Adjusting for Peer-Influence in Propensity Scoring When Estimating Treatment Effects

Matthew O. Jackson* Zhongjian Lin† Ning Neil Yu‡

Draft: August 2020

ABSTRACT

Analyses of treatments, experiments, policies, and observational data, are confounded when people's treatment outcomes and/or participation decisions are influenced by those of their friends and acquaintances. This invalidates standard matching techniques as estimation tools. For instance, the vaccination decisions of a person's peers affect the person's choice to vaccinate and the probability that the person is exposed to a disease (violating the usual Stable Unit Treatment Value Assumption). We account for these interferences by explicitly modeling peer interaction in treatment participation decisions, and balancing matchings to overcome correlation in outcomes. We incorporate these interaction effects into one of the most common techniques used to evaluate treatment effects: propensity score matching, and provide asymptotic results. We illustrate that peer-influenced propensity score matching gives more accurate results than standard propensity score matching in the estimation of the effectiveness of vaccinations and estimation of the impact of exercise participation on depression.

Keywords: Peer-Influenced Propensity Score Matching (PIPSM); Peer Effects; Influence Network; Vaccination; Exercise; Depression

JEL Classifications: C31; C35; C57; D85; I12

*Department of Economics, Stanford University, and the Santa Fe Institute. Email: jacksonm@stanford.edu. Jackson gratefully acknowledges support under NSF grant SES-1629446.

†Department of Economics, Emory University. Email: zhongjian.lin@emory.edu.

‡Institute for Social and Economic Research, Nanjing Audit University; Freeman Spogli Institute for International Studies, Stanford University. Email: ningyu@stanford.edu. Yu gratefully acknowledges support from the National Natural Science Foundation of China and Jiangsu Special-Term Professorship.

1. INTRODUCTION

Much of social science research, and the design of many policies and programs, depend on causal inference.[1] However, it is rare to have a completely balanced experiment in which the population of people receiving treatment exactly matches the population not receiving treatment along all of the dimensions of characteristics that can affect the outcome.[2] This is especially true in situations in which people's treatment decisions depend on their characteristics. As such, researchers use a variety of matching techniques to account for the lack of balance and recover unbiased estimates of treatment effects. The prominent method of *propensity score matching*, henceforth *PSM*, invented by Rosenbaum and Rubin (1983), balances the populations by estimating the probability that a person with a given set of characteristics would be treated, and then matches based on that probability – which is the coarsest way of matching that balances populations.[3]

Although propensity scores, and other matching methods, balance populations based on individual characteristics, people's treatment decisions can depend on more than just those characteristics. People's awareness of treatments, and their propensity to take them up and follow through, are often influenced by their friends, families, and acquaintances. The list of behaviors that are peer-influenced is long and touches on most forms of human behavior: consumption of drugs and alcohol, diet, criminal behavior, education, product purchases, vaccination, exercise, work habits, hobbies, participation in a study or program, etc.[4]

[1] Reviews include Imbens (2004), Blundell and Dias (2009), Imbens and Wooldridge (2009), DiNardo and Lee (2011), Lee (2016), Abadie and Cattaneo (2018), and Mogstad and Torgovitsky (2018), among others.

[2] Following the literature, we define policies or treatments broadly: different medical options, education programs, occupations, risky behaviors, exercise schemes, etc.; including both observational data and (randomly) controlled trials where individuals have some discretion over their participation in treatment. Even with careful experiments, if there is less than full compliance (not all targeted individuals opt in) or not all those who initially opt in finish the program, modeling treatment choices is critical. For instance, see Heckman and Navarro-Lozano (2004), Heckman and Vytlacil (2005, 2007a,b), and Abbring and Heckman (2007).

[3] It solves the 'curse of dimensionality' of earlier matching techniques that tried to more finely match on particular characteristics, of which there can be too many configurations. A search for 'propensity score matching' on the National Center for Biotechnology Information website on September 16, 2019 returned 17658 research works. For more information, see Hirano, Imbens, and Ridder (2003), Dehejia and Wahba (2002), Caliendo and Kopeinig (2008), Huber, Lechner, and Wunsch (2013), Frölich (2004), Imbens (2015), and Abadie and Imbens (2016).

[4] For a variety of applications see Manski (1993), Glaeser, Sacerdote, and Scheinkman (1996), Hoxby (2000), Brock and Durlauf (2001a), Gaviria and Raphael (2001), Sacerdote (2001), Duflo and Saez (2003), Zimmerman (2003), Arcidiacono and Nicholson (2005), Falk and Ichino (2006), Jackson (2008), Bayer, Hjalmarsson, and Pozen (2009), Bramoullé, Djebbari, and Fortin (2009), Calvó-Armengol, Patacchini, and Zenou (2009), Guryan, Kroft, and Notowidigdo (2009), Bajari, Hong, Krainer, and Nekipelov (2010), Lee, Liu, and Lin (2010), Waldinger (2011), Aral and Walker (2011), Banerjee, Chandrasekhar, Duflo, and Jackson (2013), Goldsmith-Pinkham and Imbens (2013), Bramoullé, Kranton, and D'amours (2014),

This causes two problems. First, propensity scores are no longer dependent solely on an individual's personal characteristics and so may be mis-estimated – especially given that individuals with similar characteristics tend to be friends with each other and then have correlated decisions.[5] Second, when treatment decisions are correlated across people it is also frequently true that friends' and acquaintances' treatments affect a person's treatment outcomes. This is a violation of the usual 'Stable Unit Treatment Value Assumption' (SUTVA), which presumes that an individual's outcomes are independent of other people's participation decisions (Cox, 1958; Rubin, 1974, 1980; Holland, 1986). Thus, peers can be simultaneously influencing both a person's treatment decision and the outcome, which presents two forms of interference for estimating how outcomes depend upon treatment.

As a simple example, consider estimating the effects of exercise on depression. Not only does a person's decision to exercise depend on her friends' exercise habits, but there is also evidence that having more friends affects a person's mental health (e.g., Kawachi and Berkman, 2001; Thoits, 2011; Eisenberger and Cole, 2012). This implies that a person's treatment decision is correlated with things that affect the potential outcomes of the treatment, which precludes the use of standard propensity score methods.

As an even starker example, friends' decisions to vaccinate for an infectious disease not only affect a person's decision to vaccinate, but also affect whether the person is likely to catch a disease. Thus, other people's treatment decisions interfere directly with both a person's treatment decision and treatment outcome.

To overcome these problems, we provide a technique that permits and accounts for network structure and peer effects both in participation decisions and in determining a person's outcomes. We develop the technique with respect to PSM, since that is one of the most prominent techniques (see Caliendo and Kopeinig (2008) for background). However, we note that the prevalent issue that we are pointing out, and our conceptual approach to addressing it, extend readily to other matching techniques, since there are circumstances in which other matching methods outperform propensity scoring (e.g., King and Nielsen (2019)).

In particular, we propose a new technique that we call *peer-influenced propensity*

Dahl, Løken, and Mogstad (2014), Aral and Nicolaides (2017), Booij, Leuven, and Oosterbeek (2017), Bursztyn and Jensen (2017), Feld and Zölitz (2017), and Banerjee et al. (2020). For broader background see the discussion in Durlauf and Ioannides (2010), Dishion and Tipsord (2011), Epple and Romano (2011), Sacerdote (2011, 2014), Jackson and Zenou (2015), Morse (2015), Einav, Farronato, and Levin (2016), Jackson, Rogers, and Zenou (2017), Centola (2018), Jackson (2019), and Bramoullé, Djebbari, and Fortin (2019).

[5]For background references on homophily and peer effects, see Jackson, Rogers, and Zenou (2017), Jackson (2019).

score matching, henceforth *PIPSM*, which accounts for peer interference on a person's outcomes and treatment choice. By controlling for a person's characteristics, network, and the participation decisions of others, any remaining influence on the person's treatment choice is independent noise. This societal conditional independence implies that a person's potential outcomes are independent of their *own* participation decision *conditional upon their own propensity*, which allows us to provide unbiased estimation of their own outcome given their propensity. In particular, we focus on binary treatment choice problems in which the decision is either to participate or not, and we model interdependent participation decisions as a simultaneous game of incomplete information.[6] The participation payoff of an individual is influenced by her peers' participation rate as well as the total number of participating peers (i.e., *local average* and *local aggregate* as coined by Liu, Patacchini, and Zenou (2014), and corresponds to a preference interaction in Manski (2000)). A Bayesian-Nash equilibrium of such a game is characterized by a system of simultaneous nonlinear equations, which determine each individual's probability of participating conditional on the network and all demographics – that we call her *peer-influenced propensity score* (PIPS). We provide sufficient conditions under which the equilibrium is unique. As we show in an appendix, pseudo-likelihood estimators of PIPS are consistent and asymptotically normal.

The second part of the PIPSM estimation involves matching individuals with the same propensity score (or similar scores) in order to estimate what would have happened if the individual would have had the alternative treatment. This is the basic idea behind PSM, here augmented to account for peer influence in treatment choices. In this second step, we have to account for the correlations in outcomes among friends and acquaintances that violate SUTVA and could invalidate the estimation. We prove that if the overall correlation with *all* other individuals with similar propensity scores satisfies some weak bounds in large enough samples, and an appropriately balanced matching is used, then the estimation is consistent and fast (root n). The key idea is that even though an individual's outcome may be correlated with her friends, and possibly up to some distance in the network, that correlation tends to decay with distance. People who have the same propensity but who are far away from each other in the network do not interfere substantially with each other's outcomes. Thus, a matching that has equal weight on individuals with the same propensity score, will end up matching them mostly with people who are not too close in the network (on average in a large network), and so is consistent, as the overall correlation vanishes. To prove this we apply a powerful central limit theorem that applies to general structures of correlation: spatial, network or

[6] This is easily extended to more treatment choices.

otherwise.

We provide results on the identification of PIPSM and its large sample properties, both theoretically and via simulations; and illustrate it in an application to data on depression and exercise. PIPSM eliminates a systematic bias and performs markedly better than standard PSM in all those exercises.

Although we develop the peer-influenced propensity score matching technique by estimating the peer-influenced participation decisions and incorporating those in the derivation of propensity scores, there is another method that can also be used, at least hypothetically. In small networks and settings with relatively few relevant demographic characteristics, one can augment an individual's characteristics with information about the full network and all of the society's characteristics, and then estimate propensity scores directly off of that information. This would be an equally valid approach to solving the difficulty we have identified, since all the things that ultimately determine treatment participation would be conditioned upon. Although we prove that this method would also solve the issues that we are pointing out if it could be practically implemented; it involves deriving propensity scores from very large dimensional information and quickly becomes computationally infeasible, even with relatively small societies. The advantage of PIPSM is that it significantly eases the computational problems and is practical in nontrivial sized societies.

There is, of course, previous literature on social interactions interfering with causal inference. Some of it has discussed problems that arise when the treatment selections of other individuals affect the outcome variable of an individual (Manski, 2013; Perez-Heydrich et al., 2014; Lazzati, 2015). Partial identification approaches have been used for estimation in those cases, using a functional form that incorporates that interference directly, based on some so-called shape restrictions. Instead of using an aggregate function with a shape restriction, we explicitly model the underlying treatment choices with social interactions, and work with a balanced matching approach that overcomes the problem of correlation in outcomes that is generally ignored in the literature. Balat and Han (2020) and Hoshino and Yanagi (2020) also address the problems of interference through interdependent treatment choices. They work with different settings and, most importantly, do not address propensity scoring and matching methods, which is the core of our work.[7] Hoshino and Yanagi (2020) also allow for interference between others'

[7] There are other technical differences, for example, they have fixed numbers of players, while we let the number of players go to infinity in a single large network. They work with games of complete information and impose equilibrium selection criteria to deal with multiple equilibria, while we impose restrictions on model primitives to preclude multiple equilibria in a game of incomplete information. But our propensity scoring techniques could also be used with other models of participation decisions, such as theirs.

treatment choices and a given individual's outcomes, and impose assumptions that allow that interference to be directly identified. We do not impose any assumptions on the interference structure, and get consistent estimation via balanced matchings, which overcome local correlations as we prove below.

There is also a literature on other mechanisms that interfere with treatment effects, such as some sort of correlation in group outcomes or contagion or contamination across individuals (e.g., see Ogburn and VanderWeele, 2014). There is a fruitful set of approaches for dealing with the various cases that arise, including works of Hudgens and Halloran (2008), Tchetgen and VanderWeele (2012), Manski (2013), Toulis and Kao (2013), Liu and Hudgens (2014), van der Laan (2014), Lazzati (2015), Aronow and Samii (2017), Sofrygin and van der Laan (2017), Athey, Eckles, and Imbens (2018), Basse and Airoldi (2018), and Leung (2019). In particular, Forastiere, Airoldi, and Mealli (2020) generalize PSM to a case in which SUTVA is violated, but treatment decisions are not interdependent in a game-theoretic way and influence is limited to a person's neighborhood (chosen by researchers). Our work is complementary to that literature.

The paper unfolds as follows. Section 2.1 introduces the model and defines treatment effects; Section 2.2 shows, given valid PIPS, how to identify treatment effects with PIPSM. Section 3 details one approach for researchers to identify and estimate PIPS based on a participation game. Section 4.1 runs Monte Carlo experiments, and shows that PIPSM clearly outperforms PSM in estimating vaccination effectiveness when social interactions influence vaccination, using idealized datasets; Section 4.2 studies how exercise affects depression, using both idealized data and survey data from the National Longitudinal Study of Adolescent Health. Section 5 illustrates balance tests. Section 6 provides further remarks and discusses future research.

2. A MODEL OF TREATMENT EFFECTS IN NETWORKS

2.1. Individuals, Networks, and Treatments

There are n individuals, indexed by i in $\mathcal{I} \coloneqq \{1, 2, \cdots, n\}$. An *influence network* is represented by an $n \times n$ adjacency matrix g with entries being either 0 or 1, where $g_{ij} = 1$ indicates that individual i is influenced by another individual j, and $g_{ij} = 0$ indicates that i is not influenced by j, and all diagonal entries of g are set to zero by convention. Individual i's set of *influencers* is $N_i(g) \coloneqq \{j \in \mathcal{I} : g_{ij} = 1\}$, with cardinality $d_i = \sum_{j \in \mathcal{I}} g_{ij}$.

The analysis below fully extends to weighted directed graphs (letting g_{ij} take on any value in $[0, 1]$), without any changes.

We maintain the assumption that g is exogenous to the treatment in question.

A vector $\mathbf{X}_i \in \mathbb{R}^{\dim}$, where dim is a positive integer, summarizes the covariates of

individual i (i.e., i's characteristics). $\mathbf{X}$ denotes the matrix $(\mathbf{X}_1', \mathbf{X}_2', \ldots, \mathbf{X}_n')'$.

We adopt the standard setting of independently identically distributed $\mathbf{X}_i$s across individuals, which ensures a richness of types in the sample; but we allow for g to be formed after the realization of $\mathbf{X}$. The techniques that we employ apply regardless of how g was formed; so we do not model network formation here as it is unnecessary to the analysis or identification here.

Each individual i chooses between two treatment options $Z_i \in \{0,1\}$, where $Z_i = 1$ means participating in the treatment, and $Z_i = 0$ means not participating (control). We defer the modeling of treatment choice to Section 3.

For individual i, the *treatment effect* of Z_i on Y_i, the outcome of interest, is $Y_i(1) - Y_i(0)$, where $Y_i(1)$ is the potential outcome under treatment, and $Y_i(0)$ is the potential outcome under control. The objective is to estimate the *average treatment effect*

$$\tau_{\text{ATE}} \coloneqq \mathbb{E}(Y_i(1) - Y_i(0)),$$

the *average effect of treatment on the treated*

$$\tau_{\text{ATT}} \coloneqq \mathbb{E}(Y_i(1) - Y_i(0) \,|\, Z_i = 1),$$

and the *average effect of treatment on the untreated*

$$\tau_{\text{ATU}} \coloneqq \mathbb{E}(Y_i(1) - Y_i(0) \,|\, Z_i = 0).$$

We work with the following *societal conditional independence* condition, which states that conditional on all demographics and the network, treatment choice is independent of potential outcomes.

Assumption 1 (Societal Conditional Independence (SCI)).

$$(Y_i(1), Y_i(0)) \perp Z_i \mid g, \mathbf{X}$$

for every individual $i \in \mathcal{I}$.

This is a relaxation of a standard conditional independence (CI) condition[8] that requires $(Y_i(1), Y_i(0)) \perp Z_i \mid \mathbf{X}_i$ for every i. Our relaxation of CI is in the same spirit as the relaxation by Angrist and Kuersteiner (2004, 2011) for the purpose of developing 'policy propensity scores' in time series analysis; but instead ours applies to peer effects rather than time effects. Most importantly, CI fails to hold whenever a person's outcomes and participation decisions are both influenced by things like how many friends the person has, or those friends' characteristics, or their participation decisions, etc. We emphasize that SCI makes no assumptions about the correlation between Y_i and Y_j or Z_j.

[8] The assumption is also known as 'selection on observables,' 'conditional ignorability,' etc.

We also make the usual common support assumption, here called *societal common support*, that under any realization of the network and demographics, the probability of participation is nondegenerate for each individual.

Assumption 2 (Societal Common Support). *There exists* $\eta > 0$ *such that*

$$0 < \eta < \mathbb{P}(Z_i = 1 \,|\, g, \mathbf{X}) < 1 - \eta$$

for every individual $i \in \mathcal{I}$.

2.2. Peer-Influenced Treatments

We now show that if we use *peer-influenced propensity scores* (PIPS), then we obtain independence between potential outcomes and the treatment choice. This is then enough to ensure that we have appropriate balance conditional upon the propensity scores. So, we present this result first, and then later discuss methods of estimating propensity scores.

Suppose that if a researcher observes the network and all information about all individuals' characteristics, then that uniquely identifies the expected participation decisions (up to random terms), so that the propensity (likelihood) of an individual to participate can be predicted (except for an independent error term) by knowing $(g, \mathbf{X})$. This is true in our model of participation choices that appears below, and is also true if a researcher uses some other model or method of estimating peer-influenced participation decisions. Under such a prediction, let $\pi_i(g, \mathbf{X}) \coloneqq \mathbb{E}(Z_i \,|\, g, \mathbf{X})$ denote likelihood that i participates; i.e., the peer-influenced propensity of i to participate in treatment.

We prove the following proposition, which shows that treatment outcomes are then independent of participation decisions *conditional upon* PIPS.

Proposition **1** (Societal Conditional Independence for PIPS)**.** *Under Assumptions 1 and 2, for every* $i \in \mathcal{I}$*, treatment outcomes are independent of participation decisions conditional upon PIPS:*

$$(Y_i(1), Y_i(0)) \perp Z_i \,|\, \pi_i(g, \mathbf{X}).$$

The proof is presented in Appendix A.1 and is a straightforward extension of the classical result of Rosenbaum and Rubin (1983), which derives conditional independence of outcomes and participation, conditional upon standard propensity scores (PS) in the form of $\pi_i(\mathbf{X}_i)$ under the stronger CI condition. We work with the weaker SCI condition, and use PIPS.

With Proposition 1 in hand, treatment effects can then be defined as

$$\tau_{\text{ATE}} = \mathbb{E}[\mathbb{E}(Y_i \,|\, Z_i = 1, \pi_i(g, \mathbf{X})) - \mathbb{E}(Y_i \,|\, Z_i = 0, \pi_i(g, \mathbf{X}))];$$
$$\tau_{\text{ATT}} = \mathbb{E}\{[\mathbb{E}(Y_i \,|\, Z_i = 1, \pi_i(g, \mathbf{X})) - \mathbb{E}(Y_i \,|\, Z_i = 0, \pi_i(g, \mathbf{X}))] \,|\, Z_i = 1\};$$
$$\tau_{\text{ATU}} = \mathbb{E}\{[\mathbb{E}(Y_i \,|\, Z_i = 1, \pi_i(g, \mathbf{X})) - \mathbb{E}(Y_i \,|\, Z_i = 0, \pi_i(g, \mathbf{X}))] \,|\, Z_i = 0\}.$$

Similar to the traditional PSM, PIPSM uses matching on PIPS to calculate sample analogues of the above to use as estimates of treatment effects. After providing an example, we define how those PIPS are estimated in Section 3.

3. THE DERIVATION AND ESTIMATION OF PEER-INFLUENCED PROPENSITY SCORES AND MATCHINGS

3.1. The Participation Game

We model the interdependence of treatment choice as a game of incomplete information that we call the *participation game*. The corresponding solution concept that we use is Bayesian equilibrium, which we simply refer to as an *equilibrium* in what follows.

Following the discrete choice literature (e.g., Mcfadden (1974)), we normalize the utility of not participating to 0. The utility from participating is then

$$\mathbf{X}_i^T \beta + \frac{\alpha_1}{d_i} \sum_j g_{ij} Z_j + \alpha_2 \sum_j g_{ij} Z_j - \varepsilon_i.$$

$\mathbf{X}_i^T \beta$ specifies the deterministic individual utility from participation that depends on covariates. The expression $\frac{\alpha_1}{d_i} \sum_j g_{ij} Z_j$ models the influence of the 'local average' participation rate of her influencers, which is common in discrete choice models with social interactions.[9] When $d_i = 0$, we set this component to 0. The expression, $\alpha_2 \sum_j g_{ij} Z_j$ models the 'local aggregate' influence or the total number of participating influencers (Liu, Patacchini, and Zenou, 2014). We call α_1 and α_2 the *social influence parameters*.[10] Lastly, ε_i is an idiosyncratic utility shock for individual i and independent of other variables.

This model is agnostic as to the mechanisms behind social influence. It may involve learning of the perceived benefits of treatment via friends, or exposure and awareness

[9] See Brock and Durlauf (2001a,b), Durlauf and Ioannides (2010), Lee, Li, and Lin (2014), Lin and Xu (2017), and Xu (2018). Linear-in-means models are also popular among peer effect studies with continuous outcome variables: e.g., Manski (1993), Graham (2008), Lee (2007), Bramoullé, Djebbari, and Fortin (2009), De Giorgi, Pellizzari, and Redaelli (2010), Hirano and Hahn (2010), Goldsmith-Pinkham and Imbens (2013), and Blume et al. (2015).

[10] In environments such as schools, there may be limited impact of characteristics of influencers on the participation decision, as argued by Gaviria and Raphael (2001), so $\{\mathbf{X}_j\}_{j \in N_i(g)}$ do not enter the utility specification. But one can add those if they are likely to matter and identification can be worked out from the network (see Bramoullé, Djebbari, and Fortin, 2009).

about treatment, or complementarities in behavior (e.g., using the same software as a coauthor, enjoying to exercise with a friend). Participation in risky behaviors such as smoking and binge drinking may serve as a social lubricant for some sorts of participation decisions. People may simply enjoy the company of influencers.

For simplicity, we ignore the generalizations of this specification, but the model is easily extended in applications in which some other specification is more natural. For instance, peer effects may be non-linear and involve thresholds.[11] It may also involve substitution effects rather than complementarities. Those can be built into the utility function instead of the formulation we provide above.

In an equilibrium, each individual i, after observing ε_i and knowing the network and everyone's characteristics, chooses a treatment to maximize her expected utility, breaks ties in favor of some choice (e.g., $Z_i = 0$). With a continuous distribution on idiosyncratic error values, agents face ties with probability zero, so equilibria are in pure strategies. Formally, an agent's decision rule is

$$Z_i = \mathbf{1}\left\{\mathbf{X}_i^T\beta + \frac{\alpha_1}{d_i}\sum_j g_{ij}\mathbb{E}(Z_j\,|\,g,\mathbf{X},\varepsilon_i) + \alpha_2\sum_j g_{ij}\mathbb{E}(Z_j\,|\,g,\mathbf{X},\varepsilon_i) - \varepsilon_i > 0\right\}. \tag{3.1}$$

It is not our view that treatment choices are necessarily simultaneous and part of some deductive process: decisions could be made sequentially or after discussions in social networks, and may be arrived at via some dynamic. However, a time series of participation decisions or a full record of communication will often be unavailable. This game still directly captures the interdependence of choice, in line with the extensive literature on peer influence, without jeopardizing the practicality of statistical inference. If a researcher has some other (tractable) model that predicts participation decisions and seems more natural in some application, then that can be substituted without affecting the basic logic of our results.

The following distributional assumption about the private utility shocks helps characterize the set of equilibria.

Assumption 3 (Logistic IID). *Utility shocks $\{\varepsilon_i\}_{i\in\mathcal{I}}$ are independently distributed according to the standard Logistic distribution.*

The independence of errors implies that $\pi_j(g,\mathbf{X}) := \mathbb{E}(Z_j\,|\,g,\mathbf{X}) = \mathbb{E}(Z_j\,|\,g,\mathbf{X},\varepsilon_i)$, so i's error does not impact j's decision, which gives the tractability of equilibrium.

By Equation (3.1), individual i switches from participating to not participating at $\varepsilon_i = \mathbf{X}_i^T\beta + \alpha_1\frac{1}{d_i}\sum_{j\in N_i(g)}\mathbb{E}(Z_j\,|\,g,\mathbf{X}) + \alpha_2\sum_{j\in N_i(g)}\mathbb{E}(Z_j\,|\,g,\mathbf{X})$, as ε_i increases. We sometimes abuse

[11] See Jackson and Storms (2017) for a discussion of the importance and estimation of such thresholds.

notation by suppressing arguments in $\pi_i(g, \mathbf{X})$ and calling π_i a strategy or $\pi := (\pi_i)_{i \in \mathcal{I}}$ an equilibrium.

The assumption that errors follow a standard Logistic distributional assumption is easily replaceable. One could, for instance, assume instead that errors follow a standard normal distribution and put a Probit in place of the Logit. From a practical standpoint, two approaches often generate similar (though not identical) inference. More generally, any continuous distribution with a bounded probability density function would work easily. One could instead obtain PIPS in a semiparametric manner (e.g., see Lin, 2019).

Denote the cumulative density function of the standard Logistic Distribution by $\Lambda(t) := \frac{e^t}{1+e^t}$ and probability density function by $\lambda(t) := \Lambda'(t) = \frac{e^t}{(1+e^t)^2}$. Using Equation (3.1) and Assumption 3, the *best response* correspondence for agent i in the participation game is then represented by the function $\mathrm{BR}_i : [0,1]^{\mathcal{I}} \to [0,1]$ such that

$$\mathrm{BR}_i(\pi) = \Lambda\left(\mathbf{X}_i^T \beta + \alpha_1 \frac{1}{d_i} \sum_j g_{ij} \pi_j(g, \mathbf{X}) + \alpha_2 \sum_j g_{ij} \pi_j(g, \mathbf{X})\right). \tag{3.2}$$

Let $\mathrm{BR} := (\mathrm{BR}_i)_{i \in \mathcal{I}}$. Thus, an equilibrium $\pi^*(g, \mathbf{X})$ is such that for every $i \in \mathcal{I}$,

$$\pi_i^*(g, \mathbf{X}) = \Lambda\left(\mathbf{X}_i^T \beta + \alpha_1 \frac{1}{d_i} \sum_j g_{ij} \pi_j^*(g, \mathbf{X}) + \alpha_2 \sum_j g_{ij} \pi_j^*(g, \mathbf{X})\right). \tag{3.3}$$

To ensure the existence and uniqueness of $\pi^*(g, \mathbf{X})$,[12] we assume that the number of direct influencers on any given individual is bounded above, as n goes to infinity (the network grows larger).

Assumption 4 (Bounded and Variable Degree). *There exists $M > 0$ and $m > 0$ such that $0 < \max(d_i) \leq M$, and there is variation in the out-degree such that $\mathrm{Var}(d_i | d_i > 0) > m > 0$ for all n sufficiently large.*

The bounded degree assumption is innocuous in almost all social situations as it just says that there is some finite number of people who directly influence a given person, which is obviously satisfied in contexts in which people's participation decisions are influenced by the people with whom they regularly interact. Even with social media, people may only directly follow a few hundred or thousand others out of potential hundreds of millions, and so the condition is still satisfied.[13]

[12] We avoid dealing with multiplicity of equilibria because it poses a challenge for identification. But it is worth mentioning that partial identification methods help obtain interval estimations for games with fixed numbers of players such as oligopolistic competition (Tamer, 2003; Chernozhukov, Hong, and Tamer, 2007; Tamer, 2010; de Paula and Tang, 2012; de Paula, 2013; Kline and Tamer, 2020).

[13] The condition does not rule out arbitrary numbers of indirect influencers, as the game captures effects of friends of friends, and friends of friends of friends, etc.

The positive variance ensures that people have different numbers of influencers, which is also true in most data. The variance condition is needed to rule out perfect collinearity of the local aggregate and local average terms in our utility specification. If one drops either the local average term or the local aggregate term in the specification, then the variance of out-degrees is not needed.

We also impose a standard *bounded social influence* assumption that helps ensure uniqueness of equilibrium.[14]

Assumption 5 (Bounded Social Influence). *The social influence parameters are nonnegative and bounded above:* $\alpha_1, \alpha_2 \geq 0$ *and* $\alpha_1 + \alpha_2 < \frac{4}{M}$.

This condition ensures that peer influence is nonnegative, and that the feedback from influencers' participation decisions is not so high as to induce an infinite feedback and multiple equilibria.[15]

It must be emphasized that these are only sufficient conditions and in some applications can be violated without jeopardizing the implementation of the estimation procedure that we propose for PIPS.

The following proposition shows the existence and uniqueness of equilibrium, providing a foundation for the algorithm that we use to estimate the equilibrium.

Proposition 2 (Existence and Uniqueness of Equilibrium)**.** *Under Assumptions 3 to 5,* BR *is a contraction mapping in* $[0,1]^{\mathcal{I}}$ *with respect to the metric* $\Delta(\pi, \pi') \coloneqq \max_{i \in \mathcal{I}} |\pi_i - \pi'_i|$*, and thus there exists a unique equilibrium of the participation game,* $\pi^*(g, \mathbf{X})$.

3.2. Estimating the Peer-Influenced Propensity Scores

We now turn to the question of estimating PIPS, $\pi^*(g, \mathbf{X})$.

If the data set is relatively small, and the information in $(g, \mathbf{X})$ is of low dimension, then each i's propensity to participate can be directly estimated using standard machine learning techniques (e.g., a random-forest algorithm) using the network information without having to explicitly model participation. For instance, the background data for each i could potentially include fairly rich information beyond X_i, including information about i's position in the network and information about the characteristics of i's friends' characteristics, and their friends' characteristics, etc. These are the determinants of those friends' participation decisions, and so indirectly accounts for the

[14] Different versions of Assumption 5 appear in studies of games with social interactions (Brock and Durlauf, 2001a; Glaeser and Scheinkman, 2003; Horst and Scheinkman, 2006, 2009; Lin and Xu, 2017; Xu, 2018). They consider only local average but not local aggregate.

[15] Facing some more general distribution of errors that has a bounded PDF $f_\varepsilon(\cdot)$, one could assume that $\alpha_1 + \alpha_2 < \frac{1}{M \sup_c f_\varepsilon(c)}$ in place of Assumption 5.

peer influence. However, in many, if not most, applications of interest such a calculation is computationally impractical. Thus, the researcher needs to model how peers make participation decisions so that the propensities to participate can be jointly estimated as a function of the network and characteristics. The model above provides such an approach.

In particular, Assumptions 4 and 5 not only help identify equilibrium, but also jointly ensure feasible statistical inference from interdependent observations in a large influence network. Assumption 4 ensures that as n goes to ∞, we obtain subnetworks that are distant from each other. Assumption 5 bounds social influence so that the impacts of a random shock to one individual are limited for distant individuals. Intuitively, from the best response strategy in Equation (3.2), we know that individual's choice probability depends on each of her peer's choice probability with strength bounded above by $\frac{M(\alpha_1+\alpha_2)}{4} < 1$. As the distance, dis, from her to another individual grows, the strength of dependence is bounded above by $\left[\left(\frac{M(\alpha_1+\alpha_2)}{4}\right)^{\text{dis}}\right]$, so decays exponentially. Generally speaking, when two individuals are far away from each other in the network, the interdependence of their choice probabilities becomes ignorable.[16] In particular, it follows that an analogous version of the *network decaying dependence* condition of Xu (2018) holds for our participation game.

Let $(g,\mathbf{X})_i^h$ be network and demographic information associated with the subnetwork within social distance h from individual i, and $(g,\mathbf{X})_i^{>h}$ be any possible realization of the network and demographics beyond this subnetwork.

Proposition **3.** *[Network Decaying Dependence] Under Assumptions 3 to 5, for every $i \in \mathcal{I}$, as h grows to infinity at a rate $o(n)$,*[17]

$$\sup_{(g,\mathbf{X})_i^{>h}} \left| \mathbb{P}(Z_i = 1 \,|\, g,\mathbf{X}) - \mathbb{P}(Z_i = 1 \,|\, (g,\mathbf{X})_i^h, (g,\mathbf{X})_i^{>h}) \right| \to 0 \quad \text{as} \quad n \to \infty.$$

Network decaying dependence is a network version of a 'weak dependence' condition that permits the analysis of network-generated data. It ensures that a large network contains many approximately independent subnetworks, and thus large sample results for weakly dependent data can be applied.

Thus, we next describe how to estimate our model parameters $\theta := (\beta^T, \alpha_1, \alpha_2)^T$, and PIPS π^*, which defines a unique equilibrium, via a *nested pseudo likelihood* algorithm. This sort of algorithm has been used for dynamic discrete choice models by Aguirregabiria and

[16] M bounds the degrees, and then bounds the expansion properties of the network as a function of distance, which combined with the bounded social influence condition, ensures that the indirect influences must disappear with distance. Without these conditions, behaviors would infinitely feedback through the network.

[17] Little o notation indicates that $h(n)/n \to 0$.

Mira (2002), for dynamic games by Aguirregabiria and Mira (2007), and for large network games by Lin and Xu (2017).

The nested pseudo likelihood algorithm proceeds as follows.

Initiation: Take an initial guess of PIPS, denoted by $\hat{\pi}^{(0)}$.[18]

Iteration: Given $\hat{\pi}^{(K)}$, run a Logit regression of Z_i on $\mathbf{X}_i$, $\frac{1}{d_i}\sum_{j\in N_i(g)}\hat{\pi}_j^{(K)}$, and $\sum_{j\in N_i(g)}\hat{\pi}_j^{(K)}$ to obtain parameters $\hat{\theta}^{(K+1)}$. For every $i \in \mathcal{I}$, compute

$$\hat{\pi}_i^{(K+1)} := \Lambda\Big(\mathbf{X}_i^T\hat{\beta}^{(K+1)} + \hat{\alpha}_1^{(K+1)}\cdot\frac{1}{d_i}\sum_{j\in N_i(g)}\hat{\pi}_j^{(K)} + \hat{\alpha}_2^{(K+1)}\cdot\sum_{j\in N_i(g)}\hat{\pi}_j^{(K)}\Big).$$

Termination: Iterate until the difference between two consecutive π estimates is sufficiently small, say, when $\|\hat{\pi}^{(K+1)} - \hat{\pi}^{(K)}\| < \text{tol}$. Set the nested pseudo likelihood estimates as $\hat{\theta} := \hat{\theta}^{(K+1)}$ and $\hat{\pi} := \hat{\pi}^{(K+1)}$.

The contraction mapping result of Proposition 2 essentially ensures that the nested pseudo likelihood algorithm converges if the starting point lies in a neighborhood of the true parameter (Kasahara and Shimotsu, 2012).

In Appendix A.4, we state a result about consistency of a nested pseudo-likelihood estimator of the parameters. This is an adaptation of the results of Aguirregabiria and Mira (2007), Lin and Xu (2017), and Hu and Lin (2020) to our setting.

3.3. Asymptotic Properties of the Matching

Once we have PIPS estimated, we then have to do the matching in order to estimate treatment effects. We now describe that part of the process.

We emphasize that Proposition 1 does *not* state that $(Y_i(1), Y_i(0)) \perp Z_{-\mathbf{i}}|\pi_i(g,\mathbf{X})$, *nor* does it state that $(Y_i(1), Y_i(0)) \perp (Y_j(1), Y_j(0))|\pi_i(g,\mathbf{X})$ – so it can still be that the participation decisions, and even the outcomes, of others could be correlated with and hence interfere with a given person's outcomes. For instance, in the case of vaccination the outcomes of an individual would be correlated with both the vaccination decision and infection of her friends. So, a *conditional version* of SUTVA is also violated. Nonetheless, we still develop consistent estimates of treatment effects.

In particular, denote the nonempty treatment and control groups based on a particular propensity score by $\mathcal{T}(\pi) := \{i \in \mathcal{I} : Z_i = 1, \pi_i = \pi\}$ and $\mathcal{U}(\pi) := \{i \in \mathcal{I} : Z_i = 0, \pi_i = \pi\}$, respectively. Let $\mathcal{M}_i(\pi) \neq \emptyset$ be the set of individuals who are 'matched' to individual i out of the set of individuals with $\pi_i = \pi$ (so these are the people with the opposite treatment).

[18] For example, we can obtain them from a Logit regression of Z_i on $\mathbf{X}_i$.

Define the treatment effects conditional upon π by

$$\widetilde{\tau}_{\text{ATT}}(\pi) := \frac{1}{|\mathcal{T}(\pi)|} \sum_{i \in \mathcal{T}(\pi)} \Big(Y_i - \frac{1}{|\mathcal{M}_i(\pi)|} \sum_{j \in \mathcal{M}_i(\pi)} Y_j \Big);$$

$$\widetilde{\tau}_{\text{ATU}}(\pi) := \frac{1}{|\mathcal{U}(\pi)|} \sum_{i \in \mathcal{U}(\pi)} \Big(\frac{1}{|\mathcal{M}_i(\pi)|} \sum_{j \in \mathcal{M}_i(\pi)} Y_j - Y_i \Big);$$

$$\widetilde{\tau}_{\text{ATE}}(\pi) := \frac{1}{n(\pi)} \Big[\sum_{i \in \mathcal{T}(\pi)} \Big(Y_i - \frac{1}{|\mathcal{M}_i(\pi)|} \sum_{j \in \mathcal{M}_i(\pi)} Y_j \Big) + \sum_{i \in \mathcal{U}(\pi)} \Big(\frac{1}{|\mathcal{M}_i(\pi)|} \sum_{j \in \mathcal{M}_i(\pi)} Y_j - Y_i \Big) \Big],$$

where $n(\pi) = |\{i : \pi_i = \pi\}|$.

The challenge that is faced is that $(Y_i(1), Y_i(0))$ and $(Y_j(1), Y_j(0))$s can be correlated, even conditional upon $\pi_i(g, \mathbf{X})$ and $\pi_j(g, \mathbf{X})$, and so the sums above involve correlated random variables. In particular, these can be highly correlated among friends. As in the vaccination example, if one friend gets infected with a disease, that makes it more likely that their friends will also become infected.

The key is that even though the outcome and decisions of friends can be correlated, the overall amount of total correlation in outcomes in the whole society is bounded, and an individual's outcome only strongly correlates with some subset. In order to admit the fairly arbitrary correlation patterns that can arise in networks, we work with a powerful version of a central limit theorem that bounds total correlation, but remain agnostic on where it appears.

We say that a matching $\mathcal{M}$ (or, to be rigorous, a sequence of matchings) has *asymptotically low total correlation* at a propensity score π if the overall covariance between the outcomes of individuals with the same π value is of a smaller order than the overall variance, when averaged across all pairs. Formally,

$$\sum_{i,j:\ \pi_i = \pi_j = \pi} \text{Cov}((Y_i(0) - \overline{y}^0_\pi)^2, (Y_j(0) - \overline{y}^0_\pi)^2 \,|\, \pi_i = \pi_j = \pi) = o(n^2 (\text{Var}(Y_i(0) \,|\, \pi_i = \pi))^2),$$

where $\overline{y}^0_\pi := \mathbb{E}(Y_i(0) \,|\, \pi_i = \pi)$, and

$$\sum_{i,j:\ \pi_i = \pi_j = \pi} \text{Cov}(Y_i(0), Y_j(0) \,|\, \pi_i = \pi_j = \pi) = o(n \text{Var}(Y_i(0) \,|\, \pi_i = \pi));$$

and analogous conditions hold for $Y(1)$. Thus, for instance, friends' outcomes can be highly correlated, and even friends of friends, etc., as long as their outcomes are not too correlated with the bulk of other individuals.

Under these conditions, the ATT, ATE, and ATU, conditional upon π in finite network samples, will converge to the true values as the network becomes large, with an asymptotic normal distribution and a rate proportional to one over the square root of

the number of observations, provided the matching is done appropriately.

Specifically, we say that a matching $\mathcal{M}$ is *balanced* at π, if the number of times that each j appears in different $\mathcal{M}_i(\pi)$s is the same across js that have $\pi_j = \pi$ and are in the same group (treated or untreated). An easy way to produce a balanced matching is to match each i with a particular treatment Z_i to all those with the same propensity but who have the opposite treatment.

Proposition **4.** *Consider a sequence of networks and models indexed by n, for which $(Y(0), Y(1))$ are bounded above and below with variances bounded away from 0 conditional upon a propensity score π. If the sequence of matchings $\mathcal{M}_i(\pi)$ has asymptotically low total correlation and is balanced at π, then $\widetilde{\tau}_{\mathrm{ATT}}(\pi)$, $\widetilde{\tau}_{\mathrm{ATU}}(\pi)$, and $\widetilde{\tau}_{\mathrm{ATE}}(\pi)$ converge in probability to have normal distributions around their true values (conditional on π) with standard deviations of order $1/\sqrt{\min[\pi, 1-\pi]n(\pi)}$.*

Proposition 4 tells us that any balanced matching will provide a consistent estimate of the true treatment effects, conditional upon the true propensity scores.

The balance condition on the matching is important to the result, as it ensures that all of the individuals with the same propensity score are sampled with the same rate. Otherwise, one could end up with a clustering of over-sampled matches of individuals who have correlated values. Of course, the condition does not need to hold exactly, as having some imbalance will not change the result, just as in any weighted version of the central limit theorem, as long as no weight overwhelms the others. For instance, the result extends if one samples Y_is and each appears a vanishing fraction of times out of the total.

We remark that an alternative to the definition of asymptotically low total correlation, which also provides a basis for Proposition 4, is instead to apply the conditions to $\mathrm{Cov}(Y_i(1) - Y_j(0), Y_{i'}(1) - Y_{j'}(0) | \pi_i = \pi_j = \pi_{i'} = \pi_{j'} = \pi)$. This applies directly to the differences $Y_i(1) - Y_j(0)$ (across different individuals), and a similar proof technique establishes the result.[19]

From Proposition 4 we can deduce that, if the propensity scores take on a finite set of values, then the overall sample treatment effects will converge to their true values. In

[19] One can check that the alternative condition is actually stronger than the condition that we have worked with.

particular, define sample treatment effects to be:

$$\widetilde{\tau}_{\text{ATT}} \coloneqq \frac{1}{|\mathcal{T}|}\sum_{\pi}\sum_{i\in\mathcal{T}(\pi)}\Big(Y_i - \frac{1}{|\mathcal{M}_i(\pi)|}\sum_{j\in\mathcal{M}_i(\pi)} Y_j\Big);$$

$$\widetilde{\tau}_{\text{ATU}}(\pi) \coloneqq \frac{1}{|\mathcal{U}|}\sum_{\pi}\sum_{i\in\mathcal{U}(\pi)}\Big(\frac{1}{|\mathcal{M}_i(\pi)|}\sum_{j\in\mathcal{M}_i(\pi)} Y_j - Y_i\Big);$$

$$\widetilde{\tau}_{\text{ATE}} \coloneqq \frac{1}{n}\sum_{\pi}\Big[\sum_{i\in\mathcal{T}(\pi)}\Big(Y_i - \frac{1}{|\mathcal{M}_i(\pi)|}\sum_{j\in\mathcal{M}_i(\pi)} Y_j\Big) + \sum_{i\in\mathcal{U}(\pi)}\Big(\frac{1}{|\mathcal{M}_i(\pi)|}\sum_{j\in\mathcal{M}_i(\pi)} Y_j - Y_i\Big)\Big].$$

We get the following corollary.

Corollary **1.** *Consider a sequence of networks and models indexed by n, for which π takes on a finite set of values and $(Y(0), Y(1))$s are bounded above and below with variances bounded away from 0 conditional upon each π. If the matching $\mathcal{M}_i(\pi)$ has asymptotically low total correlation and is balanced at each π, then $\widetilde{\tau}_{\text{ATT}}$, $\widetilde{\tau}_{\text{ATU}}$, and $\widetilde{\tau}_{\text{ATE}}$ converge in probability to have normal distributions around their true values with standard deviations of order $1/\sqrt{\min_{\pi}[\pi n(\pi), (1-\pi)n(\pi)]}$.*

In order to get consistency using estimated propensity scores instead of the true ones, it is then sufficient that the estimates of the propensity scores becomes accurate in a large network. Proposition 6 in Appendix A.6 provides a set of regularity conditions that ensure that the parameter estimates of the utility model become accurate. From that, if we have a finite set of distinct true equilibrium propensities π (for instance with a grid of Xs and degrees), the fraction of individuals whose estimated propensities are correctly binned within a radius of the their true π tends to 1.

Next, consider a matching that has an 'approximate balance'[20] – such as the caliper matching, $\mathcal{M}_i$ contains the others whose PIPS lies within the caliper ('propensity range' centering on own PIPS with a predetermined radius). Then we can define[21]

$$\widehat{\tau}_{\text{ATT}} \coloneqq \frac{1}{|\mathcal{T}|}\sum_{i\in\mathcal{T}}\Big(Y_i - \frac{1}{|\mathcal{M}_i|}\sum_{j\in\mathcal{M}_i} Y_j\Big);$$

$$\widehat{\tau}_{\text{ATU}} \coloneqq \frac{1}{|\mathcal{U}|}\sum_{i\in\mathcal{U}}\Big(\frac{1}{|\mathcal{M}_i|}\sum_{j\in\mathcal{M}_i} Y_j - Y_i\Big);$$

$$\widehat{\tau}_{\text{ATE}} \coloneqq \frac{1}{n}\Big[\sum_{i\in\mathcal{T}}\Big(Y_i - \frac{1}{|\mathcal{M}_i|}\sum_{j\in\mathcal{M}_i} Y_j\Big) + \sum_{i\in\mathcal{U}}\Big(\frac{1}{|\mathcal{M}_i|}\sum_{j\in\mathcal{M}_i} Y_j - Y_i\Big)\Big],$$

[20] One needs that no individual i appears in a non-vanishing fraction of all matchings.

[21] One can use other popular methods such as the *nearest neighbor matching*, the *stratification matching*, and the *kernel matching* (Caliendo and Kopeinig, 2008; Imbens and Rubin, 2015), and simply check that individuals are being evenly matched and can do the matching without replacement in order to force such balance if necessary.

The difference between the $\widehat{\tau}$ and $\widetilde{\tau}$ are in the matchings, which are off of estimated πs. Thus, as the accuracy of the estimated πs increases, the fraction of individuals who are correctly matched will tend to 1, and so will the balance of the matching, for each π in the range. Then applying Corollary 1, it follows that the estimated ATT, ATU, and ATE will converge to their true values.

We verify that such an approximate balance property holds in our empirical illustrations below.

4. ILLUSTRATIONS

Via simulations and application to a data set, we illustrate the accuracy of PIPSM and its improvement over PSM.

4.1. A Monte Carlo Illustration on a Vaccination Model

In this section, we use Monte Carlo simulations to compare PIPSM with the traditional PSM in a stylized vaccination model.

Consider testing the effectiveness of a vaccine for infectious diseases. Let Z_i denote person i's treatment, where 1 indicates that the person was vaccinated and 0 indicates otherwise, and Y_i denote whether i contracts the disease.

In such a case, there is clearly interference across the treatment of different people, as when more a person's contacts are vaccinated, then that person has a lower chance of being infected (if the vaccine is effective). Thus, $(Y_i(1), Y_i(0))$ depends on the vaccination decisions of the rest of the population, Z_{-i}; and SUTVA is clearly violated.

Moreover, people's vaccination decisions are also correlated with those of their friends and the more general population. In particular, people talk to each other and their opinions about vaccination depend on the decisions of their peers and the more general population. One could also imagine a free-rider effect in which the more peers who vaccinate, the less a given person does. While this is hypothetically possible, it does not appear in the data where the positive information effect dominates (e.g., see the discussion in Banerjee et al. (2020)). We also note that our approach can be extended to allow for free-rider effects in the equilibrium estimation. This means Z_i is correlated with Z_{-i}, as well as Y_{-i}, all conditional upon X_i, and therefore the usual CI is also violated.

Note that SCI is satisfied in this example. Once one knows the network and demographics, one can estimate the participation decisions. In particular, via an equilibrium calculation, we can deduce $\pi_i(g, \mathbf{X})$, the probability that i gets treated conditional upon the network and vector of covariates. Conditional upon $\pi_i(g, \mathbf{X})$, any remaining random element in Z_i comes from an idiosyncratic error term and not from any demographics, peer effects or other network influence. Thus, conditional upon $\pi_i(g, \mathbf{X})$,

i's treatment probability, Z_i, is independent of $(Y_i(1), Y_i(0))$. This allows us to estimate the outcomes for any individual i, and hence various average effects, despite the fact that the usual SUTVA and CI conditions are violated. The correlation between people's outcomes is then handled via a balanced matching across the network.

To illustrate the issues most clearly, we begin with a benchmark case in which CI is satisfied and there are no peer effects on either a person's treatment decision or their treatment outcomes. We then introduce the peer effects on both treatment decisions and outcomes. There, we examine two cases in which CI and SUTVA are both violated, but our SCI is satisfied. The two cases illustrate different functional forms for how a person's outcomes depend on other's treatment decisions.

We generate a random social network g of n individuals as follows. Each individual i has a degree independently drawn from $d_i \in \{0, 1, \ldots, 10\}$ with equal likelihood on each degree. Then, we randomly choose d_i of the other $n-1$ individuals for this individual to have as friends. The network is thus directed, in that j can influence i without requiring (but not precluding) that i influence j in return. The example can also be done with any other network, but this simplifies the code.

Individual i has demographic characteristics (X_{1i}, X_{2i}) drawn, respectively, from a standard normal distribution and a uniform distribution over $[-1, 1]$. The participation choices of individuals are generated based on our Bayesian game with a Logistic random utility shock:

$$Z_i = \mathbf{1}\left\{\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \alpha_1 \frac{1}{d_i}\sum_j g_{ij}\pi_j^*(g, \mathbf{X}) + \alpha_2 \sum_j g_{ij}\pi_j^*(g, \mathbf{X}) - \varepsilon_i > 0\right\},$$

where $\pi_j^*(g, \mathbf{X})$ is the equilibrium propensity of agent j. We use the true parameter values $\beta = (0, 1, 1)$ and $\alpha = (1, 0.1)$.

We examine three different data generating processes for the potential outcomes. In what follows, let (u_1, u_0) be independently drawn from the standard normal distribution.

CI is satisfied (Case 1) $Y_i(1) = u_{1i}$ and $Y_i(0) = 1 + u_{0i}$. In this case, the true ATE, ATT, and ATU are all -1.

CI and SUTVA are violated but SCI is satisfied

(Case 2.1) $Y_i(1) = u_{1i}$ and $Y_i(0) = 1 - \frac{1}{d_i}\sum_j g_{ij}Z_j + u_{0i}$. In this case, ATE is the average of $\frac{1}{d_i}\sum_j g_{ij}Z_j - 1$ of all n individuals, ATT is the average of $\frac{1}{d_i}\sum_j g_{ij}Z_j - 1$ of those in the treatment group, and ATU is the average of $\frac{1}{d_i}\sum_j g_{ij}Z_j - 1$ of those in the control group.

(Case 2.2) $Y_i(1) = u_{1i}$ and $Y_i(0) = 1 - \sum_j g_{ij}Z_j + u_{0i}$. In this case, ATE is the average of $\sum_j g_{ij}Z_j - 1$ of all n individuals, ATT is the average of $\sum_j g_{ij}Z_j - 1$ of those

in the treatment group, and ATU is the average of $\sum_j g_{ij} Z_j - 1$ of those in the control group.

These are stylized interactions, but capture the essential forms of interference that could occur from something like vaccination. A person's participation decision depends on those of her friends and the general population, and an unvaccinated person ($Y_i(0)$) has an outcome that decreases as a function of the average vaccination rate of i's friends (Case 2.1), or the aggregate vaccination of i's friends (Case 2.2).

We follow a 3×3 factorial design, for each of the three data generating processes described above, running 1000 simulations for sample sizes of 200, 400, and 800. Tables 1 and 2 report the average biases and mean squared errors of ATE, ATT, and ATU estimates based on PIPSM and PSM.

Table 1: Comparison of PIPSM and PSM When CI Is Satisfied (Case 1)

	Average Bias					
	PSM			PIPSM		
	ATE	ATT	ATU	ATE	ATT	ATU
n=200	-0.005	-0.001	-0.012	0.005	0.007	0.001
n=400	0.004	0.006	-0.001	0.001	0.003	0.000
n=800	0.001	0.001	0.003	0.003	0.005	-0.001
	Mean Squared Error					
n=200	0.053	0.078	0.063	0.058	0.088	0.066
n=400	0.025	0.035	0.032	0.029	0.043	0.032
n=800	0.012	0.019	0.014	0.013	0.021	0.014

Table 2: Comparison of PIPSM and PSM When CI and SUTVA Are Violated but SCI Is Satisfied (Case 2.1 and 2.2)

	Average Bias											
	Case 2.1						Case 2.2					
	PSM			PIPSM			PSM			PIPSM		
	ATE	ATT	ATU	ATE	ATT	ATU	ATE	ATT	ATU	ATE	ATT	ATU
n=200	-0.052	-0.082	0.012	-0.002	-0.009	0.011	-0.484	-0.719	0.012	-0.012	-0.024	0.011
n=400	-0.058	-0.087	0.003	0.001	0.001	0.001	-0.498	-0.735	0.003	0.008	0.012	0.001
n=800	-0.057	-0.086	0.005	0.001	-0.002	0.008	-0.504	-0.745	0.005	0.002	-0.001	0.008
	Mean Squared Error											
n=200	0.056	0.087	0.063	0.055	0.086	0.061	0.459	0.959	0.063	0.192	0.381	0.061
n=400	0.036	0.057	0.028	0.032	0.049	0.034	0.381	0.803	0.028	0.106	0.206	0.034
n=800	0.018	0.029	0.016	0.015	0.023	0.017	0.314	0.673	0.016	0.046	0.093	0.017

In all cases, the magnitudes of the average biases for PIPSM estimates are all well below .01 when $n = 800$. PIPSM performs similarly as the traditional PSM in the case where CI is satisfied and $n = 200$ (Table 1), but then significantly outperforms PSM when CI and SUTVA are violated (Table 2). For instance in Case 2.2, the magnitude of the bias in standard PSM estimation is at least .4 on ATE and .7 on ATT, while for PIPSM it is less than .024 even for the smallest n. The one situation in which standard PSM does well is at estimating ATU. This happens since in that case $Y_i(0)$ is directly observed and that is where the interference enters, while $Y_i(1)$ has no interference (in this example, other people's vaccinations make no difference in a person's infection rate when she is vaccinated), so $Y_i(1)$ can be estimated from any population, as it is always mean 0 in all of our cases.[22]

4.2. An Empirical Illustration: The Effect of Exercise on Depression

Again, the purpose of this illustration is to contrast the results that one obtains by using PIPSM with those from the traditional PSM, but in this case with real data rather than a simulated model. There are two main conclusions. First, participation is significantly dependent upon peers' participation. Thus, given that depression has been found to be correlated with social interactions, the traditional CI condition is violated and an approach such as PIPSM is warranted. Second, the estimates that one obtains from PIPSM differ substantially from those obtained when ignoring peer effects and using PSM.

Interestingly, the standard PSM estimates can be biased either upwards or downwards compared to the PIPSM, depending on the network structure. This is consistent with equilibrium behavior correlating participation in ways that can move a group of agents into or out of participation, since coordination with peers becomes a significant influence. This can thus change the pool of participants in treatment in either direction. We see the variation when looking across different networks in the data. Moreover, PIPSM estimates are very stable across different matching methods while PSM results vary considerably. This may be due to the fact that peer-influence correlates behaviors, and so the matching method then matters when one ignores that dependency. We also check that PIPSM better balances network characteristics of matched pairs (Section 5).

4.2.1. First, a Monte Carlo Simulation

Before analyzing the actual data, we first illustrate the differences between PIPSM and PSM in another Monte Carlo simulation. This helps provide intuition for why PSM goes wrong, and PIPSM offers consistent estimates.

[22] If $Y_i(1)$ was also interfered with, or even depended upon a person's own covariates X_i, then that would cause problems for PSM because PSM would not have the right propensity matches to estimate the $Y_i(1)$ from; but it would not cause problems for PIPSM.

Let Y_i and Z_i be the level of a person's depression and the binary decision to exercises, respectively. In this simulation, people with more friends (higher d_is) have less depression, and people who do not exercise ($Z_i = 0$) have a heightened depression level by 5. Other aspects are the same as before.

CI and SUTVA are violated but SCI is satisfied

(Case 2.3) $Y_i(1) = -d_i + u_{1i}$ and $Y_i(0) = 5 - d_i + u_{0i}$. In this case, the true ATE, ATT, and ATU are all -5.

Standard PSM gets things very wrong. The reason is that the d_is are not the same in the treated and untreated groups: people with higher d_is have more friends and tend to exercise more. Thus, the treatment group tends to have more friends, and thus lower depression, while the untreated group tends to have fewer friends. Without capturing this, the estimation is biased. PIPSM adjusts for this since the effect of friends are captured in the modeling of the participation decisions, and then the propensity matching puts people together with others who have similar friending patterns.

Table 3: Comparison of PIPSM and PSM when CI and SUTVA are violated but SCI is satisfied (Case 2.3)

	Average Bias					
	PSM			PIPSM		
	ATE	ATT	ATU	ATE	ATT	ATU
n=200	-0.962	-1.006	-0.870	-0.070	-0.076	-0.060
n=400	-0.991	-1.004	-0.962	-0.025	-0.012	-0.054
n=800	-0.991	-1.012	-0.948	-0.019	-0.018	-0.019
	Mean Squared Error					
n=200	1.439	1.780	1.315	0.286	0.573	0.372
n=400	1.273	1.443	1.241	0.158	0.297	0.198
n=800	1.116	1.224	1.049	0.070	0.135	0.108

Note that ATE, ATT and ATU estimated through PSM are all severely biased while PIPSM estimates have small and decreasing biases. The declining pattern of the PIPSM MSE is consistent with the theoretical $\sqrt{n}$ convergence rate.

4.2.2. Data Description

We now present the application to data. We use data from the *National Longitudinal Study of Adolescent Health* (Add Health) (Harris et al., 2011). Starting with the 1994-95 academic year, the study collected information on a representative school-based cluster sample of adolescents in grades 7-12 in the United States. We base our study on the in-school survey of the first study wave only in order to avoid missing data due to the use of

multiple surveys across different study waves. We use the 5 largest school networks (each school network consists of one or two sister schools).

We construct the influence network from students' responses to friend nomination questions. They were asked to identify at most 5 male friends and 5 female friends, and we assume that the treatment choice of doing exercise was influenced by the friends these students nominated.

Covariates are simple demographic characteristics summarized in Table 4. These school networks have relatively large student populations, from 1,505 to 2,213. They are well balanced in gender, with the proportions of female students ranging from 50.6% to 54.1%. White students form the majority except that the second school has Hispanic students as the majority. A student's parental education is the average of two answers to the survey questions: "How far in school did he/she go?"[23]

The treatment variable, 'exercise,' comes from the survey question: "How many times in a normal week do you work, play, or exercise hard enough to make you sweat and breathe heavily?" Students may answer "Never," "1 to 2 times," "3 to 5 times," "6 to 7 times," or "more than 7 times." Exactly the last two answers correspond to $Z_i = 1$, or the treatment choice of doing exercise more or less daily. Based on our definition of exercise, the proportions of students who exercise range from 29.5% to 55.3%.

The measure of depression is based on the survey question: "In the last month, how often did you feel depressed or blue?" Students may answer "never," "rarely," "occasionally," "often," or "everyday." Since the answers are ordinal, we dichotomize so that depression status $Y_1 = 0$ for answering "never" and $Y_1 = 1$ otherwise. The proportions of students have at least experienced some depression range from 58.7% to 71.8% across the 5 study schools.

4.2.3. PSM and PIPSM Results

We compare the estimations of the effects of exercise (participation in treatment) based on PS and PIPS (Table 5).

Among all schools except the second school, the estimates of the social influence parameter for local aggregate, α_2, are significant at the 5% level; none of those for local average, α_1, is significant.[24] This provides evidence for the existence of peer influence in

[23]Answers of 1-8 correspond to "eighth grade or less," "more than eighth grade, but did not graduate from high school," "high school graduate," "completed a GED," "went to a business, trade, or vocational school after high school," "went to college but did not graduate," "graduated from college or a university," and "professional training beyond a four-year college." If there is one answer instead of two; e.g., single-parent family, we take the single answer.

[24]The standard errors of nested pseudo likelihood estimators come from the last steps of the iterative Logit regressions; they are reasonable placeholders. There is no way for us to use sample analogs because,

Table 4: Summary Statistics

Network	1		2		3		4		5	
# of Students	2213		1558		1540		1528		1505	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Age	15.230	1.554	15.866	1.460	15.255	1.663	15.181	1.557	14.707	1.751
Gender	0.517	0.500	0.541	0.498	0.516	0.500	0.533	0.499	0.506	0.500
White	0.853	0.354	0.392	0.488	0.906	0.292	0.806	0.395	0.810	0.392
Hispanic	0.138	0.345	0.893	0.309	0.051	0.220	0.043	0.202	0.168	0.374
Black	0.009	0.092	0.087	0.283	0.028	0.165	0.143	0.350	0.037	0.188
Asian	0.015	0.123	0.017	0.128	0.027	0.161	0.033	0.180	0.024	0.153
Native	0.075	0.264	0.024	0.152	0.047	0.212	0.026	0.158	0.063	0.243
Other	0.103	0.303	0.316	0.465	0.044	0.206	0.021	0.143	0.110	0.313
Par. Edu.	4.600	1.917	5.089	2.634	5.293	1.941	6.491	1.642	5.210	2.131
Exercise	0.408	0.492	0.295	0.456	0.396	0.489	0.399	0.490	0.553	0.497
Depression	0.681	0.466	0.587	0.493	0.689	0.463	0.718	0.450	0.660	0.474

exercise participation, and supports the use of PIPS for matching. Furthermore, although many studies of peer effects find significant influence of the local average in the absence of local aggregate terms, our application shows that in certain settings the local aggregate can be more important than the local average.

Table 5: Traditional Propensity Scores and Peer-Influenced Propensity Scores

Network	1		2		3		4		5	
Model	PS	PIPS	PS	PIPS	PS	PIPS	PS	PIPS	PS	PIPS
Age	-0.160**	-0.135**	-0.090**	-0.088**	-0.125**	-0.122**	-0.147**	-0.131**	-0.159**	-0.147**
Gender	-1.326**	-1.295**	-1.314**	-1.327**	-1.231**	-1.244**	-1.289**	-1.295**	-1.313**	-1.320**
Hispanic	0.325**	0.357**	-0.655*	-0.666**	-0.250	-0.269	0.176	0.212	-0.612**	-0.487**
Black	-0.844*	-0.770	-0.175	-0.154	0.424	0.519	-0.147	-0.043	-0.479	-0.360
Asian	-0.184	-0.057	-0.350	-0.391	-0.834**	-0.814**	-0.093	-0.048	0.101	0.257
Native	0.485**	0.506**	0.983**	0.994**	0.617**	0.665**	0.309	0.316	-0.047	-0.050
Other	-0.326*	-0.359**	0.207	0.201	-0.057	0.008	-0.011	0.031	-0.311	-0.313
Par. Edu.	0.016	0.006	0.037*	0.037*	0.066**	0.056*	0.069**	0.057	0.064**	0.040
Intercept	2.577**	1.873**	1.501**	1.390*	1.727**	1.430**	2.020**	1.583**	3.059**	2.551**
α_1	–	0.000	–	0.000	–	0.000	–	0.341	–	0.000
α_2	–	0.196**	–	0.130	–	0.176**	–	0.132**	–	0.178**

*: 10% Significant; **: 5% Significant

Table 6 reports the estimated treatment effects, including ATE, ATT, and ATU.

for nested pseudo likelihood estimators, choice probabilities stem from equilibrium behavior and it is difficult to solve for expressions of their asymptotic variances. Bootstrap methods for one large social network are still under development.

Reported standard errors are the default outputs of the Matching package for R.[25] For both PSM and PIPSM, we conduct nearest neighbor matchings and caliper matchings (for radius 0.01 and 0.1). All statistically significant estimates are negative, which is consistent with the theory that exercise mitigates depression (Babyak et al., 2000; Lawlor and Hopker, 2001; Salmon, 2001; Dunn et al., 2005; Ströhle, 2009).

There are noticeable differences between PSM and PIPSM results. The bias in PSM can be either positive or negative, depending on the network. This may be due to the complementarities in participation. These results suggest that there is no easy shortcut to adjusting results without actually using PIPSM.

Table 6: Treatment Effects of Exercise on Depression

Network	1		2		3		4		5	
Model	PSM	PIPSM	PSM	PIPSM	PSM	PIPSM	PSM	PIPSM	PSM	PIPSM
Nearest Neighbor										
ATE	0.018	-0.007	-0.030**	-0.070**	-0.015	0.005	-0.039**	-0.004	-0.040**	-0.068**
ATT	0.030	-0.018	-0.043*	-0.039	-0.004	0.008	-0.083**	0.028	-0.046**	-0.101**
ATU	0.010	-0.000	-0.021	-0.083**	-0.022	0.003	-0.010	-0.025	-0.032	-0.027
Caliper with Radius 0.01										
ATE	-0.032**	-0.006	-0.062**	-0.071**	-0.011	0.018	-0.035**	-0.010	-0.044**	-0.066**
ATT	-0.042*	-0.011	-0.078**	-0.036	-0.001	0.034	-0.070**	0.022	-0.045**	-0.095**
ATU	-0.026	-0.003	-0.055**	-0.086**	-0.018	0.007	-0.011	-0.031	-0.043*	-0.032
Caliper with Radius 0.1										
ATE	-0.030**	-0.006	-0.069**	-0.072**	-0.015	0.006	-0.038**	-0.005	-0.040**	-0.068**
ATT	-0.043*	-0.016	-0.082**	-0.041	-0.003	0.012	-0.082**	0.025	-0.047**	-0.101**
ATU	-0.021	-0.000	-0.063**	-0.084**	-0.023	0.003	-0.010	-0.025	-0.032	-0.027

** 5% Significant; * 10% Significant

5. BALANCE TESTS

We now provide three different tests that the matchings under PIPSM are appropriately balanced. We implement these balance tests with regards to the empirical application of depression from the last section. These are all tests that can be done with any application of PIPSM.

The first two are checks of how closely the balance condition from Section 3.3 holds: checking that no individuals are excessively matched and that people are rarely matched to their friends. This helps ensure that overall correlation in outcomes is minimized.[26] The third examines whether the PIPS balances people well in terms of both network

[25] When we use the variance estimators of Abadie and Imbens (2016), results are qualitatively similar.

[26] We also do a direct test of the correlation of outcomes among people with similar propensity scores, which are reported in Online Appendix OA.3. Those are insignificantly different from 0. These are not a

and demographic characteristics across people with the same propensity scores, which is popular among PSM applications.

5.1. Balance in the Number of Times People are Matched

To check whether the balance condition from Section 3.3 holds, we first verify that no individual is matched more than a few times.

Regarding the first test, using the nearest neighbor matching algorithm, and caliper matchings with radius 0.01 and 0.1, the vast majority of individuals are matched no more than two times, and tiny fractions are ever sampled more than ten times out of hundreds. This is illustrated in Figure OA.1 in the online appendix.

5.2. Balance in the Number of Times People are Matched with a Friend

Second, we consider the fraction of times that an individual is matched with a direct friend. These are the individuals for whom outcomes are most likely to be interfering and hence correlated. This frequency is extremely low, again across all three matching techniques, ranging from 0.002 to 0.005. See Table OA.1 and Figure OA.2 in the online appendix.

5.3. Balance in Characteristics Across Treatment Choices

Third, along the lines of Proposition 1, we can also adapt the proof of Rosenbaum and Rubin (1983) to obtain a network version of a balancing property that examines whether the demographics and network positions of individuals with different treatment choices, but the same PIPS are similar. The following proposition shows that this balance should be a property of the PIPS, and then we verify it in our application.

Proposition **5** (Balancing Property)**.** *Under Assumptions 1 and 2, for every* $i \in \mathcal{I}$,

$$(g, \mathbf{X}) \perp Z_i \mid \pi_i(g, \mathbf{X}).$$

Accordingly, in ideal cases, PIPSM balances not only individual characteristics X_i, but also network characteristics such as local average, local aggregate, in-degree, out-degree, and centrality, between treatment and control groups. We focus on whether PIPSM successfully balances network characteristics in our application.

Table 7 reports the balance test results of caliper matchings with a radius of 0.01 for the first school network; the results for other school networks are in Tables OA.2 to OA.5 of the online appendix. There is an overwhelming advantage of PIPSM over PSM in balancing network characteristics. For the first and third school, only local

direct check of Corollary 1, because we cannot measure the covariances of the Y_is since each one only has a single observation. Nonetheless, the lack of correlations across matchings should hold, and do.

Table 7: Balance Tests for the First School Network

Covariates	Before Matching			After Matching (PSM)			After Matching (PIPSM)		
	Mean		t-test	Mean		t-test	Mean		t-test
	T	C	p-value	T	C	p-value	T	C	p-value
Age	15.050	15.354	0.000	15.053	14.927	0.001	15.133	15.048	0.207
Gender	0.343	0.637	0.000	0.352	0.364	0.040	0.363	0.366	0.789
Hispanic	0.150	0.129	0.157	0.143	0.145	0.927	0.137	0.144	0.671
Black	0.006	0.011	0.170	0.002	0.008	0.115	0.006	0.007	0.681
Asian	0.015	0.015	0.969	0.015	0.019	0.460	0.015	0.023	0.237
Native	0.092	0.064	0.019	0.083	0.085	0.879	0.078	0.068	0.418
Other	0.101	0.104	0.805	0.099	0.101	0.880	0.100	0.103	0.817
Par. Edu.	4.653	4.564	0.285	4.637	4.635	0.898	4.639	4.727	0.334
Local Average	0.385	0.351	0.000	0.383	0.374	0.217	0.377	0.387	0.108
Local Aggregate	2.064	1.726	0.000	2.064	1.763	0.000	1.965	1.951	0.788
Out-Degree	4.685	4.280	0.001	4.695	4.114	0.000	4.538	4.516	0.852
In-Degree	4.938	4.105	0.000	4.946	3.836	0.000	4.801	4.088	0.000
Centrality	0.866	0.735	0.867	0.719	0.000	0.000	0.824	0.804	0.471

Notes: 1. Matchings are subject to a caliper with radius 0.01.
2. T stands for 'Treatment' and C stands for 'Control.'

average is better balanced by PSM (which is the term that is insignificant in determining peer influence); for the second, fourth, and fifth school, PIPSM performs better for all 5 network characteristics considered. The t-tests for PSM are insignificant at the 5% level in only 6 out of $5 \times 5 = 25$ cases; for PIPSM, they are significant in only 6 out of 25 cases. These results confirm that PIPSM handles network characteristics as expected. It is also interesting to note that compared with PSM, PIPSM also improves balance for individual characteristics. This makes sense, since these interact with peer influence in participation decisions, and so failing to take into account those peer effects can over-load individual characteristics in treatment behavior, thus unbalancing traditional PSM.

6. CONCLUDING REMARKS

We have introduced a method for estimating treatment effects when individuals make peer-influenced treatment choices. We use game theory to model the interdependence of participation decisions among individuals in an influence network, and solve for participation probabilities conditional on characteristics and peer participation; i.e., PIPS. We then use the PIPS for matching and estimation of treatment effects.[27] As a

[27] A large class of inverse probability weighting methods (Robins, Hernán, and Brumback, 2000; Austin and Stuart, 2015) can use our peer-influenced propensity scores too, though some researchers call for closer scrutiny of them (Solon, Haider, and Wooldridge, 2015).

result, the identification and estimation of treatment effects based on PIPSM are robust against interference caused by peer effects in program participation. The peer influence on participation, as well as the contrast of PIPSM and traditional PSM, are well illustrated in our application, as well as PIPSM's balancing of network characteristics between treatment and control groups.

We can also imagine a brute-force attack on the problem of carrying out a PSM study of treatment effects in social networks. An econometrician can add various plausibly relevant network characteristics (such as out-degree, in-degree, Bonacich centrality, local average, and local aggregate) to individual characteristics for the purpose of estimating propensity scores. The high-dimensionality of influence networks may necessitate developing a machine learning algorithm. Without an equilibrium concept, such an algorithm may still struggle to capture the appropriate interdependencies in treatment choices.

Our model can be generalized to the case of multi-valued treatment choices. We can use either ordered or multinomial discrete choice models for treatment choice, and follow Imbens (2000) and Angrist and Kuersteiner (2011) in using the resulting propensity scores. Our model also ignores the possibility that participation is not an option for a subset of individuals. The availability of social programs may differ among individuals due to reasons such as phased rollout and intentional randomization. Often, these programs lack full compliance and no disruption, and social networks play key roles in determining treatment choice. We leave these for future work.

It is worth noting that PIPSM inherits the shortcoming common to propensity score methods that it works for a given population, and in this case a given influence network. We are estimating the treatment effects for the treated, untreated, and overall, based on those actual distributions in the population; but in another population with different demographics and different participation decisions one could end up with a very different distribution of who is treated versus untreated, and thus different balances on all treatment effects. For instance, an estimate the impact of a social distancing on infection due to a virus in one country might not be a good predictor of what would happen in another country that has a different distribution of population ages as well as a different interaction network.

REFERENCES

Abadie, A. and M. D. Cattaneo (2018). Econometric methods for program evaluation. *Annual Review of Economics 10*, 465–503.

Abadie, A. and G. W. Imbens (2016). Matching on the estimated propensity score.

Econometrica 84(2), 781–807.

Abbring, J. H. and J. J. Heckman (2007). Econometric evaluation of social programs, part III: Distributional treatment effects, dynamic treatment effects, dynamic discrete choice, and general equilibrium policy evaluation. In *Handbook of Econometrics*, Volume 6, pp. 5145–5303. Elsevier.

Aguirregabiria, V. and P. Mira (2002). Swapping the nested fixed point algorithm: A class of estimators for discrete Markov decision models. *Econometrica 70*(4), 1519–1543.

Aguirregabiria, V. and P. Mira (2007). Sequential estimation of dynamic discrete games. *Econometrica 75*(1), 1–53.

Angrist, J. D. and G. M. Kuersteiner (2004). Semiparametric causality tests using the policy propensity score. *NBER Working Paper*.

Angrist, J. D. and G. M. Kuersteiner (2011). Causal effects of monetary shocks: Semiparametric conditional independence tests with a multinomial propensity score. *Review of Economics and Statistics 93*(3), 725–747.

Aral, S. and C. Nicolaides (2017). Exercise contagion in a global social network. *Nature Communications 8*, 14753.

Aral, S. and D. Walker (2011). Creating social contagion through viral product design: A randomized trial of peer influence in networks. *Management science 57*(9), 1623–1639.

Arcidiacono, P. and S. Nicholson (2005). Peer effects in medical school. *Journal of Public Economics 89*(2-3), 327–350.

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

Athey, S., D. Eckles, and G. W. Imbens (2018). Exact p-values for network interference. *Journal of the American Statistical Association 113*(521), 230–240.

Austin, P. C. and E. A. Stuart (2015). Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. *Statistics in Medicine 34*(28), 3661–3679.

Babyak, M., J. A. Blumenthal, S. Herman, P. Khatri, M. Doraiswamy, K. Moore, W. E. Craighead, T. T. Baldewicz, and K. R. Krishnan (2000). Exercise treatment for major depression: Maintenance of therapeutic benefit at 10 months. *Psychosomatic Medicine 62*(5), 633–638.

Bajari, P., H. Hong, J. Krainer, and D. Nekipelov (2010). Estimating static models of strategic interactions. *Journal of Business & Economic Statistics 28*(4), 469–482.

Balat, J. and S. Han (2020). Multiple treatments with strategic interaction. *Journal of Econometrics*, *forthcoming*.

Banerjee, A., A. G. Chandrasekhar, E. Duflo, J. Floretta, M. O. Jackson, H. Kannan, F. Loza, A. Sanka, and A. Schrimpf (2020). Selecting the most effective nudge: Evidence from a large scale experiment on immunization. *mimeo: MIT*.

Banerjee, A., A. G. Chandrasekhar, E. Duflo, and M. O. Jackson (2013). The diffusion of microfinance. *Science 341*(6144).

Basse, G. W. and E. M. Airoldi (2018). Model-assisted design of experiments in the presence of network-correlated outcomes. *Biometrika 105*(4), 849–858.

Bayer, P., R. Hjalmarsson, and D. Pozen (2009). Building criminal capital behind bars: Peer effects in juvenile corrections. *The Quarterly Journal of Economics 124*(1), 105–147.

Blume, L. E., W. A. Brock, S. N. Durlauf, and R. Jayaraman (2015). Linear social interactions models. *Journal of Political Economy 123*(2), 444–496.

Blundell, R. and M. C. Dias (2009). Alternative approaches to evaluation in empirical microeconomics. *Journal of Human Resources 44*(3), 565–640.

Booij, A. S., E. Leuven, and H. Oosterbeek (2017). Ability peer effects in university: Evidence from a randomized experiment. *The Review of Economic Studies 84*(2), 547–578.

Bramoullé, Y., H. Djebbari, and B. Fortin (2009). Identification of peer effects through social networks. *Journal of Econometrics 150*(1), 41–55.

Bramoullé, Y., H. Djebbari, and B. Fortin (2019). Peer effects in networks: A survey. *Annual Review of Economics*, *forthcoming*.

Bramoullé, Y., R. Kranton, and M. D'amours (2014). Strategic interaction and networks. *American Economic Review 104*(3), 898–930.

Brock, W. A. and S. N. Durlauf (2001a). Discrete choice with social interactions. *The Review of Economic Studies 68*(2), 235–260.

Brock, W. A. and S. N. Durlauf (2001b). Interactions-based models. In *Handbook of Econometrics*, Volume 5, pp. 3297–3380. Elsevier.

Bursztyn, L. and R. Jensen (2017). Social image and economic behavior in the field: Identifying, understanding, and shaping social pressure. *Annual Review of Economics 9*, 131–153.

Caliendo, M. and S. Kopeinig (2008). Some practical guidance for the implementation of propensity score matching. *Journal of Economic Surveys 22*(1), 31–72.

Calvó-Armengol, A., E. Patacchini, and Y. Zenou (2009). Peer effects and social networks in education. *The Review of Economic Studies 76*(4), 1239–1267.

Centola, D. (2018). *How Behavior Spreads: The Science of Complex Contagions*. Princeton University Press.

Chandrasekhar, A. G. and M. O. Jackson (2018). A network formation model based on subgraphs. *Available at SSRN 2660381*.

Chernozhukov, V., H. Hong, and E. Tamer (2007). Estimation and confidence regions for parameter sets in econometric models. *Econometrica 75*(5), 1243–1284.

Cox, D. R. (1958). *Planning of Experiments*. Wiley.

Dahl, G. B., K. V. Løken, and M. Mogstad (2014). Peer effects in program participation. *American Economic Review 104*(7), 2049–74.

De Giorgi, G., M. Pellizzari, and S. Redaelli (2010). Identification of social interactions through partially overlapping peer groups. *American Economic Journal: Applied Economics 2*(2), 241–75.

Dehejia, R. H. and S. Wahba (2002). Propensity score-matching methods for nonexperimental causal studies. *Review of Economics and Statistics 84*(1), 151–161.

DiNardo, J. and D. S. Lee (2011). Program evaluation and research designs. In *Handbook of Labor Economics*, Volume 4, pp. 463–536. Elsevier.

Dishion, T. J. and J. M. Tipsord (2011). Peer contagion in child and adolescent social and emotional development. *Annual Review of Psychology 62*, 189–214.

Duflo, E. and E. Saez (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.

Dunn, A. L., M. H. Trivedi, J. B. Kampert, C. G. Clark, and H. O. Chambliss (2005). Exercise treatment for depression: Efficacy and dose response. *American Journal of Preventive Medicine 28*(1), 1–8.

Durlauf, S. N. and Y. M. Ioannides (2010). Social interactions. *Annual Review of Economics 2*(1), 451–478.

Einav, L., C. Farronato, and J. Levin (2016). Peer-to-peer markets. *Annual Review of Economics 8*, 615–635.

Eisenberger, N. I. and S. W. Cole (2012). Social neuroscience and health: Neurophysiological mechanisms linking social ties with physical health. *Nature Neuroscience 15*(5), 669.

Epple, D. and R. E. Romano (2011). Peer effects in education: A survey of the theory and evidence. In *Handbook of Social Economics*, Volume 1, pp. 1053–1163. Elsevier.

Falk, A. and A. Ichino (2006). Clean evidence on peer effects. *Journal of Labor Economics 24*(1), 39–57.

Feld, J. and U. Zölitz (2017). Understanding peer effects: On the nature, estimation, and channels of peer effects. *Journal of Labor Economics 35*(2), 387–428.

Forastiere, L., E. M. Airoldi, and F. Mealli (2020). Identification and estimation of treatment and interference effects in observational studies on networks. *Journal of the American Statistical Association, forthcoming*.

Frölich, M. (2004). Finite-sample properties of propensity-score matching and weighting estimators. *Review of Economics and Statistics 86*(1), 77–90.

Gaviria, A. and S. Raphael (2001). School-based peer effects and juvenile behavior. *Review of Economics and Statistics 83*(2), 257–268.

Glaeser, E. and J. A. Scheinkman (2003). Non-market interactions. In *Advances in Economics and Econometrics: Theory and Applications, Eight World Congress*. Cambridge University Press.

Glaeser, E. L., B. Sacerdote, and J. A. Scheinkman (1996). Crime and social interactions. *The Quarterly Journal of Economics 111*(2), 507–548.

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

Graham, B. S. (2008). Identifying social interactions through conditional variance restrictions. *Econometrica 76*(3), 643–660.

Guryan, J., K. Kroft, and M. J. Notowidigdo (2009). Peer effects in the workplace: Evidence from random groupings in professional golf tournaments. *American Economic Journal: Applied Economics 1*(4), 34–68.

Harris, K. M., C. T. Halpern, E. Whitsel, J. M. Hussey, J. Tabor, P. P. Entzel, and J. R. Udry (2011). The national longitudinal study of adolescent health: Research design. *http://www.cpc.unc.edu/projects/addhealth/design*.

Heckman, J. and S. Navarro-Lozano (2004). Using matching, instrumental variables, and control functions to estimate economic choice models. *Review of Economics and Statistics 86*(1), 30–57.

Heckman, J. J. and E. Vytlacil (2005). Structural equations, treatment effects, and econometric policy evaluation. *Econometrica 73*(3), 669–738.

Heckman, J. J. and E. J. Vytlacil (2007a). Econometric evaluation of social programs, part I: Causal models, structural models and econometric policy evaluation. In *Handbook of Econometrics*, Volume 6, pp. 4779–4874. Elsevier.

Heckman, J. J. and E. J. Vytlacil (2007b). Econometric evaluation of social programs, part II: Using the marginal treatment effect to organize alternative econometric estimators to evaluate social programs, and to forecast their effects in new environments. In *Handbook of Econometrics*, Volume 6, pp. 4875–5143. Elsevier.

Hirano, K. and J. Hahn (2010). Design of randomized experiments to measure social interaction effects. *Economics Letters 106*(1), 51–53.

Hirano, K., G. W. Imbens, and G. Ridder (2003). Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica 71*(4), 1161–1189.

Holland, P. W. (1986). Statistics and causal inference. *Journal of the American statistical Association 81*(396), 945–960.

Horst, U. and J. A. Scheinkman (2006). Equilibria in systems of social interactions. *Journal of Economic Theory 130*(1), 44–77.

Horst, U. and J. A. Scheinkman (2009). A limit theorem for systems of social interactions. *Journal of Mathematical Economics 45*(9-10), 609–623.

Hoshino, T. and T. Yanagi (2020). Treatment effect models with strategic interaction in treatment decisions. *Available at SSRN 3270447*.

Hoxby, C. (2000). Peer effects in the classroom: Learning from gender and race variation. *NBER Working Paper*.

Hu, Y. and Z. Lin (2020). Binary choice with misclassification and social interactions, with an application to peer effects in attitude. *Working Paper*.

Huber, M., M. Lechner, and C. Wunsch (2013). The performance of estimators based on the propensity score. *Journal of Econometrics 175*(1), 1–21.

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

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

Imbens, G. W. (2004). Nonparametric estimation of average treatment effects under exogeneity: A review. *Review of Economics and Statistics 86*(1), 4–29.

Imbens, G. W. (2015). Matching methods in practice: Three examples. *Journal of Human Resources 50*(2), 373–419.

Imbens, G. W. and D. B. Rubin (2015). *Causal Inference in Statistics, Social, and Biomedical Sciences*. Cambridge University Press.

Imbens, G. W. and J. M. Wooldridge (2009). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature 47*(1), 5–86.

Jackson, M. O. (2008). *Social and Economic Networks*. Princeton University Press.

Jackson, M. O. (2019). *The Human Network: How Your Social Position Determines Your Power, Beliefs, and Behaviors*. Pantheon Books: New York.

Jackson, M. O., B. W. Rogers, and Y. Zenou (2017). The economic consequences of social-network structure. *Journal of Economic Literature 55*(1), 49–95.

Jackson, M. O. and E. C. Storms (2017). Behavioral communities and the atomic structure of networks. *Available at Arxiv 1710.04656*.

Jackson, M. O. and Y. Zenou (2015). Games on networks. In *Handbook of Game Theory with Economic Applications*, Volume 4, pp. 95–163. Elsevier.

Kasahara, H. and K. Shimotsu (2012). Sequential estimation of structural models with a fixed point constraint. *Econometrica 80*(5), 2303–2319.

Kawachi, I. and L. F. Berkman (2001). Social ties and mental health. *Journal of Urban Health 78*(3), 458–467.

King, G. and R. Nielsen (2019). Why propensity scores should not be used for matching. *Political Analysis 27*(4), 435–454.

Kline, B. and E. Tamer (2020). Econometric analysis of models with social interactions. In *The Econometric Analysis of Network Data*, pp. 149–181. Academic Press.

Lawlor, D. A. and S. W. Hopker (2001). The effectiveness of exercise as an intervention in the management of depression: Systematic review and meta-regression analysis of randomised controlled trials. *BMJ 322*(7289), 763.

Lazzati, N. (2015). Treatment response with social interactions: Partial identification via monotone comparative statics. *Quantitative Economics 6*(1), 49–83.

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.

Lee, L.-f., J. Li, and X. Lin (2014). Binary choice models with social network under heterogeneous rational expectations. *Review of Economics and Statistics 96*(3), 402–417.

Lee, L.-f., X. Liu, and X. Lin (2010). Specification and estimation of social interaction models with network structures. *The Econometrics Journal 13*(2), 145–176.

Lee, M.-j. (2016). *Matching, Regression Discontinuity, Difference in Differences, and Beyond*. Oxford University Press.

Leung, M. P. (2019). Treatment and spillover effects under network interference. *The Review of Economics and Statistics*, 1–42.

Lin, Z. (2019). Social interactions quantiles. *Working Paper*.

Lin, Z. and H. Xu (2017). Estimation of social-influence-dependent peer pressures in a large network game. *The Econometrics Journal 20*(3), 86–102.

Liu, L. and M. G. Hudgens (2014). Large sample randomization inference of causal effects in the presence of interference. *Journal of the American Statistical Association 109*(505), 288–301.

Liu, X., E. Patacchini, and Y. Zenou (2014). Endogenous peer effects: Local aggregate or local average? *Journal of Economic Behavior & Organization 103*, 39–59.

de Paula, Á. (2013). Econometric analysis of games with multiple equilibria. *Annual Review of Economics 5*(1), 107–131.

de Paula, Á. and X. Tang (2012). Inference of signs of interaction effects in simultaneous games with incomplete information. *Econometrica 80*(1), 143–172.

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *The Review of Economic Studies 60*(3), 531–542.

Manski, C. F. (2000). Economic analysis of social interactions. *Journal of Economic Perspectives 14*(3), 115–136.

Manski, C. F. (2013). Identification of treatment response with social interactions. *The Econometrics Journal 16*(1), 1–23.

Mcfadden, D. (1974). Conditional logit analysis of qualitative choice behavior. *Frontiers in Econometrics*, 105–142.

Mogstad, M. and A. Torgovitsky (2018). Identification and extrapolation of causal effects with instrumental variables. *Annual Review of Economics 10*, 577–613.

Morse, A. (2015). Peer-to-peer crowdfunding: Information and the potential for disruption in consumer lending. *Annual Review of Financial Economics 7*, 463–482.

Newey, W. K. and D. McFadden (1994). Large sample estimation and hypothesis testing. In *Handbook of Econometrics*, Volume 4, pp. 2111–2245. Elsevier.

Ogburn, E. L. and T. J. VanderWeele (2014). Causal diagrams for interference. *Statistical Science 29*(4), 559–578.

Perez-Heydrich, C., M. G. Hudgens, M. E. Halloran, J. D. Clemens, M. Ali, and M. E. Emch (2014). Assessing effects of cholera vaccination in the presence of interference. *Biometrics 70*(3), 731–741.

Robins, J. M., M. A. Hernán, and B. Brumback (2000). Marginal structural models and causal inference in epidemiology. *Epidemiology 11*(5), 551.

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

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

Rubin, D. B. (1980). Randomization analysis of experimental data: The fisher randomization test comment. *Journal of the American statistical association 75*(371), 591–593.

Sacerdote, B. (2001). Peer effects with random assignment: Results for Dartmouth roommates. *The Quarterly Journal of Economics 116*(2), 681–704.

Sacerdote, B. (2011). Peer effects in education: How might they work, how big are they and how much do we know thus far? In *Handbook of the Economics of Education*, Volume 3, pp. 249–277. Elsevier.

Sacerdote, B. (2014). Experimental and quasi-experimental analysis of peer effects: Two steps forward? *Annual Review of Economics 6*(1), 253–272.

Salmon, P. (2001). Effects of physical exercise on anxiety, depression, and sensitivity to stress: A unifying theory. *Clinical Psychology Review 21*(1), 33–61.

Sofrygin, O. and M. J. van der Laan (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).

Solon, G., S. J. Haider, and J. M. Wooldridge (2015). What are we weighting for? *Journal of Human Resources 50*(2), 301–316.

Ströhle, A. (2009). Physical activity, exercise, depression and anxiety disorders. *Journal of Neural Transmission 116*(6), 777.

Tamer, E. (2003). Incomplete simultaneous discrete response model with multiple equilibria. *The Review of Economic Studies 70*(1), 147–165.

Tamer, E. (2010). Partial identification in econometrics. *Annual Review of Economics 2*(1), 167–195.

Tchetgen, E. J. and T. J. VanderWeele (2012). On causal inference in the presence of interference. *Statistical Methods in Medical Research 21*(1), 55–75.

Thoits, P. A. (2011). Mechanisms linking social ties and support to physical and mental health. *Journal of Health and Social Behavior 52*(2), 145–161.

Toulis, P. and E. Kao (2013). Estimation of causal peer influence effects. In *International Conference on Machine Learning*, pp. 1489–1497.

van der Laan, M. J. (2014). Causal inference for a population of causally connected units. *Journal of Causal Inference 2*(1), 13–74.

Waldinger, F. (2011). Peer effects in science: Evidence from the dismissal of scientists in Nazi Germany. *The Review of Economic Studies 79*(2), 838–861.

White, H. (2000). *Asymptotic Theory for Econometricians*. Academic Press.

Xu, H. (2018). Social interactions on large networks: A game theoretic approach. *International Economic Review 59*(1), 257–284.

Zimmerman, D. J. (2003). Peer effects in academic outcomes: Evidence from a natural experiment. *Review of Economics and Statistics 85*(1), 9–23.

APPENDIX A

A.1 Proof of Proposition 1

We show that $\Pr(Z_i = 1 \,|\, Y_i(1), Y_i(0), \pi_i(g,\mathbf{X})) = \Pr(Z_i = 1 \,|\, \pi_i(g,\mathbf{X}))$, which implies the independence of $(Y_i(1), Y_i(0))$ and Z_i conditional upon $\pi_i(g,\mathbf{X})$. This follows from these equalities:

$$\begin{aligned}
\Pr(Z_i = 1 \,|\, Y_i(1), Y_i(0), \pi_i(g,\mathbf{X})) &= \mathbb{E}(Z_i \,|\, Y_i(1), Y_i(0), \pi_i(g,\mathbf{X})) \\
&= \mathbb{E}\left[\mathbb{E}(Z_i \,|\, Y_i(1), Y_i(0), \pi_i(g,\mathbf{X}), g, \mathbf{X}) \,|\, Y_i(1), Y_i(0), \pi_i(g,\mathbf{X})\right] \\
&= \mathbb{E}\left[\mathbb{E}(Z_i \,|\, Y_i(1), Y_i(0), g, \mathbf{X}) \,|\, Y_i(1), Y_i(0), \pi_i(g,\mathbf{X})\right] \\
&= \mathbb{E}\left[\mathbb{E}(Z_i \,|\, g, \mathbf{X}) \,|\, Y_i(1), Y_i(0), \pi_i(g,\mathbf{X})\right] \\
&= \mathbb{E}\left[\pi_i(g,\mathbf{X}) \,|\, Y_i(1), Y_i(0), \pi_i(g,\mathbf{X})\right] \\
&= \pi_i(g,\mathbf{X}) \\
&= \Pr(Z_i = 1 \,|\, \pi_i(g,\mathbf{X})),
\end{aligned}$$

where the third equality is from the law of iterated expectation and the fifth from the SCI assumption. In other words, $(Y_i(1), Y_i(0))$ and Z_i are independent conditional upon $\pi_i(g,\mathbf{X})$.

A.2 Proof of Proposition 2

Start with any $\pi^{(0)}, \pi^{(1)} \in [0,1]^{\mathcal{I}}$ and let $\pi^{(2)} := \mathrm{BR}(\pi^{(0)})$ and $\pi^{(3)} := \mathrm{BR}(\pi^{(1)})$. For every $i \in \mathcal{I}$ with $d_i > 0$, it follows that

$$\begin{aligned}
&|\pi_i^{(3)} - \pi_i^{(2)}| \\
&= \left| \Lambda\left(\mathbf{X}_i^T\beta + \frac{\alpha_1}{d_i}\sum_{j\in N_i(g)}\pi_j^{(1)} + \alpha_2\sum_{j\in N_i(g)}\pi_j^{(1)}\right) - \Lambda\left(\mathbf{X}_i^T\beta + \frac{\alpha_1}{d_i}\sum_{j\in N_i(g)}\pi_j^{(0)} + \alpha_2\sum_{j\in N_i(g)}\pi_j^{(0)}\right)\right| \\
&= \lambda\left(\mathbf{X}_i^T\beta + \frac{\alpha_1}{d_i}\sum_{j\in N_i(g)}\tilde{\pi}_j + \alpha_2\sum_{j\in N_i(g)}\tilde{\pi}_j\right)\cdot\left(\frac{\alpha_1}{d_i} + \alpha_2\right)\cdot\left|\sum_{j\in N_i(g)}(\pi_j^{(1)} - \pi_j^{(0)})\right| \\
&\le \lambda\left(\mathbf{X}_i^T\beta + \alpha_1\frac{1}{d_i}\sum_{j\in N_i(g)}\tilde{\pi}_j + \alpha_2\sum_{j\in N_i(g)}\tilde{\pi}_j\right)\cdot(\alpha_1 + \alpha_2)\cdot\left|\sum_{j\in N_i(g)}(\pi_j^{(1)} - \pi_j^{(0)})\right| \\
&\le \lambda\left(\mathbf{X}_i^T\beta + \alpha_1\frac{1}{d_i}\sum_{j\in N_i(g)}\tilde{\pi}_j + \alpha_2\sum_{j\in N_i(g)}\tilde{\pi}_j\right)\cdot(\alpha_1 + \alpha_2)M\cdot\max_{j\in N_i(g)}\left|\pi_j^{(1)} - \pi_j^{(0)}\right| \\
&\le \frac{M(\alpha_1+\alpha_2)}{4}\max_{j\in\mathcal{I}}\left|\pi_j^{(1)} - \pi_j^{(0)}\right|,
\end{aligned}$$

where the second equality follows from the mean value theorem (which implies the existence of all $\tilde{\pi}_j \in [0,1]$ that makes the equality valid), and the third inequality from $\lambda(\cdot) \le \frac{1}{4}$. This inequality remains valid for the case of $d_i = 0$.

It then follows that $\Delta(\pi^{(3)}, \pi^{(2)}) \le \frac{M(\alpha_1+\alpha_2)}{4}\Delta(\pi^{(1)}, \pi^{(0)})$, where $(\alpha_1 + \alpha_2) \in (0, \frac{4}{M})$. By the contraction mapping theorem, there exists a unique fixed point for BR in $[0,1]^{\mathcal{I}}$, say π^*.

So $\pi^* = \mathrm{BR}(\pi^*)$. Since $\pi = \mathrm{BR}(\pi)$ if and only if π describes an equilibrium, π^* is the unique equilibrium of the participation game.

A.3 Proof of Proposition 3

Define $N_{(i,h)}$ as the subnetwork where any individuals in the subnetwork is within social distance of h from individual i. Let $(g,\mathbf{X})_i^h$ be network and demographic information associated with this subnetwork, $\overline{(g,\mathbf{X})_i^{>h}}$ be one realization of public information beyond it, and $\overline{\pi}_j := \mathbb{E}(Z_j \,|\, (g,\mathbf{X})_i^h, \overline{(g,\mathbf{X})_i^{>h}})$.

For every $j \in N_{(i,h)}$, it follows that $d_j > 0$, and by the mean value theorem,

$$\begin{aligned}\left|\pi_j^* - \overline{\pi}_j\right| &= \lambda\left(\mathbf{X}_i^T\beta + \alpha_1 \frac{1}{d_j}\sum_{j'\in N_j(g)} \tilde{\pi}_{j'} + \alpha_2 \sum_{j'\in N_j(g)} \tilde{\pi}_{j'}\right)\cdot\left(\frac{\alpha_1}{d_i} + \alpha_2\right)\cdot\left|\sum_{j'\in N_j(g)} (\pi_{j'}^* - \overline{\pi}_{j'})\right| \\ &\le \frac{M(\alpha_1+\alpha_2)}{4}\max_{j'\in N_j(g)} |\pi_{j'}^* - \overline{\pi}_{j'}| \\ &\le \frac{M(\alpha_1+\alpha_2)}{4},\end{aligned}$$

where $\tilde{\pi}_{j'}$ is some real number between π_{j^*} and $\overline{\pi}_{j'}$. For every $j \in N_{(i,h-1)}$, because any influencer j' of j belongs to $N_{(i,h)}$, it follows that

$$\left|\pi_j^* - \overline{\pi}_j\right| \le \left(\frac{M(\alpha_1+\alpha_2)}{4}\right)^2 \max_{j''\in N_{j'}(g), j'\in N_j(g)} |\pi_{j''}^* - \overline{\pi}_{j''}| \le \left(\frac{M(\alpha_1+\alpha_2)}{4}\right)^2.$$

By induction, for any $q \le h$,

$$\max_{j\in N_{(i,h-q)}} \left|\pi_j^* - \overline{\pi}_j\right| \le \left(\frac{M(\alpha_1+\alpha_2)}{4}\right)^{q+1}.$$

Since $i \in N_{(i,0)}$, we have $\left|\pi_i^* - \overline{\pi}_i\right| \le \left(\frac{M(\alpha_1+\alpha_2)}{4}\right)^{h+1} \to 0$ as $h \to \infty$ by Assumption 5.

A.4 Identification of θ in PIPS

Let $R_n := \left(\log\left(\frac{\pi_1^*}{1-\pi_1^*}\right), \cdots, \log\left(\frac{\pi_n^*}{1-\pi_n^*}\right)\right)^T$.

Lemma **1.** *Suppose Assumptions 3 to 6 hold. For all n sufficiently large, we identify the structural parameters θ by*

$$\theta = \mathbb{E}\left(\frac{1}{n}\overline{X}^T\overline{X}\right)^{-1} \cdot \mathbb{E}\left(\frac{1}{n}\overline{X}^T R_n\right). \tag{.1}$$

Proof. By the Logistic distribution of the random utility shock, it follows that $\log\left(\frac{\pi_i^*}{1-\pi_i^*}\right) = \overline{X}_i^T\theta$ for every $i \in \mathcal{I}$; that is, $R_n = \overline{X}\theta$. So $\overline{X}^T R_n = \overline{X}^T\overline{X}\theta$.

The closed form of θ in Equation (.1) immediately follows once we have the rank condition satisfied in Assumption 6 and $\{\varepsilon_i\}_{i\in\mathcal{I}}$ identified from the data. This is the identification strategy used for the independent heterogeneously distributed observations (see §3.2 in White, 2000). □

A.5 Proof of Proposition 4

Let[28]

$$\tau_{\text{AT}}(\pi) \coloneqq \mathbb{E}(Y_i \,|\, Z_i = 1, \pi_i(g, \mathbf{X}) = \pi) - \mathbb{E}(Y_i \,|\, Z_i = 0, \pi_i(g, \mathbf{X}) = \pi).$$

Note that

$$\begin{aligned}\widetilde{\tau}_{\text{ATT}}(\pi) &= \frac{1}{|\mathcal{T}(\pi)|} \sum_{i \in \mathcal{T}(\pi)} \left(Y_i - \frac{1}{|\mathcal{M}_i(\pi)|} \sum_{j \in \mathcal{M}_i(\pi)} Y_j \right) \\ &= \left(\frac{1}{|\mathcal{T}(\pi)|} \sum_{i \in \mathcal{T}(\pi)} Y_i \right) - \left(\frac{1}{|\mathcal{U}(\pi)|} \sum_{j \in \mathcal{U}(\pi)} Y_j \right),\end{aligned}$$

where the second equation follows from the balance condition.

By Theorem 6.1 (see Corollary 6.1) in Chandrasekhar and Jackson (2018), it follows that each of the two sums converges to their analogs in $\tau_{\text{AT}}(\pi)$ defined above. The same is true for $\widetilde{\tau}_{\text{ATU}}(\pi)$ and $\widetilde{\tau}_{\text{ATE}}(\pi)$, establishing the claim.

A.6 Consistency of Nested Pseudo-Likelihood Estimation

We need two additional assumptions for the consistency and asymptotic normality of nested pseudo-likelihood estimators.

Let $\overline{X}_i \coloneqq \left(\mathbf{X}_i^T, \frac{1}{d_i} \sum_{j \in N_i(g)} \pi_j^*, \sum_{j \in N_i(g)} \pi_j^* \right)^T$ and $\overline{X} \coloneqq \left(\overline{X}_1, \cdots, \overline{X}_n \right)^T$.

Assumption 6. *$M_n \coloneqq \mathbb{E}\left(\overline{X}^T \overline{X} / n \right)$ is uniformly positive definite.*

Assumption 6 is used for identification. Assumption 4 essentially rules out the perfect collinearity between $\frac{1}{d_i} \sum_{j \in N_i(g)} \pi_j^*$ and $\sum_{j \in N_i(g)} \pi_j^*$. With unique equilibrium, we identify the conditional choice probabilities, $\{\mathbb{E}(Z_i | g, \mathbf{X})\}_{i \in \mathcal{I}}$, from the data by the identification in infinity argument. We then identify the structural parameters θ with Assumption 6 using a strategy of Logit regression for independent heterogeneously distributed observations building upon the variation of the out-degree from Assumption 4. We summarize the identification result in Lemma 1 of Appendix A.4.

The following notation is useful in defining the last assumption needed for the consistency results. Let $\mathcal{L}(\theta, \pi) \coloneqq \frac{1}{n} \sum_{i=1}^n \mathcal{L}_i(\theta, \pi)$ be the pseudo likelihood function associated with the Logit regression of Z_i on $\mathbf{X}_i$, $\frac{1}{d_i} \sum_{j \in N_i(g)} \pi_j$, and $\sum_{j \in N_i(g)} \pi_j$, where $\pi \in \Pi \subseteq [0,1]^{\mathcal{I}}$ is a profile of arbitrary choice probabilities. Set $\mathcal{L}_0(\theta, \pi) \coloneqq \lim_{n \to \infty} \mathbb{E}[\mathcal{L}(\theta, \pi)]$. Further, let $\tilde{\theta}_0(\pi) \coloneqq \arg\max_{\theta \in \Theta} \mathcal{L}_0(\theta, \pi)$, $\phi_0(\pi) \coloneqq \text{BR}(\tilde{\theta}_0(\pi), \pi)$[29], $\tilde{\theta}_n(\pi) \coloneqq \arg\max_{\theta \in \Theta} \mathcal{L}(\theta, \pi)$, and $\phi_n(\pi) \coloneqq \text{BR}(\tilde{\theta}_n(\pi), \pi)$. Define the population nested pseudo likelihood fixed points set as $\Lambda_0 \equiv \{(\theta, \pi) \in (\Theta, \Pi) : \theta = \tilde{\theta}_0(\pi), \pi = \phi_0(\pi)\}$ and the nested pseudo likelihood fixed points set of sample size n as $\Lambda_n \equiv \{(\theta, \pi) \in (\Theta, \Pi) : \theta = \tilde{\theta}_n(\pi), \pi = \phi_n(\pi)\}$. Let $\mathcal{N}$ denote a closed neighborhood of (θ_0, π^*).

[28] Note that once we condition on π, this is just an average treatment for those with π, so we call this AT, and it is the same benchmark for convergence in this proof regardless of whether we work with $\widetilde{\tau}_{\text{ATT}}(\pi)$, $\widetilde{\tau}_{\text{ATU}}(\pi)$, or $\widetilde{\tau}_{\text{ATE}}(\pi)$.

[29] We bring in the parameter in the best response functions.

Assumption 7. *(i) The parameter space Θ for θ is compact and the true parameter θ_0 is an interior point of Θ; (ii) (θ_0, π^*) is an isolated population nested pseudo likelihood fixed point; (iii) $\mathcal{L}_0(\theta, \pi)$ is globally concave and its second derivative with respect to θ is a nonsingular matrix in a neighborhood of π^*; (iv) the operator $\phi_0(\pi) - \pi$ has a nonsingular Jacobian matrix at π^*; (v) there exist nonsingular matrices $\Omega_1(\theta_0)$ and $\Omega_2(\theta_0)$ such that*

$$\mathbb{E}\left\{\frac{\partial^2 \mathcal{L}(\theta_0, \pi^*)}{\partial\theta\partial\theta^T} + \frac{\partial^2 \mathcal{L}(\theta_0, \pi^*)}{\partial\theta\partial\pi^T} \cdot \left[I - \left(\frac{\partial \mathrm{BR}(\pi^*; \theta_0)}{\partial\pi}\right)^T\right]^{-1} \cdot \frac{\partial \mathrm{BR}(\pi^*; \theta_0)}{\partial\theta^T}\right\} \xrightarrow{p} \Omega_1(\theta_0),$$

$$\text{and} \quad \mathbb{E}\left(n \frac{\partial \mathcal{L}(\theta_0, \pi^*)}{\partial\theta} \frac{\partial \mathcal{L}(\theta_0, \pi^*)}{\partial\theta^T}\right) \xrightarrow{p} \Omega_2(\theta_0),$$

where $\Omega_1(\theta_0)$ is also negative definite.

Assumption 7 provides a series of regularity conditions that ensures that behavior is well-behaved enough in some neighborhood to ensure that parameter estimates are also well-behaved. Assumption 7(i) is a standard condition for likelihood-type estimator in the literature (see Newey and McFadden, 1994). Assumption 7(ii) is essentially an identification assumption for using the nested pseudo likelihood algorithm. It is straightforward to show that $\theta_0 = \tilde{\theta}_0(\pi^*)$ solves $\arg\max_{\theta\in\Theta} \mathcal{L}_0(\theta, \pi^*)$. Without Assumption 7(ii), $\arg\max_{\theta\in\Theta} \mathcal{L}_0(\theta, \pi^*)$ might admit multiple solutions and each of these represents a fixed point of the nested pseudo likelihood algorithm. Assumption 7(iii) guarantees that $\tilde{\theta}_0(\pi)$ is a single-valued and continuous function in a neighborhood of π^* which is required in the proof of consistency. We need Assumption 7(iv) in order to assure that there exists a sample nested pseudo likelihood fixed point close to (θ_0, π^*). The condition states that the Jacobian of the population nested pseudo likelihood operator does not have any eigenvalues in the unit circle. Assumption 7 (v) is a high-level condition for non-singular limiting matrices as n goes to infinity. Moreover, the non-degeneracy of $\Omega_1(\theta_0)$ and $\Omega_2(\theta_0)$ requires that all of the determinants of the finite counterparts are outside an open ball of zero for all n, which is essentially a rank condition.

The next proposition follows from a result of Aguirregabiria and Mira (2007), Lin and Xu (2017), and Hu and Lin (2020), and so is offered without proof.

Proposition **6.** *Under Assumptions 3 to 7, it follows that $\hat{\theta} \xrightarrow{P} \theta_0$ and*

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} \mathcal{N}\left(0, \Omega_1^{-1}(\theta_0)\Omega_2(\theta_0)\Omega_1^T(\theta_0)^{-1}\right).$$

Adjusting for Peer-Influence in Propensity Scoring When Estimating Treatment Effects

Matthew O. Jackson, Zhongjian Lin, and Ning Neil Yu

Online Appendix

OA.1. Balanced Matching in the Empirical Illustration

Table OA.1: The Fraction of Individuals Matched with an Influencer

School	Nearest Neighbor	Caliper (0.01)	Caliper (0.1)
1	0.003	0.003	0.003
2	0.002	0.002	0.002
3	0.004	0.003	0.004
4	0.003	0.003	0.003
5	0.005	0.004	0.005

Figure OA.1: Histograms for the Numbers of Times of Being Matched

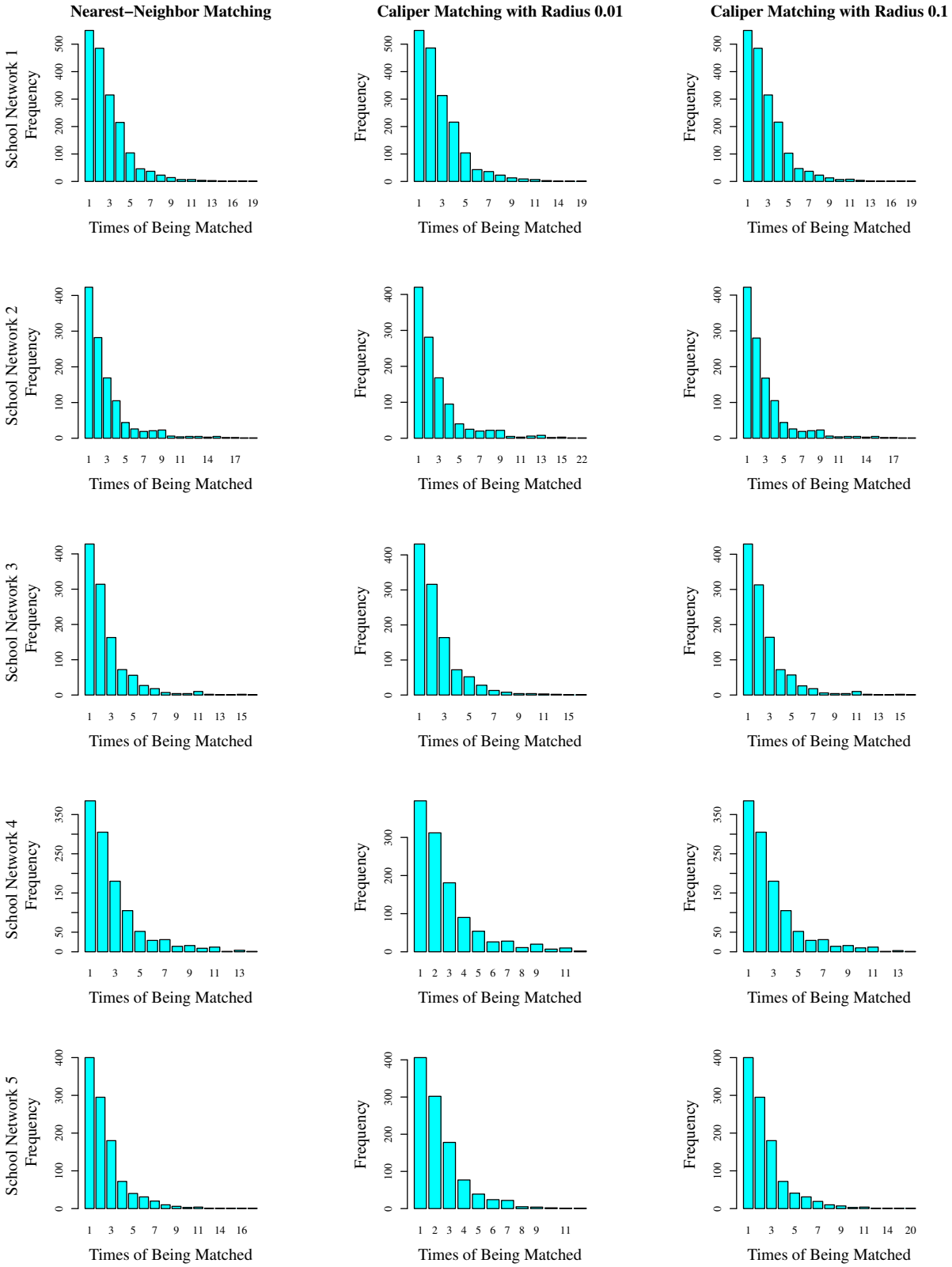
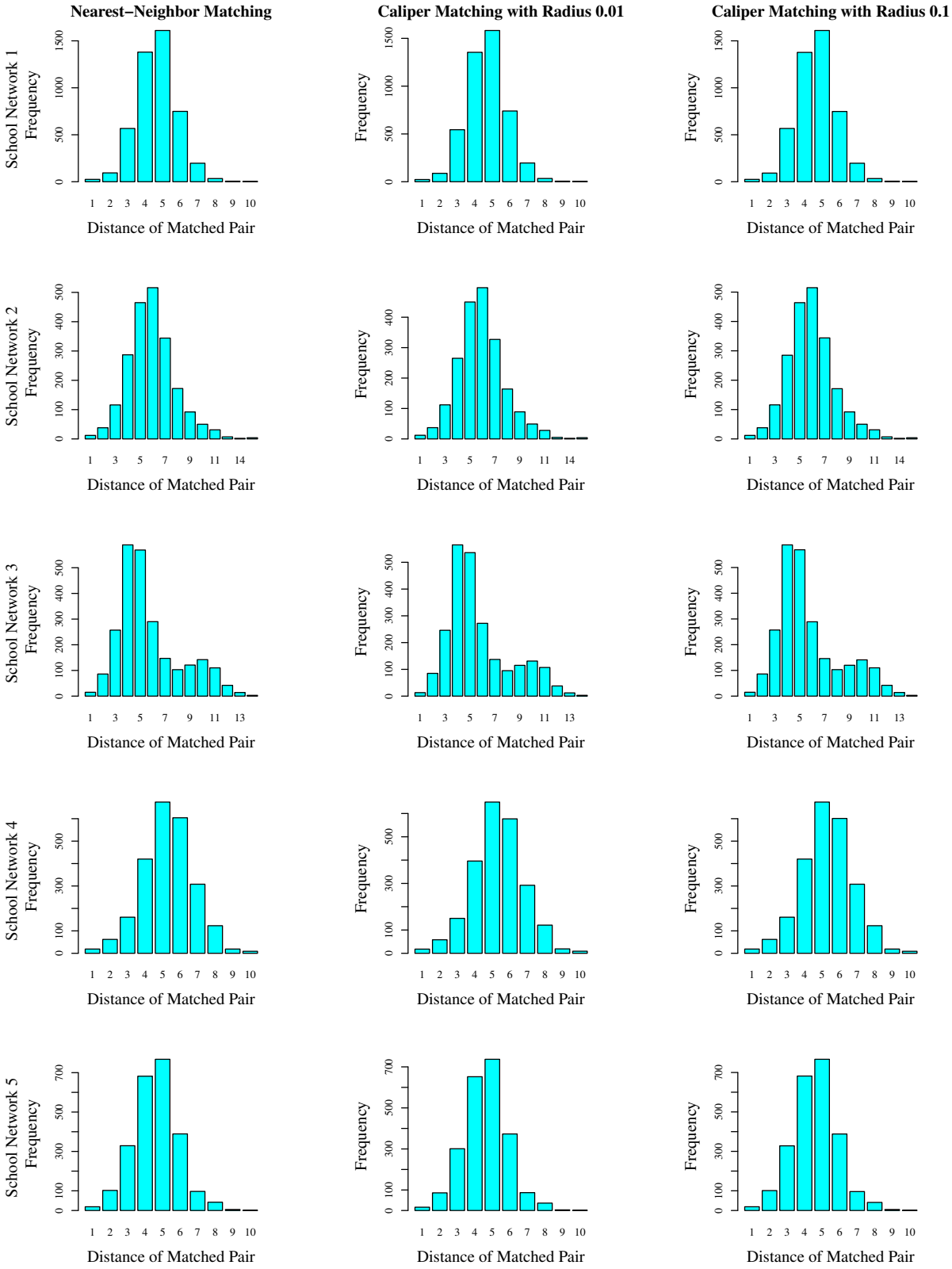


Figure OA.2: Histograms for the Distances between Matched Pair



OA.2. Balanced Characteristics in the Empirical Illustration

Table OA.2: Balance Tests for the Second School Network

Covariates	Before Matching			After Matching (PSM)			After Matching (PIPSM)		
	Mean		t-test	Mean		t-test	Mean		t-test
	T	C	p-value	T	C	p-value	T	C	p-value
Age	15.773	15.904	0.109	15.805	15.822	0.765	15.819	15.884	0.443
Gender	0.325	0.631	0.000	0.340	0.328	0.065	0.340	0.321	0.016
Hispanic	0.867	0.904	0.045	0.891	0.929	0.012	0.898	0.903	0.746
Black	0.098	0.083	0.347	0.097	0.071	0.092	0.081	0.079	0.884
Asian	0.017	0.016	0.884	0.014	0.002	0.052	0.014	0.206	0.451
Native	0.039	0.017	0.027	0.018	0.021	0.777	0.019	0.008	0.202
Other	0.349	0.303	0.083	0.359	0.369	0.487	0.356	0.348	0.787
Par. Edu.	5.288	5.006	0.054	5.262	5.168	0.415	5.279	5.271	0.959
Local Average	0.200	0.194	0.515	0.198	0.187	0.353	0.200	0.198	0.810
Local Aggregate	0.652	0.621	0.429	0.636	0.579	0.219	0.640	0.617	0.565
Out-Degree	2.120	2.198	0.524	2.087	1.925	0.263	2.104	2.087	0.898
In-Degree	2.325	2.113	0.105	2.336	1.993	0.024	2.326	2.171	0.331
Centrality	0.653	0.683	0.454	0.644	0.599	0.356	0.647	0.657	0.838

Table OA.3: Balance Tests for the Third School Network

Covariates	Before Matching			After Matching (PSM)			After Matching (PIPSM)		
	Mean		t-test	Mean		t-test	Mean		t-test
	T	C	p-value	T	C	p-value	T	C	p-value
Age	15.063	15.381	0.000	15.126	15.005	0.002	15.148	15.145	0.975
Gender	0.341	0.631	0.000	0.355	0.364	0.225	0.369	0.370	0.930
Hispanic	0.048	0.053	0.655	0.042	0.035	0.505	0.045	0.048	0.804
Black	0.035	0.024	0.225	0.021	0.022	0.896	0.025	0.031	0.572
Asian	0.015	0.035	0.011	0.012	0.014	0.696	0.013	0.008	0.480
Native	0.063	0.037	0.026	0.040	0.038	0.807	0.040	0.042	0.827
Other	0.048	0.042	0.600	0.045	0.030	0.126	0.049	0.039	0.416
Par. Edu.	5.493	5.161	0.001	5.449	5.391	0.301	5.417	5.356	0.575
Local Average	0.363	0.330	0.000	0.359	0.363	0.684	0.357	0.367	0.303
Local Aggregate	1.825	1.563	0.000	1.839	1.614	0.002	1.771	1.755	0.811
Out-Degree	4.252	4.021	0.126	4.311	3.882	0.009	4.182	4.193	0.944
In-Degree	4.488	3.866	0.001	4.589	3.702	0.000	4.437	3.932	0.006
Centrality	0.812	0.704	0.003	0.833	0.683	0.000	0.804	0.757	0.223

Table OA.4: Balance Tests for the Fourth School Network

Covariates	Before Matching			After Matching (PSM)			After Matching (PIPSM)		
	Mean		t-test	Mean		t-test	Mean		t-test
	T	C	p-value	T	C	p-value	T	C	p-value
Age	15.025	15.284	0.001	15.093	15.097	0.893	15.163	15.045	0.093
Gender	0.351	0.654	0.000	0.365	0.364	0.683	0.377	0.395	0.013
Hispanic	0.046	0.040	0.600	0.038	0.021	0.024	0.039	0.061	0.098
Black	0.130	0.151	0.225	0.128	0.136	0.673	0.133	0.163	0.161
Asian	0.036	0.032	0.638	0.029	0.020	0.191	0.038	0.030	0.455
Native	0.026	0.025	0.887	0.019	0.012	0.248	0.022	0.030	0.400
Other	0.021	0.021	0.935	0.019	0.003	0.012	0.018	0.020	0.803
Par. Edu.	6.634	6.395	0.005	6.663	6.725	0.151	6.584	6.577	0.934
Local Average	0.342	0.300	0.000	0.343	0.310	0.004	0.333	0.334	0.908
Local Aggregate	1.285	1.082	1.303	1.078	0.000	0.000	1.222	1.234	0.794
Out-Degree	3.008	2.792	0.069	3.064	2.591	0.000	2.925	2.976	0.654
In-Degree	3.087	2.740	0.028	3.113	2.517	0.001	3.023	2.657	0.037
Centrality	0.803	0.716	0.013	0.817	0.672	0.000	0.774	0.768	0.854

Table OA.5: Balance Tests for the Fifth School Network

Covariates	Before Matching			After Matching (PSM)			After Matching (PIPSM)		
	Mean		t-test	Mean		t-test	Mean		t-test
	T	C	p-value	T	C	p-value	T	C	p-value
Age	14.548	14.903	0.000	14.625	14.660	0.508	14.613	14.834	0.010
Gender	0.373	0.670	0.000	0.401	0.408	0.248	0.409	0.417	0.593
Hispanic	0.121	0.226	0.000	0.125	0.105	0.042	0.129	0.124	0.760
Black	0.034	0.040	0.510	0.035	0.025	0.160	0.036	0.036	0.989
Asian	0.024	0.024	0.973	0.025	0.008	0.007	0.023	0.027	0.561
Native	0.063	0.064	0.912	0.040	0.031	0.178	0.061	0.058	0.786
Other	0.083	0.143	0.000	0.085	0.076	0.442	0.089	0.081	0.548
Par. Edu.	5.415	4.958	0.000	5.353	5.285	0.446	5.336	5.487	0.139
Local Average	0.537	0.466	0.000	0.539	0.497	0.000	0.526	0.535	0.348
Local Aggregate	2.692	2.062	0.000	2.705	2.290	0.000	2.518	2.691	0.024
Out-Degree	4.466	3.742	0.000	4.517	3.925	0.000	4.242	4.559	0.011
In-Degree	4.529	3.664	0.000	4.527	3.785	0.000	4.398	4.175	0.231
Centrality	0.873	0.665	0.000	0.883	0.718	0.000	0.813	0.885	0.015

OA.3. Correlation Check of $Y(0)$s and $Y(1)$s

These examine the lack of overall correlation between the $Y(1)$s and $Y(0)$s of people who have similar propensity scores. The propensity scores are split evenly into 5 bins (Table OA.6) or 10 bins (Table OA.7), across their range. The correlations are then reported for each bin.

Table OA.6: Correlation of $Y(0)$s and $Y(1)$s in Bins in Empirical Illustration

	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5
School Network 1 Correlation	0.009	0.073	0.002	0.027	-0.088
p-value	0.830	0.085	0.964	0.552	0.414
Number of Pairs	585	561	468	494	88
Proportion	0.265	0.254	0.212	0.224	0.040
School Network 2 Correlation	0.004	0.140	-0.059	-0.120	0.021
p-value	0.906	0.094	0.238	0.051	0.911
Number of Pairs	700	145	402	266	32
Proportion	0.452	0.094	0.259	0.172	0.021
School Network 3 Correlation	-0.091	0.051	-0.060	-0.013	0.020
p-value	0.152	0.257	0.283	0.804	0.880
Number of Pairs	249	500	321	383	61
Proportion	0.163	0.327	0.210	0.251	0.040
School Network 4 Correlation	0.005	-0.041	-0.002	0.038	-0.109
p-value	0.913	0.463	0.976	0.463	0.174
Number of Pairs	454	323	198	370	158
Proportion	0.298	0.212	0.130	0.242	0.104
School Network 5 Correlation	0.116	-0.033	0.081	-0.019	0.043
p-value	0.238	0.279	0.024	0.982	0.295
Number of Pairs	150	318	351	350	307
Proportion	0.100	0.212	0.234	0.234	0.205

Table OA.7: Correlation of $Y(0)$s and $Y(1)$s in Bins in Empirical Illustration

	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6	Bin 7	Bin 8	Bin 9	Bin 10
School Network 1	0.073	0.059	-0.007	0.103	0.038	-0.033	0.001	0.068	-0.091	-0.079
p-value	0.287	0.253	0.892	0.123	0.579	0.600	0.991	0.331	0.458	0.755
Number of Pairs	212	373	333	227	218	248	288	204	68	18
Proportion	0.096	0.169	0.151	0.103	0.099	0.112	0.130	0.092	0.031	0.008
School Network 2	-0.070	0.100	0.070	0.400	-0.067	0.011	-0.042	-0.089	0.255	-0.395
p-value	0.739	0.680	0.954	0.005	0.177	0.772	0.092	0.486	0.307	0.182
Number of Pairs	315	370	100	41	159	241	179	84	18	13
Proportion	0.203	0.239	0.065	0.026	0.103	0.155	0.115	0.055	0.012	0.008
School Network 3	NA	-0.087	0.100	-0.019	-0.094	-0.030	0.004	-0.035	-0.003	NA
p-value	NA	0.185	0.081	0.792	0.291	0.675	0.955	0.664	0.982	NA
Number of Pairs	15	233	306	193	129	192	228	153	50	7
Proportion	0.010	0.152	0.200	0.126	0.084	0.126	0.149	0.100	0.033	0.005
School Network 4	-0.031	0.023	0.014	-0.113	-0.105	0.012	0.111	-0.044	-0.073	-0.074
p-value	0.711	0.685	0.839	0.267	0.441	0.886	0.119	0.562	0.439	0.656
Number of Pairs	148	306	221	98	56	142	197	172	116	39
Proportion	0.097	0.201	0.145	0.064	0.037	0.093	0.129	0.113	0.076	0.026
School Network 5	-0.129	0.180	-0.035	-0.022	0.070	0.091	-0.027	-0.013	-0.028	0.189
p-value	0.812	0.312	0.311	0.613	0.368	0.016	0.578	0.636	0.806	0.031
Number of Pairs	49	99	154	164	185	166	152	196	204	102
Proportion	0.033	0.066	0.103	0.110	0.124	0.111	0.102	0.131	0.136	0.068

All treatment group observations in Bin 1 and Bin 10 have $Y(1) = 1$ which leads to NAs.