar{n}_t} \sum_{ig} f_{igt}^V \right) \max_{ig} \Delta_{igt}$$
(B.31)

The discrepancy  $\max_{ig} \|\Delta_{igt}\|$  is  $o_p(1)$  by assumption (i). Conversely, by assumption (iii)  $\mathbb{E}[f_{igt}^V] < \infty$  and Lemma B.7 imply that  $\frac{1}{G_t \bar{n}_t} \sum_{ig} f_{igt}^V = O_p(1)$ . Finally, by combining the two finding we conclude that the right-hand side of (B.31) is  $o_p(1)$ .

Assumptions (i) implies that  $f(Z_{igt}, V_{2igt}^0, \theta)$  has bounded moments. Similarly, Random Sampling implies that groups are independent. Therefore, we can apply a group level law of large numbers to show that as  $(G, N) \to \infty$ ,

$$\hat{f}(\boldsymbol{\tau},\boldsymbol{\theta}) = \frac{1}{G_t \bar{n}_t} \sum_{ig} f(Z_{igt}, V_{2igt}^0, \boldsymbol{\tau}, \boldsymbol{\theta}) + o_p(1) = \mathbb{E}[f(Z_{igt}, V_{2igt}^0, \boldsymbol{\tau}, \boldsymbol{\theta})] + o_p(1)$$

Our next task is to show that the criterion function is stochastically equicontinuous, in the sense defined by Newey (1991). Let  $(\theta, \theta^*)$  be two distinct parameter values and define a uniform
bound on the derivative  $S_{igt}$  as in Lemma B.6. Then

$$\begin{aligned} \|\widehat{f}(\boldsymbol{\tau},\boldsymbol{\theta}) - \widehat{f}(\boldsymbol{\tau}^{*},\boldsymbol{\theta}^{*})\| &= \left\| \frac{1}{G_{t}\bar{n}_{t}} \sum_{ig} \frac{\partial}{\partial(\boldsymbol{\tau},\boldsymbol{\theta})} f(Z_{igt}, V_{2igt}^{0}, \widetilde{\boldsymbol{\tau}}, \widetilde{\boldsymbol{\theta}}) [(\boldsymbol{\tau},\boldsymbol{\theta})' - (\boldsymbol{\tau}^{*},\boldsymbol{\theta}^{*})'] \right\| \\ &\leq \left( \frac{1}{G_{t}\bar{n}_{t}} \sum_{ig} \sup_{(\boldsymbol{\tau},\boldsymbol{\theta},v_{2})\in\mathcal{T}\times\Theta\times\mathcal{V}_{2}} \left\| \frac{\partial}{\partial(\boldsymbol{\tau},\boldsymbol{\theta})} f(Z_{igt},v_{2},\widetilde{\boldsymbol{\tau}},\widetilde{\boldsymbol{\theta}}) \right\| \right) \|(\boldsymbol{\tau},\boldsymbol{\theta})' - (\boldsymbol{\tau}^{*},\boldsymbol{\theta}^{*})'\| \\ &\leq \left( \frac{1}{G_{t}\bar{n}_{t}} \sum_{ig} f_{igt}^{(\boldsymbol{\tau},\boldsymbol{\theta})} \right) \|(\boldsymbol{\tau},\boldsymbol{\theta})' - (\boldsymbol{\tau}^{*},\boldsymbol{\theta}^{*})'\| \end{aligned} \tag{B.32}$$

By assumption (iv)  $\mathbb{E}[f_{igt}^{(\boldsymbol{\tau},\boldsymbol{\theta})}] < \infty$  and by Lemma B.7,  $\frac{1}{G_t \bar{n}_t} \sum_{ig} f_{igt}^{(\boldsymbol{\tau},\boldsymbol{\theta})} = O_p(1)$ . This exactly fits the definition of stochastic equicontinuity. Since the parameter space is compact, the function converges point-wise and the sample-criterion is stochastically equicontinuous, then by Theorem 2.1 in Newey (1991),

$$\sup_{\boldsymbol{\tau}\in\mathcal{T}}\sup_{\boldsymbol{\theta}\in\Theta}\|\widehat{f}(\boldsymbol{\tau},\boldsymbol{\theta})-\mathbb{E}[f(Z_{igt},V_{2igt}^{0},\boldsymbol{\tau},\boldsymbol{\theta})]\| \to^{p} 0$$
(B.33)

|   |   | п. |
|---|---|----|
|   |   | н  |
|   |   |    |
|   |   | н  |
| _ | _ | J  |