Estimating peer effects and network formation models with missing links
(Preliminary)

Eyo I. Herstad*

First version: November, 2022. This version: 21st December, 2023

Abstract

Estimates of peer effects may suffer from bias if the network data has missing links. Moreover, if links are not missing at random, estimates of parameters in network formation models may also be biased. I contribute to the literature on identification of peer effects and network-formation models with partially sampled network data in three ways. My first contribution is to develop a consistent peer-effects estimator that uses link-formation probabilities from the true network-formation model. My second contribution is to develop two estimators of a network formation model that are robust to the missingness of links correlating with unobservable link-specific shocks. These are an inverse-probability-weighted likelihood estimator that uses the probabilities of observing links as weights, and a semi-parametric estimator. The first estimator requires the researcher to estimate the probability of a link being observed, while the second does not at the cost of a slower convergence rate. My third contribution is to show sharp partial identification of endogenous peer effects when there is no information on how the network is formed. The bounds from this exercise will be more informative if the researcher has information about the unobserved network. I apply my peer-effect estimator to a new dataset from two Norwegian schools that I merge with administrative data. In this data, I observe the complete network as well as a partial sample of links constructed by restricting students to only list some of their friends. Using the complete and partial sample of network links from the data, I find that my method reduces the bias in the peer-effect estimator by 65%. Finally, I demonstrate that naive estimators can lead to misleading results about household behavior in microfinance take-up by applying my peer-effect estimator to the dataset of Banerjee et al. (2013).

*Email: eyoherstad@uchicago.edu; Website: www.eyoherstad.com. I am grateful to my advisors Alexander Torgovitsky, Stephane Bonhomme and Magne Mogstad for their valuable guidance and support. I appreciate the support of the Nordic Institute for Studies in Innovation, Research and Education (NIFU), especially Ragnar Alne and Claes Lampi, for their help gathering the data needed for this project. I'd also like to thank Jiarui Liu, Wayne Gao, Sung Jae Jun, Konrad Menzel, Takuma Habu, Jonas Lieber and Myungkou Shin, as well as seminar participants at the BFI seminar and Econometric advising group at The University of Chicago for helpful comments.

1 Introduction

People do not make decisions in a vacuum. Every day, decisions are made based on the behavior of our peers, be they friends, classmates, coworkers or others we interact with. Moreover, consequences of these decisions are also affected by the behavior of our peers. There is a growing literature within economics exploring how these peer effects affect economic outcomes (Bramoullé et al., 2020) and how networks form (Graham, 2020). For example, a student's school results can be affected by his friends' classroom behavior and abilities. Households' decisions are affected by the information it obtains from the peers of its members, and the decisions of a firm can affect the behavior of the firms it interacts with.

Empirical studies of peer effects and network formation models generally have missing data on network links, and do not have good ways to correct for the missing data. A survey of applied papers in development economics found an average sampling rate of 44% of individuals in each network (Chandrasekhar and Lewis, 2016). When researchers gather information on interactions, it can be prohibitively costly, or impossible, to gather information about every agent's interaction with all other agents. Usually researchers work with partial data on the network structure, but have access to census data on covariates and outcomes for everyone (see for example Banerjee et al. (2013), Oster and Thornton (2012) and Griffith (2022)). For example, a researcher interested in educational peer effects can gather data on academic results and family background of students from the school. To obtain data on interactions between the students, the researcher would need to survey the students directly. If some students refuse to do the survey, or simply are not in the school when the survey is given, the researcher will have missing data on network links. Similarly if the survey has limited space for the students to list their friends, we will have missing data on the network links even for those students who completed the survey.

Interpreting the aggregate economic implications of peer effects depends on the structure of the network. In a school setting, this means that the way students choose their friends will decide how peer effects affect the average and variance of learning, as measured by test scores. Therefore, if a school wants to maximize average learning, minimize variance, or some combination of the two, it needs to take the links into account. Suppose there is a positive peer effect, meaning a student's test score is increased by having friends who have high test scores. Imagine a classroom with four students. Two of the students would have had high test scores with no peer effects ("high type" students), while the two others would have had low test scores ("low type" students). Suppose friendships are formed between the two high types and the two low type students. Then the high type students will amplify each other's scores through the positive peer effect. The peer effect would then allow them to achieve much better results than the low type students. If friendships are instead formed between the high and low type students, the peer effect would make the high and low type students' test scores more similar. As such the same peer effects can both increase and reduce average outcomes, and the dispersion of outcomes, depending on which individuals are connected. Understanding the aggregate economic implications of peer effect therefore requires knowledge about how agents choose who to interact with. When we do not observe all links in the network, it's therefore important to study how links form in order to understand the impact of peer effects.

In this paper I develop methods that allow researchers to identify peer effects and network formation models when they only have partial information about the network. These methods are robust to the missingness of network links potentially correlating with unobservable shocks in both the outcome equation and the network formation model. This expands the settings where researchers can credibly estimate these models, as previous work has focused on settings where network links are randomly missing (Chandrasekhar and Lewis, 2016). I will show that as the researchers obtains more information about the network, they get sharper identification of the parameters of interest. Specifically, if the researcher assumes a specification for the network formation model, my methods gives point identification of all parameters in the peer effect regression.

This paper has three main contributions to the literature on identification of peer effects and network formation models with missing data. The first contribution is constructing a point identified estimator of the parameters of a peer effect regression by specifying a network formation model. The estimator uses networks drawn independently from the same distribution as the true network, allowing for the construction of a consistent estimator using simulated instruments and covariates. Importantly, this estimator is robust to the missingness of links correlating with unobservable determinants of the outcome.

The second contribution is to derive two estimators of a network formation model that are robust to the missingness of links correlating with unobservable link-specific preference shocks. The first of these estimators models the probability of a given link being observed. I then recover the parameters of the model by solving an inverse-probability weighted likelihood, which gives consistent estimates under a conditional independence assumption. My second estimator instead puts weak conditions on the missingness process and estimates the parameters of the network formation model using a semi-parametric estimator. I then derive conditions on the sampling that allow me to estimate the necessary parameters to generate probabilities of links forming from the semi-parametric parameter estimates.

The third contribution is to show sharp partial identification of endogenous peer effects in settings where the researcher has no information about how the network forms. To achieve this I optimize over all possible realizations of the unobserved network structure. I show that this optimization has, depending on the peer effect model, either a linear or a quadratic representation, making it computationally feasible to obtain the bounds of the identified set using linear or quadratic integer programs. These bounds can be improved by restricting the set of potential configurations of the unobserved parts of the network. This approach can be computationally costly in large networks, but the researcher can easily obtain valid bounds for the endogenous peer effect by relaxing the integer constraints in the linear and quadratic programs. These bounds can be improved through the branch-and-bound algorithms used by most modern solvers.

Together these contributions for a framework allow researchers to correct for missing links based on their knowledge of the network. In the case where the unobserved network is completely unknown, the researcher can calculate the bounds as described above, which will give the smallest set of parameters consistent with the observed data. As the researcher gains more information about the network, these bounds will become more informative. Finally, if the researcher is able to specify a network formation model, they can obtain point identification despite the missing links.

To evaluate the performance of the estimators I propose, I collect data from Norwegian middle schools that allow me to observe both the true and the counterfactually observed data with missing links. This partial network data is constructed to be similar to what a researcher might have observed if they could only observe up to five male and female friendships per student.[1] This allows me to compare my consistent estimators using the partial data to the "true" effect, as estimated using the complete dataset. I find that my estimator reduces the bias in the endogenous peer effect by 65%.

The estimators are next applied to the dataset of Banerjee et al. (2013) to further assess the relevance of the bias in economic applications. I show that the uncorrected

[1] Importantly this dataset is constructed through respondents choosing which five male/female students to list, letting me construct a dataset as if I only had data on this limited list of friends.

estimators would lead researchers to mistakenly conclude that the network affects micro-finance uptake through information diffusion, when in reality the corrected estimators suggest the primary effect is through risk-sharing and joint decision making among connected households.

Several papers have previously studied ways to correct for missing link biases in peer effect regressions. Chandrasekhar and Lewis (2016) was the first paper to discuss biases due to partial network data. They develop corrections in cases where the missingness is fully at random (uncorrelated with observed or unobserved characteristics) especially in the context of regressions on network characteristics. They also develop an "analytical correction" for peer effect regressions, which involves running the peer effect regression only on a sub sample of individuals for whom we observe all the links. I add to their results by allowing the missingness of the links to be correlated with observed and unobserved characteristics in the peer effect regression, and my estimators do not require researchers to observe all the links for any individuals.[2] Chandrasekhar and Lewis (2016) also develop a GMM estimator that uses draws from the distribution of the network. This estimator is different than the one I develop in this paper as it requires the researcher take a stance on the distribution of the outcome conditional on the observed parts of the network. I will instead develop an estimator that is robust to missingness correlating with unobserved shocks in the network formation model.

Other papers have followed up the seminal contribution by Chandrasekhar and Lewis (2016) by investigating the missing link bias in different settings. Griffith (2022) studies censored networks where agents can only list a certain number of their total links. He then derives bias corrections and bounds of the bias under an assumption of *order irrelevance* implying students select which friends to list randomly, and assuming there is no endogenous peer effect. I improve upon these methods by constructing bounds for the bias in a more general setting, as well as achieving point identification by specifying the distribution of the network. Boucher and Houndetoungan (2020) develop an estimator that uses a maximum likelihood approach, estimating the network formation model and the peer effect equation jointly. This requires the researcher to make parametric assumptions about the error term in the peer effect regression. My estimator that does not require this parametric assumption.

Semi-parametric estimation of panel models is a well established literature in econometrics. My semi-parametric estimator uses similar logic to the estimators de-

[2]These results are especially important for symmetric networks, where researchers often define links as existing only if both agents respond that the link exists. In these situations, we do not know all the links for anyone, unless we observe the complete network

veloped in Manski (1987), see also Manski (1975) and Manski (1985). The work of Manski has been further developed in the literature by, for example, Abrevaya (2000), Pakes et al. (2015) and Shi et al. (2018). Unlike these papers I will work in a static network setting, which induces additional difficulties due to the structure of the unobserved heterogeneity (Graham, 2017). Applying these methods to a network setting is also done by Gao (2020)[3]. My approach is different as I do not require the estimation of the conditional expectation of each individuals degree conditional on covariates. This is especially important in the context of missing links, as we may have very few observations for each individual, which may make these conditional expectations infeasible to estimate.

Peer effects in education has for a long time been a topic of great interest to economic researchers. Sacerdote (2011) surveys the results in the literature, which fin broadly positive peer effects that vary significantly in magnitude. The majority of these papers focus on peers as defined by classmates, rather than through friendships (Boozer and Cacciola, 2001; Imberman et al., 2012; McEwan, 2003). Other papers focus instead on peer effects through friends, both on test scores (Griffith (2022) and Cohen-Cole et al. (2012)), but also on other outcomes for students such as criminal activity (Patacchini and Zenou, 2012) and health outcomes (Trogdon et al., 2008). I contribute to these literatures by offering a new identification scheme for peer effects in test scores using peer forecast errors as instruments. The dataset gathered for this paper also shows how typical forms of sampling network links lead to misleading insights about how friendships are distributed amongst students.

The rest of this paper is structured as follows. Section 2 gives an overview of the problem of missing data in peer effect and network formation models. Section 3 discusses the concrete issues for peer effect regressions and introduces my estimator of peer effects under a high level assumption about the network formation model. Section 4 introduces my two estimators of network formation models that are robust to partial data. Section 5 develops partial identification of the endogenous peer effect when the researcher has no information about the network formation model. In section 6, I apply my method to microfinance take-up in India, as well as educational peer effects in Norwegian secondary schools. Section 7 concludes.

[3]Similar approaches using maximum score methods have also been considered by Toth (2017),Kim (2018), Candelaria (2020) and Gao et al. (2022)

2 The anatomy of the problem with partial network data

In this section I will discuss how missing data on network links affect peer effect regressions, and show through a simple example how this bias behaves. As we will see, this bias is structured very differently than the classical measurement error most researchers are familiar with.

Researchers investigating peer effects are often interested in both *endogenous peer effects* and *contextual effects*. That is the feedback between an individuals outcome and their peers outcomes and the the effect on an individuals outcome from their peers covariates respectively. This motivates a specification like the one below, called the *linear-in-sums* model.

$$y_i = \alpha(Ay)_i + X_i\beta + (AX)_i\gamma + \epsilon_i \tag{1}$$

Let N be the number of individuals in the network, and define A to be a symmetric $N \times N$ matrix with $A_{i,j} = 1$ if agent i and j are connected, and $A_{i,j} = 0$ otherwise. This implies that we can write each element of Ay as $(Ay)_i = \sum_{j=1}^{N} A_{i,j}y_j$, the sum of outcomes for all individuals connected to person i. In some cases researchers are instead interested in the effects of average peer outcome and peer covariates. In these cases we can use an alternative specification where we replace A in the outcome equation with G, defined as

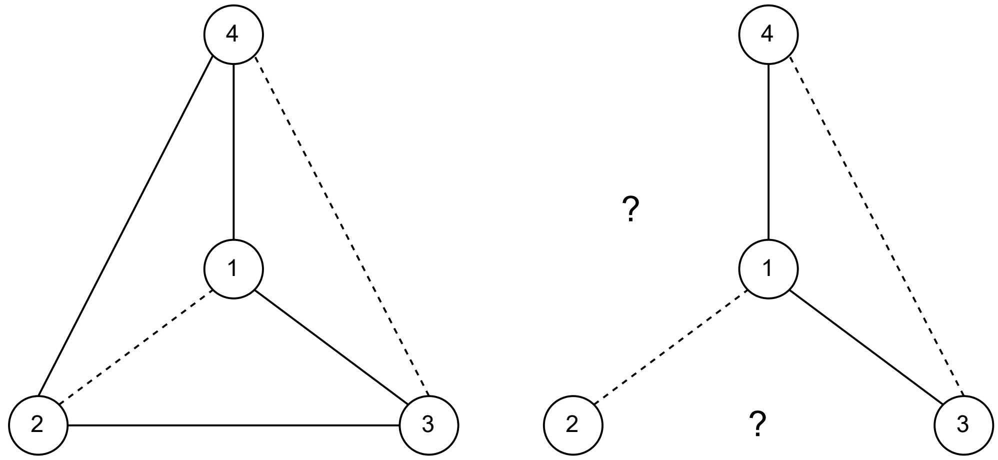
$$G_{i,j} = \begin{cases} \frac{1}{\sum_{k=1}^{N} A_{i,k}} A_{i,j} & \text{if } \sum_{k=1}^{N} A_{i,k} > 0 \\ 0 & \text{Otherwise} \end{cases}$$

This is called the *linear-in-means* model, and is generally the more studied model (Bramoullé et al., 2020). Since G is a deterministic function of A, which specification we use will not affect the results in most cases.

As discussed, to fully understand network effects researchers will also be interested in how links form. I will, when necessary, model each link $A_{i,j}$ as forming according to the following equation

$$A_{i,j} = \mathbf{1}\{W_{i,j}\theta + V_i + V_j + U_{i,j} \geq 0\} \tag{2}$$

Where $W_{i,j}$ are observed characteristics of nodes i and j, V_i is a measure of node



(a) Complete graph (b) Observed graph

Figure 1: These figures indicate the actual and observed interactions between four nodes, labeled by their numbers. Solid lines indicate two nodes are linked, while a dotted line indicate two nodes are not linked. The question marks mean the link is unobserved by the researcher.

i's "charisma", that is an intrinsic quality about i that makes them form more links. $U_{i,j}$ is a link specific preference shock. This type of model is well developed in the literature, see Graham (2017). I will go into further details about this model in section 4. This specification implies that links form independently conditional on observed characteristics and latent charisma. While I maintain this specification throughout the paper, it is not needed for all of my results. In section 3 I will instead work with the high-level assumption that the researcher has access to consistent estimates of *some* network formation model.

2.1 An example with four individuals

To show the anatomy of the biases caused by partial network data, suppose we have a dataset consisting of 4 people, with links as described in Figure 1a and outcomes

$$y_1 = 1 \quad y_2 = -3 \quad y_3 = 2 \quad y_4 = 4$$

We can write the adjacency matrix of the true network, as shown in Figure 1a, as

$$A = \begin{pmatrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \end{pmatrix} \quad \Rightarrow \quad Ay = \begin{pmatrix} 6 \\ 6 \\ -2 \\ -2 \end{pmatrix}$$

Suppose a naive researcher is interested in how individual outcomes co-vary with peer outcomes, and wants to calculate $\text{Cov}((Ay)_i, y_i)$. A simple calculation yields that $\text{Cov}((Ay)_i, y_i) = -6$. However in reality the researcher only observes the partial network as shown in figure 1b. We can write the adjacency matrix of this network as

$$A^{obs} = \begin{pmatrix} 0 & 0 & 1 & 1 \\ 0 & 0 & ? & ? \\ 1 & ? & 0 & 0 \\ 1 & ? & 0 & 0 \end{pmatrix}$$

where the question marks denote that a given link is not observed. Note that since A is symmetric, we only have two missing values. Suppose the researcher ignores the two missing links, assuming they equal zero. We would then obtain the estimate $\text{Cov}(y_i, (A^S y)_i) = 6 > -6 = \text{Cov}(y_i, (Ay)_i)$. One might think that this is simple measurement error, but the error is both more systematic and unpredictable than normal classic measurement error. Figure 2 shows a histogram of the possible values of the covariance when we two links are dropped. This distribution is not normally distributed or even centered around the true value.

How can we the researcher then obtain identification of $\text{Cov}(y_i, (Ay)_i)$ while only observing A^{obs}? Without further information, we cannot get point identification, but we can get partial identification by using our knowledge about what the network could potentially look like.

Since the network is symmetric, we have four complete adjacency matrices that are consistent with the observed data, shown below

$$\tilde{A}_1 = \begin{pmatrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} \quad \tilde{A}_2 = \begin{pmatrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \end{pmatrix} \quad \tilde{A}_3 = \begin{pmatrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} \quad \tilde{A}_4 = \begin{pmatrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \end{pmatrix}$$

The identified set for $\text{Cov}(y_i, (Ay)_i)$ would therefore be the set $\Theta = \{\text{Cov}(y_i, (\tilde{A}_k y)_i)\}_{k=1}^4$.

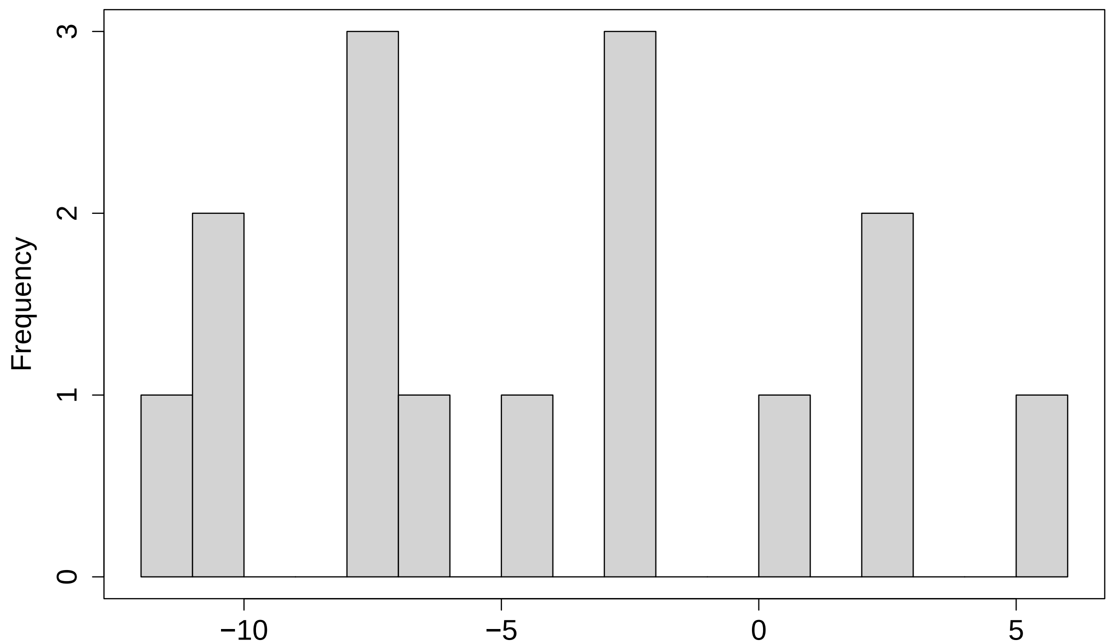


Figure 2: Histogram of possible values of $\text{Cov}(y_i, (A^S y)_i)$ when two links are dropped.

Calculating this with our outcomes gives us the values

$$\text{Cov}(y_i, (\tilde{A}_1 y)_i) = 6 \quad \text{Cov}(y_i, (\tilde{A}_2 y)_i) = -2.5 \quad \text{Cov}(y_i, (\tilde{A}_3 y)_i) = 2.5 \quad \text{Cov}(y_i, (\tilde{A}_4 y)_i) = -6$$

Since we know one of these combinations is the true network, we know $\text{Cov}(y_i, (Ay)_i) \in \Theta$. Note that with the information we have, there is no way to exclude any of the four possible networks. This means this is also the sharp identified set for $\text{Cov}(y_i, (Ay)_i)$.

While this exercise is simple in this small example, it will in general entail an integer program, either linear or quadratic, to find the largest and smallest value in Θ. This is how I will construct the partially identified sets in my paper.

2.2 Peer effect models

Usually the object researchers are interested in is not a covariance like $\text{Cov}(y_i, (Ay)_i)$. Instead they want to be able to look at the effect of friends outcomes on an agents own outcome controlling for covariates, and taking into account the inherent endogeneity between y_i and $(Ay)_i$. This motivates looking at *peer effect regressions*, like the linear-in-sums model shown in (1)

Estimating (1) by OLS will not give us consistent estimates of α. This is because the existence of the peer effects makes $\text{Cov}(\epsilon_i, (Ay)_i) \neq 0$, since a high value of ϵ_i leads to higher values of y_i, which increases the outcome of all nodes connected to i by α. This is called the *reflection problem* (Manski, 1993), which can be resolved by using $(AX)_i$ as an instrument for $(Ay)_i$ (Bramoullé et al., 2009).

Define $Z_i = ((AX)_i, X_i)$ and $T_i = ((Ay)_i, X_i)$. The IV-estimate of $\xi = (\alpha, \beta)$ is then given by solving the equations

$$\mathbb{E}[Z_i y] = \mathbb{E}[Z_i T_i]\hat{\xi}.$$

ξ is identified by the standard IV-assumptions of relevance ($\mathbb{E}[Z_i T_i]$ is invertible) and exogeneity ($\mathbb{E}[Z_i \epsilon_i] = 0$). As discussed in Bramoullé et al. (2009), this holds as long as covariates are exogenous and there is a sufficient lack of transitivity in the network.

However our researcher doesn't know the true network, and so cannot use $(AX)_i$ as an instrument. Suppose they instead wants to use $(A^S X)_i$. The problem for this researcher is then not just that the variable of interest is mismeasured, but also that his instruments are mismeasured.

We can rewrite the outcome equation as

$$y = \alpha A^S y + X\beta + \epsilon + \alpha(Ay - A^S y).$$

Defining Z_i^S as the residuals from regressing each component of $(A^S X)_i$ on X_i, and assume $\text{Cov}(Z_i^S, (A^S y)_i) \neq 0$. We can write the IV-estimate as

$$\hat{\alpha}_n^{PI} = \frac{\hat{\text{Cov}}(Z_i^S, y_i)}{\hat{\text{Cov}}(Z_i^S, (A^S y)_i)}.$$

The probability limit of $\hat{\alpha}_n^{PI}$ can be written as

$$\text{plim}_{n\to\infty} \hat{\alpha}_n^{PI} = \alpha + \frac{\text{Cov}(Z_i^S, \epsilon_i)}{\text{Cov}(Z_i^S, (A^S y)_i)} + \alpha \frac{\text{Cov}(Z_i^S, (Ay)_i - (A^S y)_i)}{\text{Cov}(Z_i^S, (A^S y)_i)} \tag{3}$$

It's standard in the literature to assume $\mathbb{E}[\epsilon_i | X, A] = 0$, where X is the matrix of all covariates for every i. This would make our instrument valid if we observed the complete graph, that is $A^S = A$. However when $A^S \neq A$, Z_i^S may not be a valid instrument. Define $S_{i,j} \in \{0, 1\}$ as an indicator for whether we observe link i, j, and S as the $N \times N$ matrix with $S_{i,j}$ as it's (i, j)'th component. Then the strengthened assumption under partial information becomes $\mathbb{E}[\epsilon_i|X, A, S] = 0$. In this case the second element of (3) equals 0 as $\text{Cov}(Z_i^S, \epsilon_i) = 0$. This is a strong restriction on the missingness of the data, and I will introduce estimators that relax this assumption later in the paper.

Let $S_N = \{i : \frac{1}{N}\sum_j^N S_{i,j} = 1\}$, that is the set of individuals for whom we observe all links. For any i in this set, $(Ay)_i = (A^S y)_i$, implying the third term of (3) equals zero. This is the "analytical correction" of Chandrasekhar and Lewis (2016). However in many settings we do not observe all links for any agents, due to a requirement of links being reciprocal. For example, in an educational setting, it's common to define a friendship as both students listing each other as friends. In this case, the set of individuals who are fully sampled becomes vanishingly small as N increases.[4]

In these settings we can still partially identify α by the same arguments as we used on the covariance earlier. Note that since Z_i^S is a valid instrument (as long as $\mathbb{E}[\epsilon_i|X, A, S] = 0$), we could obtain valid instruments of α using Z_i^S if we observed $(Ay)_i$. This insight and some algebra shows that

$$\alpha = \hat{\alpha}_n^{PI} \frac{\text{Cov}(Z_i^S, (A^S y)_i)}{\text{Cov}(Z_i^S, (Ay)_i)}$$

We can use our earlier method to find the identified set for $\text{Cov}(Z_i^S, (Ay)_i)$, which will imply an identified set for α. Manually checking the covariance for every possible value of the network is infeasible when many links are missing. To get around this, I will instead re-write the covariance into an object I can optimize over, to find the largest and smallest values of Θ. Note that $\mathbb{E}[Z_i^S] = 0$ by construction as Z_i^S is defined as the instrument residualized by the X's, including a constant. We can therefore write

$$\text{Cov}(Z_i^S, Ay) = \sum_{i=1}^N Z_i^S \sum_{j=1}^N A_{i,j} y_j = \sum_{i=1}^N \sum_{j=1}^N A_{i,j} Z_i^S y_j := \sum_{i=1}^N \sum_{j=1}^N \underbrace{c_{i,j}}_{=Z_i^S y_j} A_{i,j}.$$

which is linear in $A_{i,j}$. This means we can solve an integer linear program to obtain the largest and smallest values of the identified set. If we're instead interested in a model where outcomes depend on the average outcome of your peers, the problem

[4] For example, consider a case where we randomly sample a subset of individuals and observe the links between them. Then S_N satisfies $|S_N| \to 0$ as $N \to \infty$, as $\mathbb{P}(A_{i,j} = 1|S_i = 1, S_j = 0) \to 1$ in a dense network. That is the probability of one of your friends not being sampled approaches 1 for every individual.

becomes an integer quadratic program.[5] Furthermore, we can obtain fast valid bounds by solving the relaxed program, where we allow $A_{i,j}$ to take non-integer values in $[0, 1]$.

However these methods are all dependent on the missingness of data being conditionally independent of ϵ_i. This is a strong assumption. For example, in cases where researchers are interested in how information flows through the network, an in-depth survey may affect the types of information the agents seek out. Alternatively, students choosing to not respond to surveys may depend on their unobserved ability.

I will relax this assumption by estimating the distribution of the network, and using draws from this distribution to estimate the peer effects. This will yield point identification, and not depend on the assumption of missingness being independent of ϵ, as the estimator does not use the observed network directly when estimating the peer effect regression. A key feature of this method will be that there is heterogeneity in network formation behavior. This will ensure that there is non-zero correlation between instruments constructed by one network drawn from the distribution, and the peer outcomes constructed from a different network. Intuitively, consistency will follow from these networks having the same asymptotic distribution as the true network, though the proof is made complicated by the correlation between the true network and the observed outcomes.

Of course, for this approach to be valid it must be possible to credibly estimate the network formation model consistently even when we have missing data. I therefore develop two estimators that are consistent even when the missingness of the data correlates with unobserved link-specific preference shocks in the network formation model.

In this section I've focused on the *linear-in-sums* model for ease of exposition. However in many empirical applications researchers are instead interested in the effect of the average outcome of your peers on your own outcome, the *linear-in-means* model. As this is the more common model, I will focus on that specification for the rest of the paper.

2.3 Empirical examples

Peer effect regressions and network formation models are used in many empirical settings. Below are three examples of applications where my methods are relevant. The crucial connection between these applications are the researcher having access

[5] Unless the researcher has prior knowledge of $d_i = \sum_{j=1}^{N} A_{i,j}$, that is the number of friends of each individual. See Section 5 for details.

to data on some outcomes and covariates for the entire population, and the networks of interest being relatively large. This latter part is crucial as the asymptotic setting of my estimator is for the size of the network to grow infinitely large. Settings with many networks, such as Example 3, are valid as long as each network is growing in size.

Emprical Example 1 (Production networks)**.** *Suppose a researcher has access to tax registry data on firm covariates, as well as a separate dataset containing information on which firms trade with each other, such as the Compustat database. See for example Atalay et al. (2011), Shimamoto et al. (2022). Similarly, Lim (2018) estimates a network formation model on this data to investigate how production networks form.*

Emprical Example 2 (Educational surveys with partial responses)**.** *Many researchers are interested in both how friendships form in schools, as well as how those friendships affect educational, health and criminal outcomes. By gathering data from a school a researcher can get a lot of data about the students, but gathering friendship data requires a survey. Often, not all students will complete the survey, and which students completes the survey may not be random. For example, students who struggle in class may also struggle to fill out a long survey, leading to them not finishing the survey. There will therefore be non-randomly missing links in the network data. A good example of this setting is the well-used AdHealth dataset (see for example Griffith (2022), Patacchini and Zenou (2012), Cohen-Cole et al. (2012) and Trogdon et al. (2008)), as well as one of the empirical applications of this paper.*

Emprical Example 3 (Microfinance take-up in villages)**.** *Information diffusion and joint decision making are common traits amongst households in the developing world, especially in the context of seeding information in a network (see for example Banerjee et al. (2013) and Akbarpour et al. (2020)). Researchers are interested in how households spread information and risk between themselves and their connected households, but network data is often limited due to budgetary reasons. A peer effect framework of this problem is also studied in Chandrasekhar and Lewis (2016), and is also one of the empirical applications of this paper.*

3 Using the network formation model to achieve point identification

In this section I will show how a researcher can obtain point estimates of the peer effect regression by specifying an etwork formation model. I will work under a high-level assumption about the network formation model and will give estimators that allow us to estimate the network formation model with missing links in section 4. Note however, that other estimators, such as the methods using aggregate relational data of Breza et al. (2020) could also be used. Alternatively, researchers with access to this kind of data could use it to improve the bounds on the endogenous peer effect without specifying a network formation model using the methods discussed in section 5.

As already stated, the standard assumptions in the literature to estimate peer effect regressions with complete network information is that the network is exogenous, outcomes are non-explosive and that peer effects don't cancel out (Bramoullé et al., 2020). In my setting these assumptions can be stated as follows

Assumption 1 (Existence)**.** $|\alpha| < 1$ *and* $\alpha + \beta\gamma \neq 0$.

This assumption guarantees that the outcomes remain finite, and that the peer effects don't "cancel out".

Assumption 2 (Exogeneity of Network and Covariates)**.**
$\mathbb{E}[\epsilon_i | W, V, U, X] = 0$.

This assumption is common in the literature, though often instead stated in terms of the network A, instead of the determinants of the network, which in my setting are W, V, U. As we will see, even under this assumption identification and estimation of peer effects is non-trivial with partial network information. I discuss some ways this assumption can be relaxed in Appendix A.3.

I'll make the following assumption about the missingness of the data

Assumption 3. $S_{i,j} \sim F_S \; \forall i, j$. *Where* F_S *is potentially unknown to the researcher and may depend on other r.v. in them model.*

I will be making restrictions on F_S at various points of the paper. However importantly this assumption implies that missingness process follows the same process for everyone, and doesn't depend on the sample size.

In many of the cases economists are interested in, we want to be able to include contextual effect, as well as be robust to the sampling potentially correlating both with ϵ and U. In this setting I will only be making the following assumption

Assumption 4 (Consistent network formation model)**.** *The researcher has access to estimates $\hat{P}(A_{i,j} = 1|W_{i,j}, \hat{V}_i, \hat{V}_j)$ for all individuals i, j, and these estimates satisfy*

$$\hat{P}(A_{i,j} = 1|W_{i,j}, \hat{V}_i, \hat{V}_j) \xrightarrow{p} \mathbb{P}(A_{i,j} = 1|W_{i,j}, V_{0,i}, V_{0,j})$$

Where $V_{0,i}$ is the true charisma of individual i, and $\mathbb{P}(A_{i,j} = 1|W_{i,j}, V_{0,i}, V_{0,j})$ is the true distribution that generated the observed network.

This assumption is a high-level assumption on the researchers knowledge of the network formation model. I will discuss models that imply Assumption 4 hold in settings where the researcher has data with missing network links in Section 4.

Removing biases stemming from missing network links by specifying the distribution of the network was suggested in Chandrasekhar and Lewis (2016), who developed an OLS and GMM estimators of the effect of network characteristics on outcomes of interest. The key intuition for these solutions goes back to Rubin (1976), who suggested using distributions to "integrate out" missing data. In this section I will introduce a simulation based IV estimator that will do exactly this. This estimator will give us point identification of the parameters in the peer effect model.

If we observed the true network, we could use G^2X as an instrument. Define $Z_i = ((G^2X)_i, X_i, (GX)_i)$ and $T_i = (Gy)_i, X_i, (GX)_i$. Let $\xi = (\beta, \gamma, \alpha)$, then the IV-estimator for ξ is

$$\hat{\xi}_n^{IV} = (\frac{1}{N}\sum_i Z_i' T_i)^{-1}(\frac{1}{N}\sum_i Z_i' y_i) \xrightarrow{p} (\mathbb{E}[Z_i' T_i])^{-1}\mathbb{E}[T_i y_i]$$

That is asymptotically, the IV-estimand is the fraction of two expectations. With missing data on network links we can't calculate $\hat{\xi}^{IV}$, but by specifying the network formation model we can construct simulated instruments and covariates that will allow us to compute similar objects.

The estimator is constructed from M repetitions of two independent draws of the network. The reason to use M repetitions is to avoid instability issues that can occur when the estimator is generated from only one set of draws. Define the draw that generated the data as A_0. The first draw (A_1) is used to construct the vector of instruments Z_i^{sim}. The second draw (A_2) is used to construct the covariates, T_i^{sim}, including the endogenous peer effect. Specifically, for the $m'th$ draw of A_1, A_2 we have $Z^{sim}_{i,m} = (G^2_{1,m}X, X, G_{1,m}X)$ and $T^{sim}_{i,m} = (X, (G_{2,m}X)_i, (G_{2,m}y)_i)$. We can then define the simulated equivalent of $\hat{\xi}^{IV}$ as

$$\hat{\xi}^{SIV} = \left(\frac{1}{M}\sum_{m=1}^{M}\frac{1}{N}\sum_{i=1}^{N} Z^{sim}_{i,m}T^{sim}_{i,m}\right)^{-1}\left(\frac{1}{M}\sum_{m=1}^{M}\frac{1}{N}\sum_{i=1}^{N} Z^{sim}_{i,m}y_i\right)$$

The estimator uses simulated instruments and covariates to estimate the parameters of the peer effects. To show $\hat{\xi}^{SIV}_n$ is consistent for the true parameters I will make the following assumptions.

Assumption 5 (Regularity Assumptions for SIV)**.**

1. *X_i, ϵ_i are Lebesgue integrable i.i.d. random variables*
2. *The moments $\mathbb{E}[Z^{sim}_{i,m}T^{sim}_{i,m}]$ and $\mathbb{E}[Z^{sim}_{i,m}y]$ exist.*
3. *$E[Z^{sim}_{i,m}T^{sim}_{i,m}]$ is invertible.*

Note that Assumption 5.3 requires there to be sufficient correlation across draws of the network to generate correlation between $Z^{sim}_{i,m}, T^{sim}_{i,m}$. This can be achieved by the network formation model containing sufficient heterogeneity in link formation through observed covariates $W_{i,j}$ and the latent charisma, V_i. If instead the network is generated through an Erdős-Rényi graph, where every link has a constant probability of forming, Assumption 5.3 is unlikely to hold.

Proposition 1. *Let Assumptions 1,2, 4 and 5 hold. Then $\hat{\xi}^{SIV} \xrightarrow{p} \xi$ as $N, M \to \infty$.*

Proof. See Appendix A.5 □

Appendix A.5 also gives an estimator for over-identified IV, which lets researchers include several instruments, such as $G^2X, G^3X, G^4X, ...$ etc. Note that the asymptotic regime of the estimator is for the network to grow large. In applications with many networks, this can be interpreted as every network growing large as the sample size grows.

Define G_0 to be the true network that generated the outcomes. The key part of the consistency proof is then that as N increases, the mismeasurement $G_0y - G_{2,m}y$ equals zero in expectation due to G_0 and G_2 being drawn from the same distribution. An important complication is due to y being a function of G_0. Despite this the structure of the bias still allows us to get consistency. As Z_i is not constructed using *either* G_0 or G_2, it's correlation with G_0y and G_2y is also the same, implying the bias is asymptotically zero. A similar argument goes through for the mismeasured contextual effects. Had we instead constructed Z using G_2, it would have a different correlation with $G_2 y$ than with $G_0 y$, leading to an asymptotic bias. Similarly, as the number of draws increases, Assumption 5.2-3 guarantees the limit of the SIV estimator exists.

The main challenge in the over-identified case is to ensure that the weighting of the instruments does not use T_i^{sim}, as this would induce a different correlation between Z_i^{sim} and $G_0 y$ and $G_2 y$. I therefore suggest to either separately estimate the first stage coefficients, or to use a third draw A_3 to calculate them. We could also construct a simulated GMM estimator to achieve more efficient weights. I suggest a GMM estimator and show it's consistency in A.12.

4 Network formation models with missing network links

This section will discuss estimators that will satisfy Assumption 4 when the missingness of network links potentially correlate with unobservable preference shocks in the network formation model. Estimators for network formation that can be estimated on random subsets of network links have been developed by Chandrasekhar and Lewis (2016) and Graham (2017). Similarly, Breza et al. (2020) discuss methods when the researcher has no information on links, but has access to aggregate relational data. My estimators will instead focus on cases where the missingness of links correlates with the unobserved link-specific preference shocks.

Recall the network formation model discussed previously,

$$A_{i,j} = \mathbf{1}\{W_{i,j}\theta + V_i + V_j + U_{i,j} \geq 0\}.$$

Implying a model of transferable utility where two agents form a link if the surplus of the link is positive, and the surplus depends on a k-dimensional observable characteristics ($W_{i,j}$), the sum of their respective social influence (V_i, V_j), and random preference shock.

The model allows for some dependence between how links form through V_i. However conditional on observable characteristics and $\{V_i\}_{i=1}^N$ the network links form independently. Other network formation models allow for richer types of preferences. For example, Gao et al. (2022) give estimators for models with non-transferable utility, and Menzel (2015); De Paula et al. (2018); Graham (2016) develop estimators that allow for more strategic interactions between agents when forming links. However as these estimators are either computationally costly, do not yield point estimates or require panel data, I will focus on the model of (2).

Graham (2017) suggested two estimators for the parameters of (2) by assuming $U_{i,j}$ followed a logistic distribution, and that the sample is a random selection of agents and the links connecting them. This leads to the following maximum likelihood estimator for θ, V

$$(\hat{\theta}, \hat{V}) = \arg\max_{\theta,V} \binom{N}{2}^{-1} \sum_{i=1}^{N} \sum_{j<i} p_{\theta,V}(A_{i,j}|W_{i,j}) = \arg\max_{b,V} l_n(\theta, V)$$

Before I introduce missing links, I will consider the Assumptions needed to obtain consistent estimates of (2) when the researcher has a random sample of links. Estimation of this model is made difficult by V_i, which grows as N increases. I will make the following Assumption, which are the assumptions of Graham (2017)'s Joint Maximum Likelihood estimator (JMLE) in my notation.

Assumption 6.

1. *$\theta \in int(\Theta), supp(W) = \mathcal{W}$, where $\mathcal{W}, \Theta$ are compact subsets of $\mathbb{R}^k$.*
2. *Network formation follows 2 with $U_{i,j}$ as i.i.d. draws from a standard logistic distribution.*
3. *The support of V is $\mathcal{V}$, a compact subset of $\mathbb{R}$.*
4. *The likelihood $\mathbb{E}[l_n(b, v)|W, V_0]$ is uniquely maximized at $b = \theta, v = V$.*

Are these assumptions sufficient to guarantee consistency when we have missing data that may correlate with $U_{i,j}$? When we have missing data on network links, the likelihood a researcher would use if they naively implemented JMLE is

$$l_n^S(\theta, V) = \binom{N}{2}^{-1} \sum_{i=1}^{N} \sum_{j<i} S_{i,j} \log p_{\theta,V}(A_{i,j}|W_{i,j})$$

That is the likelihood evaluated only on the observed links. Consider the following conditional independence assumption for the sampling indicator $S_{i,j}$

Assumption 7. $S_{i,j} \perp\!\!\!\perp U_{i,j}|W_{i,j}, V_{0,i}, V_{0,j}$.

Where $V_{0,i}$ is the true value of V_i. It turns out that Assumption 7 is a sufficient condition for the consistency of the estimator with missing data on network links. Should it not hold, we can construct counterexamples that allow us to break identification. Consider the case of $S_{i,j} = \mathbf{1}\{U_{i,j} > 0\}$. Intuitively, the logistic assumption no longer holds in sample due to the sampling, which means the likelihood is misspecified, leading to inconsistent estimates. The following proposition summarizes these results.

Proposition 2 (Potential inconsistency of sampled likelihood)**.** *Let assumption 6 hold. Then assumption 7 is sufficient for the consistency of* $(\hat{\theta}^S, \hat{V}^S)$*, and if assumption 6 does not hold, there exists counterexamples where* $\hat{V}, \hat{\theta}$ *are inconsistent.*

Proof. See Appendix A.6 □

The independence relation of assumption 7 might be difficult to justify in many applications. For example, in surveys with non-response bias we may worry that the response rate depends on things that correlate with $U_{i,j}$. In settings where agents are only able to list a certain amount of their friends, they may choose who to list based on how much they like each person, which likely correlates with the preference shock $U_{i,j}$. The following section will therefore introduce an alternative estimator that allows the researcher use auxiliary variables to re-establish an independence relation similar to the one in assumption 7. Section 4.2 will discuss a semi-parametric estimator that will be consistent under more general types of missing data, like the second case discussed above.

4.1 Estimation using knowledge about sampling

While researchers may not want to use Assumption 7, they may have credible reasons to believe an expanded version, which allows for variables not included in $W_{i,j}$, to hold. Such an assumption is given below

Assumption 8.

$S_{i,j} \perp\!\!\!\perp U_{i,j} | W_{i,j}, V_{0,i}, V_{0,j}, X_i^S, X_j^S$

Assumption 8 for some auxiliary variables, X_i^S, not contained in W, to generate the desired independence. I will use these variables to generate inverse probability weights that will adjust the observed likelihood so that it behaves like the randomly sampled likelihood asymptotically. To do this I make the following assumptions

Assumption 9.

1. $P(S_{i,j}=1|W_{i,j},V_i,V_jX_i^S,X_j^S)>0$ *for all* $W_{i,j},V_i,V_j,X_i^S,X_j^S \in supp(W_{i,j},V_i,V_j,X_i^S,X_j^S)$.

2. (a) $P(S_{i,j}=1|W_{i,j},V_i,V_j,X_i^S,X_j^S)$ *is known*

   (b) $\hat{P}_N(S_{i,j}=1|W_{i,j},V_i,V_j,X_i^S,X_j^S)$ *is known and satisfies*

$$\lim_{N\to\infty}\hat{P}_N(S_{i,j}=1|W_{i,j},V_i,V_j,X_i^S,X_j^S)=P(S_{i,j}=1|W_{i,j},V_i,V_j,X_i^S,X_j^S)$$

The first part of assumption 9 ensures that there are no values of $W_{i,j},X^S,V$ which never have observed links, ensuring that our re-weighted estimator will eventually recover the true distribution of $U_{i,j}$. The second part is challenging due to the inclusion of V_i, which is an unobservable variable. In this section I will assume that $S_{i,j} \perp\!\!\!\perp V_i,V_j|W_{i,j},X_i^S,X_j^S$, implying

$$P(S_{i,j}=1|W_{i,j},V_i,V_j,X_i^S,X_j^S)=P(S_{i,j}=1|W_{i,j},X_i^S,X_j^S)$$

. See appendix A.2 for an alternative approach without this assumption.

Estimating $P(S_{i,j}=1|W_{i,j},X_i^S,X_j^S)$ can be done through knowledge of how the sampling process works. For example, suppose a researcher has data on students at a school, and observes links going between students who answered a survey. Define $S_i=1$ if individual i answered the survey, with $S_i=0$ otherwise. Then we can write $S_{i,j}=S_iS_j$, a so called induced sub-sample.[6] If each student's decision to respond to the survey is independent conditional on $W_{i,j},X_i^S,X_j^S$, then

$$P(S_{i,j}=1|W_{i,j},X_i^S,X_j^S)=P(S_i=1|W_{i,j},X_i^S,X_j^S)P(S_j=1|W_{i,j},X_i^S,X_j^S)$$

The researcher can estimate the probability of each student answering the survey to construct the probability of observing a link. If $W_{i,j}=f(W_i,W_j)$ and answering the survey does not depend on other people's covariates the expression simplifies to

$$P(S_{i,j}=1|W_{i,j},X_i^S,X_j^S)=P(S_i=1|W_i,X_i^S)P(S_j=1|W_j,X_j^S)$$

These assumptions are not necessary for the estimator, but should be thought of as an example of how a researcher can construct the probabilities of links being observed. The key point for the estimators described below is that the researcher either has prior knowledge of the probabilities or can estimate them consistently.

[6] Alternative types of sampling have similar forms. If we observe all links including a node (star sampling), we can write $S_{i,j}=S_i+S_j-S_iS_j$.

Define the re-weighted estimator as

$$l_n^C(\theta, V) = \frac{1}{n}\sum_{i<j} \frac{1}{\hat{P}_N(S_{i,j}=1|W_{i,j}, X_i^S, X_j^S)} S_{i,j}\log p_\theta(A_{i,j}|W_{i,j}, V)$$

Proposition 3 (Consistency of corrected likelihood)**.** *Assume Assumptions 6, 8 and 9 hold, then*

$$(\hat{\theta}_n^C, \hat{V}_n^C) = \arg\max_{b,v} l_n^C(b, v) \xrightarrow{p} (\theta_0, V)$$

Proof. See appendix A.6 □

The proof of Proposition 3 shows that the re-weighted likelihood converges to the randomly sampled likelihood, as the sample size grows. Consistency of the estimates then follows by the arguments used in Graham (2017).

Note that more complex maximum likelihood estimators can be adapted to have re-weighing results similar to the estimator discussed here. For example, we can construct a version of the tetrad logit estimator of Graham (2017) by considering the variable $S_{i,j,k,l}$, an indicator that the links connecting the four individuals i, j, k, l are observed, rather than $S_{i,j}$. Assuming we observed the induced subsample of links between individuals sampled in a conditionally independent way, the estimator is easily constructed in a similar manner to above, using the definition of $S_{i,j,k,l} = S_i S_j S_k S_l$.

In datasets with these kinds of missing data it's also likely that there are many individuals for whom we observe no links. This in turn implies we cannot recover V_i for such individuals. I develop a way to extrapolate these values and recover estimates of link probabilities satisfying Assumption 4 in Appendix A.11.

4.2 Semi-parametric estimation

Some forms of sampling use the edges directly to sample, which will make the independence relation of Assumption 8 unlikely to hold. Similarly, some datasets have missingness patterns that are too complicated for the researcher to credibly estimate the probability of edges being missing, or may not have sufficient data available to estimate a credible model for the sampling. I will therefore propose a second estimator in this section. I will develop an estimator based on the relative local linking decisions of agents. This estimator is based on previous work on semi-parametric estimators like the maximum rank correlation estimators of Han (1987); Abrevaya (2000), and the maximum score estimator developed in Manski (1975) and Manski

(1985). Additionally, this estimator can also be used to obtain estimates of θ, V up to scale without making parametric assumptions about U. However to correct the peer effect regression, the distributional assumptions on U will have to hold in population. Otherwise, we have no way of mapping our parameter estimates to the probability of two agents forming a link.

The key insight of the estimator stems from the following lemma, building on similar ideas in Manski (1987) and Abrevaya (2000).

Lemma 1. *Let* $\Delta A^*_{i,k,l} = (W_{i,k} - W_{i,l})\theta + V_k - V_l$ *and* $\Delta A_{i,k,l} = A_{i,k} - A_{i,l}$*. Then for* $(i, j, k, l) \in [1, ..., n]^4$

$$\Delta A^*_{i,k,l} > \Delta A^*_{j,k,l} \Rightarrow \mathbb{E}\left[\mathbf{1}\{\Delta A_{i,k,l} \geq \Delta A_{j,k,l}\}\right] > \mathbb{E}\left[\mathbf{1}\{\Delta A_{i,k,l} \leq \Delta A_{j,k,l}\}\right]$$

Proof. See Appendix A.7. □

I will refer to combinations of four individuals like i, j, k, l as a “square”. Importantly the event $\Delta A^*_{i,k,l} > \Delta A^*_{j,k,l} \Leftrightarrow (W_{i,k} - W_{i,l})\theta > (W_{j,k} - W_{j,l})\theta$ does not depend on any element of V. This allows me to construct an estimator for θ without estimating V at the same time. This feature is especially important for computational feasibility, as the objective function will be non-convex.

Note that when $\Delta A_i = \Delta A_j$, this square provides no information on θ as the inequalities in lemma 1 are not strict. However the expressions are not invariant to re-labeling the nodes. In fact for any given square there are 24 potential ways to relabel nodes to get different versions of the expressions in lemma 1. However it's easy to check that all these possible configurations will reduce to one of the following inequalities in an undirected network.

$$\begin{aligned}
A_{i,k} - A_{i,l} - (A_{j,k} - A_{j,l}) \lesseqgtr 0 \\
A_{j,i} - A_{j,l} - (A_{k,i} - A_{k,l}) \lesseqgtr 0 \\
A_{i,j} - A_{i,l} - (A_{k,j} - A_{k,l}) \lesseqgtr 0
\end{aligned}$$

These three cases will each correspond to one of the three sub-objectives defined below

$$
\begin{aligned}
s^1_{i,j,k,l}(\theta) &= \mathbf{1}\{A_{i,k} - A_{i,l} - (A_{j,k} - A_{j,l}) \geq 0\}\mathbf{1}\{(W_{i,k} - W_{i,l} - (W_{j,k} - W_{j,l}))\theta \geq 0\} \\
&\quad + \mathbf{1}\{A_{i,k} - A_{i,l} - (A_{j,k} - A_{j,l}) \leq 0\}\mathbf{1}\{(W_{i,k} - W_{i,l} - (W_{j,k} - W_{j,l}))\theta \leq 0\} \\
s^2_{i,j,k,l}(\theta) &= \mathbf{1}\{A_{j,i} - A_{j,l} - (A_{k,i} - A_{k,l}) \geq 0\}\mathbf{1}\{(W_{j,i} - W_{j,l} - (W_{k,i} - W_{k,l}))\theta \geq 0\} \\
&\quad + \mathbf{1}\{A_{j,i} - A_{j,l} - (A_{k,i} - A_{k,l}) \leq 0\}\mathbf{1}\{(W_{j,i} - W_{j,l} - (W_{k,i} - W_{k,l}))\theta \leq 0\} \\
s^3_{i,j,k,l}(\theta) &= \mathbf{1}\{A_{i,j} - A_{i,l} - (A_{k,j} - A_{k,l}) \geq 0\}\mathbf{1}\{(W_{i,j} - W_{i,l} - (W_{k,j} - W_{k,l}))\theta \geq 0\} \\
&\quad + \mathbf{1}\{A_{i,j} - A_{i,l} - (A_{k,j} - A_{k,l}) \leq 0\}\mathbf{1}\{(W_{i,j} - W_{i,l} - (W_{k,j} - W_{k,l}))\theta \leq 0\}
\end{aligned}
$$

All of which are invariant to the values of the Vs.

The full objective combines these three cases for all possible squares, and can be written as

$$
S_N(\theta) = \frac{1}{3\binom{N}{4}} \sum_{(i,j,k,l)\in P} \left(s^1_{i,j,k,l}(\theta) + s^2_{i,j,k,l}(\theta) + s^3_{i,j,k,l}(\theta) \right) \tag{4}
$$

where P is the set of all possible squares formed from our N nodes. It's worth noting that $|P|$ is very large, having $3\binom{N}{4}$ elements. When estimating θ we only need to look at informative squares, which is usually a small fraction of all possible squares. Finding the informative squares can be computationally costly in large networks. However this procedure only needs to be done once, as the set of informative squares only depends on the links, and not the values of θ.

I will use the following assumptions to show identification

Assumption 10 (Semi-parametric identification assumptions)**.**

1. *Network formation follows Equation* (2)

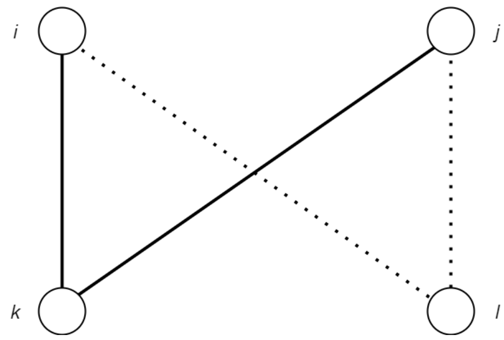
2. *There is an element of θ for which $\theta^h \neq 0$ which is normalised so that $|\theta^h| = 1$, and θ is contained in a compact subset of $\mathbb{R}^k$.*

3. *$U_{i,j}|S_{i,j} = 1$ are distributed i.i.d. for all (i, j).*

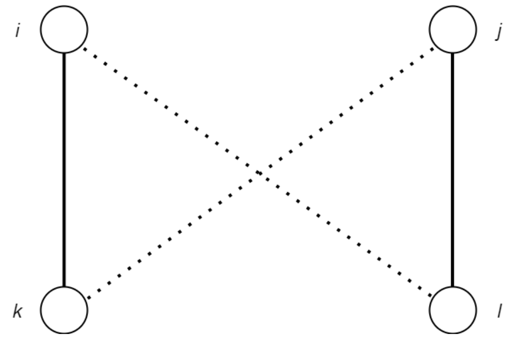
4. *$W_{i,j} \perp\!\!\!\perp U_{i,j}|S_{i,j} = 1$, and*

   (a) *The support of $W_{i,j}$ is not contained in any proper linear subspace of $\mathbb{R}^k$*

   (b) *At least one covariate of $W^c_{i,j}$ has everywhere positive Lebesgue density conditional on $(W^1_{i,j}, ..., W^{c-1}_{i,j}, W^{c+1}_{i,j}, ..., W^k_{i,j})$*



(a) No information, as $\Delta A_{i,k,l} - \Delta A_{j,k,l} = 0$



(b) Informative, as $\Delta A_{i,k,l} - \Delta A_{j,k,l} = 2$

Figure 3: An informative and uninformative case links for the θ estimation.

The first part of the assumption is that the network formation model is well specified. The second part assumes that the researcher has one variable which they know is non-zero, that is used to normalize all other variables. The third part is a conditional version of the common i.i.d. assumption in semi-parametric estimators. This is a restriction on the sampling procedure, so that $S_{i,j}$ cannot induce different distributions for $U_{i,j}$, or dependencies between the preference shocks of different linkages. As such we can think of this as a uniformity assumption about sampling. Finally the fourth part assumes the observable characteristics are independent of $U_{i,j}$, again conditionally, and that at least one of the covariates is continuous conditional on the other variables.

Proposition 4. *Let assumption 10 hold, define* $\hat{\theta}_N = \arg\max_{b \in B} S_N(b)$, *then we have*

$$\lim_{N \to \infty} \hat{\theta}_N = \theta \quad a.s.$$

Proof. See Appendix A.7 □

To understand the intuition of this result, it's useful to visualize what combinations of four agents are and are not informative for the objective. Figure 3 shows two such cases. Case (a) is uninformative since person i and person j are making the same decisions about person k and l, as both i and j choose to form a link with k but not with l. In case (b) person i chooses to be friends with person k but not person l, while j chooses the opposite. Under assumption 10 this must, in expectation, be due to $\Delta A^*_{i,k,l} > \Delta A^*_{j,k,l}$, so θ should be adjusted such that this holds in as many squares as possible.

The next step is to estimate the social influence parameters V. Both because researchers may be interested in them on their own, but also because they are needed to obtain consistent estimates of the peer effect regressions.

To do this, I define some reference person 0, and I will then aim to recover each individual's charisma relative to person 0. This means estimating

$$\Delta V_k = V_k - V_0 \qquad \forall k \neq 0$$

Define $A^*_{ik} = W_{i,k}\hat{\theta} + \Delta V_i$. Then using the logic of Manski (1987) and lemma 1, $A^*_{ik} > A^*_{i0} \Rightarrow \mathbb{E}[A_{ik}] > \mathbb{E}[A_{i0}]$.

This allows us to estimate each ΔV_k separately by finding the ΔV_k that maximizes.

$$S^k_N = \frac{1}{N-2} \sum_{i \in (1,\dots,n)/(0,k)} \mathbf{1}\left\{W_{i,k}\hat{\theta} - W_{i,0}\hat{\theta} + \Delta V_k > 0\right\} \mathbf{1}\{A_{ik} \geq A_{i0}\}$$
$$+\mathbf{1}\left\{W_{i,k}\hat{\theta} - W_{i,0}\hat{\theta} + \Delta V_k < 0\right\} \mathbf{1}\{A_{ik} \leq A_{i0}\} \qquad (5)$$

Where $\hat{\theta}$ is the estimate of θ from the previous procedure.

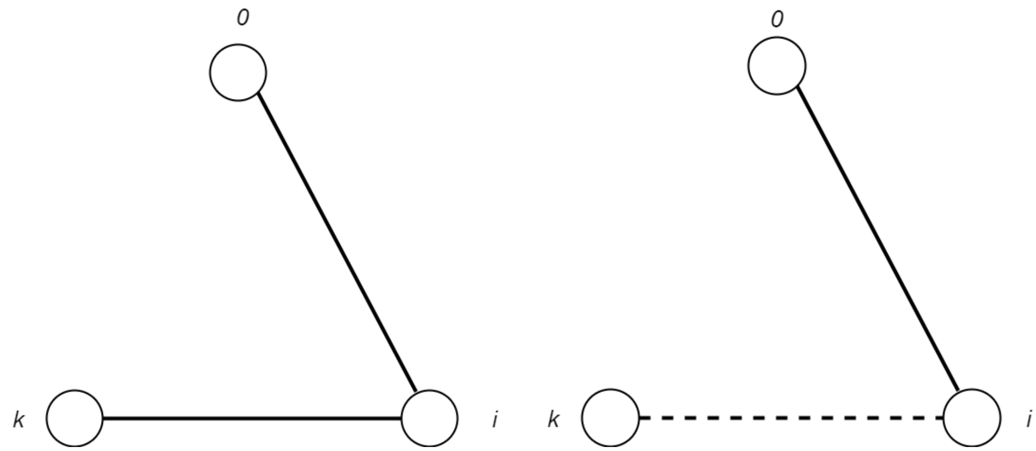
Proposition 5. *Let Assumption 10 hold. Then the estimates $\widehat{\Delta V}$ generated from solving* (5) *satisfy*

$$plim_{N \to \infty} \widehat{\Delta V}_k = \Delta V_k = V_k - V_0$$

Proof. See Appendix A.7 □

The proof of this proposition follows the same logic as the estimator for θ. Intuitively, people with large social influence relative to person 0 should have more friendships with other people than their covariates predict. As such this estimator constructs the social influence estimates through relative decisions, rather than the degree distribution of agents directly. This is because we do not necessarily have good knowledge of the total amount of connections for any node. We can again show informative and uninformative cases graphically. Figure 4 shows an informative and uninformative case for this objective. By comparing the link decision of each individual i relative to individual k and the reference person, we obtain information about the relative influence of individual k relative to 0 whenever person i makes a different decision about k than person 0. This difference in decisions has to be, correcting for the observable differences, due to k being more/less socially influential than person 0.

Note that the estimator described above only uses a small share of the total number of "triangles" in the network. Ideally we would prefer to use all the triangles



(a) No information, as $A_{ik} - A_{i0} = 0$ (b) Informative, as $A_{ik} - A_{i0} \neq 0$

Figure 4: An informative and uninformative case of links for the estimation of V.

available, to take full advantage of all the data. However to do so would mean jointly estimating every V_i at the same time. Not only is this a more computationally costly optimization, but since the objective is non-convex, we need to run the algorithm several times with different starting values to ensure we find the global optimum. To do so for the joint estimation of V would require picking "many" starting values from $\mathbb{R}^N$, which is infeasible with current software. However since my estimator estimates each V_i separately, this issue is removed.

A standard way to improve both the computational cost and the asymptotic efficiency of discontinuous objectives like those in equation (4) and (5) is to use a kernel to make the objective continuous in the parameters. This both allows easier computation (since we can use gradient based methods), and can also improve the asymptotic convergence rate of the estimates (Horowitz, 1992). Consistency of this smoothed estimator is shown in Appendix A.8, and all the estimates in the empirical application use the smoothed estimators.

In cases where we're only interested in the parameters of the network formation model, we now have all we need. However to generate the probability of two agents linking for the entire network we may also need to have a way to extrapolate the charisma parameters to unobserved parts of the data, I discuss how to do this in Appendix A.11. Separately, in cases where we estimated θ using the semi-parametric estimator, we need to recover the scale and level of U to calculate the probability of two agents linking.

4.3 Recovering the scale of U

To use the semi-parametric estimates to calculate probabilities from the network formation model, we need to recover the scale of the logit parameters. This is due to the logit parameters being scaled by the variance of $U_{i,j}$, while the parameters of the semi-parametric estimator is scaled by the continuous covariate picked by the researcher. Unlike the first estimator that used assumption 8, the semi-parametric estimator has made no conditional independence assumption between the $S_{i,j}$ and $U_{i,j}$. However to recover the scale we will need such an assumption, given below.

Assumption 11. *Let $C_{i,j}$ be a binary random variable, and $U_{i,j} \perp\!\!\!\perp S_{i,j} | C_{i,j} = 1$.*

This assumption states that there is some variable $C_{i,j}$ such that $U_{i,j}$ is indendent of $S_{i,j}$ conditional on it being equal 1. For example, in the empirical application $C_{i,j} = 1$ is the event where neither i nor j lists the total allowable amount of friends. In this case we can arguably say that the links being observed is independent of the unobserved preference shock $U_{i,j}$, as both students are able to list all friendships. Crucially we do not use links between censored and uncensored units, as the link being nominated by the censored student (rather than one of their other friends) may be due to a high preference shock in the link between the censored and uncensored student. The exact nature of $C_{i,j}$ will depend on the types of missingness faced by researchers, but it is necessary to have some way of recovering the scale of $U_{i,j}$ to be able to calculate the linking probabilities needed for the SIV estimator.

Note that I allow for $V_i, V_j \not\perp\!\!\!\perp S_{i,j} | C_{i,j} = 1$, allowing for there to be lower/higher amount of links given $C_{i,j} = 1$ than in the true network. In fact in the setting where students are only allowed to list a certain amount of friends, the distribution of V_i estimated naively with the observed data is likely very different from the overall distribution of V.

This allows us to recover the scale and level of U by running the logit regression of $A_{i,j}$ on $W_{i,j}\hat{\theta} + \hat{V}_i + \hat{V}_j$ for the set with $C_{i,j} = 1$. Consistency of this procedure is shown in Appendix A.9. Note that since this involves a logit on an estimated regressor, there can be significant finite sample biases. However implementing the bias correction procedure of Stefanski and Carroll (1985) may alleviate this problem [in progress].

The following algorithm summarizes the steps to compute the linking probabilities for the entire population with the semi-parametric estimates.

Algorithm 1.

1. *Estimate* $\hat{\theta}, \hat{V}$ *by solving the objectives* (4) *and* (5)

2. *Recover* $\hat{V}_0 = -\mathbb{E}[\Delta V_i]$.

3. *Estimate the relative scale and level* σ_1, σ_2 *by running the logit of* $A_{i,j}$ *on* $W_{i,j}\hat{\theta} + \hat{V}_i + \hat{V}_j$ *for the subset* T.

4. *Calculate* $P(A_{i,j} = 1|W_{i,j}, V_i, V_j)$ *as*

$$\frac{exp(\hat{\sigma}_1 + \hat{\sigma}_2(W_{i,j}\hat{\theta} + \hat{V}_i + \hat{V}_j))}{1 + exp(\hat{\sigma}_1 + \hat{\sigma}_2(W_{i,j}\hat{\theta} + \hat{V}_i + \hat{V}_j))}$$

Proposition 6. *Assume Assumptions 10 and 11 and 13 hold. Then the predictions* $P(A_{i,j} = 1|W_{i,j}, V_i, V_j)$ *produced by Algorithm 1 satisfy assumption 4*

Proof. See Appendix A.10 □

5 Partial identification without the network formation model

In this section I discuss identification in cases where we do not have the network formation model. I will show how to obtain partial identification by using integer programming methods to search over possible network configurations. Consider the following Assumptions

Assumption 12 (Partial identification setting)**.**

1. $S_{i,j} \perp\!\!\!\perp (\epsilon_i, \epsilon_j)\ \forall (i,j) \in [1, ..., N]^2$

2. $\gamma = 0$.

The first part of Assumption 12 assumes that sampling is independent of the outcomes, which is required for the observed instrument Z_i^S to be exogenous, and therefore valid. The second part assumes away any effects from peer covariates on the outcome (so called *contextual effects*) in the outcome equation.

Importantly, Assumption 12 does not assume that we observe all the links for some individuals. Formally, this would mean assuming that there exists a set S_N defined as

$$S_N = \{i \in (1, ..., N) : \sum_{j=1}^{N} S_{i,j} = N\}$$

Satisfying $\lim_{N\to\infty} |S_N| = \infty$. This would, together with to part 1 of Assumption 12 and Assumptions 1 and 2, allow us to use the analytical correction in Chandrasekhar and Lewis (2016) to get a consistent estimate of α, β and γ, even if $\gamma \neq 0$.

The existence of such a set is unlikely in many empirical contexts, as links are often defined based on reciprocity. For example, in an educational context friendships are commonly defined as existing if both students nominating each other as friends. Restricting attention to students for whom we know all the friendships of would therefore be a large restriction on the sample size. In empirical settings where network data is gathered through surveys of parts of the population, S_N would be the set of individuals who only list other households who are surveyed as being linked to them, which in dense network would satisfy $\lim_{N\to\infty} |S_N| = 0$ for any (constant) sampling rate less than 1.

Define

$$G^S_{i,j} = \begin{cases} \frac{1}{\sum_{k=1}^N S_{i,k} A_{i,k}} S_{i,j} A_{i,j} & \text{if } \sum_{k=1}^N A_{i,k} S_{i,k} > 0 \\ 0 & \text{otherwise} \end{cases}$$

We can then define $(G^S y)_i = \sum_{j=1}^N G^S_{i,j} y_j$, and rewrite the outcome equation to

$$y = G^S y\alpha + X\beta + \alpha(Gy - G^S y) + \epsilon$$

Like before, define Z_i^S as the residuals of regressing $((G^S)^2 X)_i$ on X_i. The feasible IV estimator using Z_i^S as an instrument for $(G^S y)_i$ then defined by

$$\begin{aligned} \hat{\alpha}_n^{PI} &= \frac{\text{Cov}(y_i, Z_i^S)}{\text{Cov}(Z_i^S, (G^S y)_i)} \qquad (6) \\ &= \alpha + \frac{\text{Cov}(Z_i^S, \alpha((Gy)_i - (G^S y)_i))}{\text{Cov}(Z_i^S, (G^S y)_i)} + \frac{\text{Cov}(Z_i^S, \epsilon)}{\text{Cov}(Z_i^S, (G^S y)_i)} + \gamma \frac{\text{Cov}(Z_i^S, (GX)_i - (G^S X)_i)}{\text{Cov}(Z_i^S, (G^S y)_i)} \end{aligned}$$

To see why Assumption 12 is needed, consider the case where it doesn't hold. In this case, both the second, third and fourth terms above could be non-zero. The first bias stems from the instrument correlating with the endogenous measurement error in $G^S y$. The second stems from the observed instrument Z_i^S being invalid, even if we observe Gy perfectly. This would similarly make the analytical correction of Chandrasekhar and Lewis (2016) invalid. The second bias stems from the contextual effects, yielding mismeasured covariates. As the residualisation of $(G^S)^2 X$ on $X, G^S X$ does not guarantee that $\text{Cov}(Z_i^S, GX) = 0$, this is an additional source of bias.

With Assumption 12, the only remaining bias is caused by the term

$$\text{Cov}(Z_i^S, \alpha((Gy)_i - (G^S y)_i))$$

This is the covariance between the instrument and the measurement error in peer outcomes. Note that the bias is zero when $\alpha = 0$, as the bias is created by the dependence between an individuals covariates and their peers outcomes, which is fully driven by the peer effect.

I will show that it's possible to construct valid bounds for the endogenous peer effect by searching over possible network configurations. Unlike the methods discussed in section 3, this does not require any additional assumptions about the network formation model.

5.1 Constructing a bound

Note that we can rewrite (6) to

$$\alpha = \hat{\alpha}_n \frac{\text{Cov}((G^S y)_i, Z_i^S)}{\text{Cov}((Gy)_i, Z_i^S)}$$

We can write out the unobserved covariance $\text{Cov}(Z_i, (Ay)_i)$ as, remembering that $\mathbb{E}[Z_i] = 0$

$$\text{Cov}((Gy)_i, Z_i) = \mathbb{E}\left[\sum_j G_{i,j} Z_i^S y_j\right]$$

The empirical counterpart to this is the double sum

$$\frac{1}{N} \sum_i \sum_j \frac{1}{\sum_{k=1}^N A_{i,k}} Z_i^S y_j A_{i,j} := \sum_i \sum_j t_i c_{i,j} A_{i,j}$$

Where I've defined $t_i = \sum_{k=1}^N A_{i,k}$, the inverse of agent i's number of peers (d_i). Since t_i is a non-linear function of $A_{i,j}$ this is no longer a linear program. However it can be re-written as a quadratic program with the 2N added variables $\{t_i\}_{i=1}^N$ and $\{d_i\}_{i=1}^N$. To see this, note that the objective is quadratic in $t_i, d_i, A_{i,j}$ in the sense that if we write $v = (A_{1,2}, A_{1,3}, ..., A_{N-1,N}, t_1, ..., t_N, d_1, ..., d_N)$, then the objective can be written as

$$\sum_i \sum_j t_i c_{i,j} A_{i,j} = v' Q v$$

For correctly defined Q. Similarly we can define N quadratic constraints of the form

$$t_i d_i = 1 \quad \Leftrightarrow v' Q_i^C v = 1$$

This means we can represent the covariance as a quadratic program with quadratic constraints.

We can bound this object by solving the problem over the missing links

$$\begin{aligned} b_{max} = \max_{A_{i,j}, t_i, d_i} \sum_i \sum_j t_i c_{i,j} A_{i,j} \qquad s.t. \\ t_i d_i = 1 \quad \forall i \\ d_i = \sum_{j=1}^{N} A_{i,j} \quad \forall i \\ A_{i,j} \in \{0,1\} \quad \forall (i,j) \\ A_{j,i} = A_{i,j} \qquad \forall (i,j) \\ A_{i,j} \in \mathcal{A} \end{aligned}$$

With the lower bound b_{min} defined for the minimum of the same problem, and $\mathcal{A}$ is a set of linear/quadratic constraints on the network structure. Achieving informative bounds is dependent on a researchers ability to impose restrictions on this set.

For example, we could assume that sampling is conditionally independent from links forming, given some X. This would imply that $\mathbb{E}[A_{i,j}|S_{i,j} = 1, X] = \mathbb{E}[A_{i,j}|S_{i,j} = 0, X]$, which is a constraint linear in $A_{i,j}$. Alternatively we may have some other information about the network, such as the aggregate relational data (ARD) of Breza et al. (2020). This type of data are answers to survey questions of the form "How many friends do you have with characteristic X?". This can be transformed into constraints of the form $\sum_{i,j} \mathbf{1}\{X_j = X\} A_{i,j} = d_{x,i}$, which is linear in $A_{i,j}$ and therefore easy to implement in the program. Similarly we could ask about the total amount of friends. Simulations in Appendix C.1 show that knowledge about the total degree of friends makes the bounds much more informative. This knowledge also makes computation of the bounds easier, as it makes the objective linear.

The bounds are computed through Integer programs, due to the restriction that $A_{i,j}$ is binary. These programs can be costly to solve, especially if the network is large and has many unobserved links. However most modern solvers for these programs solve them through a method called branch and bound. A nice property of these solvers is that it naturally generates bounds on the solution to the program while solving the problem. The solver then proceeds to improve this bound until it finds the optimum, meaning it's possible for a researcher to obtain valid bounds quickly, and better bounds the longer the researcher is willing to wait.

To see this intuitively, consider replacing the constraints $A_{i,j} \in \{0,1\}$ with $A_{i,j} \in [0,1]$. As any solution with the first set of constraints also satisfy the second set, the minimum and maximum solution with the second set must be weakly smaller/larger than the optimal values with the first set of constraints. We can therefore obtain valid bounds fast by solving the problem as if there are no integer constraints.

These bounds are sharp, but can be uninformative. Specifically, if the bounds for $\text{Cov}((Ay)_i, Z_i^S)$ contain 0, the bounds are completely uninformative, due to the inverse relation between α and $\text{Cov}((Ay)_i, Z_i^S)$. However as long as $\text{sign}(b_{min}) = \text{sign}(b_{max}) \neq 0$ we have informative bounds on α.

Define $\mathcal{A}_0 = \{A_{i,j} : S_{i,j} = 0\}$, that is the true values of the unobserved links.

Proposition 7. *Let the restrictions $\mathcal{A}$ be such that $\mathcal{A}_0 \in \mathcal{A}$, Assumptions 1, 2 and 12 hold, and $sign(b_{max}) = sign(b_{min}) \neq 0$, then*

$$\alpha \in [b_{min}^{-1}\hat{\alpha}, b_{max}^{-1}\hat{\alpha}]$$

and the bounds are sharp

Proof. See appendix A.4 □

Sharpness follows intuitively from the fact that given a network, the IV-assumptions give us a value of α that is consistent with the observed data. The added assumption of $\mathcal{A}_0 \in \mathcal{A}$ guarantees that the researcher isn't imposing any restrictions in the program that aren't satisfied by the true network, guaranteeing validity of the bounds. See Appendix C.1 for simulations showing when these bounds can be informative.

6 Empirical Applications

In this section I will go through two applications, which cover two of the most common reasons for partial network data.

The first application is a new dataset gathered from two secondary schools in Norway. The survey was constructed to get both the full network of friendships, as well as the partial data one might have gathered if one only gave students the option of listing 10 of their friends. This lets me compare the corrected partial estimates to the "true" estimates obtained from the full network data.

The second application uses data gathered through surveys in 43 Indian villages to answer how information diffusion affects take up of microfinance Banerjee et al. (2013). There I find that while the survey was administered almost randomly, the time difference between the survey and the measurement of the outcome lead to the survey correlating with take-up of micro-finance.

6.1 Peer effects in Norwegian Middle schools

To study the empirical performance of my estimators, I gathered data from two secondary schools in Norway. The schools are public schools whose students come from the local geographic area. Since private schools are not common in Norway, the student body has a very varied background. Some students have highly educated parents with large amounts of resources, while other students have parents without much formal education and relatively low income and wealth. This variation in student background makes the schools ideal for studying how peers background and skills affect learning of students, as measured by their GPA.

A common issue in network data is that individuals are only able to report up to a certain amount of connections, see for example (Griffith, 2022), Oster and Thornton (2012), Cai et al. (2015) and Kandpal and Baylis (2013). This implies a highly complicated distribution of $S_{i,j}$, especially if individuals don't list random subsets of their links. To test the performance of my estimator in this setting, I gathered two sets of friendship data through a survey. I first asked the students which classes they had friends in, and then let them select all their friends in these classes. This questions yield a complete network consisting of all the students friends, where a link between two students is said to exist if both students list each other as friends.

I then asked the students which friends they would have picked if they were only able to pick up to five male and five female friends. Using the responses to this question allows me to construct a sample similar to the kind commonly observed in the literature. I will refer to this as the restricted sample.

The restricted sample of links is constructed as follows. A link is said to exist if both students named each other as one of their top friends. Students who list the maximum allowable male or female friends have missing data on links in classes which they have friends. For classes which they earlier said they have no friends, the researcher knows all links are zero.

A comparison of how many friendships remain in the restricted dataset gives an insight into how the missingness works in these settings. On average students retain 38% of their friendships in the restricted network data. This reduction is

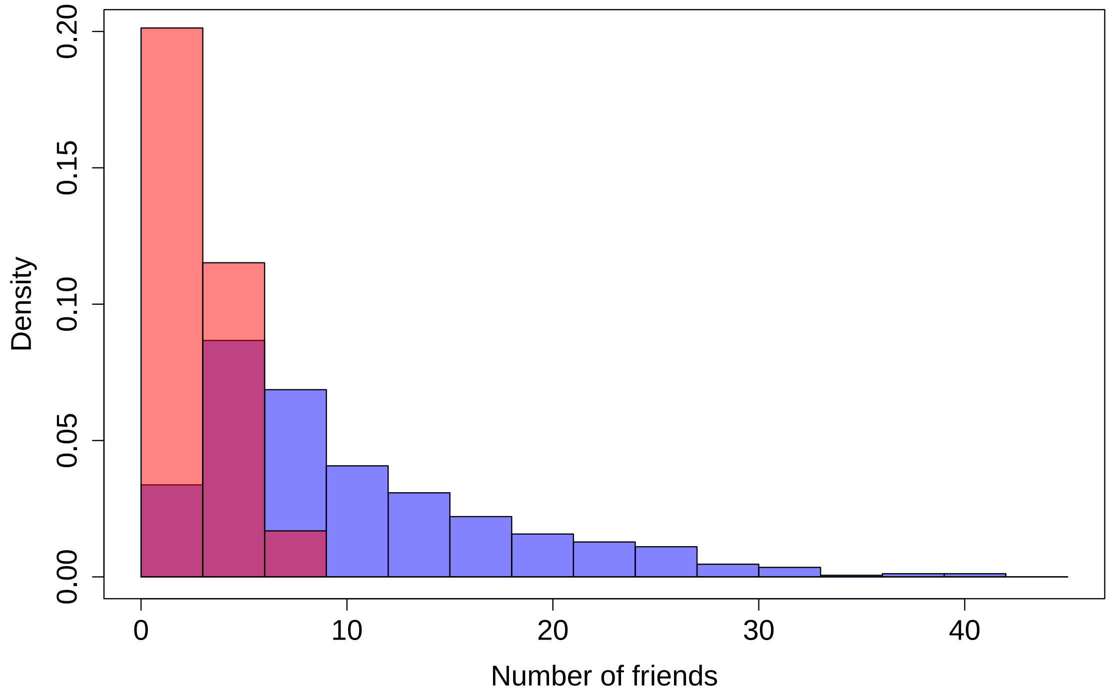


Figure 5: Histogram of the degree distribution of the complete network (in blue) and the restricted one (in red).

not just amongst the most popular students who cannot list all their friends, but also amongst the less popular students who's popular friends no longer list them as a friend. Defining a popular student as one who has more than 10 friendships in the complete network data, these student's retain 25% of their friendships in the restricted dataset, significantly less than the average student. The "unpopular" students still lose over half their friendships (retaining 46%) in the restricted network data, and the numbers are similar for the "very unpopular" students with less than five friends, who lose 55%. Students who have only one friend have a 30% chance of losing their one friend in the restricted sample. This exemplifies how the "censoring" creates missingness for all individuals in the network, not just those who have many friends.

Figure 5 shows the degree distribution for both the complete and restricted network data. As expected, the restricted data is not just a censored version of the degree distribution of the complete network. Instead the degree of all students are pushed towards zero, creating a "squeezed" distribution that appears much more homogenous than the true distribution.

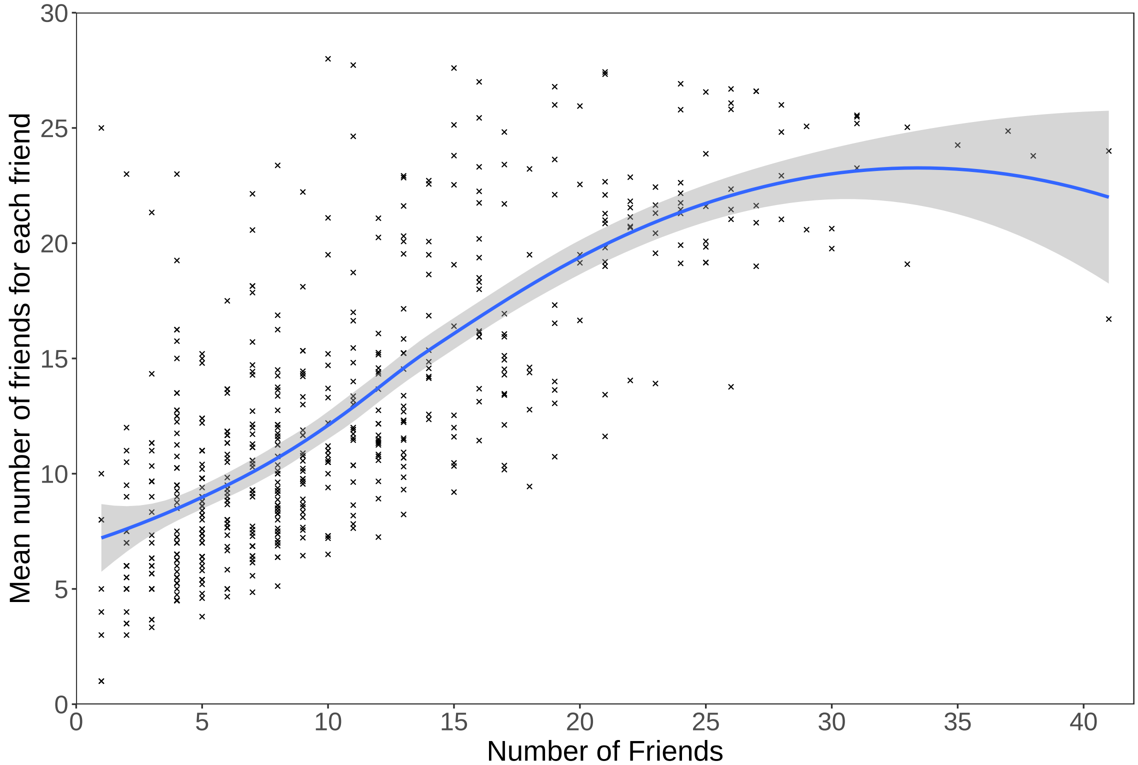
	(1)	(2)	(3)	(4)	(5)	(6)
Distance	-1	-1	-1	-1	-1	-1
$\|\Delta income\|$	0.03	-0.04	0.03	-0.03	-0.06	-0.04
$\|\Delta wealth\|$	-0.02	0.00	-0.06	-0.002	-0.00	-0.00
Same parent educ	0.70	0.14	0.00	0.52	-0.11	-0.20
Same class	21.88	16.7	16.23	17.23	21.81	18.51
Estimator	Logit	Logit	Semi-param.	Logit	Logit	Semi-param.
Sample	Complete	Complete	Complete	Restricted	Restricted	Restricted
V	No	Yes	Yes	No	Yes	Yes

Table 1: Estimates of the network formation model, all estimates are normalized by the absolute value of the coefficient on walking distance between the schools. [Inference in progress]

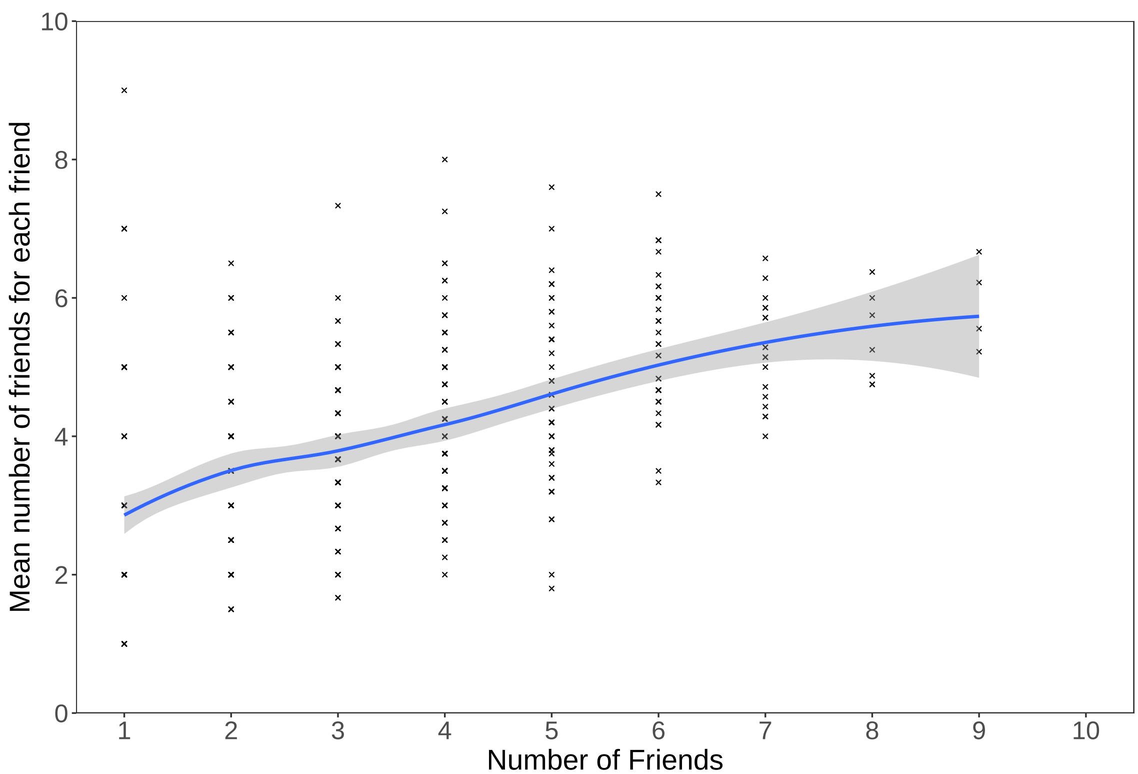
A similar tendency can be seen in Figure 6, which shows the relation a students number of friends and the average number of friends their friends have. We again see that the restricted sample greatly understates the heterogeneity in the data. The most popular students tend to have friends with similarly large number of friends on average, suggesting a large amount of clustering in friendships.

Table 1 shows estimates for a logit model and the semi-parametric estimator on both the complete and restricted network data. All coefficients have been normalized by the absolute value of the distance coefficient, to make the coefficients comparable. Ignoring the social influence parameters V causes parental education to appear to have a large effect, indicating there is significant correlation between parental education and charisma. The semi-parametric estimates for the full network indicate that the logit parameterization of $U_{i,j}$ may be misspecified, but the estimates are broadly similar. When running the estimators on the restricted data, we see that the normal logit estimator overestimates the effect of students being in the same class. The semi-parametric estimator reduces this bias somewhat, though it misses by more on the effect of having parents with the same education.

My object of interest is the peer effects in GPA measured as the average grade of students in classes they have throughout secondary school. These classes are Norwegian, English, Mathematics, Religious studies and Science. This ensures the average is taken over the same classes for different cohorts of students. The model also includes the householde income and the highest obtained education by a household member as covariates, as well as the average value of these covariates for a students friends. Our model therefore includes the effect of parental background on student GPA, as well as peer effects through the GPA and parental background of friends.



(a) Complete network data.



(b) Observed network data

Figure 6: These figures show the relation between the number of friends a student has, and the average number of friends of their friends. Subplot (a) shows the complete network data, while subplot (b) shows the restricted network data.

Estimator	Naive	Naive	SIV
Mean Peer GPA	0.80**	0.61*	0.87
	(0.158)	*(0.152)*	
HH. Income	0.01***	0.01**	0.02
	(0.004)	*(0.005)*	
HH. education	0.29***	0.07***	0.28
	(0.070)	*(0.076)*	
Mean peer HH. income	0.03**	0.01	-0.14
	(0.012)	*(0.008)*	
Mean peer HH. education	-0.5**	-0.25*	-0.5
	(0.179)	*(0.120)*	
Sample	Complete	Restricted	Restricted
N	592	592	592
Note:	*p<0.05; **p<0.01; ***p<0.001		

Table 2: Educational peer effects estimates, using both the complete and restricted network data, and the naive and corrected estimator. HH. Education is a dummy variable that equals 1 if at least one parent has a university degree. HH. income and wealth is measured in 100 000 NOK, and is the average income/wealth of the household over the last five years.

When estimating these peer effects, we may also be worried that there are unobservable characteristics of an individual that correlate with friends abilities. For example, if students unobserved ability correlates with the covariates of their friends, these covariates are not valid instruments. I will therefore instead use the forecast error of the students friends about their own grades as an instrument for peer outcomes. My survey contained questions to the students about their expected grade was in the classes they took. I can therefore construct the forecast errors of each student by taking the average error of each of the classes not used to construct average GPA. This forecast error should be independent of anything in the students information set. As long as the students have full information about their peers underlying ability, these forecast errors are exogenous. This follows from the long established ideas in labor economics of agent information sets and forecast errors, see for example Cunha and Heckman (2007). If the "optimism" of students, measured by the sign and magnitude of their forecast error, correlates with their ability the instruments are valid.

Table 2 shows the results from estimating peer effects using the forecast errors as instruments. The first column is a standard peer effects estimator using the complete network data. The second is the same estimator, but with the restricted network data. Finally the third column shows the results for the SIV using the restricted network data.

The complete network data estimates show a strong endogenous peer effect, and large effects of having highly educated parents on GPA. Interestingly, there are negative effects from having a peer group which on average has highly educated parents. This means that given the ability level of a students friends, it's better if those friends have parents with worse education. This indicates there may be stronger peer effects from friends who have high grades due to their unobserved ability rather than due to having resource rich parents.

As expected the restricted network data leads to misleading results. The estimate of the endogenous peer effect is about 25% smaller than in the complete data, and the other coefficients are similarly distorted. The effect of having had a parent with university education is underestimated, as is the negative contextual effect of having a peer group with highly educated parents. However when we correct for the restricted network data using the SIV estimator, we drastically reduce these biases. The bias in the endogenous peer effect is reduced by 65%, and the biases in the other coefficients are similarly reduced, except for the contextual effect of household income.

The results of this section indicates that the SIV and semi-parametric estimators have the ability to significantly reduce the bias of the parameters even in small samples with complicated patterns of missingness. This indicates that while researchers should be wary of the biases stemming from restricted network data, it's possible to greatly reduce them by estimating a network formation model, and using the SIV estimator. In the next section I will apply the SIV estimator to a publicly available dataset to investigate how results change when we take the missingness of network links into account.

6.2 Peer effects in microfinance take-up in Indian villages

Banerjee et al. (2013) studies how information diffusion affects microfinance take-up in Indian villages. The researchers used census data on 75 villages that a microfinance institution was planning to expand to, and surveyed a random subset of individuals to acquire detailed information about the relationships between households in the villages (the network survey). Six months later, the microfinance institution became active in the villages, and two years later they had spread to 43 of the villages.

This data was then also used in Chandrasekhar and Lewis (2016) who develop corrections for missing data on network links, specifically in terms of regressions of outcomes on network characteristics. They also develop an "analytical correction" for peer effect regressions as mentioned previously. This correction involves doing the peer effect regression only on the set S_N, for whom we arguably observe all links. They construct a link as existing if any member of a household says they're connected to a member of a different household. While this let's us observe all the links each household nominates, it potentially marks some links of the household as not-existing. To see this, consider two households i and j where i is not surveyed and j is surveyed. If i would have claimed to be linked to j, but no member of j claimed to be linked to i, the restricted data would mark $A^S_{i,j} = 0$. However in the counterfactual world where every household was surveyed, by the definition of the paper we would have $A_{i,j} = 1$. This makes it difficult to interpret the data on network links from surveyed to non-surveyed households.

The network survey was randomly given, stratified by religion and geographic sub-location. The public version of the data does not contain geographic location data, preventing me from using it. I will instead use other covariates in the model to approximate the effect of including geo-location, but this will be a limitation for all the analysis of this section. A regression of the indicator of being surveyed on other variables in the data indicate that there is some heterogeneity in survey responses. Even then, the deviations from a completely random sampling strategy seem small, as shown in Table 3 and Figure 7.

However this random assignment does not guarantee that the sampling indicators are uncorrelated with ϵ in a microfinance peer regression. Notably, the outcome was measured long after the survey was given, and the survey's detailed nature might have led to information diffusion effects that affect the outcome.[7] This appears to have happened, as sampled units have an average microfinance take-up of 19.2%, compared to a non-sampled average of 17.5%. This difference is statistically significant and persists even when we condition on covariates. Table 4 shows the results of a regression of microfinance take-up on an indicator for the household being surveyed. As we can see the effect from the sampling is remarkably stable even when we include covariates and fixed effects for villages. This result does not prove that the error term in the peer effect regression would correlate with being surveyed, especially due to the missing geo-location variable, but it is suggestive that there is some effect from a household being sampled.

These results suggests the analytical correction of Chandrasekhar and Lewis (2016) is not a good fit for this setting, as it requires $\epsilon_i \perp\!\!\!\perp S_{i,j}$. However the methods

[7] The authors of Banerjee et al. (2013) were aware of this potential issue, and attempted to reduce the problem by avoiding mentions of the firm and explicitly asking financial questions. Even then, they did ask households who they borrowed money from, which may have primed households to be more aware of their financial situation. Alternatively, they may have attempted to obtain information about why researchers approached them, and succeeded.

	(1)	(2)
Bed nr.	0.004	0.005
	(0.004)	*(0.004)*
General Caste	0.012	-0.029
	(0.02)	*(0.069)*
Minority Caste	0.003	-0.047
	(0.048)	*(0.084)*
OBC Caste	0.017*	-0.027
	(0.013)	*(0.067)*
Schedule Caste	0.032**	-0.013
	(0.015)	*(0.067)*
Schedule Tribe	0.105***	0.063
	(0.029)	*(0.072)*
Has Electricity	0.064***	0.062***
	(0.02)	*(0.021)*
Has Latrine	0.015	0.02*
	(0.013)	*(0.013)*
Hindu	0.023	0.032
	(0.037)	*(0.041)*
Muslim	0.437***	0.439***
	(0.1)	*(0.098)*
Village Leader	0.069***	0.07***
	(0.016)	*(0.016)*
Owns property	-0.018	-0.02
	(0.018)	*(0.018)*
Village FE's.	No	Yes
N	9,594	9,594
Note:	*p<0.1; **p<0.05; ***p<0.01	

Table 3: This table shows the Average Marginal Effects from a logit of an indicator for a household being sampled on economic and cultural covariates, see the text for a full list of variables included.

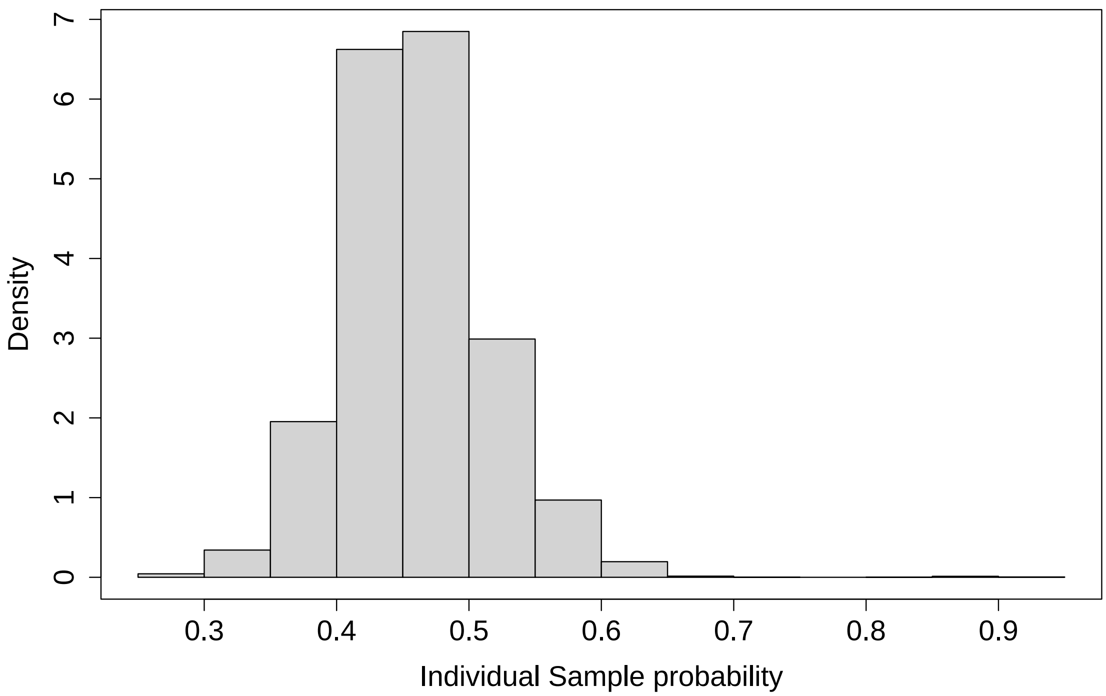


Figure 7: Histogram of the predicted probability of a household being surveyed. As we can see the probabilities are highly centered.

developed in this paper will allow us to obtain consistent estimates of the peer effects.

To estimate the model, I use the network formation model specified in Chandrasekhar and Lewis (2016), excluding the geographic distance measure as it's absent from the public version of the data. This model is the same as the model specified in Section 4, setting $V_i = 0 \; \forall i$ and estimating the model separately for each village. The covariates in the model is the difference in number of rooms, beds, electricity access and roofing materials of the households. To correct for any potential biases in estimation from the sampling process, I use the estimator discussed in section 4.1, setting X^S as the household level covariates in the network formation model as well as the cultural variables in the data set. Our assumption is then that the stratification on the missing geographic variables as well as a household refusing to respond to the survey is independent of the link-specific preference shocks conditional on X^S.

I model household microfinance take up as being driven by the economic situation of the household, measured by information about their house, the religion of the household head, as well as the sum of their connected households who use microfinance. I also include variables for if the household contains a village leader, and

	(1)	(2)	(3)
Household Surveyed	0.017**	0.017**	0.018**
	(0.008)	(0.008)	(0.008)
Village Leader		0.076***	0.076***
		(0.012)	(0.012)
room nr.		−0.007**	−0.002
		(0.004)	(0.004)
Muslim		−0.215*	−0.172
		(0.129)	(0.128)
Hindu		−0.368***	−0.289**
		(0.128)	(0.126)
Has Electricity		−0.011	0.005
		(0.017)	(0.017)
Has Latrine		−0.054***	−0.066***
		(0.010)	(0.010)
Bed nr.		−0.008**	−0.007**
		(0.003)	(0.003)
Constant	0.176***	0.549***	0.489***
	(0.006)	(0.131)	(0.132)
Rooftypes FE's	No	Yes	Yes
Village FE's	No	No	Yes
N	9,122	9,122	9,122
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 4: This table shows the results of a regression of microfinance take up on an indicator for a household being picked for the detailed survey. Columns two and three add covariates and village fixed effects respectively.

the sum of village leaders the household is connected to. As instrument I use the sum of the sum of village leaders your connected households are connected to.

Following Chandrasekhar and Lewis (2016), I interpret the coefficient on the sum of microfinance take-up amongst connected households as a risk sharing and joint-

	(1)	(2)	(3)	(4)	(5)	(6)
Sum Peer Take-up	0.003	-0.012	-0.017^{*}	-0.051^{***}	-0.019	-0.031
	(0.004)	(0.008)	(0.009)	(0.017)		
Village Leader	0.074^{***}	0.051^{***}	0.076^{***}	0.058^{***}	0.077	0.078
	(0.013)	(0.018)	(0.013)	(0.021)		
Sum Peer Leaders			0.016^{***}	0.023^{***}		0.004
			(0.006)	(0.007)		
Analytical Correction	No	Yes	No	Yes	No	No
SIV	No	No	No	No	Yes	Yes
N	9,122	4,413	9,122	4,413	9,122	9,122
Note:			*p<0.1; **p<0.05; ***p<0.01			

Table 5: Estimates of peer effects in microfinance under different specifications and estimators. The instrument for all the estimators is A^2X. All estimators include covariates for the households economic situation (access to electricity, latrine size of house) and the religion of the household head. SIV estimates calculcated with M=1000 [inference for SIV still in development.].

decision making effect amongst households. The coefficient on the sum of village leaders the household is connected to is interpreted as an information diffusion effect, as all village leaders were informed of the microfinance program.

As we can see in table 5, a researcher ignoring the missing network links would conclude that there is no endogenous peer effects, but a positive effect from being connected to one of the leaders who was informed about the microfinance program. This would imply there is some information diffusion effects but that agents do not take their peers decision about microfinance into effect when deciding if they want to join the program.

The SIV estimates, which are robust to the missingness potentially correlating with the ϵ, show that the previous interpretation is driven by biases. The information diffusion effect is very small, implying that the naive estimate is primarily driven by the endogeneity of $S_{i,j}$ and missing link bias. Instead the SIV estimates show that the network effect is primarily through the endogenous peer effect. This means that having a connected household join the program reduces a households probability of joining by 3%. This is an effect of significant magnitude given that the average probability of joining the microfinance program is around 20% in the villages. As such a significant dampener of take-up is households not joining because their connected households joined.

These results show that using estimators robust to non-randomly missing network links has direct implications for how we understand the network effects in the take-up of microfinance. Using naive estimators would make researchers mistakenly focusing on information diffusion effects, when in reality the network affects links through linked households making joint decisions about joining the microfinance program.

7 Conclusion

This paper has developed tools to identify the parameters of peer effect and network formation models with missing links. These tools are flexible enough to fit many empirical situations, yielding either partial or point identification depending on the assumptions the researcher can credibly make.

The robust estimators significantly reduce the missing link bias. By taking advantage of a unique data set containing both complete and restricted data on network links, we see that the bias is reduced by 65% when robust estimators are used. This result is encouraging for the finite sample performance of the estimators.

Partial identification of the peer effects without network formation models shows how much can be learned from the data with canonical assumptions. Simulations show that in order to obtain informative bounds, researchers need to know a lot about the structure of the unobserved parts of the network. It's therefore important for researchers to carefully consider how they gather data on network links, and how their data gathering might affect the processes they observe.

The missing link bias is empirically relevant. Using estimators not robust to the missing data leads to misleading conclusions about what drives economic decision making. A researcher that doesn't take the bias into account would have mistakingly concluded that the network effects of micro-finance diffusion is primarily driven by information diffusion effects, rather than joint decision making.

There are many exciting paths for future research based on the estimators developed in this paper. Allowing for sparse networks when estimating the network formation model would increase the settings in which the model is credible. Based on the results of Graham (2017), it seems likely that the semi-parametric estimator of θ can be consistent even if the network is asymptotically sparse. Note that the SIV estimator is consistent under both dense and sparse networks. Another important point for future research will be to create estimators robust to missing data for more general models of peer effects, to capture the potential non-linearities discussed in Sacerdote (2011).

References

Abrevaya, J. (2000). Rank estimation of a generalized fixed-effects regression model. *Journal of Econometrics*, 95(1):1–23.

Akbarpour, M., Malladi, S., and Saberi, A. (2020). Just a few seeds more: value of network information for diffusion. *Available at SSRN 3062830*.

Atalay, E., Hortacsu, A., Roberts, J., and Syverson, C. (2011). Network structure of production. *Proceedings of the National Academy of Sciences*, 108(13):5199–5202.

Bai, J. (2009). Panel data models with interactive fixed effects. *Econometrica*, 77(4):1229–1279.

Banerjee, A., Chandrasekhar, A. G., Duflo, E., and Jackson, M. O. (2013). The diffusion of microfinance. *Science*, 341(6144):1236498.

Bickel, P., Choi, D., Chang, X., and Zhang, H. (2013). Asymptotic normality of maximum likelihood and its variational approximation for stochastic blockmodels. *The Annals of Statistics*, 41(4):1922–1943.

Bonhomme, S. (2021). Teams: Heterogeneity, sorting, and complementarity. *arXiv preprint arXiv:2102.01802*.

Boozer, M. and Cacciola, S. E. (2001). Inside the'black box'of project star: Estimation of peer effects using experimental data. *Available at SSRN 277009*.

Boucher, V. and Houndetoungan, A. (2020). *Estimating peer effects using partial network data*. Centre de recherche sur les risques les enjeux économiques et les politiques.

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

Bramoullé, Y., Djebbari, H., and Fortin, B. (2020). Peer effects in networks: A survey. *Annual Review of Economics*, 12:603–629.

Breza, E., Chandrasekhar, A. G., McCormick, T. H., and Pan, M. (2020). Using aggregated relational data to feasibly identify network structure without network data. *American Economic Review*, 110(8):2454–84.

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

Candelaria, L. E. (2020). A semiparametric network formation model with unobserved linear heterogeneity. *arXiv preprint arXiv:2007.05403*.

Chandrasekhar, A. and Lewis, R. (2016). Economics of sampled networks. Technical report, Mimeo, Massachusetts Institute of Technology.

Cohen-Cole, E., Liu, X., and Zenou, Y. (2012). Multivariate choice and identification of social interactions. *Available at SSRN 2151801*.

Cunha, F. and Heckman, J. J. (2007). Identifying and estimating the distributions of ex post and ex ante returns to schooling. *Labour Economics*, 14(6):870–893.

De Paula, Á., Richards-Shubik, S., and Tamer, E. (2018). Identifying preferences in networks with bounded degree. *Econometrica*, 86(1):263–288.

Gao, W. Y. (2020). Nonparametric identification in index models of link formation. *Journal of Econometrics*, 215(2):399–413.

Gao, W. Y., Li, M., and Xu, S. (2022). Logical differencing in dyadic network formation models with nontransferable utilities. *Journal of Econometrics*.

Graham, B. S. (2016). Homophily and transitivity in dynamic network formation. Technical report, National Bureau of Economic Research.

Graham, B. S. (2017). An econometric model of network formation with degree heterogeneity. *Econometrica*, 85(4):1033–1063.

Graham, B. S. (2020). Network data. In *Handbook of Econometrics*, volume 7, pages 111–218. Elsevier.

Griffith, A. (2022). Name your friends, but only five? the importance of censoring in peer effects estimates using social network data. *Journal of Labor Economics*, 40(4):779–805.

Han, A. K. (1987). Non-parametric analysis of a generalized regression model: the maximum rank correlation estimator. *Journal of Econometrics*, 35(2-3):303–316.

Horowitz, J. L. (1992). A smoothed maximum score estimator for the binary response model. *Econometrica: journal of the Econometric Society*, pages 505–531.

Imberman, S. A., Kugler, A. D., and Sacerdote, B. I. (2012). Katrina's children: Evidence on the structure of peer effects from hurricane evacuees. *American Economic Review*, 102(5):2048–82.

Johnsson, I. and Moon, H. R. (2021). Estimation of peer effects in endogenous social networks: Control function approach. *Review of Economics and Statistics*, 103(2):328–345.

Kandpal, E. and Baylis, K. (2013). Expanding horizons: Can women's support groups diversify peer networks in rural india? *American Journal of Agricultural Economics*, 95(2):360–367.

Kim, J. S. (2018). Structural estimation of pairwise stable networks: An application to social networks in rural india. *Available at SSRN 2262092*.

Lim, K. (2018). Endogenous production networks and the business cycle. *Work. Pap.*

Manski, C. F. (1975). Maximum score estimation of the stochastic utility model of choice. *Journal of econometrics*, 3(3):205–228.

Manski, C. F. (1985). Semiparametric analysis of discrete response: Asymptotic properties of the maximum score estimator. *Journal of econometrics*, 27(3):313–333.

Manski, C. F. (1987). Semiparametric analysis of random effects linear models from binary panel data. *Econometrica: Journal of the Econometric Society*, pages 357–362.

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *The review of economic studies*, 60(3):531–542.

McEwan, P. J. (2003). Peer effects on student achievement: Evidence from chile. *Economics of education review*, 22(2):131–141.

Menzel, K. (2015). Strategic network formation with many agents. Technical report, Working papers, NYU.

Newey, W. K. and McFadden, D. (1994). Large sample estimation and hypothesis testing. *Handbook of econometrics*, 4:2111–2245.

Oster, E. and Thornton, R. (2012). Determinants of technology adoption: Peer effects in menstrual cup take-up. *Journal of the European Economic Association*, 10(6):1263–1293.

Pakes, A., Porter, J., Ho, K., and Ishii, J. (2015). Moment inequalities and their application. *Econometrica*, 83(1):315–334.

Patacchini, E. and Zenou, Y. (2012). Juvenile delinquency and conformism. *The Journal of Law, Economics, & Organization*, 28(1):1–31.

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, 63(3):581–592.

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

Shi, X., Shum, M., and Song, W. (2018). Estimating semi-parametric panel multinomial choice models using cyclic monotonicity. *Econometrica*, 86(2):737–761.

Shimamoto, D., Todo, Y., Kim, Y. R., and Matous, P. (2022). Identifying and decomposing peer effects on decision-making using a randomized controlled trial. *Empirical Economics*, 63(2):1029–1058.

Stefanski, L. A. and Carroll, R. J. (1985). Covariate measurement error in logistic regression. *The annals of statistics*, pages 1335–1351.

Toth, P. (2017). Semiparametric estimation in network formation models with homophily and degree heterogeneity. *Available at SSRN 2988698*.

Trogdon, J. G., Nonnemaker, J., and Pais, J. (2008). Peer effects in adolescent overweight. *Journal of health economics*, 27(5):1388–1399.

A Further Theoretical Results

A.1 A Random effect approach

Let $V \sim_{iid} F$, then it's straightforward to show that the likelihood of a random effect estimator of the network formation model is

$$L = \int_{\text{supp}(V)^N} \Pi_{i<j} P(A_{i,j}|\theta, v_1, ..., v_N) f_V(v_1|\theta) f_V(v_2|\theta) ... f_V(v_N|\theta) dv_1 dv_2 ... dv_N \tag{7}$$

This likelihood is intractable in most settings due to the combination of the N dimensional integral, as well as the $N(N-1)/2$ dimensional product, neither which factorizes easily.

Suppose we discretize the support of V to R points, then we can instead write the likelihood as

$$\begin{aligned}
\mathcal{L} &= \sum_{v_1} \sum_{v_2} ... \sum_{v_N} \Pi_{i<j} P(A_{i,j}|\theta, v_1, ..., v_N) \pi_V(v_1|\theta) \pi_V(v_2|\theta) ... \pi_V(v_N|\theta) \\
&= \sum_{v_1} \sum_{v_2} ... \sum_{v_N} \Pi_{i<j} P(A_{i,j}|\theta, v_i, v_j) \pi_V(v_1|\theta) \pi_V(v_2|\theta) ... \pi_V(v_N|\theta)
\end{aligned} \tag{8}$$

This is still intractable due to the n-dimensional sum, however it turns out that the form of this likelihood is very similar to the likelihood considered in Bonhomme (2021)'s work on team production[8], suggesting a variational approach might yield useful estimates of the distribution of V.

These methods consist of bypassing the intractable likelihood by instead solving the evidence lower bound, consisting of densities $q_i(v)$ that factors across i, yielding computational tractability.

While these methods have yet to have general proofs of identification, the estimator based on the variational objective 8 has been shown to be equivalent to the full likelihood in the case with fixed R and no covariates in Bickel et al. (2013). Should those results generalize, these methods may be a more tractable way of obtaining estimates of the network formation models going forward.

[8]Specifically we can think of friendships forming (or not forming) being the binary production of a team of two workers

A.2 Estimator 1 when $S \not\perp V$

When S is not independent of the fixed effects V, we cannot do the two step procedure described in the main text. Instead I suggest an iterative procedure which is guaranteed to, eventually, get consistent estimates. The procedure is designed to get around the two main issues that present themselves. The first issue is that since S and V being jointly decided, we need to find a way to feasibly estimate this new expanded problem. The second issue is that for some individuals, the fixed effects are never estimated, complicating the estimation of S.

For this appendix I will stick to the case used in the main text where $S_{i,j} = S_i S_j$, and the sampling being conditionally independent. I will further assume that $P(S_i|W_i, Z_i, V_i)$ follows a logit form, though this is not necessary for the argument.

I will get around the first issue by an iterative procedure, similar to that used by Bai (2009). To overcome the second issue, I will use the distribution of the estimated fixed effect to integrate out the unobserved parts, using assumption 11.2.

Begin with a guess for the fixed effects $\tilde{V}_0$, implying a distribution F_μ. We can then estimate a logit model for S, integrating out V for individuals with V unobserved. We can then predict $P(S|W, Z, V)$ from this logit model. Note that since the individuals we observe links for are exactly the ones we need the predicted $P(S|W, Z, V)$'s for, the lack of some fixed effects is not an issue in this step.

Conditional on our estimated $P(S|W, Z, V)$'s we can then estimate the re-weighted likelihood as before. We then take these new estimates, call them $\tilde{V}_1$ and repeat the previous step, and continue until $\tilde{V}_{t-1} = \tilde{V}_t$. Consistency follows from

1. The true V and $P(S|W, V, Z)$ are a fixed point of the algorithm

2. It is the asymptotic global minimum as conditional on having the right $P(S|W, V, Z)$, the re-weighted likelihood is minimized at θ_0, V_0, and vice versa.

A.3 Relaxing the network exogeneity assumption

Following Johnsson and Moon (2021), we could allow for some endogeneity of the network by allowing V_i to have a direct effect on y_i. A complication in this approach with missing data is that we may not be able to estimate V_i for every individual. A researcher with sufficient links observed for every individual to estimate V_i could however directly apply the estimator of Johnsson and Moon (2021), as we estimate V_i directly when we estimate the network formation model.

A.4 Proof of sharpness and validity of bounds

When we have a linear in means model, define the bounds by the solutions to the program

$$\begin{aligned}
\max_{t_i>0, d_i \in (d_i^S,\ldots,N), A_{i,j}\in\{0,1\}^M} \sum_{i=1}^{N}\sum_{j=1}^{N} t_i c_{i,j} A_{i,j} \qquad & s.t. \\
t_i d_i = 1 \qquad & \forall i \\
d_i = \sum_{j=1}^{N} A_{i,j} \qquad & \forall i \\
A_{j,i} = A_{i,j} \qquad & \forall (i,j) \\
A_{i,j} \in \mathcal{A} \qquad &
\end{aligned}$$

and the corresponding minimum of the same program.

Validity of the bounds in both cases follow directly from the true network $A_0 \in \mathcal{A}$.

Let's first define the identified set of the model. The unobservables are the parameters α, β and the unobserved parts of the network. Define this unobserved set, with some abuse of notation, as $\mathcal{A}$. The sharp identified set is the values of α, β and $\mathcal{A}$ that are consistent with the observed data and satisfies the Assumptions of the model. Those assumptions are Assumptions 1, 2 and 12. Define the set of such parameters and unobserved networks as Θ. For the program to generate sharp bounds, we must therefore show that for every $A \in \mathcal{A}$, A and it's implied parameters are in Θ

Let $\text{sign}(b_{min}) = \text{sign}(b_{max})$ and pick some $A_1 \in \mathcal{A}$ that satisfies $A_{j,i} = A_{i,j}$ (symmetry). Define d_i and t_i as above, then

$$\text{Cov}((G_1 y)_i, Z_i^S) = \sum_{i=1}^{N}\sum_{j=1}^{N} t_{i,1} c_{i,j} A_{i,j}$$

Consider the IV-estimate of α using Z_i^S if A_1 was the true network. This would be a consistent estimator of α by Assumptions 1, 2 and 12. The IV-estimator would be

$$\hat{\alpha}_{A_1} = \frac{\text{Cov}(y_i, Z_i^S)}{\text{Cov}(Z_i^S, (G_1 y)_i)} = \frac{\text{Cov}(y_i, Z_i^S)}{\text{Cov}(Z_i^S, (G^S y)_i)} \frac{\text{Cov}(Z_i^S, (G^S y)_i)}{\text{Cov}(Z_i^S, (G_1 y)_i)} = \hat{\alpha}^S \frac{\text{Cov}(Z_i^S, (G^S y)_i)}{\text{Cov}(Z_i^S, (G_1 y)_i)}$$

That is the value of the bound as defined in Proposition 7. We can similarly generate values for β that are uniquely consistent with A_1, α. By the definition of our model, this set of $A_1, \alpha_{A_1}, \beta(A_1, \alpha_{A_1})$ are consistent with the distribution of the data, implying

$$A_1, \alpha_{A_1}, \beta(A_1, \alpha_{A_1}) \in \Theta$$

Since this holds for any $A_1 \in \mathcal{A}$, it must also hold for A_{max}, A_{min} corresponding to b_{max}, b_{min}. Therefore $\alpha_{A_{max}}, \alpha_{A_{min}} \in \Theta$, so our bounds give us the sharp identified set.

In the case where $\text{sign}(b_{min}) \neq \text{sign}(b_{max})$, we cannot reject the possibility that there is, for some small positive number ϵ two networks $(A_-, A_+) \in \mathcal{A}$ such that $\text{Cov}(Z_i^S, (G_- y)_i) = -\epsilon$ and $\text{Cov}(Z_i^S, (G_+ y)_i) = \epsilon$. This in turn implies $\alpha_{A_+} = -\infty, \alpha_{A_-} = \infty$. These values are not in the identified set as they break Assumption 1, but ϵ could be picked such that $\alpha_{A_+} = 1 - \delta, \alpha_{A_-} = -1 + \delta$ for some small positive δ. These would be in the identified set, so the identified set is simply $\alpha \in (-1, 1)$.[9]

A.5 Proof of Proposition 1

Proof. Consider first the case of just-identified IV. I show the case for the linear-in-means model, all results go through with minor modifications for the linear in sums model as G is a deterministic function of A.

To simplify notation, I will start by considering the case where $M = 1$ and supress the ^{sim} superscript. Inserting for $y_i = X_i\beta + \alpha(G_2 y)_i + \gamma(G_2 X)_i + \epsilon_i + \alpha((G_0 y)_i - (G_2 y)_i) + \gamma((G_0 X)_i - (G_0 X)_i)$ in the expression for the estimator we get (ignoring m subscripts)

$$\begin{aligned}
\hat{\xi}^{SIV} &= (\frac{1}{N}\sum_{i=1}^{N} Z_i' T_i)^{-1}(\frac{1}{N}\sum_{i=1}^{N} Z_i' T_i)\xi \\
&+ (\frac{1}{N}\sum_{i=1}^{N} Z_i' T_i)^{-1}\frac{1}{N}\sum_{i=1}^{N} Z_i' \left(\epsilon_i + \alpha((G_0 y)_i - (G_2 y)_i) + \gamma((G_0 X)_i - (G_2 X)_i)\right) \\
&= \xi + (\frac{1}{N}\sum_{i=1}^{N} Z_i' T_i)^{-1}\frac{1}{N}\sum_{i=1}^{N} Z_i' \left(\alpha((G_0 y)_i - (G_2 y)_i) + \gamma((G_0 X)_i - (G_2 X)_i)\right) \\
&+ (\frac{1}{N}\sum_{i=1}^{N} Z_i' T_i)^{-1}\frac{1}{N}\sum_{i} Z_i \epsilon_i
\end{aligned}$$

To obtain identification we need the probability limit of the second and third term

[9] This argument ignores that the objective of the program is discrete, but due to the high dimensionality of the missing links, these objects will always exist unless the number of missing links are very small.

to be zero.

To show this, it's easies to write out the empirical equivalent to the expectation and do some re-ordering. We want to show that $\text{plim}_{N\to\infty}\sum_{i=1}^N Z_i\left((G_0y)_i-(G_2y)_i\right)=0$. We can write out the first part as

$$
\begin{aligned}
&= \text{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^N\sum_{k=1}^N Z_i y_k(\frac{1}{d_{i,2}}A_{ik,0}) \\
&= \text{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^N Z_i\frac{N}{d_{i,2}}\frac{1}{N}\sum_{k=1}^N y_k A_{ik,0} \\
&= \text{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^N Z_i\frac{1}{\hat{\mathbb{P}}_i(A_{i,j}=1)}\frac{1}{N}\sum_{k=1}^N y_k A_{ik,0} \\
&= \text{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^N Z_i\frac{1}{\hat{\mathbb{P}}_i(A_{i,j}=1)}\hat{\mathbb{E}}[y_k A_{ik,0}] \\
&= \mathbb{E}\left[Z_i\frac{1}{\mathbb{P}_i(A_{i,j}=1)}\mathbb{E}[y_k|A_{ik,0}=1]\mathbb{P}(A_{ik,0}=1)\right] \\
&= \mathbb{E}[Z_i\mathbb{E}[y_k|A_{ik,0}=1]]
\end{aligned}
$$

And similarly for the second term, putting them together gives

$$
\begin{aligned}
&\mathbb{E}\left[Z_i\mathbb{E}[y_k|A_{ik,0}=1]\right]-\mathbb{E}\left[Z_i\mathbb{E}[y_k|A_{ik,2}=1]\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[Z_i\mathbb{E}[y_k|A_{ik,0}=1]-Z_i\mathbb{E}[y_k|A_{ik,2}=1]|A_1,X,V,W\right]\right] \\
&= \mathbb{E}\left[Z_i\mathbb{E}\left[\mathbb{E}[y_k|A_{ik,0}=1]-\mathbb{E}[y_k|A_{ik,2}=1]|A_1,X,V,W\right]\right] \\
&= \mathbb{E}\left[Z_i\mathbb{E}\left[\mathbb{E}[y_k|A_{ik,0}=1]-\mathbb{E}[y_k|A_{ik,2}=1]|X,V,W\right]\right] \\
&= 0
\end{aligned}
$$

The first equality used iterated expectations, the second uses Assumption 4 which implies that $A_1 \perp\!\!\!\perp (A_2,A_0)|X,V,W$, and $A_1 \perp\!\!\!\perp y|X,V,W$. The final equality follows from the law of total expectation.

A similar simplified argument goes through for the contextual effects, simplified by the exogeneity of the covariates

$$
\begin{aligned}
\mathbb{E}[Z_i(G_0X-G_2X)] &= \mathbb{E}[Z_i\mathbb{E}[G_0X-G_2X|A_1,W,V,X]] \\
&= \mathbb{E}[Z_i\left(\mathbb{E}[G_0|X,W,V]-\mathbb{E}[G_2|X,W,V]\right)X]=0
\end{aligned}
$$

Finally the probability limit third term can be written out to

$$\begin{aligned}\text{plim}_{N\to\infty}(\frac{1}{N}\sum_{i=1}^{N} Z_i'T_i)^{-1}\frac{1}{N}\sum_{i} Z_i\epsilon_i &= \left(\mathbb{E}[Z_i'T_i]\right)^{-1}\mathbb{E}[Z_i\epsilon_i] \\ &= \left(\mathbb{E}[Z_i'T_i]\right)^{-1}\mathbb{E}[Z_i\mathbb{E}[\epsilon_i|Z_i]] = 0\end{aligned}$$

By Assumption 2.

However there is no guarantee that $\mathbb{E}[Z_{i,m}y_i]$, or $\mathbb{E}[Z_{i,m}T_{i,m}]$ is invertible for the draw $m = 1$. Returning to the case with $M \neq 1$, note that the sum over M sets of draws is a numerical integral over the distribution of the network. Assumption 4 and letting $M \to \infty$ then implies

$$\begin{aligned}\frac{1}{M}\sum_{m\in(1,\ldots,M)} Z_{i,m}T_{i,m} &= \sum_{(A_1,A_2)\in\text{supp}(A)} Z_{i,m}T_{i,m}P(A_1|W)P(A_2|W) := \mathbb{E}_A[Z_{i,m}T_{i,m}|W] \\ \frac{1}{M}\sum_{m\in(1,\ldots,M)} Z_{i,m}y_i &= \sum_{A\in\text{supp}(A)} Z_{i,m}y_iP(A_1|W) := \mathbb{E}_A[Z_{i,m}y_i|W]\end{aligned}$$

Where $\mathbb{E}_A$ is understood to be the expectation over $P(A|W)$. We can then use Assumption 5 to use the WLLN, yielding

$$\begin{aligned}\text{plim}_{N\to\infty}\hat{\xi}^{SIV} &= \left(\mathbb{E}[\mathbb{E}[Z_{i,m}'T_{i,m}|W]]\right)^{-1}\mathbb{E}[\mathbb{E}[Z_{i,m}'y_i|W] \\ &= \left(\mathbb{E}[Z_{i,m}'T_{i,m}]\right)^{-1}\mathbb{E}[Z_{i,m}'y_i]\end{aligned}$$

These moments all exist and $\mathbb{E}[Z_i'T_i]$ is guaranteed to be invertible by Assumption 5.

All together these results yield that

$$\text{plim}_{N\to\infty}\hat{\xi}^{SIV} = \xi$$

As desired.

In over-identified IV, the researcher will either need a third draw from the network (A_3) to compute the first stage coefficient, or alternatively first compute the first stage coefficient as

$$\hat{\delta} = \left(\frac{1}{M}\sum_{m=1}^{M}\frac{1}{N}\sum_{i=1}^{N} Z_{i,m}Z_{i,m}\right)^{-1}\left(\frac{1}{M}\sum_{m=1}^{M}\sum_{i=1}^{N} Z_{i,m}T_{i,m}\right)$$

and then performing the procedure for SIV replacing Z with $\hat{\delta}Z$, yielding a “simulated two stage least squares" (STSLS) estimator. This extra step is necessary to avoid the correlation between the predicted covariates and the error terms $(A_2 y - A_0 y)$, $(A_2 X - A_2 X)$ when the same draw is used in the first stage and to construct the covariates. With this correction the proof of identification follows exactly as above. □

A.6 Proofs for likelihood estimators

We start by proving Proposition 2. The following lemma will be useful

Lemma 2. $L_n^S \xrightarrow{p} \mathbb{E}\left[\log p_\theta(A_{i,j}|W_{i,j}, V)P(S_{i,j} = 1|A_{i,j}, W_{i,j}, V_0)|W_{i,j}, V_0\right]$

Proof. Using the arguments of Lemma 3 in the supplementary materials of Graham (2017) together with Assumption 3 yields

$$\begin{aligned}
L_n^S &\xrightarrow{p} \mathbb{E}[S_{i,j} \log p_\theta(A_{i,j}^S, W_{i,j}^S, \hat{V})|W_{i,j}, V_0] \\
&= \mathbb{E}[\mathbb{E}[S_{i,j} \log p_\theta(A_{i,j}^S, W_{i,j}^S, \hat{V})|W_{i,j}, V, A_{i,j}]|W_{i,j}, V_0] \\
&= \mathbb{E}[\mathbb{E}[S_{i,j}|W_{i,j}, V_0, A_{i,j}] \log p_\theta(A_{i,j}^S, W_{i,j}^S, \hat{V})|W_{i,j}, V_0] \\
&= \mathbb{E}\left[\log p_\theta(A_{i,j}|W_{i,j}, \hat{V})P(S_{i,j} = 1|A_{i,j}, W_{i,j}, V_0)|W_{i,j}, V_0\right]
\end{aligned}$$

□

We can now proceed to the propositions in the text, restated below for ease of reading.

Proposition (Potential inconsistency of sampled likelihood)**.** *Let assumption 6 hold. Then assumption 7 is sufficient for the consistency of $(\hat{\theta}^S, \hat{V}^S)$, and if assumption 6 does not hold, there exists counterexamples where $\hat{V}, \hat{\theta}$ are inconsistent.*

Proof. By lemma 2, we have under Assumption 8 we can write $\mathbb{P}(S_{i,j} = 1|A_{i,j}, W_{i,j}, V_0) = \mathbb{P}(S_{i,j}|W_{i,j}, V_0)$. This in turn yields

$$L^S(\theta) = \text{plim}_{n\to\infty} L_n^S = \mathbb{E}\left[\log p_\theta(A_{i,j}|W_{i,j}, V_0)P(S_{i,j} = 1|W_{i,j}, V_0)\right]$$

Note that for every value of $W_{i,j}, V_0, \theta_0$ is the unique maximizer of $\sum_{y\in\mathcal{Y}} \log p_\theta(y|x)g(y|x)$ by Assumption 6. As such $\theta_0 = \arg\max_{\theta\in\Theta} L^S(\theta)$

To construct a counterexample, simply set $S_{i,j} = \mathbf{1}\{U_{i,j} > 0\}$. The likelihood is misspecified, resulting in $\hat{\theta}_n^S \overset{p}{\not\to} \theta_0$. □

Proposition (Consistency of corrected likelihood)**.** *Assume assumption 8 and 9 hold, then*

$$(\hat{\theta}_n^C, \hat{V}_n^C) = \arg\max_{b,v} l_n^C(b, v) \xrightarrow{p} (\theta_0, V)$$

Proof. Define $w_{i,j} = \frac{1}{P(S_{i,j}=1|W_{i,j},V_0,X_i^S,X_j^S)}$. By similar steps as in lemma 2, we have

$$
\begin{aligned}
L_n^C &\xrightarrow{p} \mathbb{E}\left[w_{i,j}\log p_\theta(A_{i,j}|W_{i,j},V)P(S_{i,j}=1|A_{i,j},W_{i,j},V_0,X_i^S,X_j^S)|W_{i,j},V_0\right] \\
&= \mathbb{E}\left[w_{i,j}\log p_\theta(A_{i,j}|W_{i,j},V)P(S_{i,j}=1|W_{i,j},V_0,X_i^S,X_j^S)|W_{i,j},V_0\right] \\
&= \mathbb{E}\left[\log p_\theta(A_{i,j}|W_{i,j},V)|W_{i,j},V_0\right]
\end{aligned}
$$

Assumption 6 then allows us to apply Theorem 2 and 3 from Graham (2017) to obtain consistency of θ, V. Intuitively, the re-weighted estimator behaves asymptotically like an estimator based on a randomly drawn sample, which is exactly the setting of Graham (2017), allowing us to apply his results to obtain consistency.

□

A.7 Proofs for section 4.2

We can now prove the consistency of the semi-parametric estimator (Proposition 4). This closely follows the arguments made in Han (1987) and Abrevaya (2000). It also uses results from Manski (1987) and Newey and McFadden (1994).

Proof. The proof proceeds by checking the conditions to be able to use theorem 2.1 of Newey and McFadden (1994). The conditions are

1. $S(\theta)$ is uniquely maximized at θ_0.
2. Θ is compact.
3. $S(\theta)$ is continuous.
4. $S_N(\theta)$ converges uniformly almost surely to $S(\theta)$.

Condition 2 holds by assumption, and condition 3 holds by the same argument as Lemma 5 of Manski (1985). Condition 4 follows from Han (1987) proof of uniform convergence for monotonic semi-parametric estimators and inspection. (Proof of theorem 1, step 2.) See also Abrevaya (2000) for a similar argument.

To show that condition 1 holds, consider the expectation of the first type of contribution (s^1), the other cases follow the exact same argument with the indices switched around as needed.

$$\begin{aligned}\mathbb{E}[s^1_{i,j,k,l}(\theta)] &= \mathbb{E}[\mathbf{1}\{\Delta A_{i,k,l} \geq \Delta A_{j,k,l}\}\mathbf{1}\{\Delta W_i\theta > \Delta W_j\theta\}] \\ &+ \mathbb{E}[\mathbf{1}\{\Delta A_{i,k,l} \leq \Delta A_{j,k,l}\}\mathbf{1}\{\Delta W_{i,k,l}\theta < \Delta W_{j,k,l}\theta\}] \\ &= \mathbb{E}[\mathbb{E}[\mathbf{1}\{\Delta A_{i,k,l} \geq \Delta A_{j,k,l}\}|W]\mathbf{1}\{\Delta W_{i,k,l}\theta > \Delta W_{j,k,l}\theta\}] \\ &+ \mathbb{E}[\mathbb{E}[\mathbf{1}\{\Delta A_{i,k,l} \leq \Delta A_{j,k,l}\}|W]\mathbf{1}\{\Delta W_{i,k,l}\theta < \Delta W_{j,k,l}\theta\}] \\ &= \mathbb{E}\big[\mathbb{P}(\Delta A_{i,k,l} \geq \Delta A_{j,k,l}|W)\mathbf{1}\{\Delta W_{i,k,l}\theta > \Delta W_{j,k,l}\theta\} \\ &+ \mathbb{P}(\Delta A_{i,k,l} \leq \Delta A_{j,k,l}|W)\mathbf{1}\{\Delta W_{i,k,l}\theta < \Delta W_{j,k,l}\theta\}\big]\end{aligned}$$

Our estimate $\hat{\theta}$ is the value that maximizes this expectation. To show this is a unique maximum, pick $\tilde{\theta} \in D_\delta(\theta_0) = \{\theta : \theta \in \Theta, ||\theta - \theta_0|| < \delta\}$ s.t. $\tilde{\theta} \neq \theta_0$. By assumption 10 there must exist values $W_1, W_2 \in \mathbb{R}^K$ such that

$$\Delta W_1\tilde{\theta} > \Delta W_2\tilde{\theta} \quad \text{and} \quad \Delta W_1\theta_0 < \Delta W_2\theta_0$$

We can then define the sets B_1, B_2 as the corresponding values of ΔW for which this holds, and by Assumption 10.4b it's possble to find such sets with positive density, letting us write

$$\begin{aligned}&S(\theta_0) - S(\tilde{\theta}) \\ &= \mathbb{E}\left[\mathbb{P}(\Delta A_{i,k,l} \geq \Delta A_{j,k,l}|W)\mathbf{1}\{\Delta W_i\theta > \Delta W_j\theta\} + \mathbb{P}(\Delta A_{i,k,l} \leq \Delta A_{j,k,l}|W)\mathbf{1}\{\Delta W_i\theta < \Delta W_j\theta\}\right] \\ &- \mathbb{E}\left[\mathbb{P}(\Delta A_{i,k,l} \geq \Delta A_{j,k,l}|W)\mathbf{1}\{\Delta W_i\tilde{\theta} > \Delta W_j\tilde{\theta}\} + \mathbb{P}(\Delta A_{i,k,l} \leq \Delta A_{j,k,l}|W)\mathbf{1}\{\Delta W_i\tilde{\theta} < \Delta W_j\tilde{\theta}\}\right] \\ &\geq \mathbb{E}\Bigg[\mathbf{1}\{\Delta W_1 \in B_1\}\mathbf{1}\{\Delta W_2 \in B_2\}\mathbf{1}\{\theta_0\Delta W_1 < \theta_0\Delta W_2\} \\ &\qquad (\mathbb{P}(\Delta A_{i,k,l} \leq \Delta A_{j,k,l}|W) - \mathbb{P}(\Delta A_{i,k,l} \geq \Delta A_{j,k,l}|W))\}\Bigg] \\ &> 0\end{aligned}$$

Where the final inequality follows from lemma 1. Intuitively, in the limit any deviation from θ_0 will lead us to misclassify at least some parts of the support of W. And since the $W\theta$ is continuous, no matter how small the deviation there will always be some part with positive density that is misclassified. Since this holds for each of the s^j's, it also holds for the sum. Furthermore, since it holds for any $\delta > 0$, this proves $S(\theta)$ is uniquely maximized at θ_0. Since all the conditions now hold, we have consistency by theorem 2.1 of Newey and McFadden (1994).

The same argument yields consistency for $\widehat{\Delta V}_k$ from Proposition 5, following the observation that when $\hat{\theta}$ is consistent, $W_{i,k}\hat{\theta} - W_{i,0}\hat{\theta} \xrightarrow{p} (W_{i,k} - W_{i,0})\theta_0$.

□

A.8 Smoothing the semi-parametric estimators

As is common for estimators like my semi-parametric estimators, I will use a smoothed version which is both easier computationally and tends to have better asymptotic properties (Horowitz, 1992). Define a function $K(\cdot)$ that satisfies

1. $\lim_{x\to-\infty} K(x) = 0$, $\lim_{x\to\infty} K(x) = 1$.
2. $|K(x)| < M$ for some finite M for all $x \in (-\infty, \infty)$.

Then we can define

$$\tilde{s}^1_{i,j,k,l}(\theta) = \mathbf{1}\{A_{i,k} - A_{i,l} - (A_{j,k} - A_{j,l}) \geq 0\}K\left(\left((W_{i,k} - W_{i,l} - (W_{j,k} - W_{j,l}))\theta\right)\sigma_N^{-1}\right)$$

with $\tilde{s}^2, \tilde{s}^3$ defined equivalently. This gives the smoothed objective

$$\tilde{S}_N(\theta) = \frac{1}{3\binom{N}{4}} \sum_{(i,j,k,l)\in P} \left(\tilde{s}^1_{i,j,k,l}(\theta) + \tilde{s}^2_{i,j,k,l}(\theta) + \tilde{s}^3_{i,j,k,l}(\theta)\right)$$

Define $\tilde{\theta}$ as a solution to $\max_{\theta\in\Theta} \tilde{S}_N(\theta)$. Then we have

Proposition 8. *Assume Assumption 10 holds and let* $\lim_{N\to\infty} \sigma_N = 0$ *then* $\lim_{N\to\infty} \tilde{\theta} = \theta_0$ *almost surely.*

Proof. Follows directly by Theorem 1 in Horowitz (1992). □

A.9 Recovering the scale conditions

Consider the following Assumptions, as stated in Stefanski and Carroll (1985).

Define $n_T = |T_N|$, $n = \sum_{i,j} S_{i,j}$, and

$$\hat{r}_{i,j} = W_{i,j}\hat{\theta}_n + \hat{V}_i + \hat{V}_j$$
$$r_{i,j} = W_{i,j}\theta + V_i + V_j$$

Note that by Proposition 4 and 5, we can write $\hat{r}_{i,j} = r_{i,j} + \lambda_n e_{i,j}$, where $e_{i,j}$ is i.i.d. and $\lambda_n \to 0$ as $n \to \infty$.

We now want to apply theorem 5.1 of Stefanski and Carroll (1985) to obtain consistency of $\hat{\sigma}$. To do this we need the following conditions

Assumption 13 (Regularity conditions for scale recovery)**.** *Assumptions C.*

1. $\sum_{i=1}^{N}\sum_{j<i}(||r_{i,j}||)^2 = o(n^2)$

2. $\mathbb{E}[||e_{i,j}||] < \infty$.

Proposition 9. *Let Assumption 6 and 13.C hold, then* $\hat{\sigma} \xrightarrow{p} \sigma$.

Proof. Follows directly from Theorem 5.1 of Stefanski and Carroll (1985). □

With an expression for λ_n we could utilize theorem 1 of Stefanski and Carroll (1985) to correct for finite sample biases in σ.

A.10 Proof of Proposition 6

Note that we already have $\hat{\theta} \xrightarrow{p} \theta, \hat{\Delta V} \xrightarrow{p} \Delta V, \hat{V}_0 \xrightarrow{p} V_0$. By the continuous mapping theorem this then means

$$W_{i,j}\hat{\theta} + \hat{V}_i + \hat{V}_j \xrightarrow{p} W_{i,j}\theta + V_i + V_j$$

$(\hat{\sigma}_1, \hat{\sigma}_2) \xrightarrow{p} \sigma$ by the arguments in Stefanski and Carroll (1985), see Appendix A.9.

All we need then is to note that convergence in probability implies weak convergence, and by definition of weak convergence the expectation of any continuous and bounded function also converges.

The integrals in step 4 of algorithm 1 are bounded between 0 and 1, and continuous in V, implying that the desired probabilities converge by the definition of weak convergence.

A.11 Extrapolating charisma

If our object of interest is the network formation model itself, estimates of V and β are sufficient. However, in order to use our estimates to obtain the probabilities $P(A_{i,j} = 1)$, we need to extrapolate V to agents with no observed links. To do this I will make broad assumptions on how the charisma parameters are distributed, and how this distribution depends on $S_{i,j}$.

Let $\tilde{S}_0$ denote the set of individuals for whom we cannot estimate V, and $\tilde{S}_1$ denote the set for whom we can. I will then assume the following relation between the distribution of $V|\tilde{S}_0$ and $V|\tilde{S}_1$.

Assumption 14. *Define $f, \tilde{f}$ as the distributions of $V|\tilde{S}_0$, $V|\tilde{S}_1$ respectively. Assume there is a distribution h_μ s.t. $f = h_\mu, \tilde{f} = h_{\tilde{\mu}}$, where μ is the mean of h, which may vary across i through their covariates X_i.*

This will allow us to use the distribution of the estimated V's, to integrate out the values of the unobserved V's in $P(A_{i,j})$.

An implication of assumption 14 is that our semi-parametric estimates of ΔV_i can be written as

$$\Delta V_i = V_i - V_0 = -V_0 + \mu(X_i) + \varepsilon_i$$

Where $\varepsilon_i \sim h_0$. This let's us recover V_0, which we don't obtain from the RLD estimator, as $-\big(\mathbb{E}[\Delta V_i] - \mathbb{E}[\mu(X_i)]\big)$.[10]

More importantly though it will allow us to make statements about the probability of links forming between individuals with unknown charisma. I will detail my approach to this in algorithm 1.

A.12 GMM approach

Some researchers may prefer a GMM approach to estimating the peer effects. This section introduces an estimator and gives proof of consistency and normality, following Newey and McFadden (1994). Suppose we draw two networks A_1, A_2 from $P(A|W)$, and calculate

$$m_i(\xi, A_m^1, A_m^2) = (Z_i(A_1^m, X))'(y_i - \mathbb{E}[y_i|A_2^m, X; \xi])$$

Where $\xi = (\alpha, \beta, \gamma)$ and $\mathbb{E}[y_i|A_2^m, X; \xi]$ is calculated from the peer effect regression. We can then estimate the parameters from

$$\hat{\xi}^{SGMM} = \arg\min_{\xi} \frac{1}{M} \sum_{m \in (1,\dots,M)} \left(\frac{1}{N} \sum_i m_i(\xi, A_m^1, A_m^2) \right) W \left(\frac{1}{N} \sum_i m_i(\xi, A_m^1, A_m^2) \right)$$

where $\mathbb{E}[y_i|A_2^m, X] = (\mathbf{I} - \alpha A_2)^{-1}(X\beta + \gamma A_2 X)$ is the predicted value of y given covariates X and network A_2^m, and $Z_i(A_1^m, X)$ is the vector of instruments given network A_1^m and X. I will make the following set of assumptions

[10] Note that we cannot distinguish between a constant term in $\mu(X)$ and V_0. It is in general impossible to distinguish between a constant shift in the mean of V and the constant term in 2. Recovering the true level parameter requires the extra step discussed in section 4.3.

Assumption 15.

1. *X_i, ϵ_i are i.i.d.*

2. *There is only one value $\tilde{\xi}$ s.t. $\mathbb{E}[m_i(\tilde{\xi}, A^1, A^2)] = 0$.*

3. *$\xi \in \Xi$, where Ξ is compact*

4. *$\mathbb{E}[\sup_{\xi\in\Xi} m(\xi, A_1, A_2, X)] < \infty$.*

These assumptions are all regular in the non-linear GMM literature. Part 2 of the assumption is the strongest, and stems from the difficulty of proving global identification of non-linear GMM estimators. A condition for local identification would be $\mathbb{E}[Z\nabla_x i(y_i - \mathbb{E}[y_i|A_2, X, \xi])]$ having rank equal to it's columns, which will hold for most sets of instruments.

Proposition 10. *Let Assumptions 1, 2, 4, 5.2 and 15 hold, then as $N \to \infty$ we have $\hat{\xi}^{SGMM} \xrightarrow{p} \xi_0$*

Proof. The result follows from theorem 2.6 in Newey and McFadden (1994). The proof proceeds by verifying the four conditions given in the theorem. First, we check that $\mathbb{E}[m_i(\xi, A_m^1, A_m^2)] = 0$ when $\xi = \xi_0$. Defining the draw of A that generated the observed data as A_0, write

$$y_i = (I - \alpha A_0)^{-1}(X_i\beta + (AX)_i) + ((I - \alpha A_0)^{-1}\epsilon)_i = \mathbb{E}[y_i|A_0, X, \xi_0] + \tilde{\epsilon}_i$$

Inserting this into m_i we get

$$\begin{aligned} m_i(\xi, A_1^m, A_2^m) &= Z_i(A_1^m, X)(\mathbb{E}[y_i|A_0, X, \xi_0] - \mathbb{E}[y_i|A_2^m, X, \xi] + \tilde{\epsilon}_i) \\ &= Z_i(A_1^m, X)(\mathbb{E}[y_i|A_0, X, \xi_0] - \mathbb{E}[y_i|A_2^m, X, \xi]) + Z_i(A_1^m, X)\tilde{\epsilon}_i \end{aligned}$$

The first part of this expression stems from us not knowing A_0, and therefore being unable to construct the true conditional expectation of y. The second part is the traditional GMM part which we want to minimize to obtain consistent estimates. Focusing on the first part at $\xi = \xi_0$, and taking expectations over i (i.e. the distribution of X) yields

$$
\begin{aligned}
&\mathbb{E}[Z_i(A_1^m, X)(\mathbb{E}[y_i|A_0, X, \xi_0] - \mathbb{E}[y_i|A_2^m, X, \xi_0])] \\
&\qquad = \mathbb{E}\left[\mathbb{E}\left[Z_i(A_1^m, X)(\mathbb{E}[y_i|A_0, X, \xi_0] - \mathbb{E}[y_i|A_2^m, X, \xi_0])|A_1^m, X\right]\right] \\
&\qquad = \mathbb{E}\left[Z_i(A_1^m, X)\mathbb{E}\left[(\mathbb{E}[y_i|A_0, X, \xi_0] - \mathbb{E}[y_i|A_2^m, X, \xi_0])|A_1^m, X\right]\right] \\
&\qquad = \mathbb{E}\left[Z_i(A_1^m, X)\mathbb{E}\left[(\mathbb{E}[y_i|A_0, X, \xi_0] - \mathbb{E}[y_i|A_2^m, X, \xi_0])|X\right]\right] \\
&\qquad = \mathbb{E}\left[Z_i(A_1^m, X)\left(\mathbb{E}\left[\mathbb{E}[y_i|A_0, X, \xi_0]|X\right] - \mathbb{E}\left[\mathbb{E}[y_i|A_2^m, X, \xi_0]|X\right]\right)\right] \\
&\qquad = 0
\end{aligned}
$$

The expectation of the second term can be written out as

$$
\begin{aligned}
\mathbb{E}[Z_i(A_1^m, X)\tilde{\epsilon}_i] &= \mathbb{E}[\mathbb{E}[Z_i((I - \alpha A_0)^{-1}\epsilon)_i|X, V, W, U_0]] \\
&= \mathbb{E}[Z_i(I - \alpha A_0)^{-1}\mathbb{E}[\epsilon_i|X, V, W, U_0]] \\
&= 0
\end{aligned}
$$

Where we use assumption 2 as well as the draw of U that generated Z_i being independent from U_0.

Condition 2 of theorem 2.6 of Newey and McFadden (1994) holds by 15.3, condition 3 holds by the continuity of the inverse, and condition 4 holds by Assumption 15.4. Note that as this holds for any set of draws, it also holds when we use many draws to compute the moment, so we have $\hat{\xi}^{GMM} \to \xi_0$. □

It's worth noting some differences between $\hat{\alpha}^{SGMM}$ and the GMM based estimator suggested in Chandrasekhar and Lewis (2016). Importantly, the SGMM estimator makes no requirements on the distribution of the outcomes, or the conditional distribution of the observed network given the outcomes. This is because the estimator only uses the network formation model, *not* the observed network. Investigating the potential bias-efficiency trade-off between the bias from incorporating parts of the observed network, and the potential efficiency gain from reducing the dimensionality of the integral in the SGMM is left to future research.[11]

[11] Thank you to Konrad Menzel for pointing out the existence of this trade-off.

B Counterfactual analysis

In the full network setting, we can recover estimates of ϵ once we've obtained estimates of all the parameters in the outcome equation. This means we can discuss the effects of various network configurations on aggregate outcomes.

In a setting where some of the links are missing, this is no longer the case. This is because even as the sample grows, our estimate of ϵ will be contaminated by the missing data. To see this, note that our partial data estimate $\hat{\epsilon}^S$ can be written out as

$$\hat{\epsilon}^S = y - \hat{\alpha} G^S y - X\hat{\beta} - \hat{\gamma} G^S X \xrightarrow{p} \epsilon + \alpha(Gy - G^S y) + \gamma(GX - G^S X) \neq \epsilon \text{ if } \gamma, \alpha \neq 0$$

In this section I will describe some Counterfactuals that can be done in the peer effect setting, and make note of which ones can be accomplished without a consistent estimate of ϵ.[12]

Consider a researcher who's interested in how educational peer effects could affect the distribution of outcomes in a given school. For any given counterfactual network, we could get the distribution of outcomes from the relation

$$y = (\mathbf{I}_n - \hat{\alpha}_n G)^{-1} \left(X\hat{\beta}_n + \hat{\epsilon} \right) \tag{9}$$

However evaluating how a change in the network affects the outcomes through this equation is made complicated by the binary nature of A as well as the non-linear nature of equation 9.

In this section I will discuss alternative methods of performing counterfactual analysis using simulations from our estimated distribution of A, as well as using quadratic integer programming techniques to find the best and worst network configurations for moments of the distribution of y.

B.1 The effect of changes in network formation behaviour

By putting various variables/parameters into the network formation model, we can derive counterfactual probabilities $\tilde{P}(A = 1|W)$. For example this could be used to answer questions such as "what if students were more willing to be friends with people from disadvantaged backgrounds" etc. We can then calculate the mean of any

[12]Note that while we cannot estimate ϵ, we may still be able to recover moments of its distribution using re-sampling methods.

function of the outcomes through

$$\mathbb{E}[y] = \mathbb{E}[\mathbb{E}[y|A]] \approx \frac{1}{M}\sum_{m=1}^{M}\mathbb{E}[y^m|A^m]$$

Where A^m is drawn according to $\tilde{P}(A = 1|W)$, where $\tilde{P}$ is a counterfactual distribution for the network. We can generate $\mathbb{E}[y^m|A^m]$ from the reduced form in the previous section,

$$\mathbb{E}[y_i^m|A^m] = (\mathbf{I}_n - \hat{\alpha}_n G^m)^{-1}\left(X\hat{\beta}_n\right)$$

where we've again used the exogeneity of the error terms.

B.2 Best and worst possible networks

A similarly interesting question might be what the best possible network configuration would be for student outcomes. Consider a planner who could re-shuffle friendships freely[13]. How would average outcomes of students differ from the realized ones?

Even without an estimate for ϵ, it can be shown that this problem can be recast into the following Mixed Integer Linear Program with Quadratic constraints (MILPQC)

$$\begin{aligned}
\max_{\tilde{y},A} \quad & \mathbf{1}_n'\tilde{y} \quad s.t. \\
& X_i\hat{\beta} = \tilde{y}_i - \frac{\hat{\alpha}}{d_i}\sum_j \tilde{y}_j A_{i,j} - \sum_j A_{i,j}\sum_k \frac{\hat{\gamma}_k}{d_i}X_j^k \quad \forall i \\
& A_{i,j} \in \{0,1\} \quad \forall i,j \\
\sum_j & A_{i,j} = d_i \quad \forall i
\end{aligned}$$

Where X_i^k is the k'th covariate of i, and X_i is a vector of individual i's covariates. The reason this formulation works is due to our assumption that $\mathbb{E}[\epsilon|A, X] = 0$, meaning it will on average "drop out" of our outcome equation. The quadratic form of the constraint becomes more apparent when we re-write the problem into matrix form, with $\zeta = (\tilde{y}, A_{1,2}, A_{1,3}, ..., A_{N,N-2}, A_{N,N-1})$ as a vector of all the variables in

[13] By re-shuffling friendships I mean that we maintain the same amount of friendships for each student, but we're free to who's friends with whom.

the MILPQC.

$$\begin{aligned}\max_{\zeta} \quad & (\mathbf{1}_n, 0, ..., 0)'\zeta \\ \zeta Q_i \zeta + (e_i, F_i)\zeta &= b_i \quad \forall i \\ D\zeta &= d \\ A_{i,j} &\in \{0,1\} \quad \forall i,j\end{aligned}$$

where e_i is a vector of length N of zero's with a 1 in the i'th element, Q_i is a matrix with $\frac{\hat{\alpha}}{d_i}$ in the row's corresponding to $\{A_{i,j}\}_{j=1,..,N}$ in the i'th column. F_i is a vector with $\sum_k \frac{\hat{\gamma_k}}{d_i} X_j^k$ in the elements corresponding to $\{A_{i,j}\}_{j=1,..,N}$.

D is the matrix form of the adding up constraints on the network. A similar form can be written for the variance, except the objective becomes quadratic, and we require a consistent estimate of ϵ. While the quadratic objective is more complicated to solve, though computation is simplified by the variance being convex in y.

C Simulations

C.1 Bounds

Let I be short for the bounds being informative. I will compare two cases, one where the researcher has knowledge of the number of friends for each student, and another where the researcher has no information beyond the symmetry of the network.

I will compare the OLS estimator, the complete information IV estimator, the naive partial IV estimator, the average lower and upper bounds given that the bounds were informative (LB|I, UB|I), the share of bounds that were informative, and the share of the links that are missing. Each row correspond to a different percentage of individuals being dropped. The setting is 3 networks with 300 individuals each, with a true peer effect of 0.5. I then select 1,5,10 or 15% of individuals and drop all links involving them. The results can be seen in the tables below.

Missing persons (%)	OLS	naive-IV	LB |I	UB | I	% I	Missing links (%)	% I unknown d
1	0.46	0.39	0.45	0.5	1	0.02	0
5	0.24	0.2	0.29	0.76	1	0.1	0
10	0.16	0.13	0.21	140.49	0.3	0.19	0
15	0.11	0.09			0	0.28	0

Table 6: Simulation results for the bounds of a linear-in-means model. The columns compare the OLS estimator, the the naive IV estimator, the lower and upper bounds given that the bounds were informative (LB|I, UB|I), the share of bounds that were informative, the share of the links that are missing, and the share of informative bounds when we do not know d_i.

Missing persons (%)	OLS	naive-IV	LB |I	UB | I	% I	Missing links (%)	% I unknown d
1	0.46	0.39	0.49	0.51	1	0.02	
5	0.24	0.2	0.42	0.6	1	0.1	
10	0.15	0.12	0.36	0.83	1	0.19	
15	0.11	0.09	0.31	1.48	1	0.28	
20	0.09	0.07	0.27	11.15	0.96	0.36	

Table 7: Simulation results for the bounds of a linear-in-means model, where the peer effects are estimated on the sample with $\sum_{j=1}^{N} S_{i,j} > 0$. The columns compare the OLS estimator, the the naive IV estimator, the lower and upper bounds given that the bounds were informative (LB|I, UB|I), the share of bounds that were informative, the share of the links that are missing, and the share of informative bounds when we do not know d_i.