Estimating Peer Effects Using Partial Network Data

Vincent Boucher[†] and Elysée Aristide Houndetoungan[‡]

September 2023

Abstract

We study the estimation of peer effects through social networks when researchers do not observe the entire network structure. Special cases include sampled networks, censored networks, misclassified links, and aggregated relational data. We assume that researchers can obtain a consistent estimator of the distribution of the network. We show that this assumption is sufficient for estimating peer effects using a linear-in-means model. We provide an empirical application to the study of peer effects on students academic achievement using the widely used Add Health database and show that network data errors have a first-order downward bias on estimated peer effects.

JEL Codes: C31, C36, C51
Keywords: Social networks, Peer effects, Missing variables, Measurement errors

---

[†] Boucher: Department of Economics, Université Laval, CRREP, CREATE, CIRANO; email: `vincent.boucher@ecn.ulaval.ca`. [‡] Houndetoungan: Cy Cergy Paris Université and Thema; email: `aristide.houndetoungan@cyu.fr`. An R package, including all replication codes is available at: `https://github.com/ahoundetoungan/PartialNetwork`.

Acknowledgements: We would like to thank Isaiah Andrews, Yann Bramoullé, and Bernard Fortin for their helpful comments and insights, as always. We would also like to thank Eric Auerbach, Arnaud Dufays, Stephen Gordon, Chih-Sheng Hsieh, Arthur Lewbel, Tyler McCormick, Angelo Mele, Francesca Molinari, Onur Özgür, Eleonora Patacchini, Xun Tang, and Yves Zenou for helpful comments and discussions. Thank you also to the participants of the Applied/CDES seminar at Monash University, the Economic seminar at the Melbourne Business School, the Econometric workshop at the Chinese University of Hong Kong, and the Centre of Research in the Economics of Development workshop at the Université de Namur. This research uses data from Add Health, a program directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by Grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Special acknowledgment is given to Ronald R. Rindfuss and Barbara Entwisle for assistance in the original design. Information on how to obtain Add Health data files is available on the Add Health website (`http://www.cpc.unc.edu/addhealth`). No direct support was received from Grant P01-HD31921 for this research.

1 Introduction

There is a large and growing literature on the impact of peer effects in social networks.[1] However, since eliciting network data is expensive (Breza et al., 2020), relatively few data sets contain comprehensive network information, and existing ones are prone to data errors. Despite some recent contributions, existing methodologies for the estimation of peer effects with incomplete or erroneous network data either focus on a specific kind of sampling or errors or they are highly computationally demanding.

In this paper, we propose a unifying framework that allows for the estimation of peer effects under the widely used linear-in-means model (e.g. Manski (1993); Bramoullé et al. (2009)) when the researcher does not observe the entire network structure. Our methodology is computationally attractive and flexible enough to cover cases in which, for example, network data are sampled (Chandrasekhar and Lewis, 2011; Liu, 2013; Lewbel et al., 2022c), censored (Griffith, 2019), missclassified (Hardy et al., 2019), or summarized by aggregated relational data (ARD; Breza et al. (2020); Alidaee et al. (2020)). Our central assumption is that the researcher is able to estimate a network formation model using some partial information about the network structure. Leveraging recent contributions on the estimation of network formation models, we show that this assumption is sufficient to identify and estimate peer effects.

We propose two estimators. First, we present a computationally attractive estimator based on a simulated generalized method of moment (SGMM). The moments are built using draws from the (estimated) network formation model. We study the finite sample properties of our SGMM estimator via Monte Carlo simulation and compare its performance with the benchmark case when true network is perfectly observed. We show that the estimator performs very well, even when a large fraction of the links are missing. Second, we present a flexible likelihood-based (Bayesian) estimator

---

[1] For recent reviews, see Boucher and Fortin (2016), Bramoullé et al. (2020), Breza (2016), and De Paula (2017).

allowing us to exploit the entire structure of the data-generating process. Although the computational cost is greater, we exploit recent computational advances in the literature, e.g. Mele (2017); Hsieh et al. (2019), and show that the estimator can be successfully implemented on common-sized data sets. In particular, we apply our estimator to study peer effects on academic achievement using the widely used Add Health database. We find that data errors have first-order downward bias on the estimated endogenous effect.

Our SGMM estimator is inspired by the literature on error-in-variable models with repeated observations.[2] Using a network formation model, we obtain a consistent estimator of the distribution of the true network. We then use this estimated distribution to obtain different draws from the distribution of the network. Our moment condition can be seen as a biased-corrected version of the instrumental strategy proposed by Bramoullé et al. (2009), in which we substitute the true network with the draws from the estimated distribution. We show that our moment conditions are asymptotically valid and that the estimator is consistent for a finite number of draws from the estimated distribution of the network. This property implies a significant reduction in the computational cost of the method relative to methods based on the integration of the moment conditions (e.g. Chandrasekhar and Lewis (2011)).

Importantly, our SGMM strategy requires only the (partial) observation of a *single* cross-section, as opposed to, for example, the approach of Zhang (2020). The presence of this feature is because of two main properties of the model. First, we can consistently estimate the distribution of the mismeasured variable (i.e. the network) using a single (partial) observation of the variable. Second, in the absence of measurement error, valid instruments for the endogenous peer variable are available (Bramoullé et al., 2009).

We explore the finite sample properties of our instrumental variable estimator using Monte Carlo simulations. We consider the case in which a large fraction of the

---

[2] See Bound et al. (2001) for a review and Chen et al. (2011) for a review focused on nonlinear models.

links are unobserved. When there are no unobserved group-level fixed effects, the loss in precision due to missing links is very small, even a very large fraction of links are missing (up to 75%). With group-level unobservable fixed effects, the loss in precision is larger, but the estimator still performs very well when up to 25%-50% of the links are missing.

Our Bayesian estimator is based on likelihood and therefore uses more information about the structure of the model, leading to more precise estimates. In the context of this estimator, the estimated distribution for the network acts as a prior distribution, and the inferred network structure is updated through a Markov chain Monte Carlo (MCMC) algorithm. Our approach relies on data augmentation (Tanner and Wong, 1987), which treats the network as an additional set of parameters to be estimated. This approach saves us from integrating over the $2^{N(N-1)}$ potential networks compatible with the data, which would create a serious computational issue. In particular, our MCMC builds on recent developments from the empirical literature on network formation (e.g. Mele (2017); Hsieh et al. (2019, 2020)). We show that the computational cost of our estimator is reasonable and that it can easily be applied to standard data sets.

We study the impact of errors in adolescents' friendship network data for the estimation of peer effects in education (Calvó-Armengol et al., 2009). We show that the widely used Add Health database features many missing links—around 30% of the within-school friendship nominations are coded with error—and that these data errors strongly bias the estimated peer effects. Specifically, we estimate a model of peer effects on students' academic achievement. We probabilistically reconstructs the missing links, and we obtain a consistent estimator of peer effects using both our estimators. The bias due to data errors is qualitatively important. Our estimated endogenous peer effect coefficient is 1.5 times larger than that obtained by assuming the data contains no errors.

This paper contributes to the recent literature on the estimation of peer effects when the network is either not entirely observed or observed with noise. In particular, our framework is valid when network data are either sampled, censored, misclassified, or consist of aggregate relational data.[3] We unify these strands in the literature and provide a flexible and computationally tractable framework for estimating peer effects with incomplete or erroneous network data.

Sampled networks and censoring: Chandrasekhar and Lewis (2011) show that models estimated using sampled networks are generally biased. They propose an analytical correction as well as a two-step general method of moment (GMM) estimator. Liu (2013) shows that when the interaction matrix is not row-normalized, instrumental variable estimators based on an out-degree distribution are valid, even with sampled networks. Hsieh et al. (2018) focus on a regression model that depends on global network statistics. They propose analytical corrections to account for non-random sampling of the network (see also Chen et al. (2013)). Thirkettle (2019) also focuses on global network statistics, assuming that the researcher only observes a random sample of links. Using a structural network formation model, he derives bounds on the identified set for both the network formation model and the network statistic of interest. Lewbel et al. (2022a) develop a two-stage least square estimator for the linear in means model when some links are potentially missing. They propose valid instruments under some restrictions on the observed and true interactions matrices, or when researchers observe at least two samples of the same true network. Finally, Zhang (2020) studies program evaluation in a context in which networks are sampled locally and affected by a single type measurement error (either false positives or false negatives, but not both). Assuming that the researcher has access to two measurements of the network for each sampled unit, she presents a nonparametric estimator of the treatment and spillover effects.

Relatedly, Griffith (2019) explores the impact of imposing an upper bound to the number of links when eliciting network data, e.g. “Name your five best friends.” He

---

[3] For related literature that studies the estimation of peer effects when researchers have no network data, see Manresa (2016); De Paula et al. (2018); Souza (2014); Lewbel et al. (2022c).

presents a bias-correction method and explores the impact of censoring using two empirical applications. He finds that censoring underestimates peer effects. Griffith and Kim (2023) generalizes the approach in Griffith (2019) and present analytic bias-corrections for the reduced-form parameters in the linear-in-means and linear-in-sums models under an Expectational Equivalence assumption.

We contribute to this literature by proposing two simple and flexible estimators for the estimation of peer effects based on a linear-in-means model (as opposed to network statistics as in Hsieh and Lee (2016) and Thirkettle (2019)). Our estimators do not require many observations of the sampled network (contrary to Zhang (2020) and Lewbel et al. (2022c)). Similar to Griffith (2019) and Griffith and Kim (2023), we find—using the Add Health database—that sampling leads to an underestimation of peer effects, although we find that censoring has a negligible impact, in the context of peer effects, on academic achievement.

Our Bayesian estimator is similar in spirit to the two-stage GMM estimator proposed by Chandrasekhar and Lewis (2011), but it is computationally advantageous. Indeed, their GMM estimator is based on an unconditional moment condition, which requires integrating over the entire set of networks that are compatible in the data. In Chandrasekhar and Lewis (2011), there are $2^{N(N-1)-M}$ such networks, where M is the number of sampled pairs of individuals. Even for small networks, the computational cost is substantial whenever sampling is nontrivial. Our SGMM estimator does not suffer from this computational cost and can produce precise estimates with as little as three network simulations. While our Bayesian estimator is more computationally demanding, we exploit recent developments from the empirical literature on network formation (e.g. Mele (2017); Hsieh et al. (2019)) and show that it is computationally tractable, even when no link is sampled (e.g. with aggregated relational data (ARD)), which would otherwise require integrating over the $2^{N(N-1)}$ networks compatible with the data.

Misclassification: Hardy et al. (2019) look at the estimation of (discrete) treatment effects when the network is observed noisily. Specifically, they assume that observed links are affected by iid errors and present an expectation maximization (EM) algorithm that allows for a consistent estimator of the treatment effect. Lewbel et al. (2022b) show that when the expected number of missing links grows at rate strictly smaller than n, the 2SLS estimator in Bramoullé et al. (2009) is consistent.[4]

Our model allows for the misclassification of all links with positive probability and we do not impose restrictions on the rate of misclassification. As in Hardy et al. (2019), we use a network formation model to estimate the probability of false positives and false negatives. However, our two-stage strategy—estimating the network formation model and then the peer effect model—allows for greater flexibility. In particular, our network formation model is allowed to flexibly depend on covariates. This is empirically important, as networks typically feature homophily on observed characteristics (e.g. Currarini et al. (2010); Bramoullé et al. (2012)).

Aggregate relational data: Breza et al. (2020) propose to estimate network effects using aggregate relation data (ARD). These are obtained from such survey questions as, “How many of your friends have trait X?” They present a network formation model that can be estimated using only ARD. They show the validity of their methodology using two empirical applications in which the outcome of interest depends on the summary statistics of the network. Alidaee et al. (2020) present an alternative estimator allowing to recover nonparametrically the linking probability through ARD. Using a low-rank assumption, they present a simple penalized regression.

We show that these recent methodologies can also be applied to the study of peer effects using a linear-in-means model, which significantly expands the scope of the potential applications of these approaches.

A main contribution of this paper is that our estimators can be applied to each of the previously mentioned data issues or to their combination. Our two-step approach—first estimating the network formation and then the peer effects—is flex-

---

[4] When the growth rate is strictly smaller than $\sqrt{n}$, inference is also valid.

ible and computationally attractive. To reduce the implementation costs, we also present an easy-to-use R package—named `PartialNetwork`—for applying our estimators. The package allows replicating all simulations and empirical applications in the paper, including the estimator proposed by Breza et al. (2020). The package is available online at: `https://github.com/ahoundetoungan/PartialNetwork`. Additional implementation examples are provided in the Vignette accompanying the package. For example, we show that the implementation of our Bayesian estimator, combined with the estimator for ARD proposed by Breza et al. (2020), is straightforward and computationally tractable.

The remainder of the paper is organized as follows. In Section 2, we present the econometric model as well as the main assumptions. In Section 3, we present our SGMM estimator and study its performance via Monte Carlo simulations. In Section 4, we present our likelihood-based estimation strategy. In Section 5, we present our application to peer effects on academic achievement. Section 6 concludes the paper.

2 The Model

We now formally present our model. We first describe the *linear-in-means* model (Manski, 1993; Bramoullé et al., 2009), arguably the most widely used model for studying peer effects in networks (see Bramoullé et al. (2020) for a recent review). We then introduce our main assumption, characterizing what is known about the structure of the network.

2.1 The Linear-in-Means Model

Let $\mathbf{A}$ represent the $N \times N$ *adjacency matrix* of the network. We assume a directed network: $a_{ij} \in \{0, 1\}$, where $a_{ij} = 1$ if i is linked to j. We normalize $a_{ii} = 0$ for all i and let $n_i = \sum_j a_{ij}$ denote the number of links of i. Let $\mathbf{G} = f(\mathbf{A})$, the $N \times N$ *interaction matrix* for some function f. Unless otherwise stated, we assume that

$\mathbf{G}$ is a row-normalization of the adjacency matrix $\mathbf{A}$.[5] Most of our results hold for any function f, which preserves the independence among groups (see Assumption 2 below).

We focus on the following model:

$$\mathbf{y} = c\mathbf{1} + \mathbf{X}\boldsymbol{\beta} + \alpha\mathbf{G}\mathbf{y} + \mathbf{G}\mathbf{X}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}, \tag{1}$$

where $\mathbf{y}$ is a vector of an outcome of interest (e.g. academic achievement), c is a constant, $\mathbf{1}$ is a vector of ones, $\mathbf{X}$ is a matrix of observable characteristics (e.g. age, gender...), and $\boldsymbol{\varepsilon}$ is a vector of errors.[6] The parameter α therefore captures the impact of the average outcome of one's peers on their behavior (the endogenous peer effect). The parameter $\boldsymbol{\beta}$ captures the impact of one's characteristics on their behavior (the individual effects). The parameter $\boldsymbol{\gamma}$ captures the impact of the average characteristics of one's peers on their behavior (the contextual peer effects).

The following set of assumptions summarizes our setup.

Assumption 1. $|\alpha| < 1/\|\mathbf{G}\|$ *for some submultiplicative norm* $\|\cdot\|$.

Assumption 2. *The population is partitioned into* $M > 1$ *groups, where the size* N_r *of each group* $r = 1, ..., M$ *is bounded. The probability of a link between individuals of different groups is equal to* 0.

Assumption 3. *For each group, the outcome and individual characteristics are observed (i.e.* $(\mathbf{y}_r, \mathbf{X}_r)$, $r = 1, ..., M$, *are observed).*[7]

Assumption 4. *Exogeneity:* $\mathbb{E}[\boldsymbol{\varepsilon}_r|\mathbf{X}_r] = \mathbf{0}$ *for all* $r = 1, ..., M$.

Assumption 1 ensures that the model is coherent and that there exists a unique vector $\mathbf{y}$ compatible with (1). When $\mathbf{G}$ is row-normalized, $|\alpha| < 1$ is sufficient.

---

[5] In such a case, $g_{ij} = a_{ij}/n_i$ whenever $n_i > 0$, whereas $g_{ij} = 0$ otherwise.

[6] Note that Boucher and Bramoullé (2020) recently showed that (1) is valid (coherent, complete, and microfounded), even when y_i is binary.

[7] For any vector/matrix, we use the subscript r to denote the subvector/submatrix restricted to group r.

Assumption 2 is introduced for exposition purposes; for example, the data could consist of a collection of small villages (Banerjee et al., 2013) or schools (Calvó-Armengol et al., 2009). Our methods extend to alternative assumptions such as those proposed by Lee (2004) and Lee et al. (2010).[8] In particular, Assumption 2 implies a "many markets" asymptotic framework, meaning that the number of groups M goes to infinity as the number of individuals N goes to infinity.

Assumption 3 implies that the data is composed of a subset of fully sampled groups.[9] A similar assumption is made by Breza et al. (2020). Note that Assumption 4 does not impose restrictions on the dependence between the errors and the network structure. We detail the network formation process in the next section.

2.2 Partial Network Information

In this paper, we relax the costly assumption that the adjacency matrix $\mathbf{A}$ is observed. We assume instead that sufficient information about the network is observed so that a network formation model can be estimated.

More formally, we let $\mathcal{A}$ denote the observed information about the true network structure. That is, $\mathcal{A}$ is a function of the true network $\mathbf{A}$ and potentially of individuals' characteristics (see Example 4). We impose no particular structure on $\mathcal{A}$ but discuss important examples below (see Examples 1–4).

We assume that links are generated as follows:

$$P(a_{ij}) \equiv P(a_{ij}|\boldsymbol{\rho}_0) \propto \exp\{a_{ij}Q(\boldsymbol{\rho}_0, \mathbf{w}_{ij})\}, \tag{2}$$

where Q is some known, twice continuously differentiable function, $\mathbf{w}_{ij} = \mathbf{w}_{ij}(\mathbf{X})$ is a vector of observed characteristics for the pair ij, and $\boldsymbol{\rho}_0$ is the true value of $\boldsymbol{\rho}$, a vector of parameters to be estimated. Note that the assumption that the set of observed characteristics of the pairs $\mathbf{w}_{ij}$ are function of $\mathbf{X}$ implies that the network

---

[8] The authors assume that the adjacency matrix $\mathbf{A}$ is bounded in row- and column-sums.

[9] Contrary to Liu et al. (2017) or Wang and Lee (2013), for example.

is *exogenous*. We omit the dependence of $P(a_{ij})$ on $\mathbf{w}_{ij}$ to simplify the notation.

As will be made clear, our estimation strategy requires that the econometrician be able to draw samples from (a consistent estimator) $P(\mathbf{A})$. Thus, and for the sake of simplicity, we focus on distributions that are conditionally independent across links (i.e. $P(a_{ij}|\mathbf{A}_{-ij}) = P(a_{ij})$), as in (2), although this is not formally required.[10]

We now present our main assumption.

Assumption 5 (Partial Network Information)**.** *Given network information $\mathcal{A}$ and the parametric model (2), there exists an estimator $\hat{\boldsymbol{\rho}}$, such that $\hat{\boldsymbol{\rho}} \to_p \boldsymbol{\rho}_0$ as $N \to \infty$.*

Assumption 5 implies that, using (2), the researcher has access to "sufficient information" about the network structure to obtain a consistent estimator of the distribution of the true network $P(\mathbf{A})$. We denote such a consistent estimator by $\hat{P}(\mathbf{A}) \equiv P(\mathbf{A}|\hat{\boldsymbol{\rho}}, \mathcal{A})$. We omit the dependence on $\mathcal{A}$ to simplify the notation when this does not create confusion. Note that we can use $\hat{P}(\mathbf{A})$ to obtain the consistent estimator $\hat{P}(\mathbf{G})$ because $\mathbf{G} = f(\mathbf{A})$ for some known function f.[11]

Importantly, it should be noted that even if the econometrician has access to a consistent estimator of the distribution of the true network, that is $P(\mathbf{A})$, it *does not* imply that they could simply proxy $\mathbf{G}$ in (1) using a draw $\tilde{\mathbf{G}}$ from the distribution $\hat{P}(\mathbf{G})$. The reason is that for any vector $\mathbf{z}$, $\tilde{\mathbf{G}}\mathbf{z}$ generally does not converge to $\mathbf{G}\mathbf{z}$ as the number of individuals N goes to infinity. In other words, knowledge of $\hat{P}(\mathbf{G})$ and $\mathbf{z}$ is not sufficient to obtain a consistent estimator of $\mathbf{G}\mathbf{z}$. To see why, note that $(\mathbf{G}\mathbf{z})_i = \sum_{j=1}^{N} g_{ij} z_j$. Under Assumption 2, the set of j, such that $g_{ij} \neq 0$, is bounded;

---

[10] A prime example of a network distribution that is not conditionally independent is the distribution for an exponential random graph model (ERGM), e.g. Mele (2017). Of course, in these cases, the specification of the marginal distribution (2) must be replaced by a specification of the joint distribution $P(\mathbf{A})$. Our ability to accommodate for such model depends on what is known about the true network, i.e. $\mathcal{A}$. In particular, we need that $P(\mathbf{A}|\boldsymbol{\rho}, \mathcal{A})$ be computationally tractable and be such that Assumption 5 below holds for non-trivial $\mathcal{A}$. A specific case of ERGM that could be estimated with partial-network data is that in Boucher and Mourifié (2017), which requires asking sampled pairs of individuals about the number of links they have.

[11] Because $\mathbf{A}$ takes a finite number of values, so does $\mathbf{G}$, (thus $\mathbf{A}$ and $\mathbf{G}$ have multinomial distribution). Then, $P(\mathbf{G})$ is a continuous function of $P(\mathbf{A})$, irrespective of the function f linking $\mathbf{A}$ and $\mathbf{G}$.

thus, a consistent estimator of $P(\mathbf{G})$ is not sufficient to obtain a consistent estimator of $\mathbf{GX}$.

Assumption 5 covers a large range of cases in which networks are partially observed. We specifically discuss four leading examples in which Assumption 5 holds: *sampled networks* (Example 1), *censored networks* (Example 2), *misclassified network links* (Example 3), and *aggregated relational data* (Example 4).

Example 1 (Sampled Networks)**.** *Suppose that we observe the realizations of* a_{ij} *for a random sample of* m *pairs (e.g. Chandrasekhar and Lewis (2011)). Here* $\mathcal{A}$ *can be represented by an* $N \times N$ *binary matrix* $\mathbf{A}^{obs}$*. Conley and Udry (2010) present such a sampling scheme and ask individuals about their relationship with a random sample of other individuals; for example, "Do you know person X?"*

Such a setup is sufficient to consistently estimate flexible network formation models, such as that in Graham (2017). A simpler example would be to assume that the network formation model (2) is equal to $P(a_{ij} = 1) \propto \exp\{\mathbf{w}_{ij}\boldsymbol{\rho}\}$*. In this case, a simple logistic regression provides a consistent estimator of* $\boldsymbol{\rho}$*.*

Given this consistent estimator for $\boldsymbol{\rho}$ *and the assumed parametric model, we can compute an estimator for the distribution of the true network* $\hat{P}(\mathbf{A}) = P(\mathbf{A}|\hat{\boldsymbol{\rho}}, \mathcal{A})$*. Here,* $\hat{P}(a_{ij}) = a_{ij}^{obs}$ *for any sampled link* ij*, while* $\hat{P}(a_{ij}) \propto \exp\{a_{ij}\mathbf{w}_{ij}\hat{\boldsymbol{\rho}}\}$ *for the remaining unsampled links.*

Example 2 (Censored Network Data)**.** *As discussed in Griffith (2019), network data is often censored. This typically arises when surveyed individuals are asked to name only* $T > 1$ *links (among the* N *possible links they may have). Here,* $\mathcal{A}$ *can be represented by an* $N \times N$ *binary matrix* $\mathbf{A}^{obs}$*. We can use censored network data to estimate a network formation model, such as* $P(a_{ij} = 1) \propto \exp\{\mathbf{w}_{ij}\boldsymbol{\rho}\}$*.*

For simplicity, let us assume as in Griffith (2019) that each link has the same probability of being censored. Then, the parameters in $\boldsymbol{\rho}$ *(other than the constant) are identified from the observed ratios* $\sum_{ij} a_{ij}w_{ij}^k / \sum_{ij} w_{ij}^k$ *(for observable characteristic* *k), as these sufficient statistics are not biased by censoring. To identify the constant, note that we can compute the likelihood of the censored degree distribution (i.e.* $n_i = \sum_j a_{ij}$*). Letting* $\{\mathbf{w}_{ij}\}_j = [\mathbf{w}_{i1}, ..., \mathbf{w}_{iN}]$*, we have* $P(n_i = t|\{\mathbf{w}_{ij}\}_j; \boldsymbol{\rho})$ *for observed* $t < T$ *and* $P(n_i \geq T|\{\mathbf{w}_{ij}\}_j; \boldsymbol{\rho})$ *for observed* $t = T$*, which allows identifying the constant.*

Once such an estimator of $\boldsymbol{\rho}$ *is obtained, we can compute an estimator for the distribution of the true network* $\hat{P}(\mathbf{A}) = P(\mathbf{A}|\hat{\boldsymbol{\rho}}, \mathcal{A})$*. In particular,* $P(a_{ij}|\hat{\boldsymbol{\rho}}, a_{ij}^{obs} = 1) = 1$ *because observed links necessarily exist. Also note that for all individuals* i*, such that* $n_i < T$*, we have* $P(a_{ij}|\hat{\boldsymbol{\rho}}, a_{ij}^{obs}) = a_{ij}^{obs}$ *for all* j*, as their network data are not censored. Here, the structural model is only used to obtain the probability of links that are not observed for individuals whose links are potentially censored, i.e.* $P(a_{ij}|\hat{\boldsymbol{\rho}}, a_{ij}^{obs} = 0) \propto \exp\{a_{ij}\mathbf{w}_{ij}\hat{\boldsymbol{\rho}}\}$ *for all* ij*, such that* $n_i \geq T$*.*

Example 3 (Misclassification)**.** *Hardy et al. (2019) study cases in which networks are observed but may include misclassified links (i.e. false positives and false negatives). Here,* $\mathcal{A}$ *can be represented by an* $N \times N$ *binary matrix* $\mathbf{A}^{mis}$*. The (consistent) estimation of (2) in such a context follows directly from the existing literature on misclassification in binary outcome models, e.g. Hausman et al. (1998).*

Let q_1 *be the probability of false positives and* q_0 *be the probability of false negatives (both being elements of* $\boldsymbol{\rho}$*). The estimator for the distribution of the true network is given by* $P(a_{ij} = 1|\hat{\boldsymbol{\rho}}, a_{ij}^{mis}) = a_{ij}^{mis}(1 - \hat{q}_1) + (1 - a_{ij}^{mis})\hat{q}_0$*.*

Example 4 (Aggregated Relational Data)**.** *Aggregated relational data (ARD) are obtained from survey questions such as, "How many friends with trait 'X' do you have?" Here,* $\mathcal{A}$ *can be represented by an* $N \times K$ *matrix of integer values, where* K *is the number of traits that individuals were asked about.*

Building on McCormick and Zheng (2015), Breza et al. (2020) proposed a novel approach for the estimation of network formation models using only ARD. They assume:

$$P(a_{ij} = 1) \propto \exp\{\nu_i + \nu_j + \zeta \mathbf{z}_i'\mathbf{z}_j\}, \tag{3}$$

where ν_i, ν_j, ζ, $\mathbf{z}_i$, and $\mathbf{z}_j$ are not observed by the econometrician but follow parametric distributions. Here, parameters ν_i and ν_j can be interpreted as i and j's propensities to create links, irrespective of the identity of the other individual involved. The other component, $\zeta\mathbf{z}_i'\mathbf{z}_j$, is meant to capture homophily (like attracts like) on an abstract latent space (e.g. Hoff et al. (2002)).

Breza et al. (2020) show that it is possible to use ARD to recover the values of the variables in (3). In particular, letting $\boldsymbol{\rho} = [\{\nu_i\}_i, \{\mathbf{z}_i\}_i, \zeta]$, Alidaee et al. (2020) and Breza et al. (2019) provide sufficient conditions for the consistent estimation of $\boldsymbol{\rho}$.

Contrary to Examples 1–3, ARD does not provide information on any specific links; [12] *therefore, the predicted distribution of the true network is $P(a_{ij}|\hat{\boldsymbol{\rho}}, \mathcal{A}) = P(a_{ij}|\hat{\boldsymbol{\rho}})$, which is given by Equation (3). Here, it is worth emphasizing that the observed network information (i.e. ARD) is not very informative about the particular network structure in the data. In this sense, it could be viewed as a worst-case scenario.*

3 Simulated Generalized Method of Moment Estimators

In this section, we present an estimator based on a Simulated Generalized Method of Moments (SGMM). The key observation underlying our approach is that it is not necessary to observe the complete network structure to observe $\mathbf{y}$, $\mathbf{X}$, $\mathbf{GX}$, and $\mathbf{Gy}$. For example, one could simply obtain $\mathbf{Gy}$ from survey data, "What is the average value of your friends' y?"

However, even observing $\mathbf{y}$, $\mathbf{X}$, $\mathbf{GX}$, and $\mathbf{Gy}$, the model (1) cannot be simply estimated using simple linear regression. The reason is that $\mathbf{Gy}$ is endogenous; thus, a linear regression would produce biased estimates (e.g. Manski (1993), Bramoullé

---

[12] That is, unless ARD includes the degree distribution with some individuals reporting having no links at all.

et al. (2009)).

The typical instrumental approach to deal with this endogeneity is to use instruments based on the structural model, i.e. instruments constructed using second-degree peers (e.g. $\mathbf{G}^2\mathbf{X}$, see Bramoullé et al. (2009)). These are less likely to be found in survey data. Indeed, we could doubt the informativeness of questions such as, "What is the average value of your friends' average value of their friends' x?"

However, the above discussion is encouraging. When researchers observe $[\mathbf{y}, \mathbf{X}]$ and $[\mathbf{GX}, \mathbf{Gy}]$, information about the network is only used as a means to construct a valid instrument, e.g. following Bramoullé et al. (2009). Then, if $\mathbf{GX}$ and $\mathbf{Gy}$ are observed, the estimation of peer effects is possible using any exogeneous (strong) predictor of the true network structure, as in Kelejian and Piras (2014), König et al. (2019), and Lee et al. (2020).[13]

The problem is more challenging when $\mathbf{Gy}$ and $\mathbf{GX}$ are not observed. To simplify the notation, for the remainder of this section we define $\{\dot{\mathbf{G}}^{(r)}\}_{r=1}^{R}$, $\{\ddot{\mathbf{G}}^{(s)}\}_{s=1}^{S}$, and $\{\dddot{\mathbf{G}}^{(t)}\}_{t=1}^{T}$ as sequences of independent draws from $\hat{P}(\mathbf{G})$. We will also let $\dot{\mathbf{G}}$, $\ddot{\mathbf{G}}$, and $\dddot{\mathbf{G}}$ be single independent draws from $\hat{P}(\mathbf{G})$. Finally, for any matrix $\mathbf{B}$, we will use $\mathbf{B}_i$ to denote the ith row of $\mathbf{B}$.

Before presenting our estimation strategy, we want to emphasize that we cannot simply proxy $\mathbf{Gy}$ and $\mathbf{GX}$ using draws from $\hat{P}(\mathbf{G})$. To see why, consider the simple case for which $\boldsymbol{\gamma} = \mathbf{0}$, so that the observation of $\mathbf{GX}$ is inconsequential. We have, substituting $\mathbf{Gy}$ with $\ddot{\mathbf{G}}\mathbf{y}$,

$$\mathbf{y} = c\mathbf{1} + \mathbf{X}\boldsymbol{\beta} + \alpha\ddot{\mathbf{G}}\mathbf{y} + [\boldsymbol{\eta} + \boldsymbol{\varepsilon}],$$

where $\boldsymbol{\eta} = \alpha[\mathbf{G} - \ddot{\mathbf{G}}]\mathbf{y}$ is the error due to the approximation of $\mathbf{Gy}$ by $\ddot{\mathbf{G}}\mathbf{y}$. Importantly, the approximation error does *not* vanish as N grows because individuals

---

[13] See Proposition 1 of the Online Appendix E for a formal statement. In particular, one needs a predictor of the network such that the resulting instrument is not correlated with $\boldsymbol{\varepsilon}$ and is strongly correlated with $\mathbf{Gy}$.

belong to bounded groups (see Assumption 2).[14] Moreover, because $\mathbf{y}$ is a function of $\mathbf{G}$ (but not of $\ddot{\mathbf{G}}$), we typically have $\mathbb{E}(\mathbf{G}\mathbf{y})_i|\mathbf{Z}_i \neq \mathbb{E}(\ddot{\mathbf{G}}\mathbf{y})_i|\mathbf{Z}_i$ for some vector of instruments $\mathbf{z}_i$ and all i.

In Proposition 2 of the Online Appendix E, we show, using an argument very similar to that of Andrews et al. (2017), that the asymptotic bias induced by $\mathbb{E}\eta_i|\mathbf{z}_i \neq 0$ can be bounded, and simulations show that it can be very small for carefully constructed vectors of instruments, which can be useful in practice.

In any case, in our main result below, we show that we can include the bias directly in the moment function; this provides a consistent estimator of $\boldsymbol{\theta}$.

3.1 Main Result

Let $\tilde{\boldsymbol{\theta}} = [c, \boldsymbol{\beta}', \boldsymbol{\gamma}']'$. We maintain Assumptions 1–5 and impose the regularity Assumptions 6–9 and the identification Assumptions 10 and 11 presented in Appendix A.

Theorem 1 (SGMM)**.**

[Conditions] Let $\dot{\mathbf{Z}}^{(r)} = [\mathbf{1}, \mathbf{X}, \dot{\mathbf{G}}^{(r)}\mathbf{X}, (\dot{\mathbf{G}}^{(r)})^2\mathbf{X}, (\dot{\mathbf{G}}^{(r)})^3\mathbf{X}, ...]$ *and* $\ddot{\mathbf{V}}^{(s)} = [\mathbf{1}, \mathbf{X}, \ddot{\mathbf{G}}^{(s)}\mathbf{X}]$. *Consider also the following (simulated) moment function:*

$$\mathbf{m}_{R,S,T}(\boldsymbol{\theta}) = \frac{1}{RST}\sum_{r=1}^{R}\sum_{s=1}^{S}\sum_{t=1}^{T}\dot{\mathbf{Z}}_i^{(r)\prime}\left[(\mathbf{I} - \alpha\dddot{\mathbf{G}}^{(t)})_i\left(\mathbf{y} - (\mathbf{I} - \alpha\ddot{\mathbf{G}}^{(s)})^{-1}\ddot{\mathbf{V}}^{(s)}\tilde{\boldsymbol{\theta}}\right)\right] \tag{4}$$

[Result] For any positive integers R, S, and T, the (simulated) GMM estimator based on (4) is consistent.

Theorem 1 presents conditions for the consistency of our two-step estimator. In particular, similar to a standard simulated GMM (Gourieroux et al., 1996), consistency holds for a finite number of simulations. Our estimator therefore does not suffer

---

[14] As opposed to, for example, the context studied by Auerbach (2019).

from the curse of dimensionality faced by Chandrasekhar and Lewis (2011).[15]

To understand the intuition behind the use of the moment function (4), let us replace $\mathbf{y} = (\mathbf{I} - \alpha\mathbf{G})^{-1}(\mathbf{V}\tilde{\boldsymbol{\theta}} + \boldsymbol{\varepsilon})$ in Equation (4). In doing so, we can rewrite the term inside the triple summation in (4) as a sum of two terms. First, $\dot{\mathbf{Z}}_i^{(r)\prime}\boldsymbol{\varepsilon}_i$, which would be the standard moment function if $\mathbf{G}$ was observed (e.g. Bramoullé et al. (2009)), and second:

$$\dot{\mathbf{Z}}_i^{(r)\prime}\left[(\mathbf{I} - \alpha\dddot{\mathbf{G}}^{(t)})_i\left((\mathbf{I} - \alpha\mathbf{G})^{-1}\mathbf{V} - (\mathbf{I} - \alpha\ddot{\mathbf{G}}^{(s)})^{-1}\ddot{\mathbf{V}}^{(s)}\right)\tilde{\boldsymbol{\theta}}\right]. \tag{5}$$

That second term can be viewed as the bias-correction term of the moment function.

We show that the expectations of both terms are equal to 0 as N grows to ∞. For the first term, this is true because the simulated network is exogenous. For the second term, this is true because $\mathbf{G}$, $\dot{\mathbf{G}}$, $\ddot{\mathbf{G}}$, and $\dddot{\mathbf{G}}$ are asymptotically drawn from the same distribution.

For the special case in which $\mathbf{G}$ is observed, one could substitute $\dot{\mathbf{G}}^{(r)} = \ddot{\mathbf{G}}^{(s)} = \dddot{\mathbf{G}}^{(t)} = \mathbf{G}$ for all r, s, t, and the first term would simply be $\mathbf{Z}_i'\boldsymbol{\varepsilon}_i$, whereas the second term would be exactly equal to $\mathbf{0}$.

Note that Theorem 1 is valid whether or not $\mathbf{Gy}$ and $\mathbf{GX}$ are observed because the moment function (4) does not use information on $\mathbf{Gy}$ or $\mathbf{GX}$. It is of course possible to include this additional information if either one is observed.

For example, assume that $\mathbf{GX}$ is observed and replace $\ddot{\mathbf{V}}$ with $\mathbf{V}$ in (5). After some algebra, we obtain:

$$\dot{\mathbf{Z}}_i^{(r)\prime}\left[(\mathbf{I} - \alpha\dddot{\mathbf{G}}^{(t)})_i(\mathbf{I} - \alpha\ddot{\mathbf{G}}^{(s)})^{-1}[\alpha(\mathbf{G} - \ddot{\mathbf{G}})\mathbf{y}]\right] \tag{6}$$

and a term that is equal to zero in expectation.

We can show that the last term in brackets corresponds to the source of the

---

[15] The unconditional moment condition in Chandrasekhar and Lewis (2011) is based on the (Monte Carlo) integration of the moment condition $\mathbb{E}(\mathbf{Z}^{(s)\prime}\varepsilon|\mathbf{G})$ over $\mathbf{G}$.

asymptotic bias because of the approximation error of $\mathbf{Gy}$ by $\ddot{\mathbf{G}}\mathbf{y}$ (see Proposition 2 of the Online Appendix). Whereas Equation (6) cannot directly be substituted for (4) in Theorem 1, it is possible to construct moment functions for the particular cases for which $\mathbf{GX}$ or $\mathbf{Gy}$ is observed. These cases can be found in Corollary 1 and 2 of the Online Appendix.

3.1.1 Asymptotic Normality

Although Theorem 1 shows the consistency of the estimator, it remains silent about its asymptotic distribution. In Appendix A.3, we present how to estimate the asymptotic variance of the estimator, irrespective of the asymptotic distribution. The estimation includes the sampling uncertainty, the estimating uncertainty of $\hat{\boldsymbol{\rho}}$, and the simulating uncertainty (which decreases as R, S, and T grow). Note, however, that asymptotic normality cannot always be guaranteed because of the dependence of the estimator on $\hat{\boldsymbol{\rho}}$, estimated in the first stage (see Assumption 5).

An important special case for which asymptotic normality is achieved is when the estimator of the network formation process can be written as an m-estimator; thus, the estimator in Theorem 1 can be written as the second stage of a two-stage m-estimator, as for the network formation models in Examples 1, 2, and 3.[16]

However, while asymptotic normality holds for m-estimators, it cannot always be guaranteed. In general, researchers may not always understand perfectly the asymptotic properties of the used network formation estimator. Indeed, recall that Assumption 5 only imposes the consistency of $\hat{\boldsymbol{\rho}}$ and does not discuss the convergence rate or asymptotic framework in detail. For some network formation models, these may be difficult to formalize.

An example is latent space models based on ARD (see Example 4). Alidaee et al. (2020) propose a penalized regression, whereas Breza et al. (2020) propose a Bayesian estimator. Neither estimator can be written directly as an m-estimator, and

---

[16] See Appendix A.2 for a more detailed discussion.

the convergence rates in a many-networks asymptotic framework (see our Assumption 2) remain poorly understood.[17]

3.2 Monte Carlo Simulations

In this subsection, we study the performance of our SGMM using Monte Carlo simulations. In particular, we study a case in which links are missing at random (see Example 1). The 2SLS estimator proposed in Bramoullé et al. (2009) is not consistent, unless the expected number of missing links grows at a rate strictly less than n (Lewbel et al., 2022b). The estimator in Bramoullé et al. (2009) is however a natural benchmark since it is a special case of our SGMM estimator, when all links are observed.[18]

We present simulations for four versions of our SGMM estimator. One in which both $\mathbf{GX}$ and $\mathbf{Gy}$ are unobserved (as in Theorem 1), one in which they are both observed (as in Proposition 1 in the Online Appendix), and two intermediary cases in which either $\mathbf{GX}$ or $\mathbf{Gy}$ are observed (see Corollaries 1 and 2 in the Online Appendix). Figure 1 and 2 present the results for the endogeneous peer effect coefficient α, without and with group-level fixed effects, respectively.

Both with and without fixed effects, the estimators are centered around the true value. Precision decreases with the fraction of missing links, and when either $\mathbf{GX}$ or $\mathbf{Gy}$ is not observed. Comparing Figure 1 and Figure 2, we see that the inclusion of group-level fixed effects strongly affects precision. However, even when $\mathbf{GX}$ and $\mathbf{Gy}$ are both unobserved, and the estimator controls for unobserved group fixed effects, our SGMM estimator still achieves a reasonable precision level, even when *half* the links are missing.

Note that not observing $\mathbf{Gy}$ has a larger negative effect on precision than not

---

[17] We provide Monte Carlo simulations of the performance of our SGMM estimator for ARD in Section D.1.

[18] When all links are observed, $\dot{\mathbf{G}}^{(r)} = \ddot{\mathbf{G}}^{(s)} = \dddot{\mathbf{G}}^{(t)} = \mathbf{G}^{obs}$. See our discussion of the estimator above.

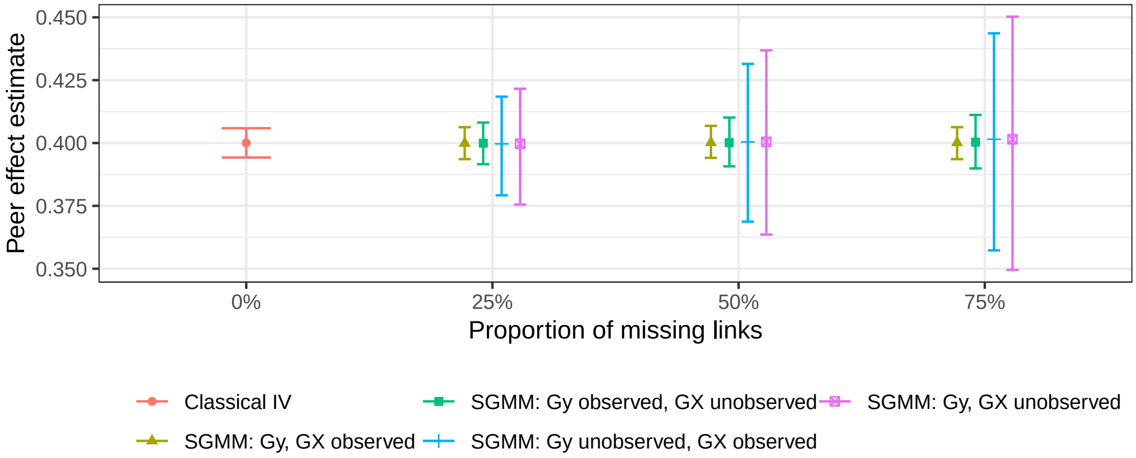

Figure 1: Estimated peer effect, no FE

Note: Dots represent estimated values of α, and bars represent 95% confidence intervals. Tables B.1 and B.2 in Appendix B present the full set of estimated coefficients. The "Classical IV" denotes the special case for which all links are observed (Bramoullé et al., 2009).
We simulate data for 100 groups of 30 individuals each. There are two observable characteristics, $x_{i1} \sim N(0,25)$ and $x_{i2} \sim Poisson(6)$. We assume that $\varepsilon_i \sim N(0,1)$. True parameter values are: $(\alpha, \boldsymbol{\beta}, \boldsymbol{\gamma}) = (0.4, 2, 1, 1.5, 5, -3)$. The network formation process follows the logistic regression: $a_{ij} \propto \exp\{\rho 1 + \rho_2|x_{i1} - x_{j1}| + \rho_3|x_{i2} - x_{j2}|\}$ where $\boldsymbol{\rho} = (0.8, -0.2, 0.1)$.

observing $\mathbf{GX}$. This is because when $\mathbf{Gy}$ is unobserved, the moment function includes the simulated value $\mathbf{y}$, which is a non-linear function of $\mathbf{G}$, i.e. the for the simulation s, $(\mathbf{I} - \alpha\ddot{\mathbf{G}}^{(s)})^{-1}\ddot{\mathbf{V}}^{(s)}\tilde{\boldsymbol{\theta}}$.

In conclusion, our SGMM performs well, even if a large fraction of links in the network are missing, provided that the baseline 2SLS estimator is precise. However, it may happen in practice (see Section 5) that even if all links were observed, the GMM approach is too imprecise to draw any conclusions. In such cases, we also provide a likelihood-based estimator, more efficient, but relying more heavily on parametric assumptions.

4 Likelihood-Based Estimation

In this section, we a likelihood-based estimator. Accordingly, greater structure must be imposed on the errors $\boldsymbol{\varepsilon}$. To clarify the exposition, we will focus on the network

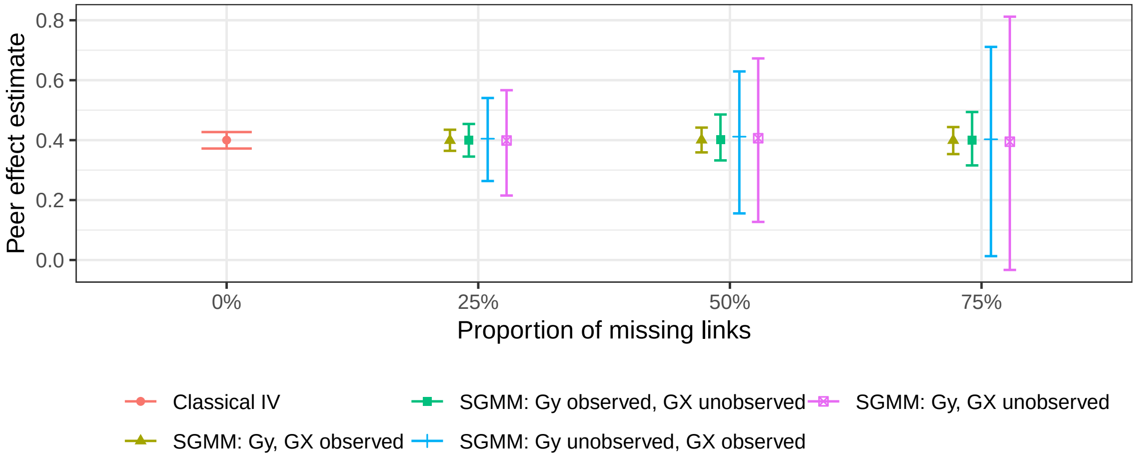

Figure 2: Estimated peer effect, FE

Note: See Figure 1.

adjacency matrix $\mathbf{A}$ rather than the interaction matrix $\mathbf{G}$. Of course, this is without any loss of generality. Given parametric assumptions for $\boldsymbol{\varepsilon}$, one can write the log-likelihood of the outcome as [19]

$$\ln \mathcal{P}(\mathbf{y}|\mathbf{A}, \boldsymbol{\theta}), \tag{7}$$

where $\boldsymbol{\theta} = [\alpha, \boldsymbol{\beta}', \boldsymbol{\gamma}', \boldsymbol{\sigma}']'$, and $\boldsymbol{\sigma}$ are unknown parameters from the distribution of $\boldsymbol{\varepsilon}$. Note that $\mathbf{y} = (\mathbf{I}_N - \alpha\mathbf{G})^{-1}(c\mathbf{1} + \mathbf{X}\boldsymbol{\beta} + \mathbf{G}\mathbf{X}\boldsymbol{\gamma} + \boldsymbol{\varepsilon})$ and $(\mathbf{I}_N - \alpha\mathbf{G})^{-1}$ exists under our Assumption 1.

If the adjacency matrix $\mathbf{A}$ is observed, then $\boldsymbol{\theta}$ could be estimated using a simple maximum likelihood estimator (as in Lee et al. (2010)) or using Bayesian inference (as in Goldsmith-Pinkham and Imbens (2013)).

Since $\mathbf{A}$ is not observed, an alternative would be to focus on the unconditional likelihood, i.e.

$$\ln \mathcal{P}(\mathbf{y}|\boldsymbol{\theta}) = \ln \sum_{\mathbf{A}} \mathcal{P}(\mathbf{y}|\mathbf{A}, \boldsymbol{\theta}) P(\mathbf{A}).$$

A similar strategy is proposed by Chandrasekhar and Lewis (2011) using a GMM estimator.

---

[19] Note that under Assumption 2, the likelihood can be factorized across groups.

One particular issue with estimating $\ln \mathcal{P}(\mathbf{y}|\boldsymbol{\theta})$ is that the summation is not tractable. Indeed, the sum is over the set of possible adjacency matrices, which contain $2^{N(N-1)}$ elements. Then, simply simulating networks from $P(\mathbf{A})$ (or rather from $\hat{P}(\mathbf{A})$) and taking the average likely lead to poor approximations. A classical way to address this issue is to use an EM algorithm (Dempster et al., 1977). Although valid, we found that the Bayesian estimator proposed in this section is less restrictive and numerically outperforms its classical counterpart.

For concreteness, we will assume that $\boldsymbol{\varepsilon} \sim \mathcal{N}(0, \sigma^2\mathbf{I}_N)$; however, it should be noted that our approach is valid for a number of alternative assumptions as long as it yields a computationally tractable likelihood. We have, for $\mathbf{G} = f(\mathbf{A})$,

$$\begin{aligned}\ln \mathcal{P}(\mathbf{y}|\mathbf{A}, \boldsymbol{\theta}) \quad = \quad & -N\ln(\sigma) + \ln|\mathbf{I}_N - \alpha\mathbf{G}| - \frac{N}{2}\ln(\pi) \\ & -\frac{1}{2\sigma^2}[(\mathbf{I}_N - \alpha\mathbf{G})\mathbf{y} - c\mathbf{1}_N - \mathbf{X}\boldsymbol{\beta} - \mathbf{G}\mathbf{X}\boldsymbol{\gamma}]' \cdot \\ & [(\mathbf{I}_N - \alpha\mathbf{G})\mathbf{y} - c\mathbf{1}_N - \mathbf{X}\boldsymbol{\beta} - \mathbf{G}\mathbf{X}\boldsymbol{\gamma}].\end{aligned}$$

Because $\mathbf{A}$ is not observed, we follow Tanner and Wong (1987) and Albert and Chib (1993), and we use data augmentation to evaluate the posterior distribution of $\boldsymbol{\theta}$. That is, instead of focusing on the posterior $p(\boldsymbol{\theta}|\mathbf{y}, \mathbf{A})$, we focus on the posterior $p(\boldsymbol{\theta}, \mathbf{A}|\mathbf{y}, \mathcal{A})$, treating $\mathbf{A}$ as another set of unknown parameters. Note that we now make the dependence on $\mathcal{A}$ explicit for clarity.

Indeed, the identification of the model rests on the a priori information of $\mathbf{A}$. A sensible prior for $\mathbf{A}$ is the consistent estimator of its distribution, i.e. $\hat{P}(\mathbf{A}|\mathcal{A}) \equiv P(\mathbf{A}|\hat{\boldsymbol{\rho}}, \mathcal{A})$. One may wish, however, to also use the information regarding the sampling uncertainty around $\hat{P}(\mathbf{A}|\mathcal{A})$. This is very similar to the inference for two-step estimators in a classical setting (e.g. for our SGMM estimator); estimation uncertainty in the first stage must be accounted for to provide valid inference in the second stage (see Section 3).

Let $\pi(\boldsymbol{\rho}|\mathcal{A})$ be the prior density on $\boldsymbol{\rho}$. How to obtain $\pi(\boldsymbol{\rho}|\mathcal{A})$, depending on whether $\hat{\boldsymbol{\rho}}$ is obtained using a Bayesian or classical setting, is discussed in Examples 5 and 6 below. Given $\pi(\boldsymbol{\rho}|\mathcal{A})$, it is possible to obtain draws from the posterior distribution $p(\boldsymbol{\rho}, \mathbf{A}|\mathbf{y}, \mathcal{A})$ using the following MCMC:

Algorithm 1. *The MCMC goes as follows for $t = 1, ..., T$, starting from any $\mathbf{A}_0, \boldsymbol{\theta}_0$, and $\boldsymbol{\rho}_0$.*

1. *Draw $\boldsymbol{\rho}^*$ from the proposal distribution $q_\rho(\boldsymbol{\rho}^*|\boldsymbol{\rho}_{t-1})$ and accept $\boldsymbol{\rho}^*$ with probability*

$$\min\left\{1, \frac{P(\mathbf{A}_{t-1}|\boldsymbol{\rho}^*, \mathcal{A})q_\rho(\boldsymbol{\rho}_{t-1}|\boldsymbol{\rho}^*)\pi(\boldsymbol{\rho}^*|\mathcal{A})}{P(\mathbf{A}_{t-1}|\boldsymbol{\rho}_{t-1}, \mathcal{A})q_\rho(\boldsymbol{\rho}^*|\boldsymbol{\rho}_{t-1})\pi(\boldsymbol{\rho}_{t-1}|\mathcal{A})}\right\}.$$

2. *Propose $\mathbf{A}^*$ from the proposal distribution $q_A(\mathbf{A}^*|\mathbf{A}_{t-1})$ and accept $\mathbf{A}^*$ with probability*

$$\min\left\{1, \frac{\mathcal{P}(\mathbf{y}|\boldsymbol{\theta}_{t-1}, \mathbf{A}^*)q_A(\mathbf{A}_{t-1}|\mathbf{A}^*)P(\mathbf{A}^*|\boldsymbol{\rho}_{t-1}, \mathcal{A})}{\mathcal{P}(\mathbf{y}|\boldsymbol{\theta}_{t-1}, \mathbf{A}_{t-1})q_A(\mathbf{A}^*|\mathbf{A}_{t-1})P(\mathbf{A}_{t-1}|\boldsymbol{\rho}_{t-1}, \mathcal{A})}\right\}.$$

3. *Draw α^* from the proposal $q_\alpha(\cdot|\alpha_{t-1})$ and accept α^* with probability*

$$\min\left\{1, \frac{\mathcal{P}(\mathbf{y}|\mathbf{A}_t; \boldsymbol{\beta}_{t-1}, \boldsymbol{\gamma}_{t-1}, \alpha^*)q_\alpha(\alpha_{t-1}|\alpha^*)\pi(\alpha^*)}{\mathcal{P}(\mathbf{y}|\mathbf{A}_t; \boldsymbol{\theta}_{t-1})q_\alpha(\alpha^*|\alpha_{t-1})\pi(\alpha_{t-1})}\right\}.$$

4. *Draw $[\boldsymbol{\beta}, \boldsymbol{\gamma}, \sigma]$ from their conditional distributions.*

As discussed, Step 1 accounts for the sampling uncertainty around the true value of $\boldsymbol{\rho}$. If the true value of $\boldsymbol{\rho}$ was known (instead of being estimated), Step 1 would not be required. Step 1 shows that the flexibility of the network formation model comes at a cost. For example, Graham (2017) and Breza et al. (2020) propose network formation models for which the number of parameters is $O(N_r)$. In turn, this large number of parameters increases the computational cost of Step 1.

Example 5 (Priors from the Asymptotic Distribution of $\boldsymbol{\rho}$)**.** *In a classical setting, and under the usual assumptions, the estimation of (2) produces an estimator $\hat{\boldsymbol{\rho}}$ of $\boldsymbol{\rho}_0$*

and an estimator of the asymptotic variance of $\hat{\boldsymbol{\rho}}$*, i.e.* $\hat{\mathbf{V}}(\hat{\boldsymbol{\rho}})$*. In this case, we define the prior density* $\pi(\boldsymbol{\rho})$ *as the density of a multivariate normal distribution with mean* $\hat{\boldsymbol{\rho}}$ *and variance–covariance matrix* $\hat{\mathbf{V}}(\hat{\boldsymbol{\rho}})$*.*

Example 6 (Priors from the Posterior Distribution of $\boldsymbol{\rho}$)**.** *In a Bayesian setting, the estimation of* $\boldsymbol{\rho}$ *from the network formation model (2) results in draws from the posterior distribution of* $\boldsymbol{\rho}$*. It is therefore natural to use such a posterior distribution as the prior distribution of* $\mathbf{A}$ *for the estimation based on (7). Performing such a sequential Bayesian updating approach comes with a well-known numerical issue.*[20]

Indeed, the evaluation of the acceptance ratio in Step 1 of Algorithm 1 below requires the evaluation of the density of $\boldsymbol{\rho}$ *at different values. Ideally, one would use the draws from the posterior distribution of* $\boldsymbol{\rho}$ *from the first step (network formation model) and perform a nonparametric kernel density estimation of the posterior distribution. However, when the dimension of* $\boldsymbol{\rho}$ *is large, the kernel density estimation may be infeasible in practice.*

This is especially true for very flexible network formation models, such as that proposed by Breza et al. (2020) for which the number of parameters to estimate is $O(N_r)$*. In such a case, it might be more reasonable to use a more parametric approach or to impose additional restrictions on the dependence structure of* $\boldsymbol{\rho}$ *across dimensions.*[21]

Detailed distributions for Steps 3 and 4 can be found in the Online Appendix. Step 2, however, requires some discussion. Indeed, the idea is the following: given the prior information $P(\mathbf{A}|\boldsymbol{\rho}_{t-1}, \mathcal{A})$, one must be able to draw samples from the posterior distribution of $\mathbf{A}$, given $\mathbf{y}$. This is not a trivial task.

In particular, there is no general rule for selecting the network proposal distribution $q_A(\cdot|\cdot)$. A natural candidate is a Gibbs sampling algorithm for each link, i.e.

---

[20] See Thijssen and Wessels (2020) for a recent discussion.

[21] For example, if we assume that the posterior distribution of $\boldsymbol{\rho}$ is jointly normal, the estimation of the mean and variance–covariance matrix is straightforward, even in a high-dimensional setting. Simulations suggest that this approach performs well in practice. See the Vignette accompagning our R package.

change only one link ij at every step t and propose a_{ij} according to its marginal distribution, i.e. $a_{ij} \sim P(\cdot|\mathbf{A}_{-ij}, \mathbf{y}, \mathcal{A})$, where $\mathbf{A}_{-ij} = \{a_{kl}; k \neq i, l \neq j\}$. In this case, the proposal is always accepted.

However, it has been argued that Gibbs sampling could lead to slow convergence (e.g. Snijders (2002), Chatterjee et al. (2013)), especially when the network is *sparse* or exhibits a high level of *clustering*. For example, Mele (2017) and Bhamidi et al. (2008) propose different blocking techniques meant to improve convergence.

Here, however, achieving Step 2 involves an additional computational issue because evaluating the likelihood ratio in Step 1 requires comparing the determinants $|\mathbf{I} - \alpha f(\mathbf{A}^*)|$ for each proposed $\mathbf{A}^*$, which is computationally intensive. In particular, taking $\mathbf{G}^* = f(\mathbf{A}^*)$ to be a row-normalization of $\mathbf{A}^*$, altering a single element of $\mathbf{A}^*$ results in a change in the entire corresponding row of $\mathbf{G}^*$. Still, comparing the determinant of two matrices that differ only in a single row is relatively fast. Moreover, when $\mathbf{G} = \mathbf{A}$, Hsieh et al. (2019) propose a blocking technique that facilitates the computation of the determinant.

In any case, note that the computational complexity of Step 2 depends strongly on $P(\mathbf{A}|\boldsymbol{\rho}_{t-1}, \mathcal{A})$, which is a function of the assumed network formation model (2) and the observed information about the network structure $\mathcal{A}$. For censored network data, for example, most of the network structure is observed (see Example 2). This implies that $P(a_{ij}|\boldsymbol{\rho}_{t-1}, \mathcal{A}) \in \{0, 1\}$ for most pairs ij. As such, few entries of $\mathbf{A}$ must be updated in Step 2. The opposite is true for ARD (see Example 4) for which all entries of $\mathbf{A}$ must be updated.

Then, the appropriate blocking technique depends strongly on $P(\mathbf{A}|\boldsymbol{\rho}_{t-1}, \mathcal{A})$ and the assumed distribution for $\boldsymbol{\varepsilon}$. For the simulations and estimations presented in this paper, we use the Gibbs sampling algorithm for each link, adapting the strategy proposed by Hsieh et al. (2019) to our setting (see Proposition 3 in the Online Appendix). This can be viewed as a *worst-case* scenario. Nonetheless, the Gibbs sampler performs reasonably well in practice, even for ARD;[22] however, we encourage researchers to try other updating schemes if Gibbs sampling performs poorly in their specific contexts. In particular, we present a blocking technique in the Online Appendix G that is also implemented in our R package `PartialNetwork`.[23]

It is important to note that the complexity of Step 2 is not limited to our Bayesian approach. Classical estimators, such as GMM estimators, face a similar challenge in requiring the integration over the entire set of networks. The strategy used here is to rely on a Metropolis–Hastings algorithm, a strategy that has also been successfully implemented in the related literature on ERGMs (e.g. Snijders (2002); Mele (2017, 2020); Badev (2018); Hsieh et al. (2019, 2020)).

Finally, note that for simple network formation models, it is possible to jointly estimate $\boldsymbol{\rho}$ and $\boldsymbol{\theta}$ within the same MCMC instead of using the two-step procedure described above. In this case, Step 1 can simply be replaced by:

1'. Draw $\boldsymbol{\rho}^*$ from the proposal distribution $q_\rho(\boldsymbol{\rho}^*|\boldsymbol{\rho}_{t-1})$ and accept $\boldsymbol{\rho}^*$ with probability
$$\min\left\{1, \frac{P(\mathbf{A}_{t-1}|\boldsymbol{\rho}^*, \mathcal{A})P(\mathcal{A}|\boldsymbol{\rho}^*)q_\rho(\boldsymbol{\rho}_{t-1}|\boldsymbol{\rho}^*)\pi(\boldsymbol{\rho}^*)}{P(\mathbf{A}_{t-1}|\boldsymbol{\rho}_{t-1}, \mathcal{A})P(\mathcal{A}|\boldsymbol{\rho}_{t-1})q_\rho(\boldsymbol{\rho}^*|\boldsymbol{\rho}_{t-1})\pi(\boldsymbol{\rho}_{t-1})}\right\}.$$

Here, $P(\mathcal{A}|\boldsymbol{\rho}^*)$ is the likelihood of the network information $\mathcal{A}$ assuming the network formation model in (2). Note that $\pi(\boldsymbol{\rho})$, the prior density on $\boldsymbol{\rho}$, no longer depends on $\mathcal{A}$ and can be chosen arbitrarily (e.g. uniform).

This approach would work well for simple models, such as the ones discussed in Examples 1 and 3. It is impractical, however, for more involved models, such as that proposed by Breza et al. (2020).

---

[22] Simulations available upon request.

[23] Available at: `https://github.com/ahoundetoungan/PartialNetwork`

5 Imperfectly Measured Networks

In this section, we assume that the econometrician has access to network data but that the data may contain errors. To show how our method can be used to address these issues, we consider a simple example where we are interested in estimating peer effects on adolescents' academic achievements.

We use the widely used AddHealth database and show that network data errors have a first-order impact on the estimated peer effects. Specifically, we focus on a subset of schools from the "In School" sample that each have less than 200 students. Table 1 displays the summary statistics.

Table 1: Summary statistics

Statistic	Mean	Std. Dev.	Pctl(25)	Pctl(75)
Female	0.540	0.498	0	1
Hispanic	0.157	0.364	0	0
Race				
White	0.612	0.487	0	1
Black	0.246	0.431	0	0
Asian	0.022	0.147	0	0
Other	0.088	0.283	0	0
Mother's education				
High	0.310	0.462	0	1
<High	0.193	0.395	0	0
>High	0.358	0.480	0	1
Missing	0.139	0.346	0	0
Mother's job				
Stay-at-home	0.225	0.417	0	0
Professional	0.175	0.380	0	0
Other	0.401	0.490	0	1
Missing	0.199	0.399	0	0
Age	13.620	1.526	13	14
GPA	2.912	0.794	2.333	3.5

Note: We only keep the 33 schools having less than 200 students from the In-School sample. The variable GPA is computed by taking the average grade for English, Mathematics, History, and Science, letting $A = 4$, $B = 3$, $C = 2$, and $D = 1$. Thus, higher scores indicate better academic achievement.

Most papers estimating peer effects that use this particular database have taken the network structure as given. One notable exception is Griffith (2019), looking at censoring: students can only report up to five male and five female friends. We also allow for censoring but show that censoring is not the most important issue with the Add Health data. To understand why, we discuss the organization of the data.

Each adolescent is assigned a unique identifier. The data includes ten variables for the ten potential friendships (maximum of five male and five female friends). These variables can contain missing values (no friendship was reported), an error code (the named friend could not be found in the database), or an identifier for the reported friends. This data is then used to generate the network's adjacency matrix $\mathbf{A}$.

Of course, error codes cannot be matched to any particular adolescent. Moreover, even in the case where the friendship variable refers to a valid identifier, the referred adolescent may still be absent from the database. A prime example is when the referred adolescent has been removed from the database by the researcher, perhaps because of other missing variables for these particular individuals. These missing links are quantitatively important as they account for roughly 30% of the total number of links (7,830 missing for 17,993 observed links). Figure 3 displays the distribution of the number of "unmatched named friends."[24]

To use the methodology developed in Section 4, we first need to estimate a network formation model using the observed network data. In this section, we assume that links are generated using a simple logistic framework, i.e.

$$P(a_{ij} = 1) \propto \exp\{\mathbf{w}_{ij}\boldsymbol{\rho}\},$$

where $\mathbf{w}_{ij}$ is built to capture homophily on the observed characteristics of i and j (see Tables C.1 and C.2 in Appendix C). We estimate the network formation model on the set of individuals for which we observe no "unmatched friends." For these

---

[24] We focus on within-school friendships; thus, nominations outside of school are not treated as "unmatched friends."

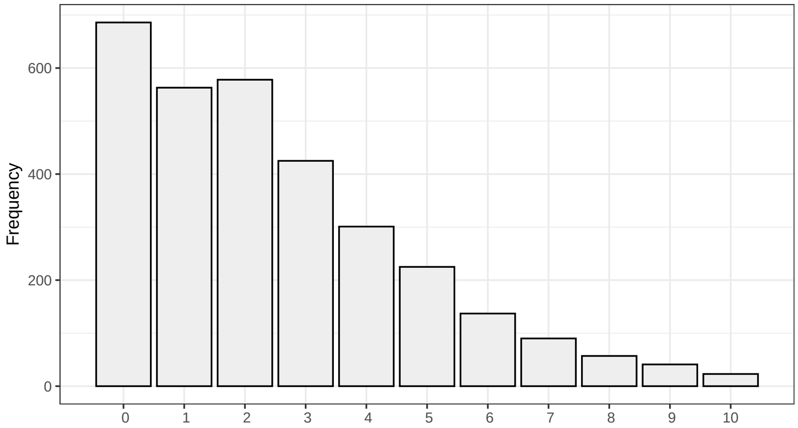

Figure 3: Frequencies of the number of missing links per adolescent

students, we know for sure that their friendship data is complete. Under a missing at-random assumption, the estimation of the explanatory variables is consistent, but the intercept is affected by a selection bias. We control for this by weighting the log-likelihood of the network following Manski and Lerman (1977). The details are presented in Appendix C.

We present the estimation results for the SGMM and for the Bayesian estimator. Figure 4 summarizes the results for the endogenous peer effect coefficient α, whereas the full set of results is presented in Appendix C. The first two estimations (*Obsv.Bayes* and *Obsv.SGMM*) assume that the observed network is the true network for both estimators. The third and fourth estimations (*Miss.Bayes* and *Miss.SGMM*) account for missing data due to error codes but not for censoring. The last two estimations (*TopMiss.Bayes* and *TopMiss.SGMM*) account for missing data due to error codes or censoring.

We first see that the SGMM is less efficient than the Bayesian. This should not be surprising since the Bayesian estimator uses more structure (in particularity homoscedastic, normally distributed errors). When we compare the estimations *Obsv.SGMM* and *Miss.SGMM*, the observed differences imply that the efficiency loss is because of the relative inefficiency of the GMM approach, and not of the missing

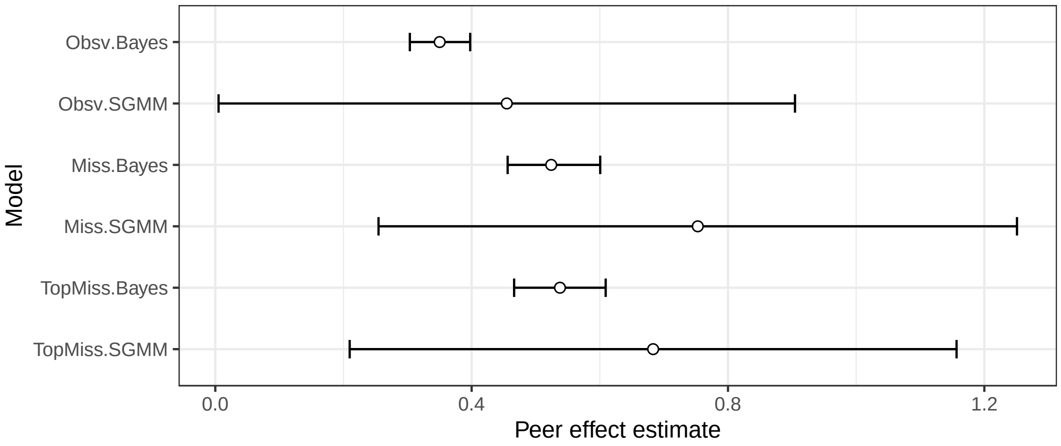

Figure 4: Peer effect estimate

Note: Dots represent estimated values (and posterior mean) of α, and bars represent 95% confidence intervals (and 95% credibility intervals). Tables C.1 and C.2 in Appendix C present the full set of estimated coefficients.

links or specifically of our SGMM estimator.[25]

Importantly, we see that the bias due to the assumption that the network is fully observed is quantitatively and qualitatively important. Using either estimator, the estimated endogenous peer effect using the reconstructed network is 1.5 times larger than thatestimated assuming the observed network is the true network.[26] Almost all of the bias is produced by the presence of error codes and not because of potential censoring.

This exercise shows that data errors are a first-order concern when using the Add Health database.[27] Not only does the bias in the endogenous peer effect coefficient α have an impact on the social multiplier (Glaeser et al., 2003), but it can also affect the anticipated effect of targeted interventions, i.e. the identity of the key player (Ballester et al., 2006).

---

[25] Recall that when the network is observed, our SGMM uses the same moment conditions as, for example, those suggested by Bramoullé et al. (2009).

[26] The difference is “statistically significant” for the Bayesian estimator.

[27] We do not argue, however, that our estimated coefficients are necessarily causal because the friendship network is likely endogenous (e.g. Goldsmith-Pinkham and Imbens (2013); Hsieh and Van Kippersluis (2018); Hsieh et al. (2020)). The estimation of peer effects with partial endogenous network data is left for future research.

Indeed, Figure 5 shows a scatter plot of the vector of centralities in the observed and reconstructed networks. The figure illustrates the effects of missing network data. First, because the reconstructed network has more links, centrality is higher on average. This is essentially the social multiplier effect. Not accounting for missing links leads to an underestimation of spillover effects. Second, some individuals, in particular those having very few links in the observed network, are in reality highly central. Therefore, targeting a policy at individuals having a high centrality in the observed network would be inefficient. In particular, Figure 6 shows that even isolated individuals and individuals interacting in isolated pairs in the observed network (having centralities of 1 and 1.35 respectively) can be, in reality, highly central. Thus, a policy based on the evaluation of an observed network, coupled with the associated endogenous peer effect coefficient α, would not only underestimate the social multiplier but would also target the wrong individuals.

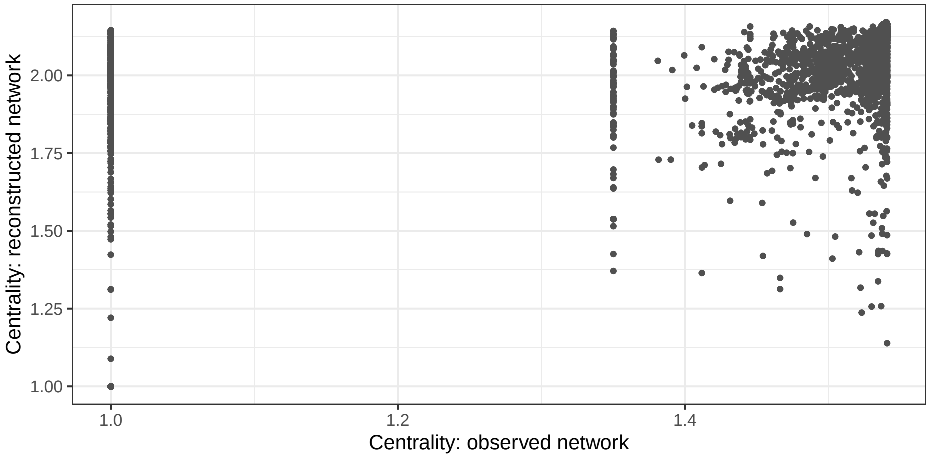

Figure 5: Centrality

Note: The centrality vector is given by $(\mathbf{I} - \hat{\alpha}\mathbf{G})^{-1}\mathbf{1}$. To compute centrality based on the observed network, we use the observed network $\mathbf{G}$ and the $\hat{\alpha}$ estimated using specification *Obsv.Bayes*. To compute centrality based on the reconstructed network, we use $\hat{\alpha}$ and $\mathbf{G}$ estimated using the specification *TopMiss.Bayes*. For both centrality vectors, we use the average vector centrality across 10 000 draws from their respective posterior distributions.

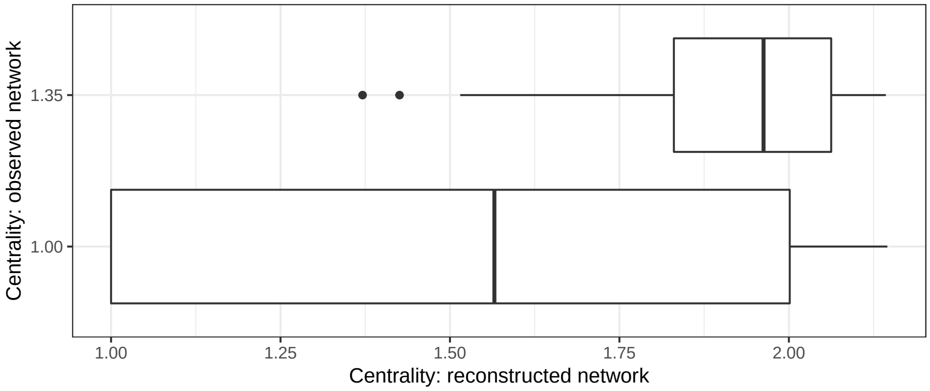

Figure 6: Centrality

Note: See note of Figure 5.

6 Conclusion

In this paper, we proposed two estimators where peer effects can be estimated without having knowledge of the entire network structure. We found, perhaps surprisingly, that even very partial information on network structure is sufficient. By specifying a network formation model, researchers can probabilistically reconstruct the true network and base the estimation of peer effects on this reconstructed network. Importantly, we provide computationally tractable and flexible estimators to do so, all of which are available in our R package `PartialNetwork`. We apply our methodology to the widely used AddHealth data and find that missing links due to noise in the data have first-order effects on the estimated peer effect coefficient. This leads to an underestimation of the social multiplier and the incorrect identification of the key players.

References

Albert, J. H. and S. Chib (1993): "Bayesian analysis of binary and polychotomous response data," *Journal of the American Statistical Association*, 88, 669–679.

Alidaee, H., E. Auerbach, and M. P. Leung (2020): "Recovering network structure from aggregated relational Data using Penalized Regression," *arXiv preprint arXiv:2001.06052*.

Andrews, D. W. (1994): "Empirical process methods in econometrics," *Handbook of Econometrics*, 4, 2247–2294.

Andrews, I., M. Gentzkow, and J. M. Shapiro (2017): "Measuring the sensitivity of parameter estimates to estimation moments," *The Quarterly Journal of Economics*, 132, 1553–1592.

Atchadé, Y. F. and J. S. Rosenthal (2005): "On adaptive Markov chain Monte Carlo algorithms," *Bernoulli*, 11, 815–828.

Auerbach, E. (2019): "Measuring differences in stochastic network structure," *arXiv preprint arXiv:1903.11117*.

Badev, A. (2018): "Nash equilibria on (un) stable networks," *arXiv preprint arXiv:1901.00373*.

Ballester, C., A. Calvó-Armengol, and Y. Zenou (2006): "Who's who in networks. Wanted: The key player," *Econometrica*, 74, 1403–1417.

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

Bhamidi, S., G. Bresler, and A. Sly (2008): "Mixing time of exponential random graphs," in *2008 49th Annual IEEE Symposium on Foundations of Computer Science*, IEEE, 803–812.

BOUCHER, V. AND Y. BRAMOULLÉ (2020): "Binary outcomes and linear interactions," Tech. rep., Aix-Marseille School of Economics, France.

BOUCHER, V. AND B. FORTIN (2016): "Some challenges in the empirics of the effects of networks," *The Oxford Handbook on the Economics of Networks*, 277–302.

BOUCHER, V. AND I. MOURIFIÉ (2017): "My friend far, far away: a random field approach to exponential random graph models," *The Econometrics Journal*, 20, S14–S46.

BOUND, J., C. BROWN, AND N. MATHIOWETZ (2001): "Measurement error in survey data," in *Handbook of Econometrics*, Elsevier, vol. 5, 3705–3843.

BRAMOULLÉ, Y., S. CURRARINI, M. O. JACKSON, P. PIN, AND B. W. ROGERS (2012): "Homophily and long-run integration in social networks," *Journal of Economic Theory*, 147, 1754–1786.

BRAMOULLÉ, Y., H. DJEBBARI, AND B. FORTIN (2009): "Identification of peer effects through social networks," *Journal of Econometrics*, 150, 41–55.

——— (2020): "Peer effects in networks: A survey," *Annual Review of Economics*, 12, 603–629.

BREZA, E. (2016): "Field experiments, social networks, and development," *The Oxford Handbook on the Economics of Networks*, 412–439.

BREZA, E., A. G. CHANDRASEKHAR, T. H. MCCORMICK, AND M. PAN (2019): "Consistently estimating graph statistics using Aggregated Relational Data," *arXiv preprint arXiv:1908.09881*.

——— (2020): "Using aggregated relational data to feasibly identify network structure without network data," *American Economic Review*, 110, 2454–84.

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

Cameron, A. C. and P. K. Trivedi (2005): *Microeconometrics: methods and applications*, Cambridge University Press.

Chandrasekhar, A. and R. Lewis (2011): "Econometrics of sampled networks," *Unpublished manuscript, MIT.[422].*

Chatterjee, S., P. Diaconis, et al. (2013): "Estimating and understanding exponential random graph models," *The Annals of Statistics*, 41, 2428–2461.

Chen, X., Y. Chen, and P. Xiao (2013): "The impact of sampling and network topology on the estimation of social intercorrelations," *Journal of Marketing Research*, 50, 95–110.

Chen, X., H. Hong, and D. Nekipelov (2011): "Nonlinear models of measurement errors," *Journal of Economic Literature*, 49, 901–37.

Chib, S. and S. Ramamurthy (2010): "Tailored randomized block MCMC methods with application to DSGE models," *Journal of Econometrics*, 155, 19–38.

Conley, T. G. and C. R. Udry (2010): "Learning about a new technology: Pineapple in Ghana," *American Economic Review*, 100, 35–69.

Currarini, S., M. O. Jackson, and P. Pin (2010): "Identifying the roles of race-based choice and chance in high school friendship network formation," *Proceedings of the National Academy of Sciences*, 107, 4857–4861.

de Castro, L., A. F. Galvao, D. M. Kaplan, and X. Liu (2019): "Smoothed GMM for quantile models," *Journal of Econometrics*, 213, 121–144.

De Paula, A. (2017): "Econometrics of network models," in *Advances in Economics and Econometrics: Theory and Applications: Eleventh World Congress (Econo-*

metric Society Monographs, ed. by M. P. B. Honore, A. Pakes and L. Samuelson, Cambridge: Cambridge University Press, 268–323.

De Paula, Á., I. Rasul, and P. Souza (2018): "Recovering social networks from panel data: identification, simulations and an application," *LACEA Working Paper Series*.

Dempster, A. P., N. M. Laird, and D. B. Rubin (1977): "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society: Series B (Methodological)*, 39, 1–22.

Glaeser, E. L., B. I. Sacerdote, and J. A. Scheinkman (2003): "The social multiplier," *Journal of the European Economic Association*, 1, 345–353.

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

Gourieroux, M., C. Gourieroux, A. Monfort, D. A. Monfort, et al. (1996): *Simulation-based econometric methods*, Oxford University Press.

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

Griffith, A. (2019): "Name your friends, but only five? The importance of censoring in peer effects estimates using social network data," *Working Paper*.

Griffith, A. and J. Kim (2023): "The Impact of Missing Links on Linear Reduced-form Network-Based Peer Effects Estimates," *Working Paper*.

Hardy, M., R. M. Heath, W. Lee, and T. H. McCormick (2019): "Estimating spillovers using imprecisely measured networks," *arXiv preprint arXiv:1904.00136*.

Hausman, J. A., J. Abrevaya, and F. M. Scott-Morton (1998): “Misclassification of the dependent variable in a discrete-response setting,” *Journal of Econometrics*, 87, 239–269.

Hoff, P. D., A. E. Raftery, and M. S. Handcock (2002): “Latent space approaches to social network analysis,” *Journal of the American Statistical Association*, 97, 1090–1098.

Horowitz, J. L. (1998): “Bootstrap methods for median regression models,” *Econometrica*, 1327–1351.

Hsieh, C.-S., S. I. Ko, J. Kovářík, and T. Logan (2018): “Non-randomly sampled networks: biases and corrections,” Tech. rep., National Bureau of Economic Research.

Hsieh, C.-S., M. D. König, and X. Liu (2019): “A structural model for the coevolution of networks and behavior,” *Review of Economics and Statistics*, 1–41.

Hsieh, C.-S. and L. F. Lee (2016): “A social interactions model with endogenous friendship formation and selectivity,” *Journal of Applied Econometrics*, 31, 301–319.

Hsieh, C.-S., L.-F. Lee, and V. Boucher (2020): “Specification and estimation of network formation and network interaction models with the exponential probability distribution,” *Quantitative Economics*, 11, 1349–1390.

Hsieh, C.-S. and H. Van Kippersluis (2018): “Smoking initiation: Peers and personality,” *Quantitative Economics*, 9, 825–863.

Johnson, C. R. and R. A. Horn (1985): *Matrix analysis*, Cambridge University Press.

Kaplan, D. M. and Y. Sun (2017): “Smoothed estimating equations for instrumental variables quantile regression,” *Econometric Theory*, 33, 105–157.

Kelejian, H. H. and G. Piras (2014): "Estimation of spatial models with endogenous weighting matrices, and an application to a demand model for cigarettes," *Regional Science and Urban Economics*, 46, 140–149.

König, M. D., X. Liu, and Y. Zenou (2019): "R&D networks: Theory, empirics, and policy implications," *Review of Economics and Statistics*, 101, 476–491.

Lee, L.-F. (2004): "Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models," *Econometrica*, 72, 1899–1925.

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

Lee, L.-f., X. Liu, E. Patacchini, and Y. Zenou (2020): "Who is the key player? A network analysis of juvenile delinquency," *Journal of Business and Economic Statistics, forthcoming.*

Lewbel, A., X. Qu, and X. Tang (2022a): "Estimating social network models With missing links," *Working Paper.*

——— (2022b): "Ignoring Measurement Errors in Social Networks," *Working Paper.*

——— (2022c): "Social networks with unobserved links," *Journal of Political Economy, Forthcoming.*

Liu, X. (2013): "Estimation of a local-aggregate network model with sampled networks," *Economics Letters*, 118, 243–246.

Liu, X., E. Patacchini, and E. Rainone (2017): "Peer effects in bedtime decisions among adolescents: a social network model with sampled data," *The Econometrics Journal*, 20, S103–S125.

Manresa, E. (2016): "Estimating the structure of social interactions using panel data," *Working paper*.

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

Manski, C. F. and S. R. Lerman (1977): "The estimation of choice probabilities from choice based samples," *Econometrica: Journal of the Econometric Society*, 1977–1988.

McCormick, T. H. and T. Zheng (2015): "Latent surface models for networks using Aggregated Relational Data," *Journal of the American Statistical Association*, 110, 1684–1695.

McFadden, D. (1989): "A method of simulated moments for estimation of discrete response models without numerical integration," *Econometrica: Journal of the Econometric Society*, 995–1026.

Mele, A. (2017): "A structural model of Dense Network Formation," *Econometrica*, 85, 825–850.

——— (2020): "Does school desegregation promote diverse interactions? An equilibrium model of segregation within schools," *American Economic Journal: Economic Policy*, 12, 228–57.

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

Onishi, R. and T. Otsu (2021): "Sample sensitivity for two-step and continuous updating GMM estimators," *Economics Letters*, 198, 109685.

Snijders, T. A. (2002): "Markov chain Monte Carlo estimation of exponential random graph models," *Journal of Social Structure*, 3, 1–40.

Souza, P. (2014): "Estimating network effects without network data," *PUC-Rio Working Paper*.

Tanner, M. A. and W. H. Wong (1987): "The calculation of posterior distributions by data augmentation," *Journal of the American Statistical Association*, 82, 528–540.

Thijssen, B. and L. F. Wessels (2020): "Approximating multivariate posterior distribution functions from Monte Carlo samples for sequential Bayesian inference," *PloS ONE*, 15, e0230101.

Thirkettle, M. (2019): "Identification and estimation of network statistics with missing link data," *Working Paper*.

Wang, W. and L.-F. Lee (2013): "Estimation of spatial autoregressive models with randomly missing data in the dependent variable," *The Econometrics Journal*, 16, 73–102.

Zhang, L. (2020): "Spillovers of program benefits with mismeasured networks," *arXiv preprint arXiv:2009.09614*.

A SGMM estimator

Throughout, otherwise stated, the subscript zero is used to denote the true value of a parameter. Consider the following regularity assumptions.

Assumption 6. *$\boldsymbol{\rho}_0 \in int(\mathcal{R})$ and $\boldsymbol{\theta}_0 \in int(\Theta)$, where $\mathcal{R}$ and Θ are both compact subsets of the Euclidean space.*

Assumption 7. *For all $r = 1, ..., R$, $s = 1, ..., S$, and $t = 1, ..., T$, $(\mathbf{I} - \alpha\dot{\mathbf{G}}^r)$, $(\mathbf{I} - \alpha\ddot{\mathbf{G}}^s)$ and $(\mathbf{I} - \alpha\dddot{\mathbf{G}}^t)$ are non-singular*

In particular, when $\mathbf{G}$ is row-normalized (so $\dot{\mathbf{G}}^r$, $\ddot{\mathbf{G}}^s$, and $\dddot{\mathbf{G}}^t$ are also row-normalized), Assumption 1 implies Assumption 7. Assumption 8 below ensures that the moment function is uniformly bounded.

Assumption 8. *$\mathbf{X}_i'$ is uniformly bounded in i. The (i, j)-th entries of $\mathbf{G}$ (so $\dot{\mathbf{G}}^r$, $\ddot{\mathbf{G}}^s$, and $\dddot{\mathbf{G}}^t$) are uniformly bounded in i and j.*

We consider an objective function for the GMM estimator of the usual form:

$$\mathcal{Q}_N(\boldsymbol{\theta}) = \left(\tfrac{1}{N}\textstyle\sum_i \mathbf{m}_i(\boldsymbol{\theta})\right)' \mathbf{W}_N \left(\tfrac{1}{N}\textstyle\sum_i \mathbf{m}_i(\boldsymbol{\theta})\right),$$

where $\mathbf{m}_i(\boldsymbol{\theta})$ is the moment function and $\mathbf{W}_N$ is positive definite.

Assumption 9. *$\mathcal{Q}_N(\boldsymbol{\theta})$ converges in probability to some $\mathcal{Q}_0(\boldsymbol{\theta})$.*

The identification condition is presented in Section A.1.1 (see in particular Assumptions 10 and 11).

A.1 Proof of Theorem 1

Let

$$\mathbf{m}_i(\boldsymbol{\theta}) = \tfrac{1}{RST}\textstyle\sum_r\sum_s\sum_t \dot{\mathbf{Z}}_i^{(r)\prime}(\mathbf{I} - \alpha\dddot{\mathbf{G}}^{(t)})_i \left(\mathbf{y} - (\mathbf{I} - \alpha\ddot{\mathbf{G}}^{(s)})^{-1}\ddot{\mathbf{V}}^{(s)}\tilde{\boldsymbol{\theta}}\right).$$

We will show that, by construction, we have $\operatorname{plim}\mathbb{E}\left(\mathbf{m}_i(\boldsymbol{\theta}_0)\right)=\mathbf{0}$ for any finite S and T, where plim is the probability limit.

To see why, let us substitute $\mathbf{y}=(\mathbf{I}-\alpha_0\mathbf{G})^{-1}(\mathbf{V}\tilde{\boldsymbol{\theta}}_0+\boldsymbol{\varepsilon})$ in the moment function. We have

$$\begin{aligned}\mathbf{m}_i(\boldsymbol{\theta}_0) &= \tfrac{1}{RST}\textstyle\sum_r\sum_s\sum_t \dot{\mathbf{Z}}_i^{(r)\prime}(\mathbf{I}-\alpha_0\dddot{\mathbf{G}}^{(t)})_i\left[(\mathbf{I}-\alpha_0\mathbf{G})^{-1}\mathbf{V}-(\mathbf{I}-\alpha_0\ddot{\mathbf{G}}^{(s)})^{-1}\ddot{\mathbf{V}}^{(s)}\right]\tilde{\boldsymbol{\theta}}_0\\ &+ \tfrac{1}{RS}\textstyle\sum_r\sum_s \dot{\mathbf{Z}}_i^{(r)\prime}(\mathbf{I}-\alpha_0\dddot{\mathbf{G}}^{(t)})_i(\mathbf{I}-\alpha_0\mathbf{G})^{-1}\boldsymbol{\varepsilon}.\end{aligned}$$

Consider the second term first. For any $r,\ s,\ t$, we have

$$\mathbb{E}\left(\dot{\mathbf{Z}}_i^{(r)\prime}(\mathbf{I}-\alpha_0\dddot{\mathbf{G}}^{(t)})_i(\mathbf{I}-\alpha_0\mathbf{G})^{-1}\boldsymbol{\varepsilon}\right)=\mathbf{0},$$

directly from Assumption 4 and the specification of the network formation process (see Assumption 5).

Consider now the first term. We need

$$\operatorname{plim}\mathbb{E}\left[\dot{\mathbf{Z}}_i^{(r)\prime}(\mathbf{I}-\alpha_0\dddot{\mathbf{G}}^{(t)})_i\left[(\mathbf{I}-\alpha_0\mathbf{G})^{-1}\mathbf{V}-(\mathbf{I}-\alpha_0\ddot{\mathbf{G}}^{(s)})^{-1}\ddot{\mathbf{V}}^{(s)}\right]\tilde{\boldsymbol{\theta}}_0\right]=\mathbf{0}.$$

Note that, for N fixed, $\mathbf{G}$, $\dot{\mathbf{G}}^{(r)}$, $\ddot{\mathbf{G}}^{(s)}$, and $\dddot{\mathbf{G}}^{(t)}$ have a finite number of possible realizations. Thus, $\mathbb{E}\left[\dot{\mathbf{Z}}_i^{(s)\prime}(\mathbf{I}-\alpha_0\dddot{\mathbf{G}}^{(t)})_i(\mathbf{I}-\alpha_0\mathbf{G})^{-1}\mathbf{V}\tilde{\boldsymbol{\theta}}_0|\mathbf{X}\right]$ can be written as a continuous function of P, the distribution of the adjacency matrix, and $\hat{P}$, a consistent estimator of that distribution.[28] Analogously, $\mathbb{E}\left[\dot{\mathbf{Z}}_i^{(s)\prime}(\mathbf{I}-\alpha_0\dddot{\mathbf{G}}^{(t)})_i(\mathbf{I}-\alpha_0\ddot{\mathbf{G}}^{(s)})^{-1}\ddot{\mathbf{V}}^{(s)}\tilde{\boldsymbol{\theta}}_0|\mathbf{X}\right]$ also can be written as a same function with the difference that P is replaced by $\hat{P}$. As both functions are continuous and because $\hat{P}$ converges in probability to P (see Assumption 5), their limits are obtained by replacing $\hat{P}$ by P and are therefore equal.

---

[28] The function would essentially composed of summations and products of terms depending on P and $\hat{P}$.

As such, we have

$$\operatorname{plim} \mathbb{E}\left[\dot{\mathbf{Z}}_i^{(s)\prime}(\mathbf{I}-\alpha_0\dddot{\mathbf{G}}^{(t)})_i(\mathbf{I}-\alpha_0\mathbf{G})^{-1}\mathbf{V}\tilde{\boldsymbol{\theta}}_0|\mathbf{X}\right] = \operatorname{plim} \mathbb{E}\left[\dot{\mathbf{Z}}_i^{(s)\prime}(\mathbf{I}-\alpha_0\dddot{\mathbf{G}}^{(t)})_i(\mathbf{I}-\alpha_0\ddot{\mathbf{G}}^{(s)})^{-1}\ddot{\mathbf{V}}^{(s)}\tilde{\boldsymbol{\theta}}_0|\mathbf{X}\right].$$

By taking the expectation of the terms in both hand sides of the previous equation, we can interchange the expectation and plim terms. Indeed, both $\mathbf{E}_{n,1} = \mathbb{E}\big[\dot{\mathbf{Z}}_i^{(s)\prime}(\mathbf{I}-\alpha_0\dddot{\mathbf{G}}^{(t)})_i(\mathbf{I}-\alpha_0\mathbf{G})^{-1}\mathbf{V}\tilde{\boldsymbol{\theta}}_0|\mathbf{X}\big]$ and $\mathbf{E}_{n,2} = \mathbb{E}\big[\dot{\mathbf{Z}}_i^{(s)\prime}(\mathbf{I}-\alpha_0\dddot{\mathbf{G}}^{(t)})_i(\mathbf{I}-\alpha_0\ddot{\mathbf{G}}^{(s)})^{-1}\ddot{\mathbf{V}}^{(s)}\tilde{\boldsymbol{\theta}}_0|\mathbf{X}\big]$ are bounded by assumption 8. The dominated convergence theorem implies that $\mathbb{E}(\operatorname{plim}\mathbf{E}_{n,1}) = \operatorname{plim}\mathbb{E}(\mathbf{E}_{n,1})$ and $\mathbb{E}(\operatorname{plim}\mathbf{E}_{n,2}) = \operatorname{plim}\mathbb{E}(\mathbf{E}_{n,2})$. As a result,

$$\operatorname{plim} \mathbb{E}\left[\dot{\mathbf{Z}}_i^{(s)\prime}(\mathbf{I}-\alpha_0\dddot{\mathbf{G}}^{(t)})_i(\mathbf{I}-\alpha_0\mathbf{G})^{-1}\mathbf{V}\tilde{\boldsymbol{\theta}}_0\right] = \operatorname{plim} \mathbb{E}\left[\dot{\mathbf{Z}}_i^{(s)\prime}(\mathbf{I}-\alpha_0\dddot{\mathbf{G}}^{(t)})_i(\mathbf{I}-\alpha_0\ddot{\mathbf{G}}^{(s)})^{-1}\ddot{\mathbf{V}}^{(s)}\tilde{\boldsymbol{\theta}}_0\right].$$

The GMM estimator

$$\hat{\boldsymbol{\theta}} = \arg\max \mathcal{Q}_N(\boldsymbol{\theta})$$

is therefore consistent under Assumptions 1–9 if no other $\boldsymbol{\theta}$ solves the moment condition (e.g. see Newey and McFadden (1994); Theorem 2.6). The next section discusses the numerical implementation and identification of the estimator.

A.1.1 Concentration and Identification

Although the GMM estimator could be solved numerically, it is helpful to concentrate the objective function around α. Let $\mathbf{R}(\alpha) = \frac{1}{RST}\sum_r\sum_s\sum_t \dot{\mathbf{Z}}^{(s)\prime}(\mathbf{I}-\alpha\dddot{\mathbf{G}}^{(t)})(\mathbf{I}-\alpha\ddot{\mathbf{G}}^{(s)})^{-1}\ddot{\mathbf{V}}^{(s)}/N$ and $\mathbf{D}(\alpha) = \frac{1}{RST}\sum_r\sum_s\sum_t \dot{\mathbf{Z}}^{(s)\prime}(\mathbf{I}-\alpha\dddot{\mathbf{G}}^{(t)})/N$ so that $\frac{1}{N}\sum_i \mathbf{m}_i(\boldsymbol{\theta}) = \mathbf{D}(\alpha)\mathbf{y} - \mathbf{R}(\alpha)\tilde{\boldsymbol{\theta}}$.

The gradient of the objective function with respect to $\tilde{\boldsymbol{\theta}}$ is

$$-2\left(\mathbf{D}(\alpha)\mathbf{y} - \mathbf{R}(\alpha)\tilde{\boldsymbol{\theta}}\right)'\mathbf{W}_N\mathbf{R}(\alpha).$$

The hessian is therefore $2\mathbf{R}'(\alpha)\mathbf{W}_N\mathbf{R}(\alpha)$, which is semi-positive definite. Hence, the following assumption ensures identification of $\tilde{\boldsymbol{\theta}}$, conditional on α.

Assumption 10. *For all α, (a) $\mathbf{R}(\alpha)$ is a full rank matrix; (b)* $\operatorname{plim} \mathbf{R}(\alpha)'\mathbf{W}_N\mathbf{R}(\alpha)$ *exists and is nonsingular.*

Condition (a) ensures that the estimator is uniquely defined: assuming that $\mathbf{R}(\alpha)$ is a full rank matrix implies that $\left(\mathbf{D}(\alpha)\mathbf{y} - \mathbf{R}(\alpha)\tilde{\boldsymbol{\theta}}\right)' \mathbf{W}_N\mathbf{R}(\alpha) = \mathbf{0}$ holds for a single value of $\tilde{\boldsymbol{\theta}}$. Condition (b) ensures the identification is conditional on α.
Under Assumption 10, the estimator of $\tilde{\boldsymbol{\theta}}$ conditionally on α is
$\hat{\tilde{\boldsymbol{\theta}}}(\alpha) = (\mathbf{R}'(\alpha)\mathbf{W}_N\mathbf{R}(\alpha))^{-1}\mathbf{R}'(\alpha)\mathbf{W}_N\mathbf{D}(\alpha)\mathbf{y}$.
The objective function can then be concentrated around α, as

$$\mathcal{Q}_N^c(\alpha) = \left(\mathbf{D}(\alpha)\mathbf{y} - \mathbf{R}(\alpha)\hat{\tilde{\boldsymbol{\theta}}}(\alpha)\right)' \mathbf{W}_N \left(\mathbf{D}(\alpha)\mathbf{y} - \mathbf{R}(\alpha)\hat{\tilde{\boldsymbol{\theta}}}(\alpha)\right)/N^2.$$

While we cannot easily ensure that this function is globally convex in α, the behavior of the function is easy to study because α is unidimensional and has a bounded support (see Assumption 1). Note that numerically minimizing the concentrated objective function implicitly checks Assumption 10. Our final assumption ensures the identification of α.

Assumption 11. *(a) $\mathcal{Q}_N^c(\alpha)$ is uniquely minimized, and (b)* $\operatorname{plim} \mathcal{Q}_N^c(\alpha)$ *is uniquely minimized at α_0.*

Again, condition (a) ensures that the estimator is well defined, whereas condition (b) is the identification condition.

A.2 Asymptotic Normality for Sequential M-Estimators

The challenge in showing the asymptotic normality of our estimator lies in our use of a sequential estimation strategy.

Here, we assume that the model can be written as a sequential m-estimator. If the first-stage estimator $\hat{\boldsymbol{\rho}}$ is an M-estimator, then it verifies the first-order conditions:

$$\frac{1}{M}\sum_{m=1}^{M}\frac{\partial Q_{1,m}(\hat{\boldsymbol{\rho}})}{\partial \boldsymbol{\rho}}=0, \tag{8}$$

for some function $Q_{1,m}$. We assume that $Q_{1,m}(\boldsymbol{\rho})$ is twice continuously differentiable in $\boldsymbol{\rho}$. Note that even if the observational unit of the first stage is at the pair level, it is always possible to take the sum or the average of the moment condition at the pair level to produce a moment condition at the group level as in (8).

Consider now the second stage, i.e. our SGMM estimator. It can be written as an m-estimator, continuously differentiable in $\boldsymbol{\rho}$ (in a neighborhood of $\boldsymbol{\rho}_0$), and twice continuously differentiable in $\boldsymbol{\theta}$:

$$\frac{1}{M}\sum_{m=1}^{M}\frac{\partial Q_{2,m}(\hat{\boldsymbol{\theta}},\hat{\boldsymbol{\rho}})}{\partial \boldsymbol{\theta}}=0, \tag{9}$$

where $Q_{2,m}(\boldsymbol{\theta},\hat{\boldsymbol{\rho}})=\mathcal{Q}_N(\boldsymbol{\theta})$ for all $\boldsymbol{\theta}$.

Note that here, $Q_{2,m}(\boldsymbol{\theta},\boldsymbol{\rho})$ is not everywhere continuous in $\boldsymbol{\rho}$. Indeed, remark that we can rewrite $\dot{\mathbf{G}}=f(\{\dot{a}_{ij}\}_{ij})=f(\{\mathbb{1}[P_{ij}(1|\hat{\boldsymbol{\rho}})\geq\dot{u}_{ij}]\}_{ij})$, where $\dot{u}_{ij}\sim_{iid}U[0,1]$, and similarly for $\ddot{\mathbf{G}}$ and $\dddot{\mathbf{G}}$. This is only a variable change and does not introduce any additional assumptions; however, it shows that, using a finite number of network draws, the moment function is not everywhere continuous in $\boldsymbol{\rho}$ because of the indicator function.

This is similar to the estimator in McFadden (1989), and asymptotic normality results (for m-estimators) can be found in Newey and McFadden (1994) and Andrews (1994). In particular Newey and McFadden (1994) Section 6 for two-steps estimators and Newey and McFadden (1994) Section 7 (in particular Theorem 7.2 and 7.3) for non-smooth objective function. Andrews (1994) (Section 3.2) also discusses non-smooth objective function and specifically provides a discussion of the MSM in McFadden (1989) (his Example 2) which features the same type of discontinuities as our SGMM estimator.[29]

A.3 Variance Estimation

Let where $\boldsymbol{\Sigma}_0 = \operatorname{plim} \boldsymbol{\Sigma}_N$ and $\boldsymbol{\Sigma}_N = \mathbb{V}(\sqrt{M}\hat{\boldsymbol{\theta}})$. We now present a simple approach to estimate $\boldsymbol{\Sigma}_0$. Our approach does not rely on the asymptotic normality and takes into account the uncertainty related to the estimation at the first stage.

Taking the first derivative of the objective function at the second stage (for finite R,S,T and conditional on $\hat{\rho}$) with respect to $\boldsymbol{\theta}$, we have

$$\frac{1}{N}\sum_i \frac{\partial \mathbf{m}_i^{\hat{P}}(\boldsymbol{\theta})'}{\partial \boldsymbol{\theta}} \mathbf{W}_N \frac{1}{N}\sum_i \mathbf{m}_i^{\hat{P}}(\boldsymbol{\theta}) = \mathbf{0}.$$

By applying the mean value theorem to $\frac{1}{N}\sum_i \mathbf{m}_i^{\hat{P}}(\boldsymbol{\theta})$, we have

$$\sqrt{M}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0) = -\left[\frac{1}{N}\sum_i \frac{\partial \mathbf{m}_i^{\hat{P}}(\hat{\boldsymbol{\theta}})'}{\partial \boldsymbol{\theta}} \mathbf{W}_N \frac{1}{N}\sum_i \frac{\partial \mathbf{m}_i^{\hat{P}}(\boldsymbol{\theta}^*)}{\partial \boldsymbol{\theta}'}\right]^{-1} \frac{1}{N}\sum_i \frac{\partial \mathbf{m}_i^{\hat{P}}(\hat{\boldsymbol{\theta}})'}{\partial \boldsymbol{\theta}} \mathbf{W}_N \frac{\sqrt{M}}{N}\sum_i \mathbf{m}_i^{\hat{P}}(\boldsymbol{\theta}_0),$$

where $\boldsymbol{\theta}^*$ is some point between $\hat{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}_0$.

Let $\mathbf{H}_N = \left[\frac{1}{N}\sum_i \frac{\partial \mathbf{m}_i^{\hat{P}}(\hat{\boldsymbol{\theta}})'}{\partial \boldsymbol{\theta}} \mathbf{W}_N \frac{1}{N}\sum_i \frac{\partial \mathbf{m}_i^{\hat{P}}(\boldsymbol{\theta}^*)}{\partial \boldsymbol{\theta}'}\right]^{-1} \frac{1}{N}\sum_i \frac{\partial \mathbf{m}_i^{\hat{P}}(\hat{\boldsymbol{\theta}})'}{\partial \boldsymbol{\theta}} \mathbf{W}_N$ and let $\mathbf{H}_0 = \operatorname{plim} \mathbf{H}_N$.

We have

$$\sqrt{M}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0) = -\mathbf{H}_N \frac{\sqrt{M}}{N}\sum_i \mathbf{m}_i^{\hat{P}}(\boldsymbol{\theta}_0).$$

Let also $\boldsymbol{\Omega}_N = \mathbb{V}(\frac{\sqrt{M}}{N}\sum_i \mathbf{m}_i^{\hat{P}}(\boldsymbol{\theta}_0))$.

We assume the following:

Assumption 12. $\operatorname{plim} \boldsymbol{\Omega}_N = \boldsymbol{\Omega}_0$ *and* $\operatorname{plim} \mathbf{H}_N = \mathbf{H}_0$ *exist and are finite matrices.*

---

[29] An alternative is to smooth the moment function (i.e. $\mathbb{1}[x \geq 0] \approx I(x/h_N)$, see e.g. Horowitz (1998)) where $h_N \to 0$ fast enough so that the approximation becomes asymptotically negligible. The selection of the bandwidth in practice is challenging. Here, we suggest applying the same function used in particular by Horowitz (1998), which is conveniently available in our R package. Specifically: $I(x) = \mathbb{1}\{-1 \leq x \leq 1\}[0.5 + \frac{105}{64}(x - \frac{5}{3}x^3 + \frac{7}{5}x^5 - \frac{3}{7}x^7)] + \mathbb{1}\{x > 1\}$, which was also used in particular by Kaplan and Sun (2017) and de Castro et al. (2019).

Under this assumption, we have

$$\boldsymbol{\Sigma}_0 = \mathbf{H}_0\boldsymbol{\Omega}_0\mathbf{H}_0'.$$

Let $\boldsymbol{\zeta}_{N,m}^{\hat{P}}$ be the sum of $\mathbf{m}_i^{\hat{P}}(\boldsymbol{\theta}_0)$ over the m-th sub-network and let $\boldsymbol{\zeta}_{N,m}^{P}$ be the same statistic where $\hat{P}$ is replaced by P. In practice, $\mathbf{H}_0$ can be estimated consistently by

$$\hat{\mathbf{H}}_0 = N\left[\sum_i \frac{\partial \mathbf{m}_i^{\hat{P}}(\hat{\boldsymbol{\theta}})'}{\partial \boldsymbol{\theta}}\mathbf{W}_N \sum_i \frac{\partial \mathbf{m}_i^{\hat{P}}(\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}'}\right]^{-1} \sum_i \frac{\partial \mathbf{m}_i^{\hat{P}}(\hat{\boldsymbol{\theta}})'}{\partial \boldsymbol{\theta}}\mathbf{W}_N.$$

On the other hand, $\boldsymbol{\Sigma}_0$ can be estimated by $\boldsymbol{\Sigma}_N$, where

$$\boldsymbol{\Sigma}_N = \mathbb{E}(\mathbb{V}(\tfrac{\sqrt{M}}{N}\textstyle\sum_m \boldsymbol{\zeta}_{N,m}^{\hat{P}}|\hat{P})) + \mathbb{V}(\mathbb{E}(\tfrac{\sqrt{M}}{N}\textstyle\sum_m \boldsymbol{\zeta}_{N,m}^{\hat{P}}|\hat{P})),$$
$$\boldsymbol{\Sigma}_N = \tfrac{M}{N^2}\,\mathbb{E}(\textstyle\sum_m \mathbb{V}(\boldsymbol{\zeta}_{N,m}^{\hat{P}}|\hat{P})) + \tfrac{M}{N^2}\,\mathbb{V}(\textstyle\sum_m \mathbb{E}(\boldsymbol{\zeta}_{N,m}^{\hat{P}}|\hat{P})).$$

The term $N^{-1}\,\mathbb{E}(\sum_m \mathbb{V}(\boldsymbol{\zeta}_{N,m}^{\hat{P}}|\hat{P}))$ is due to the error term of the model $\boldsymbol{\varepsilon}$ and the true network not being observed, whereas $N^{-1}\,\mathbb{V}(\sum_m \mathbb{E}(\boldsymbol{\zeta}_{N,m}^{\hat{P}}|\hat{P}))$ is due to uncertainty associated with the estimation of P. In practice, we can compute an estimator of $\boldsymbol{\zeta}_{N,m}^{\hat{P}}$, $m = 1, ..., M$ by replacing $\boldsymbol{\theta}_0$ by its consistent estimator $\hat{\boldsymbol{\theta}}$. Thus, we can also estimate $\sum_m \mathbb{V}(\boldsymbol{\zeta}_{N,m}^{\hat{P}}|\hat{P})$ and $\sum_m \mathbb{E}(\boldsymbol{\zeta}_{N,m}^{\hat{P}}|\hat{P})$. With several simulations from the distribution of $\hat{P}$, we finally compute $\boldsymbol{\Sigma}_N$.[30]

---

[30] Our R package offers tools to compute this variance. See also our Online Appendix.

B Full Simulation Results

Table B.1: Full simulation results without group fixed effects

Proportion of missing links	0%	25%	50%	75%
Classical IV method, Instrument matrix: $\mathbf{GX}^2$				
$\alpha = 0.4$	0.400 (0.003)			
$c = 2$	1.989 (0.172)			
$\beta_1 = 1$	1.000 (0.006)			
$\beta_2 = 1.5$	1.501 (0.008)			
$\gamma_1 = 5$	4.998 (0.025)			
$\gamma_2 = -3$	−2.998 (0.025)			
SGMM: $\mathbf{Gy}$, $\mathbf{GX}$ observed; $T = 100$				
$\alpha = 0.4$		0.400 (0.003)	0.400 (0.003)	0.400 (0.003)
$c = 2$		2.010 (0.175)	1.997 (0.177)	2.008 (0.178)
$\beta_1 = 1$		1.000 (0.006)	1.000 (0.006)	1.000 (0.006)
$\beta_2 = 1.5$		1.500 (0.007)	1.501 (0.007)	1.500 (0.008)
$\gamma_1 = 5$		5.001 (0.026)	4.999 (0.025)	4.999 (0.025)
$\gamma_2 = -3$		−3.002 (0.026)	−3.000 (0.026)	−3.001 (0.026)
SGMM: $\mathbf{Gy}$ observed, $\mathbf{GX}$ unobserved; $S = T = 100$				
$\alpha = 0.4$		0.400 (0.004)	0.400 (0.005)	0.400 (0.006)
$c = 2$		2.014 (0.225)	2.002 (0.260)	2.017 (0.286)
$\beta_1 = 1$		1.000 (0.008)	1.001 (0.011)	1.001 (0.012)
$\beta_2 = 1.5$		1.500 (0.009)	1.500 (0.010)	1.500 (0.011)
$\gamma_1 = 5$		5.001 (0.035)	4.998 (0.046)	4.995 (0.052)
$\gamma_2 = -3$		−3.002 (0.034)	−3.000 (0.039)	−3.002 (0.042)
SGMM: $\mathbf{Gy}$ unobserved, $\mathbf{GX}$ observed; $S = T = 100$				
$\alpha = 0.4$		0.400 (0.010)	0.400 (0.015)	0.401 (0.022)
$c = 2$		2.024 (0.561)	2.048 (0.802)	2.010 (1.009)
$\beta_1 = 1$		0.999 (0.021)	1.001 (0.036)	1.003 (0.051)
$\beta_2 = 1.5$		1.499 (0.020)	1.499 (0.027)	1.499 (0.034)
$\gamma_1 = 5$		5.003 (0.094)	4.997 (0.155)	4.982 (0.236)
$\gamma_2 = -3$		−3.003 (0.086)	−3.006 (0.124)	−2.998 (0.158)
SGMM: $\mathbf{Gy}$, $\mathbf{GX}$ unobserved; $R = S = T = 100$				
$\alpha = 0.4$		0.400 (0.012)	0.400 (0.018)	0.401 (0.026)
$c = 2$		2.027 (0.648)	2.066 (0.931)	2.014 (1.191)
$\beta_1 = 1$		0.999 (0.025)	1.001 (0.042)	1.003 (0.060)
$\beta_2 = 1.5$		1.499 (0.022)	1.498 (0.031)	1.499 (0.039)
$\gamma_1 = 5$		5.003 (0.111)	4.997 (0.182)	4.981 (0.277)
$\gamma_2 = -3$		−3.004 (0.100)	−3.008 (0.144)	−2.999 (0.188)

Table B.2: Full simulation results group fixed effects

Proportion of missing links	0%		25%		50%		75%	
Classical IV method, Instrument matrix: $\mathbf{GX}^2$								
$\alpha = 0.4$	0.400	(0.014)						
$\beta_1 = 1$	1.000	(0.006)						
$\beta_2 = 1.5$	1.500	(0.008)						
$\gamma_1 = 5$	5.000	(0.042)						
$\gamma_2 = -3$	-3.001	(0.046)						
SGMM: $\mathbf{Gy}$, $\mathbf{GX}$ observed; $T = 100$								
$\alpha = 0.4$			0.399	(0.018)	0.400	(0.021)	0.399	(0.024)
$\beta_1 = 1$			1.000	(0.006)	1.000	(0.007)	1.001	(0.006)
$\beta_2 = 1.5$			1.500	(0.008)	1.500	(0.009)	1.500	(0.009)
$\gamma_1 = 5$			5.002	(0.050)	4.999	(0.059)	5.002	(0.066)
$\gamma_2 = -3$			-2.997	(0.048)	-3.001	(0.051)	-2.998	(0.055)
SGMM: $\mathbf{Gy}$ observed, $\mathbf{GX}$ unobserved; $S = T = 100$								
$\alpha = 0.4$			0.400	(0.028)	0.402	(0.038)	0.400	(0.047)
$\beta_1 = 1$			1.001	(0.008)	1.000	(0.011)	1.002	(0.012)
$\beta_2 = 1.5$			1.500	(0.010)	1.500	(0.011)	1.500	(0.012)
$\gamma_1 = 5$			4.999	(0.078)	4.994	(0.107)	4.994	(0.135)
$\gamma_2 = -3$			-2.998	(0.07)	-3.002	(0.092)	-3.000	(0.104)
SGMM: $\mathbf{Gy}$ unobserved, $\mathbf{GX}$ observed; $S = T = 100$								
$\alpha = 0.4$			0.405	(0.075)	0.411	(0.116)	0.403	(0.187)
$\beta_1 = 1$			1.000	(0.023)	1.000	(0.037)	1.004	(0.051)
$\beta_2 = 1.5$			1.499	(0.023)	1.499	(0.034)	1.499	(0.046)
$\gamma_1 = 5$			4.987	(0.202)	4.968	(0.325)	4.970	(0.542)
$\gamma_2 = -3$			-3.013	(0.205)	-3.022	(0.317)	-3.003	(0.482)
SGMM: $\mathbf{Gy}$, $\mathbf{GX}$ unobserved; $R = S = T = 100$								
$\alpha = 0.4$			0.399	(0.092)	0.406	(0.136)	0.395	(0.215)
$\beta_1 = 1$			1.000	(0.027)	1.001	(0.043)	1.004	(0.060)
$\beta_2 = 1.5$			1.499	(0.026)	1.498	(0.040)	1.498	(0.053)
$\gamma_1 = 5$			5.002	(0.245)	4.981	(0.377)	4.993	(0.622)
$\gamma_2 = -3$			-3.003	(0.246)	-3.015	(0.377)	-3.000	(0.560)

C Appendix – Application

C.1 Error codes only

Each student nominates their best friends up to 5 males and 5 females. Because we know the sex of nominated friends, even when the identifier is coded with error, we associate each missing link to a male or female student. We then have two sets of network data for each student i: the set of data from i to their male schoolmates and the set of data from i to their female schoolmates. A set is considered fully observed if it has no missing values. We estimate the network formation only using the fully observed sets. The sets with partial or no observed data are inferred (even the data we do not doubt in those sets are inferred).

This approach raises a selection problem that we address by weighting each selected set, following Manski and Lerman (1977). The intuition of the weights lies in the fact that the sets with many links have lower probabilities to be selected (because error codes are more likely). The weight is the inverse of the selection probability. For a selected set $\mathcal{S}_{is}$ (of network data from i to schoolmates of sex s), the selection probability can be estimated as the proportion of sets without missing data among the sets of network data to schoolmates of sex s having the same number of links than $\mathcal{S}_{is}$.

For the Bayesian estimator, we jointly estimate the peer effect model and the network formation model (i.e. using Step 1' on Page 28). Thus, in the MCMC, $\boldsymbol{\rho}$ and the sets with partial or no network data are inferred using information from the weighted sets and the peer effect model.

C.2 Error codes and top coding

We consider the same selected sets as in the case of missing data only. However, we doubt the exactitude of a link $a_{ij} \in \mathcal{S}_{is}$ if $a_{ij} = 0$ and the number of links in $\mathcal{S}_{is}$ is five. Therefore, if the number of links in $\mathcal{S}_{is}$ is five, we adjust the weight associated with each a_{ij}. For $a_{ij} = 0$, we multiply the weight obtained in the case of missing data only by $(|\mathcal{S}_{is}| - \ell(\mathcal{S}_{is}))/(|\mathcal{S}_{is}| - \hat{\ell}(\mathcal{S}_{is}))$, and for $a_{ij} = 1$, we multiply the weight by $\ell(\mathcal{S}_{is})/\hat{\ell}(\mathcal{S}_{is})$, where $\ell(\mathcal{S}_{is})$ is the estimate of the true number of links from i to their schoolmates of sex s, $\hat{\ell}(\mathcal{S}_{is})$ if the number of links declared in $\mathcal{S}_{is}$, and $|\mathcal{S}_{is}|$ is the number of data in $\mathcal{S}_{is}$ (number of students having the sex s in the school minus one).

We denote $s = m$ for male and $s = f$ for female. Fours scenarios are possible: $\{\hat{\ell}(\mathcal{S}_{im}) < 5,\ \hat{\ell}(\mathcal{S}_{if}) < 5\}$, $\{\hat{\ell}(\mathcal{S}_{im}) = 5,\ \hat{\ell}(\mathcal{S}_{if}) < 5\}$, $\{\hat{\ell}(\mathcal{S}_{im}) < 5,\ \hat{\ell}(\mathcal{S}_{if}) = 5\}$, and $\{\hat{\ell}(\mathcal{S}_{im}) = 5,\ \hat{\ell}(\mathcal{S}_{if}) = 5\}$. In the last three cases, $\ell(\mathcal{S}_{im}) + \ell(\mathcal{S}_{if})$ is left-censored and we know the lower bound. Assuming that the number of links i follows a Poisson distribution of mean n_i^e, we estimate n_i^e using a censored Poisson regression on the declared number of links. We assume that n_i^e is an exponential linear function of i's characteristics (age, sex, ...), and we also include school-fixed effects to control for school size.

The estimate of n_i^e is $\ell(\mathcal{S}_{im}) + \ell(\mathcal{S}_{if})$, and it allows us to compute $\ell(\mathcal{S}_{im})$ and $\ell(\mathcal{S}_{if})$. For the case $\{\hat{\ell}(\mathcal{S}_{im}) = 5,\ \hat{\ell}(\mathcal{S}_{if}) = 5\}$, we assume that $\ell(\mathcal{S}_{im}) = \ell(\mathcal{S}_{if}) = 0.5(\ell(\mathcal{S}_{im}) + \ell(\mathcal{S}_{if}))$. In the other cases, as either $\ell(\mathcal{S}_{im})$ or $\ell(\mathcal{S}_{if})$ is known, the second member of $\mathcal{S}_{im}) + \ell(\mathcal{S}_{if})$ can be computed.

For the Bayesian estimator, and contrary to the case with error codes only, it is more challenging to jointly estimate the peer effect model and the network formation model in a single step. We therefore first estimate the network formation model and then the Bayesian estimator (i.e. using Step 1 in Algorithm 1). Thus, for the MCMC, the estimated distribution of $\boldsymbol{\rho}$ from the network formation model is used as a prior distribution. We then infer $\boldsymbol{\rho}$ and the network data $a_{ij} = 0$ that we are doubtful about, using information from the peer effect model and the prior distribution of $\boldsymbol{\rho}$.

Table C.1: Empirical results (Bayesian method)

	Model 1		Model 2		Model 3	
Statistic	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Peer effect model						
Peer effects	0.350	(0.024)	0.524	(0.036)	0.538	(0.037)
Own effects						
Female	0.144	(0.029)	0.135	(0.030)	0.133	(0.031)
Hispanic	−0.083	(0.042)	−0.148	(0.048)	−0.151	(0.047)
Race (White)						
Black	−0.230	(0.045)	−0.190	(0.055)	−0.189	(0.055)
Asian	−0.091	(0.089)	−0.113	(0.091)	−0.110	(0.091)
Other	0.055	(0.051)	0.039	(0.052)	0.039	(0.052)
Mother’s education (High)						
<High	−0.122	(0.039)	−0.138	(0.040)	−0.139	(0.040)
>High	0.140	(0.034)	0.123	(0.034)	0.121	(0.034)
Missing	−0.060	(0.050)	−0.069	(0.051)	−0.070	(0.051)
Mother’s job (Stay-at-home)						
Professional	0.080	(0.045)	0.075	(0.044)	0.079	(0.044)
Other	0.003	(0.035)	−0.014	(0.035)	−0.012	(0.035)
Missing	−0.066	(0.047)	−0.074	(0.048)	−0.073	(0.048)
Age	−0.073	(0.010)	−0.071	(0.010)	−0.072	(0.010)
Contextual effects						
Female	0.011	(0.049)	−0.003	(0.060)	−0.003	(0.060)
Hispanic	0.060	(0.069)	0.272	(0.102)	0.276	(0.105)
Race (White)						
Black	0.050	(0.058)	0.025	(0.073)	0.033	(0.074)
Asian	0.209	(0.186)	0.110	(0.365)	0.209	(0.385)
Other	−0.137	(0.089)	−0.044	(0.163)	−0.051	(0.167)
Mother’s education (High)						
<High	−0.269	(0.070)	−0.228	(0.141)	−0.221	(0.149)
>High	0.072	(0.059)	0.063	(0.097)	0.057	(0.102)
Missing	−0.077	(0.093)	0.107	(0.167)	0.124	(0.174)
Mother’s job (Stay-at-home)						
Professional	−0.110	(0.08)	0.102	(0.124)	0.090	(0.134)
Other	−0.101	(0.060)	−0.003	(0.100)	−0.017	(0.103)
Missing	−0.093	(0.085)	−0.075	(0.157)	−0.109	(0.165)
Age	0.066	(0.006)	0.083	(0.008)	0.086	(0.009)
SE^2	0.523		0.496		0.499	
Network formation model						
Same sex			0.310	(0.011)	0.370	(0.014)
Both Hispanic			0.416	(0.020)	0.436	(0.026)
Both White			0.312	(0.018)	0.304	(0.023)
Both Black			1.076	(0.030)	1.173	(0.038)
Both Asian			0.164	(0.034)	0.144	(0.043)
Mother’s education < High			0.226	(0.013)	0.218	(0.017)
Mother’s education > High			0.007	(0.012)	0.005	(0.014)
Mother’s job Professional			−0.116	(0.012)	−0.128	(0.016)
Age absolute difference			−0.700	(0.007)	−0.714	(0.009)
Average number of friends	3.251		4.665		5.618	

Note: Model 1 considers the observed network as given. Model 2 infers the missing links due to friendship nominations coded with error, and Model 3 infers the missing links due to friendship nominations coded with error and controls for top coding. For each model, Column "Mean" indicates the posterior mean, and Column "Std. Dev." indicates the posterior standard deviations in parentheses.

$N = 3{,}126$. Observed links = 17,993. Proportion of inferred network data: error code = 60.0%, error code and top coding = 65.0%. The explained variable is computed by taking the average grade for English, Mathematics, History, and Science, letting $A = 4$, $B = 3$, $C = 2$, and $D = 1$. Thus, higher scores indicate better academic achievement.

Table C.2: Empirical results (SGMM Method)

	Model 1		Model 2		Model 3	
Statistic	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Peer effect model						
Peer effects	0.455	(0.230)	0.753	(0.254)	0.683	(0.242)
Own effects						
Female	0.179	(0.039)	0.122	(0.036)	0.122	(0.035)
Hispanic	−0.129	(0.045)	−0.160	(0.051)	−0.160	(0.051)
Race (White)						
Black	−0.276	(0.058)	−0.172	(0.058)	−0.166	(0.059)
Asian	−0.178	(0.101)	−0.131	(0.085)	−0.124	(0.086)
Other	0.087	(0.062)	0.023	(0.061)	0.023	(0.062)
Mother's education (High)						
<High	−0.134	(0.044)	−0.121	(0.046)	−0.124	(0.047)
>High	0.109	(0.036)	0.121	(0.030)	0.122	(0.03)
Missing	−0.066	(0.053)	−0.060	(0.051)	−0.062	(0.051)
Mother's job (Stay-at-home)						
Professional	0.145	(0.055)	0.065	(0.043)	0.071	(0.043)
Other	0.043	(0.035)	−0.019	(0.031)	−0.018	(0.030)
Missing	−0.018	(0.045)	−0.072	(0.043)	−0.068	(0.043)
Age	−0.042	(0.032)	−0.072	(0.015)	−0.068	(0.016)
Contextual effects						
Female	−0.056	(0.074)	−0.014	(0.068)	−0.001	(0.068)
Hispanic	0.265	(0.121)	0.331	(0.169)	0.368	(0.175)
Race (White)						
Black	0.129	(0.125)	0.035	(0.113)	0.013	(0.108)
Asian	2.409	(1.220)	3.236	(2.359)	3.466	(2.575)
Other	−0.363	(0.180)	−0.111	(0.170)	−0.195	(0.198)
Mother's education (High)						
<High	−0.215	(0.083)	−0.206	(0.337)	−0.283	(0.355)
>High	0.168	(0.113)	−0.043	(0.139)	−0.051	(0.155)
Missing	0.240	(0.165)	−0.041	(0.280)	−0.034	(0.303)
Mother's job (Stay-at-home)						
Professional	−0.239	(0.111)	0.182	(0.142)	0.186	(0.158)
Other	−0.101	(0.072)	0.126	(0.183)	0.103	(0.198)
Missing	−0.199	(0.162)	0.247	(0.381)	0.168	(0.396)
Age	0.075	(0.033)	0.110	(0.030)	0.103	(0.029)
Network formation model						
Same sex			0.309	(0.016)	0.370	(0.015)
Both Hispanic			0.417	(0.027)	0.433	(0.025)
Both White			0.312	(0.025)	0.304	(0.023)
Both Black			1.077	(0.043)	1.171	(0.041)
Both Asian			0.165	(0.050)	0.142	(0.047)
Mother's education < High			0.226	(0.018)	0.216	(0.017)
Mother's education > High			0.009	(0.016)	0.006	(0.015)
Mother's job Professional			−0.116	(0.017)	−0.128	(0.016)
Age absolute difference			−0.701	(0.010)	−0.715	(0.009)
Average number of friends	3.251		4.664		5.613	

Note: Model 4 considers the observed network as given. Model 5 infers the missing links due to friendship nominations coded with error, and Model 6 infers the missing links due to friendship nominations coded with error and controls for top coding. For each model, Column "Mean" indicates the estimates, and Column "Std. Dev." indicates the posterior standard deviations in parentheses.
$N = 3{,}126$. Observed links = 17,993. Proportion of inferred network data: error code = 60.0%, error code and top coding = 65.0%. The explained variable is computed by taking the average grade for English, Mathematics, History, and Science, letting $A = 4$, $B = 3$, $C = 2$, and $D = 1$. Thus, higher scores indicate better academic achievement.

Figure C.1: MCMC Simulations – Peer Effect Model

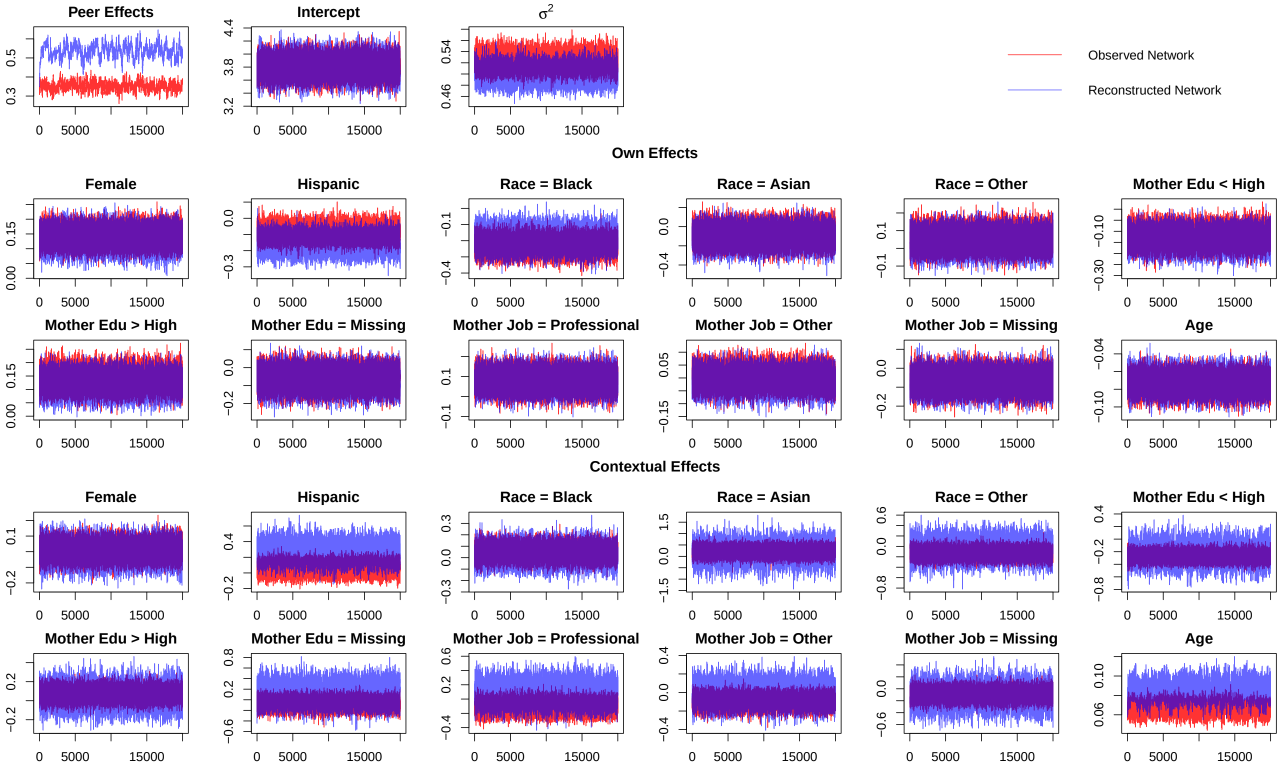

Figure C.2: MCMC Simulations – Network Formation Model