The linking effect: causal identification and estimation of the effect of peer relationship

Zheng Wang *

October 10, 2022

(For the latest draft)

Abstract

The endogeneity of network formation has been a major obstacle to the empirical study of peer influence for many important types of networks, including friendship networks, buyer-supplier networks, banking networks, etc. This paper puts forward the first causal identification strategy in the literature to study the effect of non-randomly formed peer relationships. I prove that causal identification holds under general conditions and needs neither a network formation model nor an outcome model to be specified. This is because the propensity scores of the unobserved confounders can be non-parametrically identified and estimated from the distribution of network links. Using the proposed method, I empirically estimate the causal effect of high school friendships on female students' bachelor's degree attainment. While previous literature finds that being exposed to more high-achieving boys in high school makes girls less likely to obtain a bachelor's degree, I show that this is not true when these high-achieving boys are considered friends by the girls. In fact, one additional high-achieving male friend increases the probability that a female student graduates from college by 3 p.p. Further analysis suggests that this positive influence is not a result of increased academic ability but rather comes from a significant confidence boost. These results imply that rather than shielding girls from high-achieving boys, it would be more effective to foster friendship and close interactions among them.

*European University Institute. I would like to thank Andrea Ichino, Yann Bramoulle and Fabrizia Mealli for their valuable feedbacks and guidance throughout the development of this paper. I would also like to thank participants at the EUI Microeconometrics Working Group and the AMSE PhD seminar for their helpful comments.

1 Introduction

Interest in understanding the impact of peer influence within economic and social networks has been growing rapidly in the economics literature, with an increasing emphasis on establishing causality. However, due to the lack of credible and implementable causal identification strategy for non-randomly formed networks, researchers have been deterred from empirically evaluating the causal impact of many important types of relationships, such as friendship networks, buyer-supplier networks, banking networks, etc. Understanding how agents connected through these networks are affected by the relationships among them is important for both policymakers and the agents themselves to improve their economic or social standing by cultivating certain relationships while avoiding others.

The difficulty in establishing causal identification partly comes from the absence of a causal framework specifically designed for the study of the effect of peer relationships. After all, the first step to identifying any causal effect should be to define such effect through a properly formulated framework where treatments and potential outcomes are clearly defined. In this paper, each potential relationship is a unique treatment. In other words, the existence of each network link is the subject of manipulation or intervention in a hypothetical experiment where we could assign network links at will. This contrasts with the existing literature on peer effects where the treatment is implicitly assumed to be some summary statistics of the network as a whole, such as the share of network nodes with certain characteristics [1]. In this paper, the effect of relationships is called the **linking effect**, emphasizing the fact that the treatment is the assignment of links. Explicitly viewing every pairwise relationship as a treatment opens the door to building upon existing causal inference tools for the study of the linking effect. In particular, due to the multiplicity of possible relationships for any network node, we are able to embed the study of the network linking effect in the multi-treatment causal inference framework.

This newly discovered connection between these two previously disassociated literature turns out to be highly consequential for the causal identification of the linking effect in endogenous networks. Combining well-known results in the network analysis literature[2] and

[1] In Manski (1993), this treatment is associated with the contextual peer effect.

[2] To be more specific, these results are related the Aldous-Hoover Theorem, a representation theorem for

a recent finding in the multi-treatment causal inference literature (Wang and Blei, 2019), I prove that the linking effects are identified under a set of general assumptions. These are the network variations of the standard assumptions used in traditional causal inference with observational data (Imbens and Rubin, 2015): individualistic assignment, SUTVA, unconfoundedness, and positivity. A major appeal of my identification strategy is that unconfoundedness does not need the restrictive assumption that all confounding variables are observed in the data. Instead, unconfoundedness relies solely on the assumption that there does not exist any confounder that affects one and only one network link[3]. This is because all unobserved confounders that affect more than one link, or more precisely, the propensity scores of these confounders, could be identified from the network adjacency matrix itself. Therefore, we are able to have an unconfoundedness condition that is much weaker than that required in traditional causal inference.

My identification result naturally points to the use of propensity score-based estimators. Unlike traditional propensity score estimation procedures where the probability of treatment is regressed on a set of observed pre-treatment variables, here the propensity scores of the unobserved confounders are estimated using only the observed network links. One way to operationalize the estimation is to use probabilistic factor models to capture the joint distribution of the links (Wang and Blei, 2019). This involves specifying the distributions of the unobserved confounders and the distributions of the network links conditional on the unobserved confounders. It is, however, not important which specific distributions one chooses to use, as long as the overall joint distribution of the network links is well captured. Another way to estimate the propensity scores is to use graphon estimation procedures. This is because when the network nodes are seen as randomly sampled from the superpopulation, the distribution of the network links has the same form as the distribution of a graphon, which can be estimated both non-parametrically (e.g. Zhang et al., 2017) and parametrically. With the estimated propensity scores, we can then use inverse probability weighting, subclassification, or propensity score matching to estimate the desired causal effect.

network data.

[3]This assumption is likely to hold in networks of non-trivial size because as the number of possible peer links to choose from increases, it becomes more and more difficult to conceive an individual level confounding variable that affects only one of these choices but not any other.

Thanks to these identification and estimation results, this paper will conduct one of the first empirical analyses aiming to understand the causal effect of one of the most well-known endogenous networks, friendships. Despite being the main focus of the social network literature, the impact of friendship networks has not been well-understood empirically due to the endogeneity problem. The only few existing papers that attempted to address the endogeneity issue did so by making strong assumptions both on the way friendships are formed (e.g., homophily) and the variables that affect this formation (e.g., race), subjecting the estimated results to bias when the true network formation process has a different form (Goldsmith-Pinkham and Imbens, 2013; Badev, 2018; Gagete-Miranda, 2020).

Most papers in the empirical peer effect literature circumvent this by looking at other social networks which are quasi-randomly formed. For example, Cools et al. (2019) investigates how the presence of more high-achieving male and female students in high school affects boys' and girls' bachelor's degree attainment differently. They do so by exploiting the random variations in cohort composition, a strategy commonly employed in the peer effect literature (the Danish thing, Olivetti, the military thing by Sacerdote, etc.). Cools et al. (2019) find that being exposed to more male high achievers decreases girls' likelihood of obtaining a bachelor's degree, in part by decreasing their confidence and aspiration. While these studies offer exciting findings on the effect of cohort composition, a common drawback is that the impact of social interactions cannot be separated from the influence of other factors that also vary across cohorts, such as differences in teachers' attitudes. Moreover, some of the most meaningful social interactions with long-term consequences only exist among close friends and not those who simply attend the same school during the same year. As a consequence, the patterns of peer influence among friends have largely remained unknown.

Using high school friendship data from AddHealth, the same dataset used by Cools et al. (2019) and many other studies on social networks [4], I test whether the negative impact of high-achieving male students on female students persists when these boys are considered friends by the girls. Interestingly, I find that an additional male high-achieving friend causally increases the probability that a female student obtains a bachelor's degree by 3 p.p. Further

[4] For example, Goldsmith-Pinkham and Imbens (2013); Bifulco et al. (2014); Badev (2018); Olivetti et al. (2020)

analysis suggests that this positive influence results from an increase in their confidence and not their academic ability. Indeed, having one more male high-achieving friend means the female student becomes 3.75 p.p more likely to self-report being more intelligent than their same-age peers, but no effect is found for their grades in any of the main subjects [5]. Taking these results together with the findings of Cools et al. (2019), it seems that girls are intimidated by high-achieving boys whom they do not have close relationships with, but are encouraged by those whom they see as friends. This suggests that a possible way to boost the confidence of female students and increase their chances of graduating from college is by fostering friendships with high-achieving boys in their high school.

This paper is closely related to the literature on peer effect, especially the so-called contextual peer effect defined in Manski (1993). Roughly speaking, contextual peer effect refers to the effect of peer characteristics on own outcome and is usually expressed as a parameter in a regression model. In order to give a causal interpretation to the estimated parameter, empirical researchers have taken advantage of settings with either random treatments or random peers. The former is where peer relationships are fixed and characteristics of the network nodes are randomized, while the latter is where nodal characteristics are fixed but peer relationships are randomized. In other words, the former is related to treatment spillover, while the latter is about the linking effect. Because these two cases correspond to two completely different hypothetical interventions, using one parameter to represent their effects can lead to misleading interpretations of the estimates [6]. This paper addresses the issue by developing a causal framework tailored for the study of linking effects where random peers are a special case [7]. Since in the linking effect framework the only treatment is the existence of the links, confusion on the interpretation and policy implication of the estimates is unlikely to arise.

To the best of my knowledge, Li et al. (2019) and Basse et al. (2021) are the only papers to have made the distinction between interventions on peer relationships and interventions

[5]Both the self-reported intelligence and the grades are measured one year after the friendship data was collected. The main subjects are math, science, English, and history.

[6]See Bramoullé et al. (2020) for more analysis on the problem of misinterpretation.

[7]If peer relationships are randomized, there will be no need to address the confounding (endogeneity) problem. The causal framework of the linking effect can still be used; the only difference is that there will be no need to infer the unobserved confounders and use them to correct for confounding, as randomization guarantees no confounding exists.

on peer characteristics within a formalized causal framework. However, the focus of their papers is on inference issues rather than identification, as they only consider cases where agents are assigned to groups randomly. They also focus their analysis on peer networks with a non-overlapping group structure, such as roommate networks. My framework, in contrast, allows the networks to have arbitrary structures and is suitable for analyzing both experimental and observational data.

Several econometrics papers have emerged in recent years to tackle the network endogeneity issue for Manski (1993)'s linear-in-means model. Most of these papers specify a parametric dyadic model of network formation [8]. Goldsmith-Pinkham and Imbens (2013) and Hsieh and Lee (2016) then jointly estimate the outcome model and their specified network formation model through distributional conditions. Alternatively, Arduini et al. (2015) and Johnsson and Moon (2021) take a control function approach, where they first estimate the unobserved error terms in the parametric network formation model and then plug these estimates back into the linear-in-means outcome model to control for confounding. The validity of these approaches, however, is sensitive to the correctness of the modelling assumptions. A key strong point of this paper is that no parametric assumptions are needed for the identification, except that the network links are independently distributed conditional on some unobserved variables. I show that this distributional assumption is weaker than the one assumed for dyadic models and is more general than the literature previously believed. In particular, it could accommodate strategic network formation models such as those studied in Mele and Zhu (2017); Mele (2017); Leung (2015). Moreover, it always holds under random node sampling from the superpopulation, a perspective routinely used in traditional causal inference.

The rest of the paper will be organised as follows. Section 2 gives the formal definitions of the treatment and the potential outcome and proposes a linking effect estimand to study peer influence. Section 3 provides the identification conditions of this estimand and Section 4 discusses how existing propensity score-based estimators can be adapted for its estimation. Section 5 gives simulation evidence on the bias reduction performance of the proposed identification and estimation strategy. Finally, Section 6 applies these estimators to real data to

[8] In the language of causal inference, this is equivalent to modelling treatment selection.

study the effect of high school friendship on students' bachelor's degree attainment.

2 Treatments, potential outcomes, and estimands

Suppose we are interested in a certain peer relationship network with N nodes and directed links among these nodes. N could be infinity. A link from one node to another represents the existence of a directed peer relationship. The adjacency matrix $\mathbf{D}_N$ of the network is a N by N matrix where each entry represents the existence of a link:

$$\mathbf{D}_N = \begin{bmatrix} 0 & D_1^2 & ... & D_1^N \\ D_2^1 & 0 & ... & D_2^N \\ ... & ... & ... & ... \\ D_N^1 & D_N^2 & ... & 0 \end{bmatrix},$$

In this paper, I will write $D_i^j = 1$ if there is a directed link from node j to node i. The diagonal of the adjacency matrix is 0 because we do not allow one to be their own peer. When a node is on the receiving end of the link, I call it the link receiver. When a node is on the sending side of the link, I call it a link sender. A node can act as a link receiver in one link while acting as a link sender in another and vice versa. In this paper, the outcomes of interest are measured on the link receivers, but we could just as easily measure outcomes on the link senders. When I write a pair of nodes (i, j), the first component is the link receiver, and the second component is the link sender. Whenever suitable, I also use subscripts to indicate the link receiver and superscripts as the link sender. In this paper, the treatment of interest is the (directed) linking status among pairs of network nodes. For example, for a friendship network, the treatment of interest would be the directed friendship from one person to another [9]. With two hypothetical pairwise relationships, Figure 1 highlights the hypothetical intervention, aka the treatment, that is the focus of this paper. Each relationship has three components: the receiver (R), the sender (S), and the linking status (D). In this example, the two relationships have the same receiver and sender but have

[9]Friendship doesn't need to be a reciprocal relationship, as one person consider another person as a friend doesn't necessarily mean the other way holds. This is evidenced by the friendship nominations of high school students in the Add Health data.

Figure 1: Hypothetical intervention: two counterfactuals

different linking statuses. On the left, the link from the sender to the receiver exists, but on the right, the link doesn't exist. The type of causal question this paper asks is " What would the receiver R_1's potential outcome be if it were "treated" with a link from sender S_1 (left of Figure 1), and what would the potential outcome be if it weren't " treated" with this link (right of Figure 1), and the difference between the two potential outcomes? ". In other words, what is the difference between $Y_1(D_{R_1}^{S_1} = 1)$ and $Y_1(D_{R_1}^{S_1} = 0)$? The only difference between the two hypothetical cases is the existence of the directed link from the sender to the receiver. This is why we call the linking status the "treatment".

It is important to emphasize that the hypothetical intervention this paper studies is NOT the change in the sender characteristics [10]. In this paper, link sender nodal characteristics define the multiple versions of the treatment. As an example, consider color as the nodal characteristic [11]. Figure 2 shows two hypothetical relationships between R_1 and a different sender S_2, where S_2 is red while S_1 is orange. This means the effect of $D_{R_1}^{S_1}$ on R_1 could be different from the effect of $D_{R_1}^{S_2}$ on R_1, therefore a link from S_2 should be viewed as a different treatment than a link from S_1. In the most general case, we could allow linking

[10] It is, however, the focus of the treatment spillover literature.

[11] For instance, Li et al. (2019) and Basse et al. (2021) assume the effect of linking only depends on some observed characteristic of the node chosen by the researcher ex-ante.

Figure 2: A different link sender

effects to differ in arbitrary observed and unobserved sender nodal characteristics. This is the stance taken by this paper. As a result, links from senders with different *identities* are viewed as different treatments.

Given that any link receiver could potentially receive a link from $N-1$ different link senders, and each of these links is considered a unique treatment with a unique effect on the receiver, we are in the case of multiple treatments, or multi-cause, causal inference. In other words, for any link receiver i, its treatment is a vector of $N-1$ linking status $\mathbf{D}_i := (D_i^1, ..., D_i^{i-1}, D_i^{i+1}, ..., D_i^N)$.

In single-treatment causal inference, the potential outcome of any unit could depend on the treatment status of all units in the population if no further assumption is made. The Stable Unit Treatment Unit Assumption (SUTVA) restricts the potential outcome to depend only on the unit's treatment status. Here I will make a similar assumption to only allow potential outcomes to depend on the unit's own treatment status. But because in this paper, any node i's treatment is a vector of all pairwise linking status with the other nodes, restricting the potential outcomes to only depending on their own treatment means they can depend on all pairwise linking status where i is the receiver, but couldn't depend on the linking status where i is not the receiver. I call this assumption the Linking-effect Stable

Unit Treatment Unit Assumption (L-SUTVA) to differentiate it from the usual SUTVA.

Assumption 1 (L-SUTVA).

$$Y_i(\mathbf{D}_i, \mathbf{D}_{-i}) = Y_i(\mathbf{D}_i, \tilde{\mathbf{D}}_{-i})$$

for any $(\mathbf{D}_{-i}, \tilde{\mathbf{D}}_{-i})$ and any i, where $\mathbf{D}_{-i} = (\mathbf{D}_1, ..., \mathbf{D}_{i-1}, \mathbf{D}_{i+1}, ..., \mathbf{D}_N)$.

Under L-SUTVA, the potential outcome can be written as $Y_i(\mathbf{D}_i)$ or $Y_i(D_i^1, D_i^2, ..., D_i^N)$. In traditional causal inference, SUTVA is sometimes called the no-interference assumption. This paper studies the effect of relationships, which suggests units must interact or interfere in some way. At first sight, the two may seem to be at odds. The reason why L-SUTVA is perfectly compatible with the study of linking effect lies in the definition of treatment. Recall what SUTVA says is that the *treatment status* of one unit does not interfere with another unit's *potential outcome*. In particular, it doesn't require the non-existence of network structure among the units. Whether SUTVA is likely to hold depends on the treatment and potential outcomes being studied. In this paper, since the treatment is the relationship, the no-interference implied by L-SUTVA means that one's potential outcome is only affected by one's own relationships. L-SUTVA helps reduce the space of possible potential outcomes and makes it easier to identify and estimate causal estimands. In this paper, I will always assume that L-SUTVA holds. However, L-SUTVA might not be realistic in some situations. In the future, I will extend the analysis by relaxing L-SUTVA to allow some interference.

2.1 Estimands

With the perspective that relationships are multiple treatments, we could flexibly define many interesting estimands. In this section, I will focus on the most straightforward set of estimands, which, loosely speaking, looks at the effect of an additional link. Several other possible estimands are outlined in Section A1.

τ_i^j is defined as the following contrast of i's potential outcomes:

$$\tau_i^j = Y_i(D_i^j = 1, \mathbf{D}_i^{-j} = \bar{\mathbf{d}}_i^{-j}) - Y_i(D_i^j = 0, \mathbf{D}_i^{-j} = \bar{\mathbf{d}}_i^{-j})$$

where $\mathbf{D}_i^{-j} = (D_i^1, ..., D_i^{j-1}, D_i^{j+1}, ..., D_i^N)$, and $\bar{\mathbf{d}}_i^{-j}$ is the corresponding vector of the ***realised*** or ***observed*** treatment assignment for i after taking out d_i^j. τ_i^j contrasts link receiver i's potential outcome when it receives treatment (a link) from link sender j with its potential outcome when it doesn't receive the link from j while keeping the links (or the lack of links) from other link senders fixed at their realized value.

Based on τ_i^j, we can proceed to define an average linking effect for links with j as the sender:

$$\begin{aligned}\tau^j &:= \mathbb{E}_i[Y_i(D_i^j = 1, \mathbf{D}_i^{-j} = \bar{\mathbf{d}}_i^{-j}) - Y_i(D_i^j = 0, \mathbf{D}_i^{-j} = \bar{\mathbf{d}}_i^{-j})] \\ &:= \frac{1}{N}\sum_{i=1}^{N} Y_i(D_i^j = 1, \mathbf{D}_i^{-j} = \bar{\mathbf{d}}_i^{-j}) - Y_i(D_i^j = 0, \mathbf{D}_i^{-j} = \bar{\mathbf{d}}_i^{-j})\end{aligned}$$

This is simply taking the average of the linking effect τ_i^j overall link receivers.

Next, instead of looking at the average linking effect from one link sender j, we could look at the average linking effect from link senders with some attributes $A = a$.

$$\begin{aligned}\tau^a &:= \mathbb{E}_{(i,j):A^j=a}[Y_i(D_i^j = 1, \mathbf{D}_i^{-j} = \bar{\mathbf{d}}_i^{-j}) - Y_i(D_i^j = 0, \mathbf{D}_i^{-j} = \bar{\mathbf{d}}_i^{-j})] \\ &:= \frac{1}{N}\sum_{i=1}^{N}\Big(\frac{1}{\sum_{j=1}^{N} A^j = a}\sum_{A^j=a}\big(Y_i(D_i^j = 1, \mathbf{D}_i^{-j} = \bar{\mathbf{d}}_i^{-j}) - Y_i(D_i^j = 0, \mathbf{D}_i^{-j} = \bar{\mathbf{d}}_i^{-j})\big)\Big)\end{aligned}$$

Finally, we could restrict our attention to link receivers with certain attributes $R = r$, where I use R to denote the attributes of interest for the link receivers. This can be easily done by only averaging over the link receivers with $R = r$:

$$\begin{aligned}\tau_r^a &:= \mathbb{E}_{(i,j):R_i=r,A^j=a}[Y_i(D_i^j = 1, \mathbf{D}_i^{-j} = \bar{\mathbf{d}}_i^{-j}) - Y_i(D_i^j = 0, \mathbf{D}_i^{-j} = \bar{\mathbf{d}}_i^{-j})] \\ &:= \frac{1}{\sum_{i=1}^{N} R_i = r}\sum_{i=1}^{N}\Big(\frac{1}{\sum_{j=1}^{N} A^j = a}\sum_{A^j=a}\big(Y_i(D_i^j = 1, \mathbf{D}_i^{-j} = \bar{\mathbf{d}}_i^{-j}) - Y_i(D_i^j = 0, \mathbf{D}_i^{-j} = \bar{\mathbf{d}}_i^{-j})\big)\Big)\end{aligned}$$

Under L-SUTVA, these estimands are well-defined and can be interpreted as the all-or-nothing effect in the following sense. Take τ^a as an example, it can be interpreted as the **expected** contrast between the **average** potential outcome of assigning a sender-j link to *everyone* in the node set and the **average** potential outcome of assigning a sender-j link to *no one* in the node set, where this j is **chosen randomly** (hence the expected contrast) with equal probability from the set of link senders with attribute $A^j = a$. The interpretation of τ_r^a is similar to that of τ^a, except that instead of looking at all link receivers in the node set, now we only look at link receivers with $R_i = r$.

Note that estimands for studying the direct linking effect can also be defined without the assumption of L-SUTVA. In this case, we could simply modify the potential outcome function to include the entire adjacency matrix $(D_i^j, \mathbf{D}_{-(i,j)})$. But we can no longer interpret the estimands as the all-or-nothing effect. This is because when we simultaneously change $(D_1^j, D_2^j, ..., D_N^j)$ for a given sender j, $\mathbf{D}_{-(i,j)}$ is no longer at its observed value. Instead, the estimands need to be interpreted as the **expected** treatment effect of j on a **randomly chosen** link receiver i, again keeping the other links at their realized value. The difference is that in the second interpretation, in every hypothetical experiment, intervention is only done on one link, and the average linking effect τ^j is the average from repeated experiments where a different link is modified each time. This is similar to the $EATE$ in Sävje et al. (2021) and the τ defined in Forastiere et al. (2021).

3 Identification

At the center of causal identification is the treatment assignment mechanism. In experimental studies, the assignment mechanism is known, but in observational studies, the assignment mechanism is unknown, and assumptions must be imposed on it to make causal discoveries. This section lists and discusses the assumptions needed to identify linking effects for the situation where observational data is used, and quasi-randomness is absent. As in traditional causal inference, these assumptions include individualistic assignment mechanism, unconfoundeness, and positivity.

3.1 Individualistic Assignment Mechanism

In the study of linking effects, the individualistic assignment mechanism assumption takes the form of conditionally independent links.

Assumption 2 (Individualistic Assignment Mechanism)**.** There exists sequences of random variables (vectors) $\{\mathbf{U}_i\}_{1\leq i\leq n}$ and $\{\mathbf{V}_i\}_{1\leq i\leq n}$ such that equation (1) holds.

$$Pr(\mathbf{D}_N = \mathbf{d}_N|\mathbf{U}_1, ..., \mathbf{U}_N, \mathbf{V}_1, ..., \mathbf{V}_N) = \prod_{i=1}^{N}\prod_{j\neq i}^{N} Pr(D_i^j = d_i^j|\mathbf{U}_i, \mathbf{V}_j) \tag{1}$$

We can think of $\mathbf{U}_i$ as link receiver level variables and $\mathbf{V}_j$ as link sender level variables. For any node i, $\mathbf{U}_i$ and $\mathbf{V}_i$ could share some common components. For example, for a high school friendship network, the ambition of student i could affect from whom they receive links and at the same time to whom they send links. Note that Assumption 2 does not impose $\mathbf{U}_i$ and $\mathbf{U}_k$ to be independent for any $i \neq k$ (same for $\mathbf{V}_i$ and $\mathbf{V}_k$). For example, if i and k have a shared class, their friendship patterns could depend on the attitude of their shared teacher.

It is important to point out that Assumption 2 is different from the assumption underlying dyadic regressions. Dyadic regressions, such as those analyzed in Graham (2020), usually assumes that linking decisions are independent conditional on some observed attributes X and unobserved latent attributes ϵ satisfying $\mathbb{E}[\epsilon|X = 0]$:

$$Pr(\mathbf{D}_N = \mathbf{d}_N|X_1, X_2, ..., X_N, \epsilon_1, \epsilon_2, ..., \epsilon_N) = \prod_{i=1}^{N}\prod_{j\neq i}^{N} Pr(D_i^j = d_i^j|X_i, X_j, \epsilon_i, \epsilon_j) \tag{2}$$

The goal of dyadic regressions is usually to estimate the parameters associated with the observed covariates to understand the role of these covariates in determining linking probabilities, such as those in estimating the gravity models studying the association between GDP and trade flow. Running a dyadic regression requires one to impose a functional form of the pairwise linking probability: $Pr(D_i^j = 1|X_i, X_j, \epsilon_i, \epsilon_j) = f(X_i, X_j, \epsilon_i, \epsilon_j)$ for some known f. This functional form differentiates the assumption of individualistic assignment mechanism and the assumptions underlying dyadic regressions. In contrast, this paper aims to identify and estimate the causal parameters of the outcome equation. Assumptions on the linking equation, or selection into treatment, are only used to correct for confounding. Identifying such causal effects does not require knowing the functional form of the linking equation. Therefore, there is no need to estimate parameters associated with the observed attributes.

Not surprisingly, the assumptions underlying dyadic regressions are more restrictive than Assumption 2. For example, consider a case where friendship formation probability is determined by gender only. On the one hand, a dyadic regression

$$Pr(D_i^j = 1|Gender_i, Gender_j, \epsilon_i, \epsilon_j) = logistic(|Gender_i - Gender_j| + \epsilon_i + \epsilon_j)$$

makes the following assumptions: (i) The gender of i and j only matters for friendship formation up to their absolute difference. (ii) IF $Gender_k$ for $k \neq i, j$ also affects the link formation between i and j, it can only affect $Pr(D_i^j = 1$ through the additive terms $\epsilon_i$and ϵ_j. (iii) The link function has the logistic form. On the other hand, with Assumption 2, U_i and V_j could include the gender of any nodes in the population, not just that of node i and j. In addition, the gender of these nodes could affect linking formation in any arbitrary way.

The discussion above means that Assumption 2 will hold if one believes dyadic regression is an appropriate way to model the linking equation for their research question. But Assumption 2 holds in much broader situations. In fact, if we view the network data as generated by random node sampling from an infinite super population[12], Assumption 1 will always hold. The idea is that random node sampling guarantees the network to be *vertex exchangeable*, and vertex exchangeable network can always be represented by equation (1) by the Aldous-Hoover Theorem (Crane, 2018). More details are provided in Section A3.1.

When the interest of analysis is in the finite sample itself, the data-generating process needs to be modeled. Here we view the network links as generated through a game. I already showed that if we assume network links are formed by only dyadic considerations as implied by dyadic regressions, the Individualistic Assignment Mechanism Assumption will be satisfied. More interestingly, Assumption **??** could hold even with games with strategic interactions. For example, when incorporating network externalities, Mele (2017) and Mele and Zhu (2017) showed that with their the dynamic network formation model, the stationary equilibrium link distribution has the form of equation (1) (Graham, 2020). Moreover, Assumption **??** could also hold when the network formation game is characterized by strategic

[12]Random sampling from an infinite super population is a perspective commonly adopted in the literature outside of network analysis. See Imbens and Rubin (2015) and Hernán and Robins (2020) for more discussions on the super population and finite population perspectives in causal inference.

interactions with incomplete information, as the one analyzed in Leung (2015). The idea is that when the objective is the expected utility, i's linking probability will be a function of equilibrium beliefs about others' linking decisions, conditioning on the observed attributes of all agents in the network. For this reason, equilibrium linking decisions are functions of the exogenous attributes only, allowing a representation of the form of equation (1). More details are given in Section A3.2.

[Give the network examples when the assumption could hold and when it for sure doesn't]

3.2 Unconfounded Assignment Mechanism

Assumption 3 (No single link confounder)**.** All confounders affect more than one link probability.

Proposition 1 (Unconfoundedness)**.** Let $\mathbf{Y}_i^{pot}$ be the vector of i's potential oucomes. Under Assumption 2 and Assumption 3, the following holds:

$$Pr(D_i^j = 1|\mathbf{U}_i, \mathbf{V}_j, \mathbf{Y}_i^{pot}) = Pr(D_i^j = 1|\mathbf{U}_i, \mathbf{V}_j)$$

for $\mathbf{U}_i, \mathbf{V}_j$ defined in equation (1).

Proof. First, suppose there exists another variable U_i' that is a cause of $\mathbf{Y}_i^{pot}$ and a cause of more than one of the links that i potentially receives, say D_i^1 and D_i^2. Then if we omit U_i' in the conditioning set of

$$\begin{aligned}
&Pr(\mathbf{D}_N = \mathbf{d}_N|\mathbf{U}_1, ..., \mathbf{U}_N, \mathbf{V}_1, ..., \mathbf{V}_N) \\
=&Pr(\mathbf{D}_1 = \mathbf{d}_1, ..., D_i^1 = d_i^1, D_i^2 = d_i^2, ..., D_i^N = d_i^N, ..., \mathbf{D}_N = \mathbf{d}_N|\mathbf{U}_1, ..., \mathbf{U}_N, \mathbf{V}_1, ..., \mathbf{V}_N)
\end{aligned}$$

D_i^1 and D_i^2 couldn't be conditionally independent (without conditioning on $\mathbf{U}_i'$). This is a contradiction to our starting point, which is that conditioning on $\mathbf{U}_1, ..., \mathbf{U}_N, \mathbf{V}_1, ..., \mathbf{V}_N$ makes all links independent (equation (2)). In other words, the variables $\mathbf{U}_1, ..., \mathbf{U}_N, \mathbf{V}_1, ..., \mathbf{V}_N$ by definition make all links independent, and the existence of $\mathbf{U}_i'$ is in contradiction of that definition.

With similar logic, suppose there exists a variable V_j' that is a cause of $\mathbf{Y}_i^{pot}$ and is a cause of more than one of the links that j potentially sends, say D_1^j and D_2^j. Then if we omit V_i' in the conditioning set of

$$
\begin{aligned}
&Pr(\mathbf{D}_N = \mathbf{d}_N | \mathbf{U}_1, ..., \mathbf{U}_N, \mathbf{V}_1, ..., \mathbf{V}_N) \\
=&Pr(D_1^2 = d_1^2, ..., D_1^j = d_1^j, ..., D_1^N = d_1^N, ..., D_2^1 = d_2^1, ..., D_2^j = d_2^j, ..., D_2^N = d_2^N, ..., \\
&D_N^1 = d_N^1, ..., D_N^j = d_N^1, ..., D_N^{N-1} = d_N^{N-1} | \mathbf{U}_1, ..., \mathbf{U}_N, \mathbf{V}_1, ..., \mathbf{V}_N)
\end{aligned}
$$

D_1^j and D_2^j couldn't be conditionally independent (without conditioning on $\mathbf{V}_i'$). This again is a contradiction to our starting point, which is that conditioning on $\mathbf{U}_1, ..., \mathbf{U}_N, \mathbf{V}_1, ..., \mathbf{V}_N$ makes all links independent (Equation 2).

By Assumption 3, which states confounders that only affect one link don't exist, we have effectively ruled out the existence of any confounders that affect the formation of any link. This means

$$
\begin{aligned}
&Pr(D_i^j = 1 | \mathbf{U}_1, ..., \mathbf{U}_N, \mathbf{V}_1, ..., \mathbf{V}_N, \mathbf{Y}_i^{pot}) \\
=&Pr(D_i^j = 1 | \mathbf{U}_1, ..., \mathbf{U}_N, \mathbf{V}_1, ..., \mathbf{V}_N) \\
=&Pr(D_i^j = 1 | \mathbf{U}_i, \mathbf{V}_j)
\end{aligned}
$$

The last equation comes from equation (2).

□

The intuition is that $\mathbf{U}_i, \mathbf{V}_j$ must include all the causes of link formation. Otherwise, the Individualistic Assignment Mechanism Assumption wouldn't have held. Some of these causes will confound the outcome variable; some will not. In theory, we only need to condition on the confounders, but the insight is that since we don't know which are the confounders, conditioning on all of these causes will for sure address confounding. It is for this reason that I call $\mathbf{U}_i, \mathbf{V}_j$ the ***sufficient confounders***. Notice that here the assumption of individualistic assignment implies no confounding from multiple-link confounders, while in the conventional causal inference, these are two independent assumptions.

Next, I prove that the unconfoundedness condition also holds conditional on the propensity score based on $\mathbf{U}_i, \mathbf{V}_j$. The propensity score $e(\mathbf{U}_i, \mathbf{V}_j)$ is defined as $e(\mathbf{U}_i, \mathbf{V}_j) := Pr(D_i^j = 1 | \mathbf{U}_i, \mathbf{V}_j)$, where the probability distribution is over the random sampling of both link receiver and link sender.

Lemma 1. $e(\mathbf{U}_i, \mathbf{V}_j)$ is a balancing score, that is:

$$Pr(D_i^j = 1 | \mathbf{U}_i, \mathbf{V}_j, e(\mathbf{U}_i, \mathbf{V}_j)) = Pr(D_i^j = 1 | e(\mathbf{U}_i, \mathbf{V}_j))$$

Lemma 2 (Unconfoundedness given $e(\mathbf{U}_i, \mathbf{V}_j)$)**.**

$$Pr(D_i^j = 1 | Y_i^{pot}, e(\mathbf{U}_i, \mathbf{V}_j)) = Pr(D_i^j = 1 | e(\mathbf{U}_i, \mathbf{V}_j))$$

This result is similar to the propensity score property result in the traditional causal inference, where unconfoundedness holds given the propensity score. The proof of Lemma 2 is given in Section A4.2.

3.3 Unconfoundedness and identification of estimands

Lemma 3. $Pr(D_i^j = 1 | \mathbf{U}_i, \mathbf{V}_j)$ from the conditional independence condition (1) is identified, $\forall i, j$

This is proved in Diaconis and Janson (2007). Since the propensity score $e(\mathbf{U}_i, \mathbf{V}_j) = Pr(D_i^j = 1 | \mathbf{U}_i, \mathbf{V}_j)$, the propensity score is also identified $\forall i, j$.

Assumption 4 (Positivity)**.** $0 < Pr(D_i^j = 1 | \mathbf{U}_i, \mathbf{V}_j) < 1$ for all $i \neq j \leq n$

Proposition 2. Under Assumptions 1-4, the direct linking effect is identified. For link receivers with characteristics a and link senders with characteristics r, this means

$$\begin{aligned}\tau_a^r = \mathbb{E}\left[\mathbb{E}_{(i,j):R_i=r,A^j=a}[Y_i^{obs} | e(\mathbf{U}_i, \mathbf{V}_j), D_i^j = 1]\right] \\ - \mathbb{E}\left[\mathbb{E}_{(i,j):R_i=r,A^j=a}[Y_i^{obs} | e(\mathbf{U}_i, \mathbf{V}_j), D_i^j = 0]\right]\end{aligned}$$

This is proved in Section A4.3.

4 Estimation

The estimation of the linking effect involves two steps. The first step is to estimate the propensity scores. Unlike in traditional causal inference, the propensity scores estimated in the first step are functions of unobserved latent variables. Therefore, the traditional propensity score estimation methods won't apply here. In Section 4.1, I show how techniques developed in the graphon estimation literature in network analysis and the multiple treatment literature in causal inference can be used for propensity score estimation. The second step is to use the estimated propensity scores to estimate the linking effects. Here many established methods from traditional causal inference can be used, such as inverse probability weighting (IPW), propensity score matching, and propensity score subclassification. In Section 4.2, I will illustrate how the inverse probability weighting method can be used to estimate the linking effects. Propensity score matching and subclassification can be adapted similarly as shown in Section A2.

4.1 1st-step estimation: propensity scores

4.1.1 Graphon Estimation

As discussed in Section A3.1, the linking probability in a graphon and the propensity score e_{ij} are, in fact, exactly the same. This means we could use the many statistical methods in graphon estimation to estimate the propensity scores. Here I briefly discuss how the neighborhood smoothing method proposed by Zhang et al. (2017) works. Compared to other graphon estimation methods, such as stochastic block models (Olhede and Wolfe, 2014), it has the advantage of not making restrictive assumptions on how links are formed.

First let's define a probability slice as $e(\mathbf{U}_i, \cdot) = (e(\mathbf{U}_i, \mathbf{V}_1), e(\mathbf{U}_i, \mathbf{V}_2), ..., e(\mathbf{U}_i, \mathbf{V}_N))$. The main idea is that for any link receiver i, if we could find other link receivers with similar probability slices as i, we could then use the realized treatment assignment of these link receivers to estimate $(e(\mathbf{U}_i, \mathbf{V}_1), e(\mathbf{U}_i, \mathbf{V}_2), ..., e(\mathbf{U}_i, \mathbf{V}_N))$. Specifically, let $\mathcal{N}_i := \{i' : e(\mathbf{U}_{i'}, \cdot) \approx e(\mathbf{U}_i, \cdot)\}$ be the neighbourhood of link receiver i. Then an estimator for $e_i^j :=$

$e(\mathbf{U}_i, \mathbf{V}_j)$ would be

$$\tilde{e}_i^j = \frac{\sum_{i' \in \mathcal{N}_i} D_{i'}^j}{|\mathcal{N}_i|} \tag{3}$$

To define the neighborhood, we first need a definition of similarity, or equivalently the distance, between probability slices. Zhang et al. (2017) uses the d^2 distance:

$$d(i,i') = ||e(\mathbf{U}_i, \cdot) - e(\mathbf{U}_{i'}, \cdot)||_2 = \left\{ \int_v |e(\mathbf{U}_i,) - e(\mathbf{U}_{i'}, v)|^2) \right\}^{1/2}$$

Then

$$\begin{aligned} d(i,i')^2 &= \int_v e(u_i, v)e(u_i, v) + \int_v e(u_{i'}, v)e(u_{i'}, v) - 2\int_v e(u_i, v)e(u_{i'}, v) \\ &= \int_v (e(u_i, v) - e(u_{i'}, v))e(u_i, v) + \int_v (e(u_{i'}, v) - e(u_i, v))e(u_{i'}, v) \\ &\leq \left| \int_v (e(u_i, v) - e(u_{i'}, v))e(u_{\tilde{i}}, v) \right| + \left| \int_v (e(u_i, v) - e(u_{i'}, v))e(u_{\tilde{i}'}, v) \right| + 2e_N \\ &\leq \max_{k \neq i, i'} 2 \left| \int_v (e(u_i, v) - e(u_{i'}, v))e(u_k, v) \right| + 2e_N \end{aligned}$$

where $\tilde{i}$ and $\tilde{i}'$ are such that $|u_{\tilde{i}} - u_i| \leq e_N$ and $|u_{\tilde{i}'} - u_{i'}| \leq e_N$, and e_N depends on n and is the error rate. Zhang et al. (2017) shows that such $\tilde{i}$ and $\tilde{i}'$ can be found with high probability.

The first part of $\max_{k \neq i, i'} 2 \left| \int_v (e(u_i, v) - e(u_{i'}, v))e(u_k, v) \right|$ can be estimated by

$$\tilde{d}(i,i') = \max_{k \neq i, i'} \frac{|(\mathbf{D}_i - \mathbf{D}_{i'})\mathbf{D}_k'|}{n}.$$

Intuitively, neighbourhood $\mathcal{N}_i$ should include i' with small $\tilde{d}(i,i')$. Zhang et al. (2017) defines $\mathcal{N}_i$ as

$$\mathcal{N}_i = \{i' \neq i : \tilde{d}(i,i') \leq q_i(m)\}$$

where $q_i(m)$ is the m'th quantile of $\{i' \neq i : \tilde{d}(i,i')\}$. Zhang et al. (2017) showed that with $m = C(n^{-1} logn)^{1/2}$ for any constant $C \in (0,1]$, if the propensity score function $e(\cdot,\cdot)$ is

Piecewise-Lipschitz, then $\tilde{e}_i^j$ is consistent for $e(\mathbf{U}_i, \mathbf{V}_j)$.[13] Auerbach (2021) uses a similar idea but bounds the distance $d(i, i')$ with something else.

4.1.2 Factor models

From the previous section we have seen that any node exchangeable network can be expressed by the following generative factor model.

$$\mathbf{U}_i, \mathbf{V}_j \sim U[0,1], \quad i, j = 1, ..., N \tag{4}$$

$$D_i^j | \mathbf{U}_i, \mathbf{V}_j \sim \phi(\mathbf{U}_i, \mathbf{V}_j), \quad i, j = 1, ..., N, \phi \in \Phi \tag{5}$$

Estimating propensity scores using factor models require us to specify a functional form for ϕ. One possible choice is

$$\gamma_i \mathbf{U}_i, \beta_j \mathbf{V}_j \sim \mathcal{N}(0,1), \quad i, j = 1, ..., N \tag{6}$$

$$\phi(\mathbf{U}_i, \mathbf{V}_j) = logit(\gamma_i \mathbf{U}_i + \beta_j \mathbf{V}_j), \quad i, j = 1, ..., N \tag{7}$$

Using different parameters γ_i and β_j for $\mathbf{U}_i$ and $\mathbf{V}_j$ allows the sufficient confounders to affect linking probabilities in different ways when they act as the link receiver and when they act as the link sender. As an example, let $i = 1$ and $j = 2$ and $\gamma_1 \neq \beta_1$, $\gamma_2 \neq \beta_2$, $prob(D_1^2 = 1 | U_1, U_2) = logit(\gamma_1 U_1 + \beta_2 U_2)$ and $prob(D_2^1 = 1 | U_1, U_2) = logit(\beta_1 U_1 + \gamma_2 U_2)$ will be different, allowing different linking probabilities for the directed links within the same pair of nodes. Note that since the goal is to infer the propensity scores instead of $\mathbf{U}_i, \mathbf{V}_j$, we don't need to separately estimate the parameters γ_i and β_j. Therefore there is no need to separately specify a model for these parameters [14] [15] [16].

[13] Definition of Piecewise-Lipschitz: For any $\delta, L > 0$, let $\mathcal{F}_{\delta;L}$ denote a family of piecewise-Lipschitz functions m: $[0,1]^2 \to [0,1]$ such that (i) there exists an integer $K \geq 1$ and a sequence $0 = x_0 < \cdots < x_K$ satisfying $min_{0 \leq s \leq K-1}(x_{s+1} - x_s) \geq \delta$, and (ii) both $|e(u_1, v) - e(u_2, v)| \leq L|u_1 - u_2|$ and $|e(u, v_1) - e(u, v_2)| \leq L|u_1 - u_2|$ hold for all $u, u_1, u_2 \in [x_s, x_{s+1}]$, $v, v_1, v_2 \in [x_t, x_{t+1}]$ and $0 \leq s, t \leq K - 1$.

[14] From here we can see that for any value of U, γ and β, we can find another U', γ' and β' such that $\gamma U = \gamma U'$ and $\beta U = \beta U'$. This means $U_1, ..., U_N$ are not identifiable. In fact, they are identified up to a measure preserving transformation (**?**).

[15] We could also specify $\mathbf{U}_i$ as a vector of any length.

[16] When coding the model, this also means we only need to specify the models for α_i and η_i where $\alpha_i := \gamma_i \mathbf{U}_i$ and $\eta_j := \beta_j \mathbf{V}_j$.

To operationalize the use of factor models, I follow the deconfounding procedure proposed by Wang and Blei (2019). The deconfounder is a procedure proposed by Wang and Blei (2019) to address confounding in the setting of multiple treatments. It can be used in our setting because each link can be viewed as a treatment.

Applying the deconfounder to estimate the propensity scores involves three steps. In the first step, we need to randomly select a portion of links in the adjacency matrix and set them to 0. This is called masking. The resulting new adjacency matrix is our training data. The masked links are our validation data. In the second step, we need to pick a factor model and fit the factor model with the training data. In the third step, we need to use the validation data to test whether the factor model fits the data well enough. If the test is passed, then we proceed to the estimation with the estimated propensity scores. If the test is not passed, then we will pick another factor model, and step two is repeated until we find a factor model that passes the test[17].

4.2 2nd-step estimation: treatment effect

Once the propensity scores of link formation are estimated, we could use the many propensity score-based methods commonly used in the treatment effect estimation literature to estimate the linking effects of interest. In this section, I will use an inverse probability weighting estimator to illustrate how these propensity score-based methods can be adapted to the current setting. The most basic IPW estimator is the Horvitz–Thompson estimator. The augmented inverse probability weighting (AIPW) could be used to include covariates in the outcome model. AIPW is commonly referred to as the doubly robust estimator because it is consistent if either the propensity score is correctly estimated or the outcome model is correctly specified.

The IPW estimator for the linking effect

$$\tau_r^a := \mathbb{E}_{(i,j):R_i=r,A^j=a}[Y_i(D_i^j = 1, \mathbf{D}_i^{-j} = \bar{\mathbf{d}}_i^{-j}) - Y_i(D_i^j = 0, \mathbf{D}_i^{-j} = \bar{\mathbf{d}}_i^{-j})]$$

[17] The idea of using a statistical test on validation data to see if propensity scores are accurately estimated can also be used for the neighborhood smoothing estimator, or any other graphon estimator. In fact, a similar idea was used in Zhang et al. (2017) to compare the performance of different graphon estimators.

is

$$\frac{1}{\sum_{i=1}^{N} R_i = r} \cdot \frac{1}{\sum_{j=1}^{N} A^j = a} \Big(\sum_{i:R_i=r} \sum_{j:A^j=a} \frac{D_i^j \cdot Y_i^{obs}}{e(\mathbf{U}_i, \mathbf{V}_j)} - \sum_{i:R_i=r} \sum_{j:A^j=a} \frac{(1 - D_i^j) \cdot Y_i^{obs}}{1 - e(\mathbf{U}_i, \mathbf{V}_j)} \Big) \tag{8}$$

where $e(\mathbf{U}_i, \mathbf{V}_j)$ is substituted with its estimate since the true propensity score is unknown. Same as the conventional IPW estimator, the IPW estimator in equation 8 is unbiased for the linking effect τ_r^a. The proof is detailed in Section A4.4. A regression model (9) can be used incorporate additional controls, where each pairwise observation is weighted based on their propensity score.

$$Y_i = \alpha + \beta D_i^j + \theta Controls + \epsilon_i^j \tag{9}$$

5 Simulation

In this section, I conduct simulation exercises with synthetic data to assess the performance of the proposed linking effect estimators. I will generate the synthetic data according to the data generation model (10); one is a version of the homophile model, and the other is a statistical block model. Then I use a factor model to estimate the propensity scores. These propensity scores are then used in the second stage estimation with three different estimators: the inverse probability weighting (IPW) estimator, the nearest matching estimator, and the subclassification estimator. Finally, I will compute the bias and the mean absolute error (MAE) of the estimates relative to the true effect and compare them with the bias and MAE of the naive OLS estimator that ignores confounding.

$$
\begin{aligned}
\epsilon_i^c &\sim \mathcal{N}(0,1) \\
\epsilon_i^b &\sim U[0,1] \\
X_i &\sim Bernoulli(0.6) \\
C_i &\sim U[0,1] \\
\eta_{ij} &\sim U[0,1] \\
D_{ij} &= \mathbb{1}\{g(C_i, C_j) \geq \eta_{ij}\}, \quad g = g1, g2 \\
Y_i^c &= \alpha^c + \mathbf{D}_i\beta^c + \gamma^c C_i + \delta^c X_i + \epsilon_i^c \\
Y_i^b &= \mathbb{1}\{logit(\alpha^b + \mathbf{D}_i\beta^b + \gamma^b C_i + \delta^b X_i) \geq \epsilon_i^b\} \\
&\quad \text{where } logit(s) = \frac{1}{1 + exp(-s)}
\end{aligned} \tag{10}
$$

where $(\alpha^c, \gamma^c, \delta^c) = (0.5, 4, 1)$, $(\alpha^b, \gamma^b, \delta^b) = (-4, 4, 1)$. $\beta^{c(b)} = (\beta_1^{c(b)}, ...\beta_j^{c(b)}, ..., \beta_N^{c(b)})$ is a vector of parameters relating to the causal effect of a link from sender j. I set $\beta_j^c = X_j/2$ for all j. $\beta_j^b = X_j/2$ for all j. g_1 is specified in equation (11) and g_2 is specified in Section A6.1, equation (23).

$$
g_1 : P_i^j = 1/5\big(1 + exp(-(-6 + 2.5C_1 + 1.5C_j + |C_i - C_j|))\big) \tag{11}
$$

In this simulation exercise, I consider both continuous and binary outcome variables, which are denoted by Y^c and Y^b, respectively. The network links are generated through a binomial process with success probability specified according to two different link generation processes, g_1 as in equation (11) and g_2 as in equation (23). g_1 incorporates both degree heterogeneity and homophily. On the one hand, it is an increasing function in C_i and C_j. On the other hand, the probability of linking increases as the difference in C_i and C_j becomes smaller between the link receiver and the link sender. g_2 corresponds to a stochastic block model. The details of g_2 and its corresponding simulation result is detailed in Section A6.1. Both g_1 and g_2 generate directed networks. In our setup C_i is the confounder. It enters both the outcome and link formation equations and is unobserved to the econometrician. C_i, X_i,

ϵ_i^c,ϵ_i^b and η_{ij} are independent of each other, for all $i, j = 1, ...N$.

The mean degree distribution of g_1 from the simulated datasets is given in Table 1. As the network size increases, the degree increases. This is because the linking probability doesn't change as the network grows in our link generation model. This means the more nodes there are in the network, the more link senders there are, and thus the more links a link receiver will have.

Table 1: Mean degree distribution for simulated g1 networks

	0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
N=100	0	0	0	0	0	0.1	0.8	1	1.1	1.9	4.2
N=300	0	0	0.2	1	1	1.6	2	2.6	3.3	4.7	10.2
N=500	0	0.2	1	1.5	2	2.7	3.2	4.1	5.5	7.4	15.7

Note: This table reports the mean degree distribution of the simulated networks. For each size N=100,300,500, and for each simulated network of that size, I caculate the deciles of the number of links each link receiver receives, and average over all the 500 simulated networks of that size.

For the continuous outcome, I estimate the linking effects with the linear OLS regression (12), and the binary outcome is estimated with the logistic regression (13). I run these regressions separately for link senders with $X_j = 0$ and link senders with $X_j = 1$ to study the effects of these link senders separately. For the propensity score-based methods, the regressions are weighted with weights based on propensity scores that correct for confounding. For the naive OLS, the regressions are unweighted, thus not correcting for any confounding. The target estimand in this simulation exercise is ATT. This choice is reflected in the regression weights.

$$Y_i^c = \mu_i = \rho_0 + \rho_1 D_i^j + v_i \tag{12}$$

$$Pr(Y_i^b = 1) = \frac{1}{1 + exp\big(-(\rho_2 + \rho_3 D_i^j)\big)} \tag{13}$$

Table 2 compares the bias and MAE for the three propensity score-based estimators and the naive ols estimator. The propensity score used in this table is estimated using the factor model specified in equations (14)-(16). Comparing this factor model to the one in Section 4.1.2, Z_i can be seen as $\gamma_i \mathbf{U}_i$ and $\mathbf{V}_j$ can be seen as $\beta_j \mathbf{V}_j$ where $\mathbf{U}_i$ for $i = 1, ..., N$ are vectors of length two. The number of matches for the matching estimator is 1, and the number of subclasses for the subclassification estimator is 8. The rows under X_0 are the estimates for the linking effect of a link from a sender with $X_j = 0$, whose true effects are 0 on both the binary and the continuous outcomes. The rows under X_1 are the estimates for the linking effect of a link from a sender with $X_j = 1$, whose true effect is 0.5 on the continuous outcome. The true effect of an additional link from a sender with $X_j = 1$ on the binary outcome depends on the number of other links from senders with $X_j = 1$ because the true data generation process is non-linear. It is therefore calculated from the data generation process for each observation and then averaged over all observations.

$$Z_i = (z_{1i}, z_{2i}) \sim \mathcal{N}(0,1) \times \mathcal{N}(0,1), \quad i = 1, ..., N \tag{14}$$

$$K_j = (k_{1j}, k_{2j}) \sim \mathcal{N}(0,1) \times \mathcal{N}(0,1), \quad j = 1, ..., N \tag{15}$$

$$D_i^j | Z_i, K_j \sim Bernoulli\big(logit(Z_i + K_j)\big), \quad i, j = 1, ..., N \tag{16}$$

From Table 2, we can see that the estimators based on the propensity scores estimated by the factor model offer significant bias reduction compared to the naive ols estimator. The inverse probability weighting estimator performs the best among the three propensity score-based estimators. Compared to the naive ols estimator, the inverse probability weighting estimator reduces 90% - 97% of the biases for the binary outcome and 51% - 83% of the biases for the continuous outcome. As the network becomes larger, the bias reduction increases. An interesting observation from the table is that the bias from the naive ols estimator increases as the network becomes larger. This is because as the network becomes larger, the number of links for link receivers increases. This will lead to increasingly larger accumulated linking effects from all the other links being attributed to the effect of the link under consideration as in equation (8). This phenomenon doesn't happen if confounding is corrected because, in this case, the other links are independent of the link under consideration. As we see from the first three columns, the bias from the propensity score-based estimators continues to decrease as N increases despite the increasing bias from the naive ols estimator. Table (16) in Section A6 shows similar results for the statistical block model for network formation.

In Section A6, I also show the biases and MAEs of propensity score-based estimators using the factor model estimated propensity scores concerning the estimators using the true propensity scores (Table 19 and Table 20). Finally, I show simulation results when I increase the number of matches (from 1 to 3 to 5) and the number of subclasses (from 8 to 10 to 12) as the size of the network increases. The results from these different comparisons stay similar to the ones shown in Table 2.

6 Empirical Application

Almost everyone would agree that friendship is one of the most important social networks in a person's life. After all, one does not simply spend time with their friends; they also share information, receive their help, value their opinions, mimic their actions, and learn from their experiences. But it would be much more difficult to get everyone to agree on the direction and extent to which a person would be affected by their friends. The social network literature has long been interested in understanding the pattern of peer influence among friends for outcomes including risky behavior, smoking habits, obesity, education level, labor outcomes, fertility, etc. However, due to the obstacle posed by endogenous friendship formation, these questions remain largely unanswered, at least not in ways where the endogeneity issue is adequately accounted for.

Thanks to the theoretical results developed in this paper, I am able to make one of the first steps toward uncovering the true impacts of friendship. With the AddHealth data, I will be investigating the patterns of peer influence among high school friends in the U.S. Specifically, I look at how students' probability of graduating from college is affected by having more high-achieving friends, and whether this effect differs by both the gender of themselves and the gender of the high-achieving friend [18]. The analysis is inspired by the recent paper by Cools et al. (2019), which also uses the AddHealth data and finds that being exposed to more high-achieving males in one's high school decreases the likelihood that a

[18] A high-achieving student is defined as a student who has at least one residential parent with a postgraduate degree. This is the same definition used in Cools et al. (2019)

Table 2: Simulation results for g_1

				IPW	Matching	Sub	Naive ols
Yb	Bias	X_0					
			N=100	0.077445	0.096851	0.093895	0.132864
			N=300	0.051917	0.086736	0.091974	0.1705
			N=500	0.033176	0.084199	0.087752	0.184117
		X_1					
			N=100	0.078718	0.094838	0.092844	0.132779
			N=300	0.04753	0.083476	0.087602	0.166086
			N=500	0.034532	0.085371	0.089037	0.185265
	MAE	X_0					
			N=100	0.102707	0.137418	0.111374	0.142808
			N=300	0.054298	0.087305	0.091974	0.1705
			N=500	0.03435	0.084199	0.087752	0.184117
		X_1					
			N=100	0.09447	0.11907	0.103271	0.137679
			N=300	0.050589	0.0838	0.087611	0.166086
			N=500	0.036061	0.085371	0.089037	0.185265
Yc	Bias	X_0					
			N=100	0.494439	0.591209	0.583515	0.802809
			N=300	0.261106	0.483215	0.498769	0.9596
			N=500	0.173155	0.512221	0.533149	1.144263
		X_1					
			N=100	0.454354	0.534451	0.539381	0.765181
			N=300	0.257676	0.47056	0.493571	0.95376
			N=500	0.174887	0.507023	0.533092	1.142917
	MAE	X_0					
			N=100	0.518549	0.62927	0.595459	0.806389
			N=300	0.26409	0.483215	0.498769	0.9596
			N=500	0.176396	0.512221	0.533149	1.144263
		X_1					
			N=100	0.466298	0.558012	0.542142	0.765513
			N=300	0.258652	0.47056	0.493571	0.95376
			N=500	0.176375	0.507023	0.533092	1.142917

Note: This table reports for the g_1 model the bias and the mean absolute error (MAE) of the inverse probability weighting estimator, the nearest neighbour matching estimator with replacement, the subclassification estimator and the narive ols estimator, compared to the true linking effects, for link sender with $X_j = 0$ and $X_j = 1$ separately. The number of matches for the matching estimator is 1, the number of subclasses for the subclassification estimator is 8. All the estimates are for the average treatment effect for the treated.

female student obtaining a bachelor's degree. It also finds that this negative effect could be partly explained by a decrease in the girls' confidence and aspirations, as well as their grades in math and science. But do high-achieving male *friends* also have this negative impact on girls? At the end of the day, interactions and social influence among close friends could be very different from those among students who simply attend the same school and might not have close and friendly interactions.

The results indicate that the effect of friendship could indeed be very different from the effect of cohort peers. Noticeably, an additional male high flyer friend increases the probability of a female student obtaining a bachelor's degree by 3 percentage points. Heterogeneity analysis reveals that this positive effect of male high flyer friendship is mainly driven by female students with below median ability as measured by their PVT score. Evidence also suggests that the effect mainly comes from a confidence boost instead of a tangible influence on their GPA.

6.1 Data

The data is from the National Longitudinal Study of Adolescent to Adult Health (Add Health) [19]. It is a longitudinal study of a nationally representative sample of adolescents in grades 7-12 in the United States during the 1994-95 school year (Wave I). In total, 172 schools were sampled. The Wave I data consists of an in-school questionnaire for all students in the sampled schools, followed by an in-home interview conducted for only a sample of these students. Out of the 172 schools, 16 are the so-called saturated schools, where all students who answered the in-school questionnaire were selected for the in-home interview. The sample of students who answered the Wave I in-home interview was interviewed again during the 1995-1996 school year (Wave II), another time in 2001-2002 (Wave III), again in

[19] This research uses data from Add Health, funded by grant P01 HD31921 (Harris) from the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD), with cooperative funding from 23 other federal agencies and foundations. Add Health is currently directed by Robert A. Hummer and funded by the National Institute on Aging cooperative agreements U01 AG071448 (Hummer) and U01AG071450 (Aiello and Hummer) at the University of North Carolina at Chapel Hill. Add Health was designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill. The Add Health Parent Study/Parents (2015-2017) data collection was funded by a grant from the National Institute on Aging (RO1AG042794) to Duke University, V. Joseph Hotz (PI) and the Carolina Population Center at the University of North Carolina at Chapel Hill, Kathleen Mullan Harris (PI).

2007-2008 (Wave IV), and most recently in 2016-2018 (Wave V).

For my empirical analysis, information on educational attainment is taken from the Wave IV data, when respondents were between 26-32 years old. They were asked to give their highest level of education achieved by the time of the interview. As in Cools et al. (2019), I define a dummy variable for bachelor's degree attainment equal to 1 if the respondent had obtained a four-year college degree or more and 0 otherwise. Some other secondary outcome variables are also used in this analysis. These include Wave II information on students' grades, willingness and confidence in going to college, and self-assessment of their intelligence compared to their peers.

Friendship information comes from the Wave I in-home interviews. During the interview, students were asked to nominate at most five of their female friends and five of their male friends from their school's and the sister school's roaster. Students' pre-treatment information comes from Wave I. This includes background information on the students and their parents. On the students' side, I use data on their gender, age, race, whether they were born in the US, and their PVT score [20]. On the parents' side, I use data on the residential mother and father's education level, whether they worked for pay for more than 10 hours per week at the time interview was conducted, whether they were born in the U.S., and the annual family income. The exact definitions of all variables are detailed in Table 11, along with the definitions used in Cools et al. (2019). In order to compare the results with the CFP paper, I further restrict the data following their procedure, keeping only those in grades 7-12 during Wave I, except those with less than 20 students.

6.2 Estimation of propensity scores and the linking effects

The first step of estimating the linking effect is to estimate the propensity scores from the adjacency matrix. When students were interviewed for the AddHealth data, they were only allowed to nominate their friends within the same school. This means that for each school s, we have a network represented by an adjacency matrix $\mathbf{D}_s$ with N_s nodes. The N_s nodes include every student on the school roaster. In each school, a sample of n_s students who

[20] A Picture Vocabulary Test (PVT) was administered by the interviewer during the Wave I in-home interview. PVT measures an individual's verbal ability.

were also in the school roaster was selected for the in-home interview and therefore asked to nominate their friends from the N_S students listed on the roaster. For each i of the sampled students and each student j on the school roaster, $D^j_{s,i}$ is recorded as 1 if i nominates j as their friend and is recorded as 0 if j is nominated by i as a friend. I remove any column j of the adjacency matrix $\mathbf{D}_s$ if j was not nominated by any sampled student i. For the $N_s - n_s$ students who were not sampled for the in-home interview, their adjacency matrix entries are missing, which prevents us from estimating their propensity scores of linking. This is not a problem for our analysis for two reasons. First, since they were not selected for the in-home interviews, their information on outcome variables would also be missing, meaning they wouldn't have been included in the analysis anyway. Second, the propensity scores of linking of the sampled students can still be estimated through factor models, even though they can no longer be estimated by graphon estimators. The factor model I use for this empirical analysis is the same as specified in (10).

After the propensity scores of linking are estimated for all the sampled students in each school, we are ready to estimate the linking effects of interest. In this empirical analysis, I use the augmented inverse probability weighting estimator. Specifically, I run the propensity score re-weighted pairwise regression specified in (17) for each characterization of the link receivers and the link senders c. For example c_1 could correspond to female link receivers and male high-achieving senders.

$$Y_{c,s,i} = \beta_{c,s} + \beta_{c,s} D^{c,s,j}_{c,s,i} + \rho_{c,s} \mathbf{X}_{c,s,i} + \epsilon^{c,s,j}_{c,s,i} \tag{17}$$

where $Y_{c,s,i}$ and $\mathbf{X}_{c,s,i}$ are respectively the outcome and covariates of student i that satisfies restriction c in school s. $D^j_{c,s,i}$ is a dummy variable that equals to 1 if student i nominates j as their friend where both i and j are from school s and satisfies restriction c in school s. Each pairwise observation is weighted according to its propensity of linking and its linking status. Here I estimate the treatment effect of treated (ATT), which means the weights are generated according to (18).

$$w_{c,s,i}^{c,s,j} = \begin{cases} 1 & \text{if } D_{c,s,i}^{c,s,j} = 1 \\ \frac{p_{c,s,i}^{c,s,j}}{1-p_{c,s,i}^{c,s,j}} & \text{if } D_{c,s,i}^{c,s,j} = 0 \end{cases} \tag{18}$$

where $w_{c,s,i}^{c,s,j}$ is the pairwise weight and $p_{c,s,i}^{c,s,j}$ is the estimated propensity of linking from j to i. Note that using the propensity score weighted regressions to estimate the linking effects does not mean we assume the true effect is linear and additive with respect to the covariates. Just like in traditional causal inference, regressions are only used as a way of estimation.

Wang and Blei (2019) suggested using a test statistics to assess the adequacy of propensity score estimation. This test statistics is based on the idea that well-estimated propensity scores should have good predictive power for the validation data. Following their procedure, our estimated propensity scores for each school network pass the test and perform well.

Traditionally, the adequacy of the estimated propensity scores is assessed by balance tests, where the difference in pre-treatment variables between the treated group and the control group is calculated using the propensity score-adjusted sample. This method is not directly applicable to our context. First of all, since each link sender is associated with a unique treatment, ideally, we would compare for each link sender the pre-treatment characteristics of the students who were treated by this link sender and the students who were not treated by this link sender. Because in our networks of finite size, each link sender only has a few treated students, this comparison suffers from finite sample bias. We could, however, average the differences in pre-treatment variables between treated and control students over all link senders. The second issue is that our propensity scores are based on the unobserved sufficient confounders that do not correspond directly to any observed variables. Since the propensity scores were not estimated using any observed pre-treatment variables, there is no guarantee that any selected pre-treatment variable will be balanced across the treated and the control groups. Nonetheless, we could still evaluate the balance for some variables we believe are part of the confounders.

Balance tests could be conducted by running a pairwise regression similar to 17, except that the covariates will become the outcome variables. Table 3 shows the result of a balance test for some pre-treatment variables. According to Currarini et al. (2009) race is a strong predictor of friendship formation, and (Carrell et al., 2013) suggests the same for ability. Table 3 shows that the balance for the ability variables (column 3 and column 4) and the race variable of being black are improved.

Table 3: Balance test

	Pre-treatment variable:											
	Male	US born	PVT	PVT +	M C+	F C +	Income	Age	M nHH	F nHH	Black	Hispanic
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Original	0.002	−0.002***	0.688***	0.025***	0.023***	0.022***	0.037***	−0.993***	−0.004***	−0.006***	−0.014***	−0.001
	(0.002)	(0.001)	(0.041)	(0.002)	(0.001)	(0.001)	(0.003)	(0.043)	(0.001)	(0.001)	(0.001)	(0.001)
IPW	0.001	−0.001	0.449***	0.016***	0.017***	0.012***	0.036***	−0.637***	−0.002**	−0.006***	−0.008***	0.001
	(0.002)	(0.001)	(0.044)	(0.002)	(0.001)	(0.002)	(0.003)	(0.048)	(0.001)	(0.001)	(0.001)	(0.001)

Note: This table reports the average differences between the treated and the control across all link senders. The first row is the balance test for the origianl sample. The second row is the balance test for the sample re-weighted by the propensity scores according to inverse probability weighting method. Standard errors are estimated with subsample bootstrapping with 900 subsamples drawn randomly. At each bootstrap, 90% of the individuals (nodes) within each school are sampled without replacement. The pre-treatment variables from column 1 to column 12 are: whether the ego is male, born in US, their PVT score, whether their PVT score is above the population median, whether their mother has colllege degree or above, whether their father has colllege degree or above, their annual family income (log), their age in months, whether their mother is not in the household, whether their father is not in the household, whether the respondent is black, and whether the respondent is hispanic. *p<0.1; **p<0.05; ***p<0.01

6.3 Results

Table 4 reports the estimated effect of friendships from different types of link sender on bachelor degree attainment (column 1) and some intermediate outcomes recorded during Wave II interviews. Each row corresponds to a characterization of the friendship based on the character of the receiver and the sender. The receiver characteristic is shown before the underbar "_", and the sender characteristic is shown after. "F" and "M" refer to the gender female and male, respectively. "H" and "L" refer to whether the individual is a high achiever or non-high achiever (low achiever), respectively. For example, "F_FL" means the linking effect is estimated for female link receivers and female non-high achiever link senders.

Table 4 shows a close to 3 p.p increase in female students' likelihood of obtaining a bachelor's degree by having an additional male high-achieving friend. For male students, an extra male high-achieving friend means an increase of 4.6 p.p in the probability of graduating from college. Looking at the last three columns of the table, it seems that the positive effect of a male high-achieving friend on both female and male students could be attributed to an

Table 4: Effect of friendship on bachelor degree attainment and confidence

	Dependent variable:			
	Bachelor Degree (p.p)	Want (p.p)	Will (p.p)	Intelligence (p.p)
	(1)	(2)	(3)	(4)
F_FL	0.354*	−0.171	−0.819***	−0.389
	(0.191)	(0.241)	(0.227)	(0.242)
F_ML	0.336	−0.361	−0.797**	−0.602
	(0.313)	(0.381)	(0.373)	(0.439)
F_FH	1.877	2.377*	0.737	1.364
	(1.262)	(1.245)	(1.062)	(1.345)
F_MH	2.981***	1.602	2.370***	3.748***
	(0.978)	(1.116)	(0.858)	(1.324)
M_FL	−0.041	0.144	−0.026	−0.623*
	(0.279)	(0.269)	(0.270)	(0.336)
M_ML	−0.068	0.058	−0.553**	−0.816***
	(0.227)	(0.204)	(0.247)	(0.253)
M_FH	2.801	0.930	−1.919	−1.652
	(1.906)	(1.818)	(1.764)	(1.773)
M_MH	4.645***	1.361	0.821	4.539***
	(1.526)	(1.544)	(1.314)	(1.153)

Note: This table reports the estimated effects of high school friendship on students' bachelor degree attainment (column 1), and their intermediate outcomes (column 2-4). The dependent variable in Column (2) is a dummy variable recording whether the student reported a scale 5 (1 is the lowest and 5 is the highest) on the the extent of how much they want to go to college (Wave II). The dependent variable in Column (3) is a dummy variable recording whether the student reported a scale 5 (1 is the lowest and 5 is the highest) on the likelihood that they will go to college (Wave II). The dependent variable in Column (4) is a dummy variable recording whether the student reported a scale 5 or 6 (1 is the lowest and 6 is the highest) on their intelligence compared to other people of their age (Wave II). The estimands are all ATT. Each row corresponds to a characterisation of the friendship, based on the character of the receiver and the sender. Receiver characteristics is shown before the underbar _, and sender characteristics is shown after. "F" and "M" are used to refer to the gender female and male respectively. "H" and "L" are used to refer to whether the individual is a high flyer or non-high flyer (low flyer) respectively. For example, "F_FL" means the linking effect is estimated for female link receivers and female non-high flyer link senders. The regressions reported in all columns include cohort dummies, whether the student was born in the US, their PVT score, whether their PVT score is above the population median PVT score, whether their mother's and father's highest degree is high school, some college, college, or post college, whether their mother's and father's highest education level is missing, the student's log family income, whether family is missing, the age of the student during Wave I, whether the student's mother and father were in the household, dummies for whether the student is black, hispanic, white, asian and indian. Standard errors are estimated with subsample bootstrapping with 900 subsamples drawn randomly. At each bootstrap, 90% of the individuals (nodes) within each school are sampled without replacement. *p<0.1; **p<0.05; ***p<0.01

Table 5: Heterogeneous effects of friendship on bachelor degree attainment and intelligence

	Dependent variable:			
	Bachelor degree		Intelligence	
	PVT Median -	PVT Median +	PVT Median -	PVT Median +
	(1)	(2)	(3)	(4)
F_FL	0.510	−0.347	−1.121***	−0.026
	(0.362)	(0.424)	(0.429)	(0.451)
F_ML	0.818	−1.038*	−0.139	−1.823**
	(0.634)	(0.540)	(0.836)	(0.903)
F_FH	5.912***	−0.782	6.552**	−1.115
	(2.294)	(2.656)	(3.054)	(2.531)
F_MH	3.649*	0.492	10.283***	−2.967
	(2.084)	(1.952)	(2.585)	(2.620)
M_FL	−0.780	0.329	−1.916**	1.550***
	(0.649)	(0.581)	(0.802)	(0.568)
M_ML	0.670	−0.260	−2.051***	0.736
	(0.451)	(0.423)	(0.474)	(0.462)
M_FH	0.305	3.377	−5.573	−0.074
	(5.056)	(2.160)	(3.607)	(2.236)
M_MH	4.231	8.267***	−2.164	11.954***
	(3.005)	(2.511)	(2.911)	(2.225)

Note: This table reports the estimated heterogeneous effects of high school friendship on students' bachelor degree attainment and self-assessed intelligence. Column (1) and (3) reports results for ego whose PVT score is below population median PVT score. Column (2) and (4) reports results for ego whose PVT score is above population median PVT score. The estimands are all ATT. Each row corresponds to a characterisation of the friendship, based on the character of the receiver and the sender. Receiver characteristics is shown before the underbar _, and sender characteristics is shown after. "F" and "M" are used to refer to the gender female and male respectively. "H" and "L" are used to refer to whether the individual is a high flyer or non-high flyer (low flyer) respectively. For example, "F_FL" means the linking effect is estimated for female link receivers and female non-high flyer link senders. The regressions reported in all columns include cohort dummies, whether the student was born in the US, their PVT score, whether their PVT score is above the population median PVT score, whether their mother's and father's highest degree is high school, some college, college, or post college, whether their mother's and father's highest education level is missing, the student's log family income, whether family is missing, the age of the student during Wave I, whether the student's mother and father were in the household, dummies for whether the student is black, hispanic, white, asian and indian. Standard errors are estimated with subsample bootstrapping with 900 subsamples drawn randomly. At each bootstrap, 90% of the individuals (nodes) within each school are sampled without replacement. *p<0.1; **p<0.05; ***p<0.01

increase in their confidence. In particular, an additional male high-achieving friend increases the probability that a female student reports having a high likelihood of going to college and being more intelligent than their same-age peers during the Wave II interview, one year after friendship information was recorded. As for male students, their self-assessment of being more intelligent than their same-age peers is also increased.

Female egos are also slightly more likely to graduate from college when they have an additional female friend who is not a high achiever. However, this effect disappears if we separately look at the effect on low-ability and high-ability female students. As shown in Table 5, estimates for both ability groups of female students are not significantly different from 0. Moreover, the positive effect of male high achiever friends seems to only exist for low-ability female students and high-ability male students, with an increase in the probability of going to college by about 3.6 p.p and 8.3 p.p, respectively. These positive effects are also found in their self-assessment of being more intelligent than their peers. However, is this positive impact on self-assessment of intelligence due to a confidence boost or an increase in academic performance? To answer this question, I look at the effect of friendship on egos' grades during Wave II. Table 6 and Table 7 show that across all four academic subjects, none of the grades of low-ability female students were increased by having an additional male high-achieving friend. As for male high-ability students, their English grade was improved by 0.196 points on average (lowest 1, highest 5) by having an additional male high-achieving friend, but none of the grades of the other subjects were improved.

7 Conclusion

This paper develops a new conceptual framework to study the causal effects of peer relationships. It allows researchers to define flexible estimands, including the commonly used linear-in-means/sum peer effect parameter. Moreover, it provides identification conditions for endogenously formed networks without the need for natural experiments. This identification is based on an adapted version of the traditional unconfoundedness condition. Thanks to the nature of network data, I show that confounders can be inferred from the adjacency matrix. Therefore identification does not require the traditional assumption that

Table 6: Heterogeneous effects of friendship on English and Math grades

	Dependent variable:			
	English grade		Math grade	
	PVT Median -	PVT Median +	PVT Median -	PVT Median +
	(1)	(2)	(3)	(4)
F_FL	-0.004	-0.002	0.002	-0.007
	(0.007)	(0.008)	(0.007)	(0.008)
F_ML	-0.035^{**}	0.037^{**}	0.002	-0.020
	(0.016)	(0.017)	(0.014)	(0.019)
F_FH	0.050	0.014	0.232^{***}	0.022
	(0.045)	(0.028)	(0.053)	(0.025)
F_MH	-0.025	0.050	0.039	-0.012
	(0.036)	(0.032)	(0.037)	(0.041)
M_FL	0.021	-0.006	-0.044^{**}	-0.008
	(0.015)	(0.010)	(0.019)	(0.008)
M_ML	0.009	0.00001	-0.025^{***}	0.004
	(0.008)	(0.006)	(0.009)	(0.008)
M_FH	0.159^{**}	-0.00003	0.080^{*}	0.061
	(0.079)	(0.067)	(0.043)	(0.045)
M_MH	-0.039	0.196^{***}	0.060	0.002
	(0.055)	(0.053)	(0.075)	(0.043)

Note: This table reports the estimated heterogeneous effects of high school friendship on students English and Math grades (Wave II). Column (1) and (3) reports results for ego whose PVT score is below population median PVT score. Column (2) and (4) reports results for ego whose PVT score is above population median PVT score. The estimands are all ATT. Each row corresponds to a characterisation of the friendship, based on the character of the receiver and the sender. Receiver characteristics is shown before the underbar _, and sender characteristics is shown after. "F" and "M" are used to refer to the gender female and male respectively. "H" and "L" are used to refer to whether the individual is a high flyer or non-high flyer (low flyer) respectively. For example, "F_FL" means the linking effect is estimated for female link receivers and female non-high flyer link senders. The regressions reported in all columns include cohort dummies, whether the student was born in the US, their PVT score, whether their PVT score is above the population median PVT score, whether their mother's and father's highest degree is high school, some college, college, or post college, whether their mother's and father's highest education level is missing, the student's log family income, whether family is missing, the age of the student during Wave I, whether the student's mother and father were in the household, dummies for whether the student is black, hispanic, white, asian and indian. Standard errors are estimated with subsample bootstrapping with 900 subsamples drawn randomly. At each bootstrap, 90% of the individuals (nodes) within each school are sampled without replacement. *p<0.1; **p<0.05; ***p<0.01

Table 7: Heterogeneous effects of friendship on History and Science grades

	Dependent variable:			
	History grade		Science grade	
	PVT Median -	PVT Median +	PVT Median -	PVT Median +
	(1)	(2)	(3)	(4)
F_FL	−0.008	−0.013	-0.025^{***}	−0.011
	(0.006)	(0.008)	(0.007)	(0.013)
F_ML	0.046^{***}	0.014	−0.014	0.022
	(0.016)	(0.013)	(0.017)	(0.018)
F_FH	0.081	−0.025	0.142^{**}	−0.030
	(0.058)	(0.032)	(0.062)	(0.026)
F_MH	−0.028	0.104^{***}	0.007	0.010
	(0.041)	(0.037)	(0.046)	(0.026)
M_FL	−0.035	0.030^{***}	-0.042^{**}	0.007
	(0.024)	(0.008)	(0.016)	(0.008)
M_ML	-0.025^{***}	-0.014^{*}	0.010	−0.002
	(0.008)	(0.008)	(0.013)	(0.006)
M_FH	0.091	-0.129^{**}	0.086^{**}	-0.130^{**}
	(0.059)	(0.051)	(0.042)	(0.052)
M_MH	−0.093	0.037	0.004	−0.033
	(0.061)	(0.037)	(0.078)	(0.036)

Note: This table reports the estimated heterogeneous effects of high school friendship on students History and Science grades (Wave II). Column (1) and (3) reports results for ego whose PVT score is below population median PVT score. Column (2) and (4) reports results for ego whose PVT score is above population median PVT score. The estimands are all ATT. Each row corresponds to a characterisation of the friendship, based on the character of the receiver and the sender. Receiver characteristics is shown before the underbar _, and sender characteristics is shown after. "F" and "M" are used to refer to the gender female and male respectively. "H" and "L" are used to refer to whether the individual is a high flyer or non-high flyer (low flyer) respectively. For example, "F_FL" means the linking effect is estimated for female link receivers and female non-high flyer link senders. The regressions reported in all columns include cohort dummies, whether the student was born in the US, their PVT score, whether their PVT score is above the population median PVT score, whether their mother's and father's highest degree is high school, some college, college, or post college, whether their mother's and father's highest education level is missing, the student's log family income, whether family is missing, the age of the student during Wave I, whether the student's mother and father were in the household, dummies for whether the student is black, hispanic, white, asian and indian. Standard errors are estimated with subsample bootstrapping with 900 subsamples drawn randomly. At each bootstrap, 90% of the individuals (nodes) within each school are sampled without replacement. *p<0.1; **p<0.05; ***p<0.01

all confounders have been observed. My identification result suggests the use of propensity score-based estimators, which means estimation can be easily and flexibly done with existing statistical packages.

Empirical application on the effect of high school friendship on students' likelihood to obtain a bachelor's degree reveals interesting results not known in the literature before. Contrary to the findings in Cools et al. (2019), female students, particularly those with low ability, benefit from having high-achieving male friends. They become more confident in their intelligence and are more likely to graduate from college, despite no evidence that their grades improved. As for male students, it is the high-ability ones that benefit from having high-achieving male friends. Their English grade improves significantly, and they become more confident and are more likely to graduate from college.

Additional analysis and results for the study of linking effect

Zheng Wang
European University Institute

October 10, 2022

(For the latest draft)

A1 Extensions

A1.1 Treatment defined over all links

Suppose we are interested in the comparison between two configurations, such as c^1 and c^2. A configuration is a rule C that the treatment vector has to satisfy. For example, c^1 could be 2 female and 1 male and c^2 be 1 female and 2 male. Assume L-SUTVA holds, for any node i let us denote the set of treatments that satisfies configuration c as $D_i^c = \{\mathbf{D}_i | C(\mathbf{D}_i) = c\}$, where $\mathbf{D}_i = (D_{i1}, ...D_{ij}, ..., D_{iN})$. For any $d^{c_1} \in D^{c_1}$ and $d^{c_2} \in D^{c_2}$. we can define an estimand $m_i^{c_1,c_2}$:

$$m_i^{d^{c_1},d^{c_2}} = Y_i(d^{c_1}) - Y_i(d^{c_2})$$

*European University Institute. I would like to thank Andrea Ichino, Yann Bramoulle and Fabrizia Mealli for their valuable feedbacks and guidance throughout the development of this paper. I would also like to thank participants at the EUI Microeconometrics Working Group and the AMSE PhD seminar for their helpful comments.

For any configuration c, use $|D^c|$ to denote the number of elements in the set D^c and the expectation $\mathbb{E}_c$ as the expectation over the set D^c with uniform probability. Average over the set of treatments that satisfy the configuration rules, we can define the treatment effect of configuration c^1 v.s. c^2 on node i as:

$$\begin{aligned} m_i^{c_1,c_2} &= \mathbb{E}_{c_1}[Y_i(d^{c_1})] - \mathbb{E}_{c_2}[Y_i(d^{c_2})] \\ &:= \frac{1}{|D^{c_1}|} \sum_{d^{c_1} \in D^{c_1}} Y_i(d^{c_1}) - \frac{1}{|D^{c_2}|} \sum_{d^{c_2} \in D^{c_2}} Y_i(d^{c_2}) \end{aligned}$$

Finally, by averaging over the set of egos, we can easily define the average treatment effect of configuration c^1 v.s. c^2 as:

$$\begin{aligned} m^{c_1,c_2} &= \mathbb{E}_i \left[\mathbb{E}_{c_1}[Y_i(d^{c_1})] - \mathbb{E}_{c_2}[Y_i(d^{c_2})] \right] \\ &:= \frac{1}{N} \sum_{i=1,\dots,N} \big(\frac{1}{|D^{c_1}|} \sum_{d^{c_1} \in D^{c_1}} Y_i(d^{c_1}) - \frac{1}{|D^{c_2}|} \sum_{d^{c_2} \in D^{c_2}} Y_i(d^{c_2}) \big) \end{aligned}$$

Lemma 4 (Unconfoundedness when treatment is defined over all links)**.**

$$Pr(D_i = d^c | Y_i^{pot}, \mathbf{U}_i, \mathbf{V}_1, ..., \mathbf{V}_N) = Pr(D_i = d^c | \mathbf{U}_i, \mathbf{V}_1, ..., \mathbf{V}_N)$$

and

$$Pr(D_i = d^c | Y_i^{pot}, e(\mathbf{U}_i, \mathbf{V}_1), ..., e(\mathbf{U}_i, \mathbf{V}_N)) = Pr(D_i = d^c | e(\mathbf{U}_i, \mathbf{V}_1), ..., e(\mathbf{U}_i, \mathbf{V}_N))$$

Proof. The first half of the proof is identical to that of Proposition 1. For the last part, instead we have

$$\begin{aligned} &Pr(D_i = d^c | \mathbf{U}_1, ..., \mathbf{U}_N, \mathbf{V}_1, ..., \mathbf{V}_N, \mathbf{Y}_i^{pot}) \\ =&Pr(D_i = d^c | \mathbf{U}_1, ..., \mathbf{U}_N, \mathbf{V}_1, ..., \mathbf{V}_N) \\ =&Pr(D_i = d^c | \mathbf{U}_i, \mathbf{V}_1, ..., \mathbf{V}_N) \end{aligned}$$

The first equation holds because we have ruled out any confounders that affect any of the links, which means there are no confounders to affect all of i's links. The second equation comes from Equation (2). $\square$

Assumption 5 (Overlap for all links)**.** $0 < Pr(\mathbf{D}_i = d^{c_1}|\mathbf{U}_i, \mathbf{V}_1, ..., \mathbf{V}_N) < 1$

Proposition 3. Under assumption 1,2,3 and 5, m^{c_1,c_2} is identified:

$$
\begin{aligned}
m^{c_1,c_2} = \mathbb{E}\Bigg[\mathbb{E}_i\left[\mathbb{E}_{d^{c_1}}[Y_i(\mathbf{D}_i = d^{c_1})|e(\mathbf{U}_i, \mathbf{V}_1), ..., e(\mathbf{U}_i, \mathbf{V}_N), \mathbf{D}_i = d^{c_1}]\right]\Bigg] \\
- \mathbb{E}\Bigg[\mathbb{E}_i\left[\mathbb{E}_{d^{c_2}}[Y_i(\mathbf{D}_i = d^{c_2})|e(\mathbf{U}_i, \mathbf{V}_1), ..., e(\mathbf{U}_i, \mathbf{V}_N), \mathbf{D}_i = d^{c_2}]\right]\Bigg]
\end{aligned}
$$

Proposition (3) is proved in Section A4.5.

Notice here in order to estimate this estimand, we need to condition not just on the single pairwise propensity score $e(\mathbf{U}_i, \mathbf{V}_j)$, but rather on the vector of propensity scores $e(\mathbf{U}_i, \mathbf{V}_1), ..., e(\mathbf{U}_i, \mathbf{V}_N)$. To gain some intuition, first recall that in the main analysis, the hypothetical intervention was on a single pair, and the estimand is the average of potential outcomes under repeated hypothetical interventions over different pairs each time. Here the hypothetical intervention, however, is on all the relationships of node i, thus the need to condition on the propensity scores of all relationships being formed.

Finally, note that as N goes to infinity, the overlap condition will fail to hold. To see why, write the generalised propensity score as the product of individual pairwise propensity score:

$$
\begin{aligned}
&Pr(\mathbf{D}_i = d^{c_1}|\mathbf{U}_i, \mathbf{V}_1, ..., \mathbf{V}_N) \\
&= \prod_{j=1}^{N}\left(Pr(D_i^j = 1|\mathbf{U}_i, \mathbf{V}_j)\right)^{d_i^{c_1}}\left(1 - Pr(D_i^j = 1|\mathbf{U}_i, \mathbf{V}_j)\right)^{1-d_i^{c_1}}
\end{aligned}
$$

Since $0 < Pr(D_i^j = 1|\mathbf{U}_i, \mathbf{V}_j) < 1$, this product goes to 0 as N goes to infinity, causing the overlap condition to fail. However, if our target estimand is ATT, this shouldn't be a problem: the existence of treated nodes means the propensity score is larger than 0.

A1.2 Alternative estimands

In the main analysis the treatment effect of sender-j relationship on receiver i's potential outcome is defined as the following contrast of potential outcomes:

$$\tau_i^j = Y_i(D_i^j = 1, \mathbf{D}_i^{-j} = \bar{\mathbf{d}}_i^{-j}) - Y_i(D_i^j = 0, \mathbf{D}_i^{-j} = \bar{\mathbf{d}}_i^{-j})$$

where all the non-sender-j relationships of receiver i are fixed at their observed level. This is only one of the many ways we can define the pair level estimand. In fact, for any i, j and $\mathbf{d}_i^{-j}$ we could define

$$\tilde{\tau}_i^j(\mathbf{d}_i^{-j}) = Y_i(D_i^j = 1, \mathbf{D}_i^{-j} = \mathbf{d}_i^{-j}) - Y_i(D_i^j = 0, \mathbf{D}_i^{-j} = \mathbf{d}_i^{-j}) \tag{19}$$

In this case, we could define an average linking effect for link receivers with characteristic $R_i = r$ and link senders with characteristic $A^j = a$ by averaging the pair level treatment effects over the probability distribution of the linking status of i's other (than j) relationships:

$$\tilde{\tau}_r^a = \mathbb{E}_{(i,j):R_i=r,A^j=a} \sum_{\mathbf{d}_i^{-j} \in \mathfrak{D}^j} \tilde{\tau}_i^j(\mathbf{d}_i^{-j}) Prob(\mathbf{D}_i^{-j} = \mathbf{d}_i^{-j}) \tag{20}$$

where $\mathfrak{D}^j = \cup_i \mathbf{d}_i^{-j}$ and i is a representative node randomly drawn from the population of senders satisfying $R_i = r$, [1]. Next I will prove that $\tilde{\tau}_r^a$ is identified.

[1] This estimand is similar to the kind of estimands usually defined in the literature of treatment interference, e.g. Forastiere et al. (2021). The difference is that in the treatment inference literature the "direct" or main estimand is defined by averaging over the treatments of interfering units, while here we average over the non-focal links of the same receiver.

Proof.

$$
\begin{aligned}
&\mathbb{E}_{(i,j):R_i=r,A^j=a}\big[\sum_{\mathbf{d}_i^{-j}\in\mathfrak{D}^j} Y_i(D_i^j=1,\mathbf{D}_i^{-j}=\mathbf{d}_i^{-j})Prob(\mathbf{D}_i^{-j}=\mathbf{d}_i^{-j})\big]\\
&=\mathbb{E}\Big[\mathbb{E}_{(i,j):R_i=r,A^j=a}\big[\sum_{\mathbf{d}_i^{-j}\in\mathfrak{D}^j} Y_i(D_i^j=1,\mathbf{D}_i^{-j}=\mathbf{d}_i^{-j})Prob(\mathbf{D}_i^{-j}=\mathbf{d}_i^{-j})|\mathbf{U}_i,\mathbf{V_1},...,\mathbf{V_N}\big]\Big]\\
&=\mathbb{E}_{(i,j):R_i=r,A^j=a}\Big[\sum_{\mathbf{d}_i^{-j}\in\mathfrak{D}^j}\mathbb{E}\big[Y_i(D_i^j=1,\mathbf{D}_i^{-j}=\mathbf{d}_i^{-j})Prob(\mathbf{D}_i^{-j}=\mathbf{d}_i^{-j})|\mathbf{U}_i,\mathbf{V_1},...,\mathbf{V_N}\big]\Big]\\
&=\mathbb{E}_{(i,j):R_i=r,A^j=a}\Big[\sum_{\mathbf{d}_i^{-j}\in\mathfrak{D}^j}\mathbb{E}\big[Y_i(D_i^j=1,\mathbf{D}_i^{-j}=\mathbf{d}_i^{-j})|\mathbf{U}_i,\mathbf{V_1},...,\mathbf{V_N}\big]\\
&\qquad\times Prob(\mathbf{D}_i^{-j}=\mathbf{d}_i^{-j}|\mathbf{U}_i,\mathbf{V_1},...,\mathbf{V_N})\Big]\\
&=\mathbb{E}_{(i,j):R_i=r,A^j=a}\Big[\sum_{\mathbf{d}_i^{-j}\in\mathfrak{D}^j}\mathbb{E}\big[Y_i(D_i^j=1,\mathbf{D}_i^{-j}=\mathbf{d}_i^{-j})|\mathbf{U}_i,\mathbf{V_1},...,\mathbf{V_N},D_i^j=1,\mathbf{D}_i^{-j}=\mathbf{d}_i^{-j}\big]\\
&\qquad\times Prob(\mathbf{D}_i^{-j}=\mathbf{d}_i^{-j}|\mathbf{U}_i,\mathbf{V_1},...,\mathbf{V_N})\Big]\\
&=\mathbb{E}_{(i,j):R_i=r,A^j=a}\Big[\sum_{\mathbf{d}_i^{-j}\in\mathfrak{D}^j}\mathbb{E}\big[Y_i^{obs}|\mathbf{U}_i,\mathbf{V_1},...,\mathbf{V_N},D_i^j=1,\mathbf{D}_i^{-j}=\mathbf{d}_i^{-j}\big]\\
&\qquad\times Prob(\mathbf{D}_i^{-j}=\mathbf{d}_i^{-j}|\mathbf{U}_i,\mathbf{V_1},...,\mathbf{V_N})\Big]
\end{aligned}
$$

The first equation comes from the law of iterated expectations, the second equation is due to linearity of expectations, the third equation is due to the independence between potential outcome and linking probability conditional on $(\mathbf{U}_i,\mathbf{V_1},...,\mathbf{V_N})$ (same d-separation argument as before), the fourth equation comes from the unconfoundedness of $Y_i(D_i^j=1,\mathbf{D}_i^{-j}=\mathbf{d}_i^{-j})$ conditional on $(\mathbf{U}_i,\mathbf{V_1},...,\mathbf{V_N})$ (4), and the fifth equation holds because when $D_i^j=1$ and $\mathbf{D}_i^{-j}=\mathbf{d}_i^{-j}$, $Y_i(D_i^j=1,\mathbf{D}_i^{-j}=\mathbf{d}_i^{-j})=Y_i^{obs}$. This means if $(\mathbf{U}_i,\mathbf{V_1},...,\mathbf{V_N})$ were observed, or equivalently if $\{e(\mathbf{U}_i,\mathbf{V}_1),...,e(\mathbf{U}_i,\mathbf{V}_N)\}$ were observed,

$$
\mathbb{E}_{(i,j):R_i=r,A^j=a}\big[\sum_{\mathbf{d}_i^{-j}\in\mathfrak{D}^j} Y_i(D_i^j=1,\mathbf{D}_i^{-j}=\mathbf{d}_i^{-j})Prob(\mathbf{D}_i^{-j}=\mathbf{d}_i^{-j})\big]
$$

is identified, and can be estimated with observed data. The same proof holds for $\mathbb{E}_{(i,j):R_i=r,A^j=a}\big[\sum_{\mathbf{d}_i^{-j}\in\mathfrak{D}^j} Y_i($ $0,\mathbf{D}_i^{-j}=\mathbf{d}_i^{-j})Prob(\mathbf{D}_i^{-j}=\mathbf{d}_i^{-j})\big]$. This means estimand $\tilde{\tau}_r^a$ is identified. $\square$

Figure 3: Indirect linking effect

A1.3 Other types of linking effect to explore in the future

A1.3.1 Indirect linking effect

As shown in Figure 3, we can define an indirect effect that contrasts i's potential outcome when some link sender j is linked to one of i's existing direct peer and its potential outcome when j is not linked to one of i's existing direct peer, while keeping i's existing peers fixed at the realised value. This requires the relaxation of L-SUTVA and is similar to the study of spillover effects in traditional setting (Forastiere et al., 2021).

A1.3.2 Triangle reinforced linking effect

The triangle reinforced linking effect contrasts i's potential outcome when its direct peer j also sends a link to one of i's other existing direct peer and its potential outcome when j is not linked to one of i's existing direct peer, while keeping i's existing peers fixed at the realised value. This could be used to study whether direct linking effect could be reinforced by an additional indirect link. If the underlying mechanism for the peer effect is information flow, then triangle reinforced effect shouldn't exist. It also requires the relaxation of L-SUTVA to

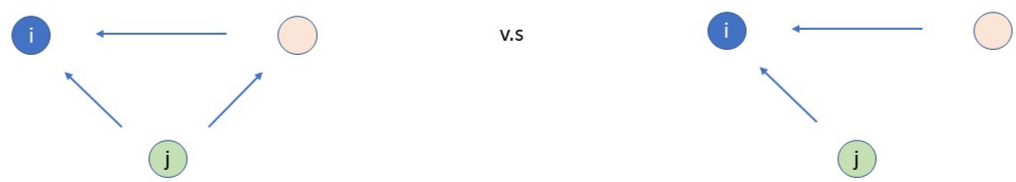

Figure 4: Triangle reinforced linking effect

allow for interference.

A1.4 Small networks

When networks are small, the estimation of propensity scores might be difficult, even if we have a large number of such small networks. This is because the estimation of propensity score is based on each single network. If the individual network is small, there is very little information for the inference of sufficient confounders and their propensity scores.

In this case, we could still make causal discovery based on additional assumptions. The idea is to assume that the effective treatment is some characteristic of the node, instead of the identity of the node. Let $Y_i^g(\cdot)$ denote the potential outcome of link receiver i in network g, this assumption is formalised as Assumption 6.

Assumption 6. For some function $l : \{0,1\}^N \to \mathbb{R}^M$

$$\begin{aligned} Y_i^g(D_{i1}, D_{i2}, ..., D_{in}) &= Y_i^g(l(D_{i1}, D_{i2}, ..., D_{iN})) \\ &= Y_i^g(l_1, ..., l_M) \end{aligned}$$

Function $l(\cdot)$ defines the effective treatment. For example if $l(\cdot) = \sum_{j=1,...,n} D_{ij}X_j$ where X is a dummy variable, assumption 6 means i's links affect i's potential outcome only through the total number of links with characteristics X. Similarly, if $l(\cdot) = \frac{\sum_{j=1,...,n} D_{ij}X_j}{\sum_{j=1,...,n} D_{ij}}$, assumption 6 means i's links affect i's potential outcome only through the share of i's links with characteristics X. Note that here we do not assume that the potential outcome is a linear function of $l(\cdot)$ as in the linear-in-means and linear-in-sum models. In both examples, we have $M = 1$, but this is not necessary. For example, $l(D_{i1}, D_{i2}, ..., D_{iN}) = (\sum_{j=1,...,n} D_{ij}X_j^1, \sum_{j=1,...,n} D_{ij}X_j^2)$ means the effective treatment is the total number of links with characteristics X^1 and the total number of links with characteristics X^2.

Next I show that under Assumption 6, causal identification and estimation of linking effect could be achieved by inferring sufficient confounders that render the distribution of effective treatment conditionally independent, as long as $M \geq 2$.

Definition A1.1. Let N^g be the number of nodes in network g, and $N = \sum_{g=1} N^g$. $o_1, ..., o_N$ and $q_1, ..., q_M$ are two vectors of random variables that satisfy the following condition:

$$Pr(l_{i1}, ..., l_{iM}|o_i, q_1, ..., q_M) = \prod_{m=1}^{M} Pr(l_{im}|o_i, q_m) \quad i = 1, ..., N$$

Effectively $l_{i1}, ..., l_{iM}$ is the multiple treatment vector of link receiver i, and since M is a fixed number, we are in the standard case studied in Wang and Blei (2019). Therefore $o_1, ..., o_N$ and $q_1, ..., q_M$ are sufficient confounders in the sense that after conditioning on them, treatment $(l_{i1}, ..., l_{iM})$ is independent of the potential outcome $Y_i(l_{i1}, ..., l_{iM})$.

Assumption 6 makes it possible to identify and estimate linking effects when networks are small. The intuition is that since nodes from different networks all share the same set of possible treatment $l_1, ..., l_M$. we could pool the link receivers across networks together to infer the sufficient confounders and their propensity scores. Note that in this case the estimators from the statistical network analysis literature, such as the neighbourhood smoothing estimator, won't work. But the factor models can still be used to estimate the propensity scores.

Finally, note that if this assumption doesn't hold, we will get biased causal estimates. This is because the sufficient confounders are defined as variables that make the supposedly effective treatments conditionally independent. If treatments are in fact at a more disaggregated level, these sufficient confounders are no long 'sufficient'.

A2 Alternative 2nd-step treatment effect estimatiors

As mentioned earlier, the inverse probability weighting estimator described in Section 4.2 is not the only 2nd-step estimator we could use to estimate the linking effect. Two of the popular ones in the causal inference literature are propensity score matching and propensity score subclassification. Here I will explain in detail how subclassification works and omit the details for matching. The case of propensity score matching is similar to subclassification. The only difference is that instead of dividing pairs into blocks based on similarity of propensity scores, we will find for each pair its M-nearest neighbour(s) in terms of their propensity scores. As in traditional propensity score matching, we could do both matching with replacement or without replacement. Next I will start with a simple example to illustrate the steps of subclassfication. Then I will provide formal justification of the subclassfication estimator.

A2.1 An example of subclassification estimator

In this example there are 8 link receivers with characteristic $R = r$ (labelled 1 to 8) and 7 link senders with characteristic $A = a$ (labelled a to g). The treatment assignment for the link receivers is given in Table 8. Here I omit the link receivers with characteristic $R \neq r$ and the link senders with characteristic $A \neq a$ because they are not needed for the estimand τ_r^a . Note that the matrix in Table 8 is not an adjacency matrix itself, but the intersection of a selection of rows and columns from the underlying adjacency matrix.

The matrix of propensity scores are shown in Table 9. These propensity scores are fictional and are only meant for illustration purpose. In particular, they are not estimated using any of the methods from Section **??**. The observed outcomes of the link receivers are: $Y_1 = Y_2 = Y_4 = Y_7 = 1$, and $Y_3 = Y_5 = Y_6 = Y_8 = 0$.

The main idea of subclassification is that if we divide the estimated propensity scores into small intervals, or subclasses, units within the same subclass will have similar estimated propensity scores and therefore can be viewed as having the same potential outcome distributions due to unconfoundedness. Here a unit is a link receiver link sender pair. Then, within

Table 8: Example treatment assignment

	a	b	c	d	e	f	g
1	0	0	0	1	0	0	0
2	0	0	1	0	0	1	0
3	1	0	0	1	0	0	0
4	0	0	0	0	0	0	1
5	0	0	1	0	0	0	1
6	1	0	0	0	1	1	0
7	0	0	0	0	0	0	0
8	1	0	1	0	0	0	0

Table 9: Example propensity scores

	a	b	c	d	e	f	g
1	0	0.1	0	0.11	0.33	0	0
2	0	0	0.5	0	0	0.33	0.16
3	0.25	0	0	0.67	0	0.25	0
4	0.15	0.33	0	0.33	0.1	0	0.27
5	0	0	0.2	0.2	0	0	0.3
6	0.33	0	0	0	0.6	0.56	0
7	0	0.2	0.3	0	0	0	0.1
8	0.5	0	0.1	0	0.3	0	0

the same subclass, the average of the missing potential outcomes for the treated units can be unbiasedly estimated by the observed outcomes of the control (untreated) units. Going back to the data above, I divide the propensity scores into three subclasses: $b_1 = (0, 0.3)$, $b_2 = [0.3, 0.5)$, $b_3 = [0.5, 1)$, with the assumption that uncounfoundedness holds within each subclass. Note that some pairs have an estimated propensity score of 0, which violates the positivity condition, so I leave them out in the data analysis. This means the estimator is now unbiased for the average effect only for those pairs within positive treatment probability.[2]

This leads to the classification of link receiver link sender pairs as shown in Table 10.

[2]In fact, in subclassfication analysis, researchers often leave out units with too low or too high propensity scores, even if they are not exactly 0 or 1. This is because with finite sample, there are often too few treated units within the subclass of very low propensity scores and two few control units within the subclass of very high propensity scores.

Table 10: Subclassification of pairs

	(0,0.3)	[0.3,0.5)	[0.5,1)
D_i^j=1	(3,a)	(6,a)	(8,a)
	(5,c)	(2,f)	(2,c)
	(8,c)	(5,g)	(3,d)
	(1,d)		(6,e)
	(4,g)		
D_i^j=0	(4,a)	(4,b)	(7,e)
	(1,b)	(7,c)	
	(7,b)	(4,d)	
	(5,d)	(1,e)	
	(4,e)	(8,e)	
	(3,f)		
	(2,g)		
	(7,g)		
number of pairs	13	8	5

The estimator is then:

$$\frac{13}{13+8+5}\left(\frac{Y_3+Y_5+Y_8+Y_1+Y_4}{5}-\frac{Y_4+Y_1+Y_7+Y_5+Y_4+Y_3+Y_2+Y_7}{8}\right)$$
$$+\frac{8}{13+8+5}\left(\frac{Y_6+Y_2+Y_5}{3}-\frac{Y_4+Y_7+Y_4+Y_1+Y_8}{5}\right)+\frac{5}{13+8+5}\left(\frac{Y_8+Y_2+Y_3+Y_6}{4}-Y_7\right)$$

Notice that the outcome of the same link receiver could be used multiple times, such as Y_4. They can appear both in the treated group and the control group, across multiple subclasses of propensity scores. This is because the propensity score is based on the pair, while the outcome is based on the link receiver only, and the same link receiver could appear in multiple pairs.

Note that unconfoundedness given propensity scores doesn't imply pairs with the same propensity scores have the same (u_i, u_j). Instead, it means that the treated units and control unis have the same distribution of u_i, u_j, and that treated units and and control units have the same distribution of potential outcomes.

A2.2 Subclassification formally

For exposition purpose, let's focus on the estimand

$$\tau_a^r = \mathbb{E}\left[\mathbb{E}_{(i,j):R_i=r,A^j=a}[Y_i^{obs}|e(\mathbf{U}_i,\mathbf{V}_j),D_i^j=1]\right] - \mathbb{E}\left[\mathbb{E}_{(i,j):R_i=r,A^j=a}[Y_i^{obs}|e(\mathbf{U}_i,\mathbf{V}_j),D_i^j=0]\right]$$

Suppose we decide to divide the propensity scores into B subclasses and assume the propensity scores within the same subclass are roughly constant, then τ_a^r can also be written as

$$\begin{aligned}\tau_a^r &= \frac{1}{B}\sum_{b=1}^{B}\frac{N_b}{N}\,\mathbb{E}_{(i,j):R_i=r,A^j=a}[Y_i^{obs}|(i,j)\in b, D_i^j=1] \\ &- \frac{1}{B}\sum_{b=1}^{B}\frac{N_b}{N}\,\mathbb{E}_{(i,j):R_i=r,A^j=a}[Y_i^{obs}|(i,j)\in b, D_i^j=0] \\ &= \frac{1}{B}\sum_{b=1}^{B}\tau_{a,b}^r\end{aligned}$$

where N_b is the number of (i,j) pairs in subclass $b \in B$, and $\tau_{a,b}^r = \frac{N_b}{N}(\mathbb{E}_{(i,j):R_i=r,A^j=a}[Y_i^{obs}|(i,j)\in b, D_i^j=1] - \mathbb{E}_{(i,j):R_i=r,A^j=a}[Y_i^{obs}|(i,j)\in b, D_i^j=0])$. To estimate $\tau_{a,b}^r$, we can simply compare the sample mean of the outcomes of the link receiver in treated pairs ($D_i^j=1$) and the sample mean of the outcomes of the link receiver in control pairs ($D_i^j=0$) belong to the subclass b. Alternatively, we could use linear regressions to estimate $\tau_{a,b}^r$ for all $b \in B$, thanks to the equivalence between $\tau_{a,b}^r$ and β_b of the following regression function:

$$Y_i = \alpha_b + \beta_b D_i^j + \epsilon_i^j$$

where observation is at the pair level. Within each subclass b, D_i^j is as good as random and independent of potential outcome. This means $\mathbb{E}[\epsilon_i^j|D_i^j]=0$, and that $\tau_{a,b}^r=\beta_b$:

$$\begin{aligned}\tau_{a,b}^r &= \mathbb{E}_{(i,j):R_i=r,A^j=a}[Y_i^{obs}|(i,j)\in b, D_i^j=1] - \mathbb{E}_{(i,j):R_i=r,A^j=a}[Y_i^{obs}|(i,j)\in, D_i^j=0] \\ &= \mathbb{E}_{(i,j):R_i=r,A^j=a}[\alpha_b+\beta_b+\epsilon_i^j|(i,j)\in b, D_i^j=1] - \mathbb{E}_{(i,j):R_i=r,A^j=a}[\alpha_b+\epsilon_i^j|(i,j)\in b, D_i^j=0] \\ &= \beta_b\end{aligned}$$

Expressing $\tau^r_{a,b}$ as a regression coefficient allows the easy incorporation of additional covariates into the analysis. Including pre-treatment predictors of the outcome in the regression could help reduce the bias coming from the variation of propensity scores within the same subclass, as well as increasing estimation precision, the same as in the conventional subclassification method Imbens and Rubin (2015).

A3 Discussion of Assumption (2)

A3.1 Super Population

We are interested in the super population if the estimands of interest are functions of the infinite population, for example the contrast in the mean potential outcomes for all units in the infinite population, including the ones not sampled. Assumption 2 is automatically satisfied if the sample network $\mathbf{D}_N$ is viewed as constructed by uniform random sampling of nodes from an infinite super population network with infinite number of nodes, where a link is recorded in the sample if it exists in the super population network. Under this construction, the randomness in link formation, or in other words, the assignment mechanism, solely comes from random sampling.

To see why random node sampling from super population implies Assumption 2, we proceed in 3 steps. First step, based on the definition in Crane (2018), Assumption 2 is equivalent to $\mathbf{D}_N$ being vertex exchangeable. Second step, under the Aldous-Hoover theorem, the equivalence of the De-Finetti theorem for network data, the distribution of vertex exchangeable network links can *always* be represented by some graphon process:

Definition A3.1 (Graphon (Crane, 2018))**.** Function $\phi \in \Phi : [0,1] \times [0,1] \to [0,1]$ has 0 diagonal. Fix any $\phi \in \Phi$ and draw $w_1, w_2, ...$ i.i.d. Uniform[0,1]. Given $w_1, w_2, ...$, assign D_i^j conditionally independently with probabilities

$$Pr(D_i^j = 1 | w_1, w_2, ...; \phi) = \phi(w_i, w_j) \tag{21}$$

This way of constructioning $\mathbf{D}$ is called a **graphon process**.

Therefore random node sampling guarantees that there exists i.i.d. $\{w_i\}_{1\leq i\leq n}$ such that

$$Pr(\mathbf{D}_N = \mathbf{d}_N | w_1, w_2, ..., w_N) = \prod_{i=1}^{N}\prod_{j\neq i}^{N} Pr(D_i^j = d_i^j | w_i, w_j) \tag{22}$$

Third step, comparing Equation (22) to Equation (1), we can see the difference is that here $\{w_i\}_{1\leq i\leq n}$ are i.i.d., while in Equation (1) $\{\mathbf{U}_i\}_{1\leq i\leq n}$ can be dependent of each other. This means if Equation (22) holds, Equation (1) will also hold.

In conclusion, when the sample network is constructed by random node sampling from an infinite super population network, the assumption of individualistic assignment mechanism must be true. This is similar to the case of conventional causal inference where random sampling from super population guarantees that the assignment mechanism is individualistic Imbens and Rubin (2015). Note that only random node sampling guarantees Assumption 2. Other sampling schemes, such as random link sampling, do not enjoy this property. An example of link sampling is in the study of co-authorship network where article is the sampling unit instead of the authors being the sampling unit.

A3.2 Finite Population

In Leung (2015)'s network formation model, i's linking decision could depend on the anticipated network structure. Network nodes simultaneously form directed links to maximise expected utility given their beliefs about the state of the network. Because the objective is the expected utility, i's linking probability will be a function of equilibrium beliefs about others' linking decisions, conditioning on the observed attributes of all agents in the network. For this reason, equilibrium linking decisions are functions of the exogenous attributes only. As such, the pairwise linking decision can be expressed as

$$D_i^j = h(Z_i, Z_j, \theta_{ij})$$

where Z_i includes both i's equilibrium beliefs about the the state of the network and i's observed exogenous attributes. Observed exogenous attributes are assumed to be common knowledge. Leung (2015) assumes that θ_{ij} are unobserved node or pairwise attributes that are private information and satisfy $\theta_{ij} \perp\!\!\!\perp V_{kl}$ for $i \neq k$. This allows θ_{ij} to be correlated with V_{il}, which means by just conditioning on Z_i and Z_j we couldn't yet write the probability distribution of the entire network links as a conditionally independent process in the form of equation (1). But if we partition θ_{ij} into $(v_{1,i}, v_{2,ij})$ where $v_{1,i}$ are unobserved shocks to link formation common to more than one sender j, and $v_{2,ij}$ are mutually independent pairwise shocks. The idea is that we could always separate out variables that cause correlations among θ_{ij} and V_{il} for $j \neq l$, and put them in $v_{1,i}$. Then

$$D_i^j = h(Z_i, Z_j, \theta_{ij})$$

becomes

$$D_i^j = h(Z_i, Z_j, v_{1,i}, v_{2,ij}) = \tilde{h}(\mathbf{U}_i, \mathbf{V}_j, v_{2,ij})$$

where $\mathbf{U}_i = (Z_i, v_{1,i})$. Conditioning on $\mathbf{U}_i, \mathbf{V}_j$, the probability distribution of network links then becomes exactly as in equation (1). Therefore, network formation games with network externalities as specified in Leung (2015) satisfy the Individualistic Assignment Mechanism Assumption.

A4 Proofs

A4.1 Proof of Lemma 1

Proof. LHS:

$$Pr(D_i^j = 1|\mathbf{U}_i, \mathbf{V}_j, e(\mathbf{U}_i, \mathbf{V}_j)) = Pr(D_i^j = 1|\mathbf{U}_i, \mathbf{V}_j) = e(\mathbf{U}_i, \mathbf{V}_j)$$

The first equality holds because $e(\mathbf{U}_i, \mathbf{V}_j)$ is a function of $\mathbf{U}_i, \mathbf{V}_j$, the second equality holds from the definition of $e(\mathbf{U}_i, \mathbf{V}_j)$.

RHS:

$$
\begin{aligned}
Pr(D_i^j = 1|e(\mathbf{U}_i, \mathbf{V}_j)) &= \mathbb{E}[D_i^j|e(\mathbf{U}_i, \mathbf{V}_j)] = \mathbb{E}[\mathbb{E}[D_i^j|\mathbf{U}_i, \mathbf{V}_j, e(\mathbf{U}_i, \mathbf{V}_j)]|e(\mathbf{U}_i, \mathbf{V}_j)] \\
&= \mathbb{E}[\mathbb{E}[D_i^j|\mathbf{U}_i, \mathbf{V}_j]|e(\mathbf{U}_i, \mathbf{V}_j)] = \mathbb{E}[e(\mathbf{U}_i, \mathbf{V}_j)|e(\mathbf{U}_i, \mathbf{V}_j)] \\
&= e(\mathbf{U}_i, \mathbf{V}_j)
\end{aligned}
$$

□

A4.2 Proof of Lemma 2

Proof.

$$
\begin{aligned}
& Pr(D_i^j = 1|Y_i^{pot}, e(\mathbf{U}_i, \mathbf{V}_j)) \\
&= \mathbb{E}[D_i^j = 1|Y_i^{pot}, e(\mathbf{U}_i, \mathbf{V}_j)] \\
&= \mathbb{E}\left[\, \mathbb{E}[D_i^j = 1|Y_i^{pot}, \mathbf{U}_i, \mathbf{V}_j, e(\mathbf{U}_i, \mathbf{V}_j)] \middle| Y_i^{pot}, e(\mathbf{U}_i, \mathbf{V}_j)\right]
\end{aligned}
$$

The inner expectation is equal to $\mathbb{E}[D_i^j = 1|\mathbf{U}_i, \mathbf{V}_j, e(\mathbf{U}_i, \mathbf{V}_j)]$ by unconfoundedness given $\mathbf{U}_i, \mathbf{V}_j$. And by the balancing property of the propensity score, this is $\mathbb{E}[D_i^j = 1|e(\mathbf{U}_i, \mathbf{V}_j)]$. Therefore the last expression is

$$
\begin{aligned}
& \mathbb{E}\left[\, \mathbb{E}[D_i^j = 1|e(\mathbf{U}_i, \mathbf{V}_j) \middle| Y_i^{pot}, e(\mathbf{U}_i, \mathbf{V}_j)\right] \\
&= \mathbb{E}[D_i^j = 1|e(\mathbf{U}_i, \mathbf{V}_j)] \\
&= Pr(D_i^j = 1|e(\mathbf{U}_i, \mathbf{V}_j))
\end{aligned}
$$

□

A4.3 Proof of proposition 2

Proof.

$$
\begin{aligned}
\tau_a^r &= \mathbb{E}_{(i,j):R_i=r,A^j=a}[Y_i(D_i^j=1,\mathbf{D}_i^{-j}=\bar{\mathbf{d}}_i^{-j})-Y_i(D_i^j=0,\mathbf{D}_i^{-j}=\bar{\mathbf{d}}_i^{-j})] \\
&= \mathbb{E}\left[\mathbb{E}_{(i,j):R_i=r,A^j=a}[Y_i(D_i^j=1,\mathbf{D}_i^{-j}=\bar{\mathbf{d}}_i^{-j})|\mathbf{U}_i,\mathbf{V}_j]\right] \\
&- \mathbb{E}\left[\mathbb{E}_{(i,j):R_i=r,A^j=a}[Y_i(D_i^j=0,\mathbf{D}_i^{-j}=\bar{\mathbf{d}}_i^{-j})|\mathbf{U}_i,\mathbf{V}_j]\right] \\
&= \mathbb{E}\left[\mathbb{E}_{(i,j):R_i=r,A^j=a}[Y_i(D_i^j=1,\mathbf{D}_i^{-j}=\bar{\mathbf{d}}_i^{-j})|\mathbf{U}_i,\mathbf{V}_j,D_i^j=1]\right] \\
&- \mathbb{E}\left[\mathbb{E}_{(i,j):R_i=r,A^j=a}[Y_i(D_i^j=0,\mathbf{D}_i^{-j}=\bar{\mathbf{d}}_i^{-j})|\mathbf{U}_i,\mathbf{V}_j,D_i^j=0]\right] \\
&= \mathbb{E}\left[\mathbb{E}_{(i,j):R_i=r,A^j=a}[Y_i(D_i^j=1,\mathbf{D}_i^{-j}=\bar{\mathbf{d}}_i^{-j})|e(\mathbf{U}_i,\mathbf{V}_j),D_i^j=1]\right] \\
&- \mathbb{E}\left[\mathbb{E}_{(i,j):R_i=r,A^j=a}[Y_i(D_i^j=0,\mathbf{D}_i^{-j}=\bar{\mathbf{d}}_i^{-j})|e(\mathbf{U}_i,\mathbf{V}_j),D_i^j=0]\right] \\
&= \mathbb{E}\left[\mathbb{E}_{(i,j):R_i=r,A^j=a}[Y_i^{obs}|\mathbf{U}_i,\mathbf{V}_j,D_i^j=1]\right]-\mathbb{E}\left[\mathbb{E}_{(i,j):R_i=r,A^j=a}[Y_i^{obs}|\mathbf{U}_i,\mathbf{V}_j,D_i^j=0]\right] \\
&= \mathbb{E}\left[\mathbb{E}_{(i,j):R_i=r,A^j=a}[Y_i^{obs}|e(\mathbf{U}_i,\mathbf{V}_j),D_i^j=1]\right]-\mathbb{E}\left[\mathbb{E}_{(i,j):R_i=r,A^j=a}[Y_i^{obs}|e(\mathbf{U}_i,\mathbf{V}_j),D_i^j=0]\right]
\end{aligned}
$$

Here the expectations are always over the node sampling distribution. The second equation is from law of iterated expectations. The third and fourth are from unconfoundedness given both $\mathbf{U}_i,\mathbf{V}_j$ and $e(\mathbf{U}_i,\mathbf{V}_j)$. The fifth and the last equalities are from no multiple versions of treatment assumption. Since $e(\mathbf{U}_i,\mathbf{V}_j)$ is identified $\forall i,j$ by Lemma 3, τ_r^a is identified.

□

A4.4 Proof of unbiasedness of IPW estimator

Proof. With slight abuse of notation for simplicity, in the following I will write $\mathbb{E}_{(i,j):R_i=r,A^j=a}[\cdot]$ as $\mathbb{E}[\cdot]$.

Here I will only prove that

$$
\mathbb{E}\left[\frac{1}{\sum_{i=1}^N R_i=r}\cdot\frac{1}{\sum_{j=1}^N A^j=a}\sum_{i:R_i=r}\sum_{j:A^j=a}\frac{D_i^j\cdot Y_i^{obs}}{e(\mathbf{U}_i,\mathbf{V}_j)}\right]=\mathbb{E}[Y_i(D_i^j=1,\mathbf{D}_i^{-j}=\bar{\mathbf{d}}_i^{-j}].
$$

The case for

$$\mathbb{E}\Big[\frac{1}{\sum_{i=1}^{N} R_i = r} \cdot \frac{1}{\sum_{j=1}^{N} A^j = a} \sum_{i:R_i=r} \sum_{j:A^j=a} \frac{(1-D_i^j) \cdot Y_i^{obs}}{1-e(\mathbf{U}_i, \mathbf{V}_j)}\Big] = \mathbb{E}[Y_i(D_i^j = 0, \mathbf{D}_i^{-j} = \bar{\mathbf{d}}_i^{-j}]$$

can be similarly proved.

$$\begin{aligned}
&\mathbb{E}\Big[\frac{1}{\sum_{i=1}^{N} R_i = r} \cdot \frac{1}{\sum_{j=1}^{N} A^j = a} \sum_{i:R_i=r} \sum_{j:A^j=a} \frac{D_i^j \cdot Y_i^{obs}}{e(\mathbf{U}_i, \mathbf{V}_j)}\Big] \\
=&\mathbb{E}\big[\frac{Y_i^{obs} \cdot D_i^j}{e(\mathbf{U}_i, \mathbf{V}_j)}\big] \\
=&\mathbb{E}\big[\frac{Y_i(D_i^j = 1, \mathbf{D}_i^{-j} = \bar{\mathbf{d}}_i^{-j}) \cdot D_i^j}{e(\mathbf{U}_i, \mathbf{V}_j)}\big] \\
=&\mathbb{E}\Big[\mathbb{E}\big[\frac{Y_i(D_i^j = 1, \mathbf{D}_i^{-j} = \bar{\mathbf{d}}_i^{-j}) \cdot D_i^j}{e(\mathbf{U}_i, \mathbf{V}_j)}\big|\mathbf{U}_i, \mathbf{V}_j\big]\Big]
\end{aligned}$$

The first equation is due to i.i.d. sampling of the nodes, the second equation holds because $Y_i^{obs} = Y_i(D_i^j = 1, \mathbf{D}_i^{-j} = \bar{\mathbf{d}}_i^{-j})$ when $D_i^j = 1$, the third equation is from iterated expectations.

Then the inner expectation can be re-written as

$$\begin{aligned}
&\mathbb{E}\big[\frac{Y_i(D_i^j = 1, \mathbf{D}_i^{-j} = \bar{\mathbf{d}}_i^{-j}) \cdot D_i^j}{e(\mathbf{U}_i, \mathbf{V}_j)}\big|\mathbf{U}_i, \mathbf{V}_j\big] \\
=&\frac{\mathbb{E}[D_i^j|\mathbf{U}_i, \mathbf{V}_j] \cdot \mathbb{E}[Y_i(D_i^j = 1, \mathbf{D}_i^{-j} = \bar{\mathbf{d}}_i^{-j}|\mathbf{U}_i, \mathbf{V}_j]}{e(\mathbf{U}_i, \mathbf{V}_j)} \\
=&\frac{e(\mathbf{U}_i, \mathbf{V}_j) \cdot \mathbb{E}[Y_i(D_i^j = 1, \mathbf{D}_i^{-j} = \bar{\mathbf{d}}_i^{-j}|\mathbf{U}_i, \mathbf{V}_j]}{e(\mathbf{U}_i, \mathbf{V}_j)} \\
=&\mathbb{E}[Y_i(D_i^j = 1, \mathbf{D}_i^{-j} = \bar{\mathbf{d}}_i^{-j}|\mathbf{U}_i, \mathbf{V}_j]
\end{aligned}$$

where the first equation holds because D_i^j and $Y_i(D_i^j = 1, \mathbf{D}_i^{-j} = \bar{\mathbf{d}}_i^{-j}$ are independent conditional on $\mathbf{U}_i, \mathbf{V}_j$, by unconfoundedness 1. Therefore

$$\begin{aligned}
&\mathbb{E}\Big[\mathbb{E}\big[\frac{Y_i(D_i^j = 1, \mathbf{D}_i^{-j} = \bar{\mathbf{d}}_i^{-j}) \cdot D_i^j}{e(\mathbf{U}_i, \mathbf{V}_j)}|\mathbf{U}_i, \mathbf{V}_j\big]\Big] \\
=&\mathbb{E}\Big[\mathbb{E}[Y_i(D_i^j = 1, \mathbf{D}_i^{-j} = \bar{\mathbf{d}}_i^{-j}|\mathbf{U}_i, \mathbf{V}_j]\Big] \\
=&\mathbb{E}[Y_i(D_i^j = 1, \mathbf{D}_i^{-j} = \bar{\mathbf{d}}_i^{-j}]
\end{aligned}$$

□

A4.5 Proof of Proposition 3

Proof. $m^{c_1,c_2} = \mathbb{E}_i\big[\mathbb{E}_{d^{c_1}}[Y_i(d^{c_1})]\big] - \mathbb{E}_i\big[\mathbb{E}_{d^{c_2}}[Y_i(d^{c_2})]\big]$. Here I will only prove that $\mathbb{E}_i\big[\mathbb{E}_{d^{c_1}}[Y_i(d^{c_1})]\big]$ is identified. The identification of $\mathbb{E}_i\big[\mathbb{E}_{d^{c_2}}[Y_i(d^{c_2})]\big]$ follows similarly.

$$\begin{aligned}
&\mathbb{E}_i\big[\mathbb{E}_{d^{c_1}}[Y_i(d^{c_1})]\big] = \mathbb{E}_i\big[\mathbb{E}_{d^{c_1}}[Y_i(\mathbf{D}_i = d^{c_1})]\big] \\
&= \mathbb{E}\Bigg[\mathbb{E}_i\big[\mathbb{E}_{d^{c_1}}[Y_i(\mathbf{D}_i = d^{c_1})|\mathbf{U}_i, \mathbf{V}_1, ..., \mathbf{V}_N]\big]\Bigg] \\
&= \mathbb{E}\Bigg[\mathbb{E}_i\big[\mathbb{E}_{d^{c_1}}[Y_i(\mathbf{D}_i = d^{c_1})|\mathbf{U}_i, \mathbf{V}_1, ..., \mathbf{V}_N, \mathbf{D}_i = d^{c_1}]\big]\Bigg] \\
&= \mathbb{E}\Bigg[\mathbb{E}_i\big[\mathbb{E}_{d^{c_1}}[Y_i(\mathbf{D}_i = d^{c_1})|Pr(\mathbf{D}_i = d^{c_1}|\mathbf{U}_i, \mathbf{V}_1, ..., \mathbf{V}_N), \mathbf{D}_i = d^{c_1}]\big]\Bigg] \\
&= \mathbb{E}\Bigg[\mathbb{E}_i\big[\mathbb{E}_{d^{c_1}}[Y_i(\mathbf{D}_i = d^{c_1})|e(\mathbf{U}_i, \mathbf{V}_1), ..., e(\mathbf{U}_i, \mathbf{V}_N), \mathbf{D}_i = d^{c_1}]\big]\Bigg]
\end{aligned}$$

The first equation comes from the law of iterated expectations. The second equation follows the unconfoundedness condition in Lemma 4. The third equation comes from the balancing property of generalised propensity scores. The last equation holds because

$$\begin{aligned}
&Pr(\mathbf{D}_i = d^{c_1}|\mathbf{U}_i, \mathbf{V}_1, ..., \mathbf{V}_N) \\
=&\prod_{j=1}\big(Pr(D_i^j = 1|\mathbf{U}_i, \mathbf{V}_j)\big)^{d_i^{c_1}}\big(1 - Pr(D_i^j = 1|\mathbf{U}_i, \mathbf{V}_j)\big)^{1-d_i^{c_1}} \\
=&\prod_{j=1}\big(e(\mathbf{U}_i, \mathbf{V}_j)\big)^{d_i^{c_1}}\big(1 - e(\mathbf{U}_i, \mathbf{V}_j)\big)^{1-d_i^{c_1}}
\end{aligned}$$

□

A5 Empirical application supplementary material

A6 Additional simulation results

A6.1 Details of network formation model g_2

The second network link generation process $Prob(D_i^j = 1) = g_2(C_i, C_j)$ is a slgihtly more complicated version of a statistical block model. The linking probabilities are asymmetric, that is $g_2(C_i, C_j) \neq g_2(C_j, C_i)$. For any node i and j, the probability of i receiving a link from j is in general higher if i) C_i is larger and ii) C_j is slightly higher than C_j. If we think of C as the ability of the node, this is a model where higher ability nodes receive more friendships, but only from nodes who are slightly more able than themselves. This might be because they don't like people who are less able than them, and admire people who are more able, but become jealous of people who are too much more able than themselves.

$$g_2 : P_i^j = \begin{cases} 0.05 & \text{if } C_i \in [0.1, 0.2) \ \& \ C_j \in (0.2, 0.21] \\ 0.1 & \text{if } C_i \in [0.2, 0.3) \ \& \ C_j \in (0.3, 0.31] \\ 0.15 & \text{if } C_i \in [0.3, 0.4) \ \& \ C_j \in (0.4, 0.41] \\ 0.2 & \text{if } C_i \in [0.4, 0.5) \ \& \ C_j \in (0.5, 0.51], \text{ or if, } C_i \in [0.5, 0.6) \ \& \ C_j \in (0.6, 0.61] \\ 0.25 & \text{if } C_i \in [0.6, 0.7) \ \& \ C_j \in (0.7, 0.71], \text{ or if, } C_i \in [0.7, 0.8) \ \& \ C_j \in (0.8, 0.81] \\ 0.3 & \text{if } C_i \in [0.8, 0.9) \ \& \ C_j \in (0.9, 0.91], \text{ or if, } C_i \in [0.9, 1] \ \& \ C_j \in (0.99, 1] \\ 0.01 & \text{if } C_i \in [a, a + 0.1) \ \& \ C_j \in [a, a + 0.1) \text{ for } a = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8 \\ & \text{or if, } C_i \in [0, 0.1) \ \& \ C_j \in [0, 0.05) \\ 0 & \text{otherwise} \end{cases} \tag{23}$$

Table 11: Variable definitions for CFP friendship re-analysis

Variable	Definition in the original papers	Definition in this paper
Post college education for parents	Dummy variable equal to 1 if the respondent reports that the highest level of education attained by their residential father and residential mother has a post-college education, and 0 otherwise. If a student either does not have a residential father/mother or the information is missing, that parent's level of education is imputed using the other parent's education [a].	Same definition. The difference is that in-home data is used instead. If the in-home data is missing, in-school data is used. This is because for saturated schools, data from in-home interviews have less missing values than data from the in-school survey.
log family income	log of total household income (thousands). If family income is missing, family income is set to the mean value for the school and a dummy is included for missing family income.	Same. In addition, for families with 0 annual family income, their income is replaced with 0.1, in order for the log income to take real values.
Grade	Grade point average is calculated based on self-reported student grades in math, science, english, and history from the Wave I in-home survey where A=4, B=3, C=2, and D or lower=1.	Same. Note: If the respondent didn't take the subject, I code the grade as missing.
MaleFrac (FemaleFrac) high	They are the fraction of male and female high flyers (those with at least one post-college parent) in the grade and school.	Same
Bachelor degree	Dummy variable equal to 1 if the respondent has completed a bachelor's degree (four-year college) and 0 otherwise.	Same
LFP	Dummy variable equal to 1 if the respondent is currently working at least 10 hours per week, is on sick leave or temporarily disabled, is on maternity/paternity leave, or is unemployed and looking for work, and is equal to zero otherwise.	Same
Ever married	Dummy variable equal to one if the respondent resported they have ever been married	Same
Children	Total number of (non-deceased) biological children they have.	Same

[a]For example, if the residential father's education is missing, but the residential mother has a high-school education, they impute a value for father post-college by taking the average value of father post-college among students of the same gender within the school who also have a residential mother with a high-school education. If there are no students with equivalent mother's education and non-missing information on father's education, they impute father post-college using the value of father post-college among all students in the school who have a residential mother with a high-school education.

Table 12: Naive OLS estimates for the effect of friendship

	Bachelor Degree (p.p)	Want (p.p)	Will (p.p)	Intelligence (p.p)
F_FL	0.638***	0.195	−0.109	0.214
	(0.163)	(0.208)	(0.206)	(0.209)
F_ML	1.150***	0.525	0.284	0.175
	(0.342)	(0.379)	(0.363)	(0.441)
F_FH	2.984***	4.441***	2.600**	0.955
	(1.118)	(0.987)	(1.065)	(1.699)
F_MH	2.052	1.474	1.429	3.451**
	(1.435)	(1.344)	(1.120)	(1.741)
M_FL	0.473*	0.152	0.147	−0.697**
	(0.282)	(0.272)	(0.283)	(0.314)
M_ML	0.499**	−0.058	−0.253	−0.327
	(0.202)	(0.189)	(0.217)	(0.244)
M_FH	4.145**	1.971	−3.514	0.021
	(1.777)	(2.013)	(2.480)	(2.833)
M_MH	3.262**	1.765	2.540**	4.102***
	(1.561)	(1.378)	(1.123)	(1.145)

Note: This table reports the naive OLS estimated effects of high school friendship on students' bachelor degree attainment (column 1), and their intermediate outcomes (column 2-4). The dependent variable in Column (2) is a dummy variable recording whether the student reported a scale 5 (1 is the lowest and 5 is the highest) on the the extent of how much they want to go to college (Wave II). The dependent variable in Column (3) is a dummy variable recording whether the student reported a scale 5 (1 is the lowest and 5 is the highest) on the likelihood that they will go to college (Wave II). The dependent variable in Column (4) is a dummy variable recording whether the student reported a scale 5 or 6 (1 is the lowest and 6 is the highest) on their intelligence compared to other people of their age (Wave II). Each row corresponds to a characterisation of the friendship, based on the character of the receiver and the sender. Receiver characteristics is shown before the underbar _, and sender characteristics is shown after. "F" and "M" are used to refer to the gender female and male respectively. "H" and "L" are used to refer to whether the individual is a high flyer or non-high flyer (low flyer) respectively. For example, "F_FL" means the linking effect is estimated for female link receivers and female non-high flyer link senders. The regressions reported in all columns include cohort dummies, whether the student was born in the US, their PVT score, whether their PVT score is above the population median PVT score, whether their mother's and father's highest degree is high school, some college, college, or post college, whether their mother's and father's highest education level is missing, the student's log family income, whether family is missing, the age of the student during Wave I, whether the student's mother and father were in the household, dummies for whether the student is black, hispanic, white, asian and indian. Standard errors are estimated with subsample bootstrapping with 900 subsamples drawn randomly. At each bootstrap, 90% of the individuals (nodes) within each school are sampled without replacement. *p<0.1; **p<0.05; ***p<0.01

Table 13: Effect of friendship on long-term outcomes

	LFP	Num Children	Married
	(1)	(2)	(3)
F_FL	0.002	0.003	0.007***
	(0.002)	(0.005)	(0.002)
F_ML	0.001	−0.031***	−0.001
	(0.004)	(0.008)	(0.004)
F_FH	−0.016	−0.081***	0.046***
	(0.012)	(0.031)	(0.012)
F_MH	0.024	−0.064***	0.006
	(0.015)	(0.018)	(0.010)
M_FL	0.010***	0.004	0.009***
	(0.003)	(0.008)	(0.003)
M_ML	0.003	−0.003	−0.003
	(0.003)	(0.006)	(0.003)
M_FH	−0.055**	0.024	−0.015
	(0.024)	(0.031)	(0.016)
M_MH	−0.034**	−0.057***	0.025*
	(0.013)	(0.021)	(0.013)
Note:	*p<0.1; **p<0.05; ***p<0.01		

Note: This table reports the estimated effects of high school friendship on students' long term outcomes measured in Wave IV. The dependent variable in Column (1) is a dummy variable recording whether the respondent was part of the labour force. The dependent variable in Column (2) is the number of children the respondent. The dependent variable in Column (3) is a dummy variable recording whether the respondent has ever been married. The estimands are all ATT. Each row corresponds to a characterisation of the friendship, based on the character of the receiver and the sender. Receiver characteristics is shown before the underbar _, and sender characteristics is shown after. "F" and "M" are used to refer to the gender female and male respectively. "H" and "L" are used to refer to whether the individual is a high flyer or non-high flyer (low flyer) respectively. For example, "F_FL" means the linking effect is estimated for female link receivers and female non-high flyer link senders. The regressions reported in all columns include cohort dummies, whether the student was born in the US, their PVT score, whether their PVT score is above the population median PVT score, whether their mother's and father's highest degree is high school, some college, college, or post college, whether their mother's and father's highest education level is missing, the student's log family income, whether family is missing, the age of the student during Wave I, whether the student's mother and father were in the household, dummies for whether the student is black, hispanic, white, asian and indian. Standard errors are estimated with subsample bootstrapping with 900 subsamples drawn randomly. At each bootstrap, 90% of the individuals (nodes) within each school are sampled without replacement. *p<0.1; **p<0.05; ***p<0.01

Table 14: Heterogeneous effects of friendship on desire and likelihood to go to college

	Dependent variable:			
	Want		Will	
	PVT Median -	PVT Median +	PVT Median -	PVT Median +
	(1)	(2)	(3)	(4)
F_FL	-2.100^{***}	0.603	-1.256^{***}	-0.328
	(0.508)	(0.385)	(0.350)	(0.407)
F_ML	0.216	-0.462	-1.396^{*}	0.091
	(0.798)	(0.876)	(0.774)	(0.697)
F_FH	5.494^{***}	-0.387	2.600	-0.410
	(2.089)	(1.708)	(3.457)	(1.584)
F_MH	4.506^{*}	-0.722	5.365^{***}	-0.441
	(2.393)	(1.548)	(1.816)	(1.260)
M_FL	0.153	-0.893^{*}	0.394	-0.890^{*}
	(0.767)	(0.496)	(0.746)	(0.522)
M_ML	1.010^{**}	-0.525	-0.352	-0.884^{**}
	(0.436)	(0.437)	(0.480)	(0.436)
M_FH	1.805	0.648	-5.984^{**}	0.746
	(2.537)	(1.928)	(2.426)	(2.587)
M_MH	-4.820^{**}	10.053^{***}	-3.905	9.147^{***}
	(2.295)	(2.411)	(2.912)	(2.593)

Note: This table reports the estimated heterogeneous effects of high school friendship on students' desire and likelihood of going to college. The dependent variable in Column (1) and Column (2) is a dummy variable recording whether the student reported a scale 5 (1 is the lowest and 5 is the highest) on the the extent of how much they want to go to college (Wave II). The dependent variable in Column (3) and Column (4) is a dummy variable recording whether the student reported a scale 5 (1 is the lowest and 5 is the highest) on the likelihood that they will go to college (Wave II). Column (1) and (3) reports results for ego whose PVT score is below population median PVT score. Column (2) and (4) reports results for ego whose PVT score is above population median PVT score. The estimands are all ATT. Each row corresponds to a characterisation of the friendship, based on the character of the receiver and the sender. Receiver characteristics is shown before the underbar _, and sender characteristics is shown after. "F" and "M" are used to refer to the gender female and male respectively. "H" and "L" are used to refer to whether the individual is a high flyer or non-high flyer (low flyer) respectively. For example, "F_FL" means the linking effect is estimated for female link receivers and female non-high flyer link senders. The regressions reported in all columns include cohort dummies, whether the student was born in the US, their PVT score, whether their PVT score is above the population median PVT score, whether their mother's and father's highest degree is high school, some college, college, or post college, whether their mother's and father's highest education level is missing, the student's log family income, whether family is missing, the age of the student during Wave I, whether the student's mother and father were in the household, dummies for whether the student is black, hispanic, white, asian and indian. Standard errors are estimated with subsample bootstrapping with 900 subsamples drawn randomly. At each bootstrap, 90% of the individuals (nodes) within each school are sampled without replacement. *p<0.1; **p<0.05; ***p<0.01

Table 15: Mean degree distribution for simulated g2 networks

	0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
N=100	0	0	0	0	0	0.2	0.8	1	1.1	1.9	4.3
N=300	0	0	0.2	1	1	1.5	2	2.6	3.3	4.7	10.2
N=500	0	0.3	1	1.5	2	2.7	3.2	4.1	5.5	7.4	15.7

Note: This table reports the mean degree distribution of the simulated networks. For each size N=100,300,500, and for each simulated network of that size, I caculate the deciles of the number of links each link receiver receives, and average over all the 500 simulated networks of that size.

Table 16: Simulation results for g_2

				IPW	Matching	Sub	Naive ols
Yb	Bias	X_0					
			N=100	0.083264	0.096999	0.099578	0.135588
			N=300	0.048002	0.084757	0.087946	0.167218
			N=500	0.036638	0.088748	0.091242	0.186257
		X_1					
			N=100	0.074813	0.087677	0.089541	0.126791
			N=300	0.04956	0.086555	0.08927	0.167975
			N=500	0.035027	0.085468	0.089596	0.184303
	MAE	X_0					
			N=100	0.103605	0.134378	0.112983	0.143077
			N=300	0.050245	0.085861	0.087946	0.167218
			N=500	0.037016	0.088748	0.091242	0.186257
		X_1					
			N=100	0.094632	0.114537	0.10228	0.13395
			N=300	0.052468	0.087344	0.089483	0.167975
			N=500	0.03631	0.085468	0.089596	0.184303
Yc	Bias	X_0					
			N=100	0.465683	0.529408	0.561574	0.779459
			N=300	0.2608	0.470754	0.494971	0.956848
			N=500	0.186274	0.526728	0.546989	1.148476
		X_1					
			N=100	0.456105	0.536314	0.54596	0.76797
			N=300	0.263195	0.482857	0.495973	0.954869
			N=500	0.177633	0.512275	0.537143	1.136513
	MAE	X_0					
			N=100	0.489148	0.601876	0.575155	0.784664
			N=300	0.262791	0.470754	0.494971	0.956848
			N=500	0.187562	0.526728	0.546989	1.148476
		X_1					
			N=100	0.465316	0.555527	0.548736	0.768234
			N=300	0.263612	0.482857	0.495973	0.954869
			N=500	0.179018	0.512275	0.537143	1.136513

Note: This table reports for the g_2 model the bias and the mean absolute error (MAE) of the inverse probability weighting estimator, the nearest neighbour matching estimator with replacement, the subclassification estimator and the narive ols estimator, compared to the true linking effects, for link sender with $X_j = 0$ and $X_j = 1$ separately. The number of matches for the matching estimator is 1, the number of subclasses for the subclassification estimator is 8. All the estimates are for the average treatment effect for the treated.

Table 17: True Propensity Score vs True Effects for g_1

				IPW	Matching	Sub
Yb	Bias	X_0				
			N=100	-0.00618	0.00017	-0.00134
			N=300	0.002273	0.002384	0.006239
			N=500	-0.00044	-0.00046	0.004332
		X_1				
			N=100	-0.00422	-0.00806	-0.00598
			N=300	-0.00232	-0.00379	0.00157
			N=500	0.000664	0.00104	0.005368
	MAE	X_0				
			N=100	0.080231	0.097323	0.073752
			N=300	0.026164	0.029931	0.023283
			N=500	0.013928	0.018599	0.013309
		X_1				
			N=100	0.065956	0.085406	0.061449
			N=300	0.025165	0.029701	0.023626
			N=500	0.016689	0.018747	0.016217
Yc	Bias	X_0				
			N=100	0.005795	0.011732	0.04541
			N=300	-0.00335	-0.00898	0.026149
			N=500	0.000463	-0.00044	0.033232
		X_1				
			N=100	-0.02251	-0.05712	-0.01709
			N=300	-0.01018	-0.01714	0.018337
			N=500	-0.0004	-0.00359	0.031436
	MAE	X_0				
			N=100	0.259116	0.281393	0.208147
			N=300	0.101655	0.104463	0.079084
			N=500	0.069394	0.070489	0.056754
		X_1				
			N=100	0.219053	0.215143	0.159517
			N=300	0.086806	0.088471	0.066411
			N=500	0.058299	0.058505	0.049941

Note: This table reports for the g_1 model the bias and the mean absolute error (MAE) of the inverse probability weighting estimator, the nearest neighbour matching estimator with replacement, and the subclassification estimator using true peopensity scores, compared to the true linking effects, for link sender with $X_j = 0$ and $X_j = 1$ separately. The number of matches for the matching estimator is 1, the number of subclasses for the subclassification estimator is 8. All the estimates are for the average treatment effect for the treated.

Table 18: True Propensity Score vs True Effects for g_2

				IPW	Matching	Sub
Yb	Bias	X_0				
			N=100	0.000634	0.000271	0.003285
			N=300	-0.00094	0.00024	0.004268
			N=500	0.000682	0.000218	0.004587
		X_1				
			N=100	-0.00704	-0.01052	-0.00802
			N=300	0.000635	-0.00019	0.004349
			N=500	-0.00077	-0.00149	0.004058
	MAE	X_0				
			N=100	0.077348	0.094386	0.067707
			N=300	0.02476	0.030613	0.02275
			N=500	0.015357	0.019648	0.014188
		X_1				
			N=100	0.068187	0.088944	0.063137
			N=300	0.024323	0.029334	0.022893
			N=500	0.016496	0.017966	0.014982
Yc	Bias	X_0				
			N=100	0.002911	-0.00577	0.02452
			N=300	-0.00701	-0.00254	0.028533
			N=500	0.003169	-0.00064	0.031012
		X_1				
			N=100	-0.01186	-0.01944	-0.00279
			N=300	-0.00522	-0.01368	0.020637
			N=500	-0.00502	-0.00856	0.027283
	MAE	X_0				
			N=100	0.270586	0.274076	0.199427
			N=300	0.100294	0.103225	0.080769
			N=500	0.068604	0.071309	0.055641
		X_1				
			N=100	0.211424	0.224245	0.162451
			N=300	0.084764	0.091543	0.068273
			N=500	0.062315	0.059699	0.049605

Note: This table reports for the g_2 model the bias and the mean absolute error (MAE) of the inverse probability weighting estimator, the nearest neighbour matching estimator with replacement, and the subclassification estimator using true peopensity scores, compared to the true linking effects, for link sender with $X_j = 0$ and $X_j = 1$ separately. The number of matches for the matching estimator is 1, the number of subclasses for the subclassification estimator is 8. All the estimates are for the average treatment effect for the treated.

Table 19: Using estimated propensity scores vs true propensity scores for g_1

				IPW	Matching	Sub
Yb	Bias	X_0				
			N=100	0.083621	0.096681	0.095233
			N=300	0.049644	0.084352	0.085735
			N=500	0.033614	0.084658	0.083421
		X_1				
			N=100	0.082936	0.102901	0.098823
			N=300	0.049853	0.087267	0.086033
			N=500	0.033867	0.08433	0.083669
	MAE	X_0				
			N=100	0.08604	0.13704	0.096486
			N=300	0.0501	0.085676	0.085735
			N=500	0.033836	0.084672	0.083421
		X_1				
			N=100	0.084721	0.124468	0.09916
			N=300	0.050546	0.087735	0.086033
			N=500	0.034166	0.08433	0.083669
Yc	Bias	X_0				
			N=100	0.488645	0.579477	0.538106
			N=300	0.26446	0.492194	0.47262
			N=500	0.172692	0.512657	0.499917
		X_1				
			N=100	0.476862	0.591576	0.556474
			N=300	0.26786	0.4877	0.475234
			N=500	0.17529	0.510615	0.501656
	MAE	X_0				
			N=100	0.488645	0.635377	0.541816
			N=300	0.264799	0.492258	0.47262
			N=500	0.172846	0.512657	0.499917
		X_1				
			N=100	0.476916	0.619324	0.556625
			N=300	0.267956	0.4877	0.475234
			N=500	0.175635	0.510615	0.501656

Note: This table reports for the g_1 model the bias and the mean absolute error (MAE) of the inverse probability weighting estimator, the nearest neighbour matching estimator with replacement, and the subclassification estimator using factor model estimated propensity scores, compared to the linking effects estimated using the true propensity socres, for link sender with $X_j = 0$ and $X_j = 1$ separately. The number of matches for the matching estimator is 1, the number of subclasses for the subclassification estimator is 8. All the estimates are for the average treatment effect for the treated.

Table 20: Using estimated propensity scores vs true propensity scores for g_2

				IPW	Matching	Sub
Yb	Bias	X_0				
			N=100	0.08263	0.096729	0.096293
			N=300	0.048945	0.084517	0.083678
			N=500	0.035957	0.088529	0.086655
		X_1				
			N=100	0.081855	0.098198	0.097559
			N=300	0.048925	0.086743	0.084921
			N=500	0.035799	0.086957	0.085537
	MAE	X_0				
			N=100	0.086134	0.138018	0.097608
			N=300	0.049154	0.086261	0.083678
			N=500	0.036224	0.088529	0.086655
		X_1				
			N=100	0.084542	0.120627	0.098128
			N=300	0.049115	0.086914	0.084921
			N=500	0.036091	0.086957	0.085537
Yc	Bias	X_0				
			N=100	0.462772	0.535179	0.537054
			N=300	0.267808	0.473296	0.466438
			N=500	0.183105	0.527369	0.515976
		X_1				
			N=100	0.467964	0.555751	0.548748
			N=300	0.268415	0.496538	0.475336
			N=500	0.182649	0.520831	0.50986
	MAE	X_0				
			N=100	0.462914	0.601572	0.537354
			N=300	0.267808	0.473863	0.466438
			N=500	0.183854	0.527369	0.515976
		X_1				
			N=100	0.46845	0.574089	0.549009
			N=300	0.268415	0.496538	0.475336
			N=500	0.182914	0.520831	0.50986

Note: This table reports for the g_2 model the bias and the mean absolute error (MAE) of the inverse probability weighting estimator, the nearest neighbour matching estimator with replacement, and the subclassification estimator using factor model estimated propensity scores, compared to the linking effects estimated using the true propensity socres, for link sender with $X_j = 0$ and $X_j = 1$ separately. The number of matches for the matching estimator is 1, the number of subclasses for the subclassification estimator is 8. All the estimates are for the average treatment effect for the treated.

Table 21: Matching and Subclassification with increasing matches and subclasses vs True Effects for g_1

				Matching	Sub
Yb	Bias	X_0			
			N=100	0.096851	0.093895
			N=300	0.088153	0.091161
			N=500	0.083725	0.085986
		X_1			
			N=100	0.094838	0.092844
			N=300	0.082749	0.08666
			N=500	0.084929	0.087267
	MAE	X_0			
			N=100	0.137418	0.111374
			N=300	0.088258	0.091161
			N=500	0.083725	0.085986
		X_1			
			N=100	0.11907	0.103271
			N=300	0.082933	0.086674
			N=500	0.084929	0.087267
Yc	Bias	X_0			
			N=100	0.591209	0.583515
			N=300	0.483253	0.493814
			N=500	0.508767	0.522097
		X_1			
			N=100	0.534451	0.539381
			N=300	0.467582	0.488238
			N=500	0.507616	0.521799
	MAE	X_0			
			N=100	0.62927	0.595459
			N=300	0.483253	0.493814
			N=500	0.508767	0.522097
		X_1			
			N=100	0.558012	0.542142
			N=300	0.467582	0.488238
			N=500	0.507616	0.521799

Note: This table reports for the g_1 model the bias and the mean absolute error (MAE) of the nearest neighbour matching estimator with replacement and the subclassification estimator with factor model estimated propensity scores, compared to the true linking effects, for link sender with $X_j = 0$ and $X_j = 1$ separately. The number of matches for the matching estimator is 1 for networks with $N = 100$, 3 for networks with $N = 300$, 5 for networks with $N = 500$. The number of subclasses for the subclassification estimator is 8 for networks with $N = 100$, 10 for networks with $N = 300$, 12 for networks with $N = 500$. All the estimates are for the average treatment effect for the treated.

Table 22: Matching and Subclassification with increasing matches and subclasses vs True Effects for g_2

				Matching	Sub
Yb	Bias	X_0			
			N=100	0.096999	0.099578
			N=300	0.083618	0.086948
			N=500	0.088433	0.089523
		X_1			
			N=100	0.087677	0.089541
			N=300	0.086064	0.08834
			N=500	0.085855	0.087839
	MAE	X_0			
			N=100	0.134378	0.112983
			N=300	0.083826	0.086948
			N=500	0.088433	0.089523
		X_1			
			N=100	0.114537	0.10228
			N=300	0.086486	0.08858
			N=500	0.085855	0.087839
Yc	Bias	X_0			
			N=100	0.529408	0.561574
			N=300	0.470578	0.489446
			N=500	0.525757	0.535877
		X_1			
			N=100	0.536314	0.54596
			N=300	0.47555	0.490552
			N=500	0.513496	0.526042
	MAE	X_0			
			N=100	0.601876	0.575155
			N=300	0.470578	0.489446
			N=500	0.525757	0.535877
		X_1			
			N=100	0.555527	0.548736
			N=300	0.47555	0.490552
			N=500	0.513496	0.526042

Note: This table reports for the g_2 model the bias and the mean absolute error (MAE) of the nearest neighbour matching estimator with replacement and the subclassification estimator with factor model estimated propensity scores, compared to the true linking effects, for link sender with $X_j = 0$ and $X_j = 1$ separately. The number of matches for the matching estimator is 1 for networks with $N = 100$, 3 for networks with $N = 300$, 5 for networks with $N = 500$. The number of subclasses for the subclassification estimator is 8 for networks with $N = 100$, 10 for networks with $N = 300$, 12 for networks with $N = 500$. All the estimates are for the average treatment effect for the treated.

References

Arduini, Tiziano, Eleonora Patacchini, and Edoardo Rainone, "Parametric and Semiparametric IV Estimation of Network Models with Selectivity," 2015.

Auerbach, Eric, "Identification and Estimation of a Partially Linear Regression Model using Network Data: Inference and an Application to Network Peer Effects," *arXiv:2105.10002 [econ]*, May 2021. arXiv: 2105.10002.

Badev, Anton, "Nash Equilibria on (Un)Stable Networks," *arXiv:1901.00373 [cs]*, December 2018. arXiv: 1901.00373.

Basse, Guillaume, Peng Ding, Avi Feller, and Panos Toulis, "Randomization tests for peer effects in group formation experiments," *arXiv:1904.02308 [stat]*, February 2021. arXiv: 1904.02308.

Bifulco, Robert, Jason M. Fletcher, Sun Jung Oh, and Stephen L. Ross, "Do high school peers have persistent effects on college attainment and other life outcomes?," *Labour Economics*, August 2014, *29*, 83–90.

Bramoullé, Yann, Habiba Djebbari, and Bernard Fortin, "Peer Effects in Networks: a Survey," *Annual Review of Economics*, 2020.

Carrell, Scott E., Bruce I. Sacerdote, and James E. West, "From Natural Variation to Optimal Policy? The Importance of Endogenous Peer Group Formation," *Econometrica*, May 2013, *81* (3), 855–882.

Cools, Angela, Raquel Fernández, and Eleonora Patacchini, "Girls, Boys, and High Achievers," Technical Report w25763, National Bureau of Economic Research, Cambridge, MA April 2019.

Crane, Harry, *Probabilistic foundations of statistical network analysis*, CRC Press, 2018.

Currarini, Sergio, Matthew O. Jackson, and Paolo Pin, "An Economic Model of Friendship: Homophily, Minorities, and Segregation," *Econometrica*, 2009, *77* (4), 1003–1045.

Diaconis, Persi and Svante Janson, "Graph limits and exchangeable random graphs," *arXiv preprint arXiv:0712.2749*, 2007.

Forastiere, Laura, Edoardo M. Airoldi, and Fabrizia Mealli, "Identification and

Estimation of Treatment and Interference Effects in Observational Studies on Networks," *Journal of the American Statistical Association*, April 2021, *116* (534), 901–918.

Gagete-Miranda, Jessica, "An aspiring friend is a friend indeed: school peers and college aspirations in Brazil," *Manuscript*, 2020, p. 46.

Goldsmith-Pinkham, Paul and Guido W. Imbens, "Social Networks and the Identification of Peer Effects," *Journal of Business & Economic Statistics*, July 2013, *31* (3), 253–264.

Graham, Bryan S., "Network data," in "Handbook of Econometrics," Vol. 7, Elsevier, 2020, pp. 111–218.

Hernán, MA and JM Robins, *Causal Inference: What If*, Boca Raton: Chapman & Hall/CRC, 2020.

Hsieh, Chih-Sheng and Lung Fei Lee, "A Social Interactions Model with Endogenous Friendship Formation and Selectivity," *Journal of Applied Econometrics*, March 2016, *31* (2), 301–319. 00118.

Imbens, Guido W. and Donald B. Rubin, *Causal Inference in Statistics, Social, and Biomedical Sciences*, Cambridge University Press, April 2015.

Johnsson, Ida and Hyungsik Roger Moon, "Estimation of Peer Effects in Endogenous Social Networks: Control Function Approach," *The Review of Economics and Statistics*, May 2021, *103* (2), 328–345.

Leung, Michael P., "Two-step estimation of network-formation models with incomplete information," *Journal of Econometrics*, September 2015, *188* (1), 182–195.

Li, Xinran, Peng Ding, Qian Lin, Dawei Yang, and Jun S. Liu, "Randomization Inference for Peer Effects," *Journal of the American Statistical Association*, October 2019, *114* (528), 1651–1664.

Manski, Charles F., "Identification of Endogenous Social Effects: The Reflection Problem," *The Review of Economic Studies*, 1993, *60* (3), 531–542.

Mele, Angelo, "A Structural Model of Homophily and Clustering in Social Networks," *SSRN Electronic Journal*, 2017.

_ **and Lingjiong Zhu**, "Approximate Variational Estimation for a Model of Network Formation," *SSRN Electronic Journal*, 2017.

Olhede, Sofia C. and Patrick J. Wolfe, "Network histograms and universality of block-model approximation," *Proceedings of the National Academy of Sciences*, October 2014, *111* (41), 14722–14727. Publisher: Proceedings of the National Academy of Sciences.

Olivetti, Claudia, Eleonora Patacchini, and Yves Zenou, "Mothers, Peers, and Gender-Role Identity," *Journal of the European Economic Association*, February 2020, *18* (1), 266–301.

Sävje, Fredrik, Peter M Aronow, and Michael G Hudgens, "Average treatment effects in the presence of unknown interference," *The Annals of Statistics*, 2021, *49* (2), 673–701. Publisher: Institute of Mathematical Statistics.

Wang, Yixin and David M. Blei, "The Blessings of Multiple Causes," *Journal of the American Statistical Association*, October 2019, *114* (528), 1574–1596.

Zhang, Yuan, Elizaveta Levina, and Ji Zhu, "Estimating network edge probabilities by neighbourhood smoothing," *Biometrika*, December 2017, *104* (4), 771–783.