

Ignoring Measurement Errors in Social Networks

Arthur Lewbel

Xi Qu

Xun Tang

Boston College

Shanghai Jiao Tong University

Rice University

February 2023

Abstract

We consider peer effect estimation in social network models where some network links are incorrectly measured. We show that if the number or magnitude of mismeasured links does not grow too quickly with the sample size, then standard instrumental variables estimators that ignore these measurement errors remain consistent, and standard asymptotic inference methods remain valid. These results hold even when the link measurement errors are correlated with regressors or with structural errors in the model. Simulations and real data experiments confirm our results in finite samples. These findings imply that researchers can ignore small numbers of mismeasured links in networks.

JEL classification: C31, C51

Keywords: Social networks, peer effects, misclassified or missing links, mismeasured network.

1 Introduction

In many social and economic environments, an individual's behavior or outcome (such as a consumption choice or a test score) depends not only on his or her own characteristics, but also on the behavior and characteristics of other individuals. Call such dependence between two individuals a *link*, and call individuals with such links *friends*. A *social network* consists of a group of linked individuals. Each individual may have a different set of friends in the network, and each individual may assign heterogeneous weights to his or her links. The structure of a social network is fully characterized by a square *adjacency matrix*, which lists all links (with possibly heterogeneous weights) among the individuals in the network.

Much of the econometric literature on social networks focuses on disentangling and estimating various social or network effects, based on observed outcomes and characteristics

of network members. These structural parameters include the effects on each individual's outcome by (i) the individual's own characteristics (*direct effects*) and possibly group characteristics (*correlated effects*), (ii) the characteristics of the individual's friends (*contextual effects*) and (iii) the outcomes of the individual's friends (*peer effects*). Standard methods of identifying and estimating these structural network effect parameters assume that the adjacency matrix of links among individuals in the sample is perfectly observed.

1.1. Our contribution. We consider the case where some network links are misclassified, or generally measured with errors. Here we provide good news for empirical researchers, by showing that relatively small amounts of measurement error in the network can be safely ignored in estimation. More precisely, we show that instrumental variable estimators like Bramoullé, Djebbari and Fortin (2009), and their standard errors, remain consistent and valid, despite the presence of misclassified, unreported, or mismeasured links, as long as the number and size of these measurement errors grows sufficiently slowly with the sample size. Moreover, these results hold even when the measurement errors are correlated with the regressors, or with the model errors. In Section 1.2 below, we give examples of how such measurement errors arise in real applications. In Section 4 we provide detailed conditions for these errors to grow at these sufficiently slow rates).

It may not be surprising that measurement errors growing at sufficiently slow rates are asymptotically negligible, but it is also not automatic. Slow measurement error rates could still blow up an estimator if the stochastic order of quadratic terms in the errors of the parameter estimates can not be bounded. What we show is that, in the case of two-stage least squares (2SLS) estimators of network models, minimal and standard regularity conditions suffice to bound these terms.

1.2. Motivation. There are many reasons why network links can be mismeasured in practice. In some data sets, links are imputed from measures of proximity or similarity of individuals (e.g., use of distance as a link in gravity models of trade). Such imputations are generally imperfect, resulting in measurement errors in the magnitudes of links.

Mismeasurement may also arise because links that are observed in one context may be irrelevant for outcomes under study in another. For instance, two people who are observed as linked on a social media platform may be connected there for business or political reasons, but have no effect on each other's personal outcomes (or vice versa). Or in a school setting, some reported friends and not others may be study partners who affect academic performance.

Even in data sets where all observable links are directly relevant for observed outcomes, link data may contain a variety of reporting or recording errors. For example, studies that focus on links within groups, such as within classrooms or villages, may not report links across groups (e.g., friendships with people in other schools). In this case, measurement

errors are caused by unrecorded links between the groups.

Another example is a panel data model of social networks, where a slowly evolving network is only observed in some intermittent time periods, and is assumed to stay fixed in between those periods. In this case, the measurement errors are due to the unobserved formation (or dissolution) of new (or existing) links between those observation periods.

A third example is when sample-collecting surveys limit the number of links (such as the number of friends) that a respondent can report, thus leading to missing links for popular individuals. Yet another example is when the link data collected from surveys have recall or response errors. For instance, two individuals may report different responses to the question of whether they are friends, leading to uncertainty in how an undirected link between them should be recorded. Or more simply, surveyors may make occasional mistakes in recording responses.

The main finding of our paper is as follows: if the size of measurement errors in the reported adjacency matrix is relatively small (i.e., grows slowly with the network size n), then asymptotic theory that ignores these measurement errors provides a good approximation for estimation and inference. Furthermore, in all four examples mentioned above, we provide specific, intuitive conditions under which this “small error” property holds (Section 4).

1.3. The Model. With a sample of n individuals, let $Y_n = (y_1, \dots, y_n)' \in \mathbb{R}^n$ be a vector of individual outcomes; let $\iota_n = (1, \dots, 1)'$ and $\epsilon_n = (\epsilon_{n,1}, \dots, \epsilon_{n,n})'$ be n -dimensional column vectors; Let $X_n = (x_1, \dots, x_n)'$ be an n -by- K matrix that consists of n vectors of exogenous regressors $x_i \in \mathbb{R}^K$ for $i \leq n$. Let G_n^* be an actual n -by- n adjacency matrix (a.k.a. network) that lists the actual links for peer effects and contextual effects.¹ Let $G_{n,ij}^*$ denote the element in row i and column j of G_n^* . We have $G_{n,ij}^* > 0$ if i and j are linked for peer effects and $G_{n,ij}^* = 0$ otherwise. For each i , let $G_{n,ii}^* = 0$ by convention in the literature. Note that $G_{n,ij}^*$ can be binary (with $G_{n,ij}^* \in \{0, 1\}$ indicating the absence or presence of a link), or continuous and non-negative (with $G_{n,ij}^* \in \mathbb{R}_+$ signifying the strength of the link). Throughout the paper, we maintain that $\min_i \sum_{j=1}^n G_{n,ij}^* > 0$ with probability one. This means there are no isolated individuals in the network, or equivalently no rows of zeros in G_n^* almost surely. This condition is standard in the literature.

We assume a linear social network model:

$$Y_n = \alpha_0 \iota_n + \lambda_0 G_n Y_n + X_n \beta_0 + G_n X_n \gamma_0 + \epsilon_n, \quad (1)$$

¹We can extend the results of this paper to allow the peer and contextual effects to operate through different adjacency matrices – say, G_n^* and C_n^* respectively – provided one of the two conditions hold: either (a) the data contain two distinctive noisy measures for G_n^* and C_n^* respectively with each satisfying the condition of “small order” measurement errors (Assumption 1), or (b) the differences between G_n^* and C_n^* are small and the data contains a single noisy network measure with small measurement errors in the sense of Assumption 1.

where G_n can be either the original adjacency matrix G_n^* , or normalized versions of G_n^* . For example, a row-normalized G_n is defined by $G_{n,ij} = G_{n,ij}^* / \left(\sum_{j'=1}^n G_{n,ij'}^* \right)$. Row normalization is common; our results hold with or without such normalization.

The parameters in equation (1) are as follows: $\lambda_0 \in \mathbb{R}$ is a scalar peer effect, $\beta_0 \in \mathbb{R}^K$ is a vector of direct effects, $\gamma_0 \in \mathbb{R}^K$ is a vector of contextual effects, and $\alpha_0 \in \mathbb{R}$ is the structural intercept. If individuals are divided into groups (such as villages or classrooms), then what are known as correlated effects are group-level fixed effects, i.e., elements of β_0 where the corresponding element of X_n is a group membership indicator.

Our goal is to estimate $\theta_0 \equiv (\alpha_0, \lambda_0, \beta_0', \gamma_0')'$. If Y_n, X_n, G_n^* (and hence G_n) were perfectly observed, the structural model would take the form of a linear regression of Y_n on a constant and the regressors $G_n Y_n, X_n,$ and $G_n X_n$. However, even if X_i is uncorrelated with ε_j for all i and j , making X_n and $G_n X_n$ strictly exogenous, this regression could not be consistently estimated by ordinary least squares, because of the endogeneity of $G_n Y_n$. Instead, one can use an instrument-based, 2SLS estimator using friends of friends of i to construct instruments for $G_n Y_n$ (see, e.g., Lee (2007) and Bramoullé, Djebbari and Fortin (2009)). For example, $G_n^2 X_n$ can be instruments for $G_n Y_n$. To implement this 2SLS estimator, one needs perfect measures of G_n^* so that the regressors $G_n Y_n$ and $G_n X_n$, and instruments such as $G_n^2 X_n$, can all be constructed without errors.

1.4. Estimation with misclassified links. Instead of observing $Y_n, X_n,$ and the true adjacency matrices G_n^* , we assume that what is observed is $Y_n, X_n,$ and a mismeasured adjacency matrix H_n^* . The differences $H_n^* - G_n^*$ are the measurement errors in links. Like G_n^* , the matrix H_n^* by convention has zeros on the diagonal.

For a given pair of individuals i and j , if $G_{n,ij}^*$ equals zero or one, misclassification of that link corresponds to $H_{ij}^* = 1 - G_{ij}^*$. More generally, measurement error in a link occurs whenever $H_{ij}^* \neq G_{ij}^*$. The measurement errors can be any combination of misclassified links and incorrectly weighted links.

We investigate the asymptotic properties of 2SLS estimation of (1) when the mismeasured adjacency matrix H_n^* is observed instead of the true unknown matrices G_n^* . So instead of a 2SLS regression of Y_n on $G_n Y_n, X_n,$ and $G_n X_n$, using as instruments $G_n^2 X_n, X_n,$ and $G_n X_n$, we consider 2SLS regression of Y_n on $H_n Y_n, X_n,$ and $H_n X_n$, using as instruments $H_n^2 X_n, X_n,$ and $H_n X_n$. Note this means that both some regressors and some instruments are mismeasured, and that the measurement errors in regressors and instruments are correlated. Moreover, we do not impose any uncorrelation or conditional independence conditions on the measurement errors. Those conditions are frequently used in the literature of measurement errors. For example, we allow the measurement errors in $H_n^* - G_n^*$ to be arbitrarily correlated with $X_n, Y_n,$ and ε_n .

We find that if the magnitude of measurement errors grows at a rate slower than \sqrt{n} ,

then the 2SLS estimator remains \sqrt{n} -consistent and asymptotically normal, and the usual formulas for inference and standard errors remain valid. As a result, under these conditions researchers can safely ignore the presence of misclassified or mismeasured links, because the estimator and inference based on H_n^* instead of G_n^* remain consistent and valid.

We also find that if the magnitude of measurement errors grows at a rate faster than \sqrt{n} but slower than n , then the 2SLS estimator is still consistent. However, in this case, the rate of convergence of the coefficient estimators is less than \sqrt{n} (due to a bias term that shrinks at a slower rate than \sqrt{n}), so the usual standard error formulas would no longer apply.

1.5 Outline. Section 2 is a short literature review. Section 3 formally presents our results for 2SLS estimation of mismeasured networks. Section 4 provides a few empirical examples where the order of measurement errors in networks are sufficiently small. This is followed by some simulation results (Section 5) and an empirical illustration (Section 6). Proofs are in the appendix.

2 Literature Review

Social network models typically allow an individual’s outcome to depend on his or her own characteristics, contextual influences from peers’ characteristics, and peer effects from peer outcomes. The traditional linear-in-means model (which assumes everyone is linked with everyone else with equal weights, either within groups or in the whole network) suffers from the “reflection problem” as pointed out by Manski (1993). This identification problem can be solved in models with more complicated social interaction structures. Lee (2007) uses conditional maximum likelihood and instrumental variable methods to estimate peer and contextual effects in a spatial autoregressive social interaction model, assuming links are perfectly observed in the data. Bramoullé, Djebbari and Fortin (2009) and Lin (2010) provide specific conditions on observed network structure in order to identify peer effects in social interaction models, using characteristics of friends of friends as instruments.

Given results like these, the model described in the introduction has been widely used to estimate peer effects in a variety of settings. Examples are studies of peer influence on students’ academic performance, sport and club activities, and delinquent behaviors (Hauser et al., 2009; Calvó-Armengol et al., 2009; Lin, 2010; Lee et al., 2010; Liu et al., 2014; Boucher et al., 2014; Patacchini and Zenou, 2012). These models all assume that the network structure is correctly measured in the data.

Regarding selection and comparison of adjacency matrices, LeSage and Pace (2009) use the Bayesian posterior distribution to choose among models with different adjacency matrices. Empirical research may also report estimates using different link weights as robustness

checks. These practices are feasible in, e.g., spatial econometric models, where link weights are assumed to be a function of observable geographic information, as in gravity models of trade. Errors in constructing such links would fit in our framework. There is also a small literature on identification and estimation of peer effects when networks are unobserved. Examples include de Paula et al. (2018) and Lewbel et al. (2021).

The issue of potentially misclassified links is acknowledged and discussed in Patacchini and Venanzoni (2014), Liu et al. (2014), and Lin (2015) among others. But these papers do not provide a formal analysis of the asymptotic impact of mismeasured links on the performance of standard estimators. Griffith (2021) studied the impact on inference when misclassification in the adjacency matrix occurs because of binding caps on the number of self-reported links. Our results fill a void in the literature by analyzing how ignoring small amounts of general measurement errors in the adjacency matrix affects the consistency of standard estimators and the validity of inference.²

3 2SLS Estimation With Mismeasured Links

We derive the asymptotic properties of a 2SLS estimator for the model in (1) when the matrix with measurement errors H_n^* is used in place of the actual, unknown G_n^* . This means the regressors $G_n Y_n$, $G_n X_n$ and instruments $G_n^2 X_n$ are replaced by $H_n Y_n$, $H_n X_n$ and $H_n^2 X_n$ in the estimator.

Write equation (1) as

$$Y_n = R_n \theta_0 + \epsilon_n = \tilde{R}_n \theta_0 + \tilde{\epsilon}_n,$$

where $R_n \equiv (\iota_n, G_n Y_n, X_n, G_n X_n)$ is a matrix of regressors that depend on the true network, $\tilde{R}_n \equiv (\iota_n, H_n Y_n, X_n, H_n X_n)$ is its observed proxy, θ_0 is the true value of θ , and $\tilde{\epsilon}_n \equiv \epsilon_n - \lambda_0 \Delta_n Y_n - \Delta_n X_n \gamma_0$ with $\Delta_n \equiv H_n - G_n$.

Let $\tilde{V}_n \equiv (\iota_n, H_n^2 X_n, X_n, H_n X_n)$ denote an n -by- $(3K + 1)$ matrix of instruments. This \tilde{V}_n is an observable proxy for the (unknown) actual instrument $V_n \equiv (\iota_n, G_n^2 X_n, X_n, G_n X_n)$.

²Referring to potential omission of friends, Patacchini and Venanzoni (2014) say that, “in the large majority of cases (more than 94%), students tend to nominate best friends who are students in the same school and thus are systematically included in the network (and in the neighborhood patterns of social interactions)”. Liu et al. (2014) report that “less than 1% of the students in our sample show a list of ten best friends, less than 3% a list of five males and roughly 4% a list of five females. On average, they declare that they have 4.35 friends with a small dispersion around this mean value (standard deviation equal to 1.41), and in the large majority of cases (more than 90%) the nominated best friends are in the same school.” Lin (2015) says, “this nomination constraint only affects a small portion of our sample, as less than 10% of the sample have listed five male or female friends. Therefore, this restriction should not have a significant impact on the results.” This last speculation is precisely what our first set of results establishes: that consistency of estimates will not be effected if the number of omitted (and hence misclassified) links is sufficiently small.

The 2SLS estimator is:

$$\hat{\theta} = \left[\tilde{R}'_n \tilde{V}_n (\tilde{V}'_n \tilde{V}_n)^{-1} \tilde{V}'_n \tilde{R}_n \right]^{-1} \tilde{R}'_n \tilde{V}_n (\tilde{V}'_n \tilde{V}_n)^{-1} \tilde{V}'_n Y_n. \quad (2)$$

We show that this estimator is consistent when the measurement errors in the adjacency matrices are small in the following sense (where \sum_i is shorthand for $\sum_{i=1}^n$):

Assumption 1 $\sum_i \sum_j E(|H_{n,ij}^* - G_{n,ij}^*|) = O(n^s)$ for some $0 < s < 1$.

Assumption 1 requires the expected sum of absolute measurement errors in G_n^* to increase at a rate slower than the sample size n . This condition holds, for example, if measurement errors occur only for a subset of individuals of order $O(n^s)$ with $s < 1$, and if the magnitude and expected number of mismeasured links for each individual in the subset are bounded. See Section 4 for more examples of how this condition holds under a variety of contexts.

Denote $S_n \equiv I_n - \lambda_0 G_n$, where I_n is an n -by- n identity matrix. When S_n is non-singular, the reduced form for outcomes is:

$$Y_n = S_n^{-1}(\alpha_0 \iota_n + X_n \beta_0 + G_n X_n \gamma_0 + \epsilon_n).$$

We maintain the following regularity conditions.

Assumption 2 (i) ϵ_n is independent from X_n ; individual errors $\epsilon_{n,i}$ are independent across i , with $E(\epsilon_{n,i}) = 0$. There exists a constant $M_0 < \infty$ such that $\Pr\{\sup_{i \leq n} E(|\epsilon_{n,i}| | H_n) \leq M_0\} = 1$ for all n . (ii) G_n^* is a sequence of pre-determined, non-stochastic matrices, and S_n is non-singular for all n . The sequences $\{G_n^*\}$, and $\{S_n^{-1}\}$ are uniformly bounded in both row and column sums. The row and column sums in the sequence $\{H_n^*\}$ are uniformly bounded in probability. (iii) The elements of X_n are uniformly bounded for all n ; $V_n' V_n / n$ converges in probability to a non-singular matrix as $n \rightarrow \infty$.

Part (i) of Assumption 2 states that X_n are exogenous. Notice that we do not impose exogeneity of H_n^* , i.e., the measurement errors $H_n^* - G_n^*$ can be correlated with both ϵ_n and X_n . This is in sharp contrast to most measurement error models, which typically require measurement errors to be independent of some observed or unobserved variables for point identification and estimation. Part (ii) requires the row and column sums of G_n^* and H_n^* to be uniformly bounded, and that the reduced form of outcomes is well-defined. Invertibility of S_n holds if $\sum_j |\lambda G_{n,ij}| < 1$ for all i . In the special case of non-negative elements and row-normalization in G_n^* , $|\lambda| < 1$ is sufficient for non-singular S_n . Part (iii) requires the matrix of actual instruments to have full column rank. The assumptions above on the actual adjacency matrix G_n are standard for linear social interactions models.

Proposition 1 *Under Assumptions 1 and 2,*

$$\widehat{\theta} - \theta_0 = O_p(n^{-1/2} \vee n^{s-1}).$$

This proposition holds because we can establish the following relationship between the feasible 2SLS estimator using the noisy measure with errors H_n , and its infeasible version that uses the unobserved actual G_n :

$$\begin{aligned} \widehat{\theta} - \theta_0 &= \left[\frac{\widetilde{R}'_n \widetilde{V}_n}{n} \left(\frac{\widetilde{V}'_n \widetilde{V}_n}{n} \right)^{-1} \frac{\widetilde{V}'_n \widetilde{R}_n}{n} \right]^{-1} \frac{\widetilde{R}'_n \widetilde{V}_n}{n} \left(\frac{\widetilde{V}'_n \widetilde{V}_n}{n} \right)^{-1} \frac{\widetilde{V}'_n \widetilde{\epsilon}_n}{n} \\ &= \left[\frac{R'_n V_n}{n} \left(\frac{V'_n V_n}{n} \right)^{-1} \frac{V'_n R_n}{n} \right]^{-1} \frac{R'_n V_n}{n} \left(\frac{V'_n V_n}{n} \right)^{-1} \frac{V'_n \epsilon_n}{n} + O_p(n^{s-1}). \end{aligned} \quad (3)$$

It is worth emphasizing that the condition of small-order measurement errors in Assumption 1 does not immediately ensure the discrepancy between a feasible 2SLS estimator (which uses $\widetilde{R}_n, \widetilde{V}_n$) and its infeasible analog (which uses R_n, V_n) is $O_p(n^{s-1})$. This is because the measurement errors enter in the 2SLS estimator non-linearly through (nested) matrix inverses. The contribution of Proposition 1 is to show that this discrepancy is indeed properly bounded under the regularity conditions in Assumption 2 despite these complications.

Under the regularity conditions in Assumption 2, $(R'_n V_n)/n$ and $(V'_n V_n)/n$ both converge in probability to constant matrices with full rank $(2K + 2)$. Under the exogeneity of X_n , the term $V'_n \epsilon_n/n$ is $O_p(n^{-1/2})$ by an application of the Chebyshev's Inequality. Combining these results, we conclude that the estimation errors in (3) is $O_p(n^{-1/2} \vee n^{s-1})$. Thus the 2SLS estimator $\widehat{\theta}$, which uses $H_n^2 X_n$ as an instruments for $H_n Y_n$, is consistent when $s < 1$.

Furthermore, if $s < 1/2$, the effect of measurement errors vanishes fast enough so that it does not affect the \sqrt{n} -rate of convergence or the asymptotic distribution of the 2SLS estimator. This is formalized in the next proposition.

Proposition 2 *Under Assumptions 1 and 2, if $s < 1/2$ then*

$$\sqrt{n}(\widehat{\theta} - \theta_0) \xrightarrow{d} N(0, \Omega),$$

where Ω is the asymptotic variance of the 2SLS estimator using the actual adjacency matrix G_n ; and Ω can be consistently estimated by $\widehat{A}^{-1} \widehat{B} \widehat{A}^{-1}$, where $\widehat{A} \equiv \frac{1}{n} \widetilde{R}'_n P_n \widetilde{R}_n$ and $\widehat{B} \equiv \frac{1}{n} \widetilde{R}'_n P_n \widehat{\Sigma}_n P_n \widetilde{R}_n$, with $P_n \equiv \widetilde{V}_n \left(\widetilde{V}'_n \widetilde{V}_n \right)^{-1} \widetilde{V}'_n$ and $\widehat{\Sigma}_n$ being a diagonal n -by- n matrix whose i -th diagonal entry is the square of the i -th residual in $Y_n - \widetilde{R}_n \widehat{\theta}$.

As noted in the introduction, even slowly growing measurement errors could asymptotically corrupt $\widehat{\theta}$ if the stochastic order of quadratic terms in $\widehat{\theta} - \theta_0$ isn't bounded. The closed

form of the 2SLS estimator plays a key role in deriving our results. In our proofs, this closed form allows us to use Cauchy-Schwartz inequalities to bound the stochastic order of these errors. Key conditions we use for this are boundedness of S_n^{-1} and X_n . Without those, the estimation errors do not obey the stochastic orders we derive.

4 Examples

This section provides several examples of how Assumption 1, which requires a slow rate of growth in link measurement errors, may hold in a range of empirical contexts.

Example 1. (Partitioning groups) Suppose the sample consists of individual units from many known, mutually exclusive groups (e.g., individual households from many villages, or individual students from many schools). Sometimes, data on links within each group is collected (e.g., kinship relationships between households within each village, or friendships between students within each school), while information about links that might exist between pairs of individuals in different groups is not collected. In such cases, all non-zero cross-group links are misclassified as zeros.

In this setting, Assumption 1 holds under intuitive conditions, regardless of whether group sizes grow with the sample size or not. A sufficient condition for Assumption 1 would be that the probability of a non-zero link across groups diminishes faster than some rate as the number of groups in the sample (denoted by M) increases.

First, consider a scenario where there are M groups, and each group m contains a finite, constant number of members n_m , which for simplicity is assumed to be the same for all groups, so $n_m = c \in \mathbf{N}_{++}$ for all m , and the sample size by construction is $n = cM$.³ In this case, the asymptotic experiment lets $M \rightarrow \infty$ with n_m fixed at c for all m . Let the probability of friendship (i.e., a nonzero link) between individuals from two groups be $q_n = O(n^{-\delta_1})$ for $\delta_1 > 0$. Suppose the sample correctly reports all links within groups, but fails to report any information about links that may exist across groups. The order of the expected number of misclassified links in this sample is then $M \times (M - 1) \times q_n = O(M^{2-\delta_1})$. Therefore, Assumption 1 holds as long as $\delta_1 \in (1, 2)$, i.e., the probability of cross-group friendships diminishes fast enough as the number of groups increases.

Next, consider an alternative scenario where the asymptotic experiment allows the group size to increase as the number of groups $M \rightarrow \infty$. Let the size of each group grow at an order of $O(M^\zeta)$ for $\zeta > 0$ so that the order of the sample size n is $O(M^{1+\zeta})$. As before, let

³The result here can be immediately extended to allow for heterogeneous group sizes, provided $n_m < c$ is uniformly bounded by a finite constant c for all m . In that case, the sample size is $n = \sum_m n_m \leq cM = O(M)$.

the probability of a link between individuals from different groups be $q_n = O(n^{-\delta_2})$. Again, suppose the sample correctly reports all links within each group but misses all links between different groups. The order of the expected number of misclassified links in the sample is then $M \times (M - 1) \times O(M^\zeta) \times q_n = O(M^{2+\zeta-(1+\zeta)\delta_2})$. Hence Assumption 1 holds as long as $\delta_2 \in (\frac{1}{1+\zeta}, 1 + \frac{1}{1+\zeta})$.

Example 2. (Panel data) Suppose the sample consists of L cross-sectional individual units, each of which is observed for T time periods. The sample size is $n = LT$. For example, the sample could report weekly test scores of L students over the course of T weeks. Let the structural social effects θ_0 be fixed over time $t = 1, 2, \dots, T$ and assume the structural errors $\epsilon_{i,t}$ are i.i.d. across $i \leq L$ and t . The panel data model fits in the structural form in (1), with $Y_n \equiv (Y'_{n,1}, Y'_{n,2}, \dots, Y'_{n,T})'$ where each $Y_{n,t}$ is a column vector that stacks L individual outcomes at time t . The other arrays X_n and ϵ_n are defined similarly. In this case, G_n is a block-diagonal matrix, with the t -th diagonal block $G_{n,t}$ being an L -by- L adjacency matrix that contains all the links in the network at time t .

Measurement errors in G_n occur if the adjacency matrices $G_{n,t}$ evolve over time, but the researcher only gets to observe them occasionally, i.e., over a strict subset of time periods $\mathcal{T}_{obs} \subset \{1, 2, \dots, T\}$, and assumes the network structure remains constant between those intermittent periods of observation. For example, $\mathcal{T}_{obs} = \{1\}$ means that the researcher only measures the adjacency matrix correctly once, as $G_{n,1}$, in the first period, but then (incorrectly) assumes it stays constant at $G_{n,t} = G_{n,1}$ for all $t = 2, \dots, T$. In this case, the magnitude of measurement errors is determined by the number of existing friendships that are dissolved, or change in strength, and by new links that are created in the subsequent periods $t \geq 2$. For another example, consider the case of weekly test scores above. Suppose the network is only observed once per semester. Then \mathcal{T}_{obs} only contains the number of semesters of observations, while T is the number of weeks for which we observe test scores, and measurement error arises because $G_{n,t}$ is held fixed for all weeks within each semester.

First, consider a large- L , small- T setting, where the asymptotic experiment lets $L \rightarrow \infty$ while holding T fixed at a constant integer. In this case, $n = TL = O(L)$. Suppose $\mathcal{T}_{obs} = \{1\}$, and suppose the probability of dissolving an existing friendship or creating a new one in each subsequent period $t \geq 2$ is $\psi_n = O(n^{-\delta_3})$ for $\delta_3 > 0$. The order of the expected number of mismeasured links is then $(T - 1) \times L \times (L - 1) \times \psi_n = O(n^{2-\delta_3})$, and Assumption 1 holds if $\delta_3 \in (1, 2)$.

Next, consider an alternative large- L , large- T setting, where the asymptotic experiment lets $L \rightarrow \infty$ and $T \rightarrow \infty$ at the same time. Suppose the number of time periods with no network measurement, i.e. $T - \#(\mathcal{T}_{obs})$, grows at rate $O(T^{\xi_1})$ for $\xi_1 \in (0, 1)$. This means the adjacency matrix is correctly measured with high frequency in the sense that the number of time periods with incorrectly imputed network measures grows slower than

T . Let's characterize the relative order of individual units as $L = O(T^{\xi_2})$ for $\xi_2 > 0$ so that $n = LT = O(T^{1+\xi_2})$. As before, let the probability for dissolving existing friendships or creating new ones at $t \geq 2$ be $\psi_n = O(n^{-\delta_4})$ for $\delta_4 > 0$. In this case, the order of the expected number of misclassified links in the full sample is then $L \times (L - 1) \times O(T^{\xi_1}) \times \psi_n$. It then follows that Assumption 1 holds if $\xi_1 + \xi_2 > 1$ and $\delta_4 \in \left(\frac{\xi_1 + \xi_2 - 1}{1 + \xi_2}, \frac{\xi_1 + 2\xi_2}{1 + \xi_2}\right)$.⁴ That is, Assumption 1 holds if the probability of link changes over time is sufficiently low, while the cross-sectional dimension in the panel data grows fast enough relative to the number of time periods.⁵

Example 3. (Caps on self-reported links from surveys) Suppose the sample consists of n individuals in a single, large network. Researchers who collect link information through survey responses sometimes specify a *cap* on the number of links that may be reported by each individual. For example, a questionnaire may ask each student in a class to name *up to* five friends. In this case, link measurement errors are caused by censoring due to the cap when it is binding. For example, a student who had seven friends but could only report five would result in two links that are mismeasured as zero. The order of these errors depends on whether (and how fast) the cap increases with the sample size, as well as the link formation probability.

Let $d_{n,i}$ denote the degree (the total number of friends) an individual i actually has in the sample (which may be more than the number reported). Assume there exists a finite integer \bar{d} such that $P\{d_{n,i} \leq \bar{d}\} = 1$ for all i and n . That is, the total number of friends an individual may actually have is bounded, regardless of the sample size. This reflects the reality that link formation and maintenance are costly in terms of individual time and energy. Furthermore, let κ_n denote a sequence of specified caps on the maximum number of reported links in the sample-collecting survey; this sequence of caps increases with the sample size n , possibly at a very slow rate such as $O(\log n)$. For each individual i , the number of missing links due to the binding cap is then $(d_{n,i} - \kappa_n)_+$, where $(\cdot)_+ \equiv \max\{\cdot, 0\}$. Under the specified conditions, $E[(d_{n,i} - \kappa_n)_+] = o(1)$. It then follows that Assumption 1 is satisfied, because the expected magnitude of overall measurement errors grows at a rate slower than the sample size n .⁶

⁴To see this, note that $O(L^2) \times O(T^{\xi_1}) \times \psi_n = O(T^{2\xi_2 + \xi_1 - \delta_4(1 + \xi_2)}) = O(n^{(2\xi_2 + \xi_1)/(1 + \xi_2) - \delta_4})$. Imposing inequalities to ensure this order is $O(n^s)$ for $s \in (0, 1)$ implies the range of conformable δ_4 .

⁵Our benchmark analysis assumes i.i.d. time-varying errors, which is restrictive in a panel data setting. However, our results generalize to allow some degree of error dependence in the usual way, since the estimator takes the form of linear two-stage least squares.

⁶Assumption 1 can also be satisfied under weaker conditions, provided the right-tail probability mass of $d_{n,i}$ diminishes sufficiently fast relative to the sample size and to the cap on self-reported links. In a model of dyadic link formation, establishing this result formally would require characterizing the magnitude of errors in the normal approximation of a binomial distribution.

Example 4. (Recall or coding errors in survey responses) Samples collected from survey responses are sometimes subject to recall errors (i.e., respondents have incorrect memory of past events or status) and coding errors (i.e., data analysts make mistakes while coding or processing raw responses). These measurement errors may grow at a slow rate relative to the sample size, especially if there is economies of scale in the quality control of data collection, or if the survey provides multiple noisy, proxy measures of the some links.

To illustrate, consider a sample network G_n of n members, where the probability of forming a friendship between any two members is $\pi_n = O(n^{-\delta_5})$ with $\delta_5 > 0$. Suppose the survey responses provides two *independent* measures of G_n (e.g., two responses about the same undirected link), denoted as H_n and W_n respectively, and that each of these two noisy measures misses each actual, existing link in G_n independently at a rate of $\phi_n = O(n^{-\nu})$ for $\nu > 0$. Suppose the data analyst records the (i, j) -th entry of the network as $\max\{H_{n,ij}, W_{n,ij}\}$. Then the order of the expected measurement errors, i.e., the total number of nonzero links recorded as zero, is given by $n \times (n - 1) \times \pi_n \times \phi_n^2 = O(n^{2-\delta_5-2\nu})$. Therefore, Assumption 1 holds as long as $\delta_5 \in (1 - 2\nu, 2 - 2\nu)$.

5 Simulation

We investigate the performance of the 2SLS estimator with mismeasured links using simulated data. The structural equation in our data-generating process (DGP) is (1), where X_n consists of two regressors: the first is independently drawn from $\{-1, 1, 2\}$ with equal probability, and the second is from $N(0, 1)$. The error terms ε_i are i.i.d. from $N(0, 1)$. Links in G_n^* are independent draws from a Bernoulli distribution with success probability $p_n = \mu/n$ for some constant $\mu < \infty$. By this construction, the expected number of friends for each individual is μ . Let G_n be a row-normalization of G_n^* .

We generate misclassified links using $H_{n,ij}^* = G_{n,ij}^* \cdot e_{1i} + (1 - G_{n,ij}^*) \cdot e_{2i}$ for $i \neq j$, where e_{1i} and e_{2i} are Bernoulli random variables with success probabilities $1 - \tau_{1i}$ and τ_{2i} respectively. Therefore, τ_{1i} is the misclassification probability that $H_{n,ij}^* = 0$ when the true $G_{n,ij}^* = 1$, and τ_{2i} is the misclassification probability that $H_{n,ij}^* = 1$ when $G_{n,ij}^* = 0$. We set $\tau_{1i} = \rho_i n^{s-1}$ and $\tau_{2i} = 100\rho_i n^{s-2}$, where $\rho_{n,i} = (\sum_{j=1}^n G_{n,ij}^*/\mu + |\varepsilon_{n,i}|)/3$. For each individual i , the probability of misclassification increases in the number of individual i 's friends $\sum_{j=1}^n G_{n,ij}^*$, and in the magnitude of i 's unobserved error $|\varepsilon_{n,i}|$. This construction makes the measurement errors both endogenous (correlated with the model errors) and correlated with the actual row-normalized G_n .

We set the model parameters to be $\alpha = 1$; $\lambda = 0.4$; $\beta = (1.5, 2)'$ and $\gamma = (0.9, 0.6)'$; choose $\mu = 20$; and experiment with the rates in measurement errors $s = 0.1, 0.3, 0.5$, and

0.7. We experiment with sample sizes $n = 200, 500,$ and 1000 . For each value of s and n , we simulate $T = 200$ samples and calculate the mean squared error, the bias, the sample standard deviation, and the asymptotic standard error of the estimator. We also report the total number of misclassified links based on the average over $T = 200$ simulated samples.

Results are summarized in Table 1. We observe several patterns:

1. The 2SLS estimates of all parameters converge at \sqrt{n} rate. The mean-squared errors decrease proportionately as the sample size increases.
2. Consistent with our asymptotic theory, the 2SLS estimator using the misclassified adjacency matrix H_n works almost as well as its infeasible analog using the actual G_n when the measurement error rate is $s < 0.5$. This suggests that the sample sizes we consider are large enough for the asymptotic approximations to apply. Note that with our DGP the estimates in Table 1 with $s < 0.5$ have error rates where the expected number of misclassified links is less than n .
3. The average standard errors are mostly close to the standard deviations for all values of s . This conforms with our asymptotic theory, because the problem with inference for larger values of s is that the bias in the estimator shrinks at rate n^{s-1} . Similarly, with $s \geq 0.5$, the parameter estimates deteriorate primarily due to the bias rather than the variance.
4. With both the true and mismeasured adjacency matrices, the mean-squared errors are much smaller for the direct effects β than for the peer and contextual effects λ and γ . The mean squared errors are also much lower for the discrete regressor effects β_1 and γ_1 than for the continuous regressor effects β_2 and γ_2 .

Table 1. 2SLS Estimators with Misclassified Links

		$n = 200$				$n = 500$				$n = 1000$			
		m.s.e.	bias	std	a.s.e.	m.s.e.	bias	std	a.s.e.	m.s.e.	bias	std	a.s.e.
True		Mis.#	0				0				0		
α		3.880	-0.114	1.971	2.197	1.519	0.031	1.235	1.310	0.762	0.065	0.873	0.887
λ		0.336	0.025	0.581	0.654	0.131	-0.010	0.362	0.386	0.068	-0.019	0.260	0.264
β_1		0.003	0.005	0.058	0.058	0.001	-0.003	0.036	0.036	0.001	-0.000	0.027	0.026
β_2		0.005	0.008	0.072	0.073	0.002	0.001	0.048	0.045	0.001	-0.000	0.032	0.032
γ_1		0.802	-0.029	0.898	1.006	0.301	0.019	0.549	0.597	0.165	0.030	0.406	0.410
γ_2		1.571	-0.040	1.256	1.348	0.561	0.020	0.750	0.796	0.278	0.032	0.528	0.545
$s = 0.1$		Mis.#	66				81				88		
α		4.100	-0.058	2.029	2.254	1.576	0.033	1.258	1.325	0.780	0.070	0.883	0.894
λ		0.365	0.008	0.605	0.672	0.135	0.010	0.368	0.391	0.070	-0.020	0.263	0.266
β_1		0.003	0.004	0.058	0.058	0.001	-0.003	0.036	0.036	0.001	-0.000	0.027	0.026
β_2		0.005	0.008	0.072	0.073	0.002	0.001	0.048	0.045	0.001	-0.000	0.032	0.032
γ_1		0.877	-0.015	0.938	1.033	0.307	0.015	0.556	0.604	0.168	0.030	0.410	0.413
γ_2		1.610	-0.012	1.272	1.382	0.574	0.019	0.760	0.805	0.284	0.033	0.533	0.549
$s = 0.3$		Mis.#	193				278				351		
α		4.599	0.058	2.149	2.388	1.678	0.014	1.2985	1.367	0.833	0.083	0.911	0.912
λ		0.405	-0.023	0.638	0.712	0.144	-0.002	0.380	0.403	0.074	-0.023	0.271	0.272
β_1		0.004	0.003	0.059	0.059	0.001	-0.003	0.035	0.037	0.001	-0.000	0.027	0.026
β_2		0.005	0.009	0.073	0.074	0.002	0.001	0.048	0.046	0.001	-0.000	0.032	0.032
γ_1		0.949	0.018	0.977	1.094	0.334	-0.005	0.579	0.622	0.179	0.031	0.423	0.421
γ_2		1.756	0.041	1.328	1.461	0.598	-0.005	0.775	0.830	0.305	0.035	0.552	0.560
$s = 0.5$		Mis.#	556				968				1401		
α		5.620	0.136	2.373	2.773	1.995	0.0670	1.414	1.519	1.060	0.133	1.023	0.982
λ		0.498	-0.032	0.706	0.828	0.172	-0.012	0.416	0.449	0.093	-0.035	0.303	0.293
β_1		0.004	0.001	0.062	0.060	0.001	-0.003	0.036	0.037	0.001	-0.000	0.028	0.026
β_2		0.005	0.011	0.073	0.075	0.002	0.001	0.049	0.046	0.001	-0.000	0.033	0.032
γ_1		1.174	-0.022	1.086	1.272	0.408	-0.015	0.640	0.691	0.218	0.032	0.467	0.453
γ_2		2.157	0.021	1.472	1.691	0.732	-0.021	0.857	0.921	0.376	0.041	0.614	0.602
$s = 0.7$		Mis.#	1605				3346				5572		
α		17.93	0.433	4.223	4.212	4.581	0.253	2.131	2.075	1.812	0.157	1.340	1.291
λ		1.549	-0.095	1.244	1.252	0.395	-0.0470	0.628	0.613	0.158	-0.025	0.398	0.385
β_1		0.004	0.002	0.066	0.064	0.002	-0.004	0.038	0.039	0.001	-0.001	0.028	0.027
β_2		0.006	0.009	0.076	0.081	0.003	-0.001	0.052	0.048	0.001	0.001	0.033	0.033
γ_1		3.643	-0.050	1.913	1.898	0.894	-0.058	0.946	0.934	0.374	-0.069	0.610	0.589
γ_2		6.452	0.011	2.547	2.533	1.545	-0.047	1.245	1.250	0.649	-0.056	0.806	0.786

Note: m.s.e (mean squared error), bias, std (standard deviation) are calculated using the empirical distribution of 200 estimates. "a.s.e." is the average of standard errors in $T = 200$ samples.

6 Application

Lin and Lee (2010) model teenage pregnancy rates, using the following model (where the subscript of sample size n is suppressed):

$$Teen_i = \lambda \sum_{j=1}^n G_{ij} Teen_j + \alpha + Edu_i \beta_1 + Inco_i \beta_2 + FHH_i \beta_3 + Black_i \beta_4 + Phy_i \beta_5 + \varepsilon_i,$$

where $Teen_i$ is the teenage pregnancy rate in county i , which is the percentage of pregnancies among females 12-17 years old, and G_{ij} is the (i, j) -th entry in the row-normalization of the original link matrix G^* , where $G_{ij}^* = 1$ if counties i and j are neighboring counties. Edu_i is the education service expenditure (in units of \$100), $Inco_i$ is median household income (divided by 1000), FHH_i is the percentage of female-headed households, $Black_i$ is the proportion of black population and Phy_i is the number of physicians per 1000 population, all in county i .⁷

The sample size is $n = 761$. Among all the $761 \times 760 = 578,360$ entries (diagonal are zero) in the original network G_n^* , there are 4,606 non-zero links. We treat the adjacency matrix reported in the sample as the true network, artificially introduce misclassified links, and then evaluate how this affect the 2SLS estimates. We generate misclassified links using $H_{ij}^* = G_{ij}^* \cdot e_{1i} + (1 - G_{ij}^*) \cdot e_{2i}$, where e_{1i} and e_{2i} are Bernoulli with success probabilities $\tau_{1i} = \rho_i n^{s-1}$ and $\tau_{2i} = 100 \rho_i n^{s-2}$ respectively. We set $\rho_i = (y_i / \bar{y})^2$, so that for each individual i misclassification is more likely to happen the larger the magnitude of the observed outcome y_i .

We report 2SLS estimates using HX and H^2X as instruments. Unlike our structural model, Lin and Lee (2010) assume contextual effects, i.e., γ in equation (1), are zero so that GX does not appear as regressors. It would therefore have been possible to just use HX as instruments for estimation. Nonetheless, to illustrate our proposition, we use both HX and H^2X as instruments.

Table 2 reports results based on 1000 Monte Carlo replications for each value of s . Results are reported where the model is estimated both with and without row normalization.

Consistent with our propositions, when the misclassification rate is low ($s < 0.5$), the 2SLS parameter and standard error estimates using the mis-measured H_n are very similar to those based on G_n . The same is true for estimation based on matrices H_n^* and G_n^* that are not row-normalized. When s increases, the bias and inaccuracy of the estimators increases, as expected. In particular, the parameter estimates (especially λ) become quite biased when $s \geq 0.5$ (which, by our theory, is when bias shrinks at a slower rate than variance).

⁷The data are collected from 761 counties in Colorado, Iowa, Kansas, Minnesota, Missouri, Montana, Nebraska, North Dakota, South Dakota, and Wyoming. Data details are in Lin and Lee (2010).

Table 2. Estimation Results with Different Misclassification Rates

	λ	α	$100\beta_1$	β_2	β_3	β_4	β_5	Mis. #
Row-normalized adjacency matrices $G_{ij} = G_{ij}^* / \left(\sum_j G_{ij}^*\right)$ and $H_{ij} = H_{ij}^* / \left(\sum_j H_{ij}^*\right)$								
True	0.4813 (0.079)	6.1911 (1.469)	-0.9824 (0.651)	-0.1871 (0.040)	0.7347 (0.063)	0.1267 (0.057)	-0.4956 (0.188)	0
$s = 0.1$	0.4897 (0.081)	6.1085 (1.480)	-0.9910 (0.651)	-0.1878 (0.040)	0.7355 (0.063)	0.1289 (0.057)	-0.4980 (0.188)	111
$s = 0.3$	0.5132 (0.085)	5.8759 (1.512)	-1.0086 (0.652)	-0.1895 (0.040)	0.7375 (0.063)	0.1341 (0.057)	-0.5049 (0.188)	418
$s = 0.5$	0.6017 (0.099)	4.9578 (1.626)	-1.0542 (0.654)	-0.1943 (0.040)	0.7422 (0.063)	0.1465 (0.057)	-0.5227 (0.189)	1578
$s = 0.7$	0.8138 (0.139)	2.7629 (1.985)	-1.1726 (0.660)	-0.2092 (0.040)	0.7589 (0.064)	0.1683 (0.057)	-0.5535 (0.191)	5948
Original adjacency matrices G_{ij}^* and H_{ij}^* without normalization								
True	0.0239 (0.009)	10.840 (1.261)	-1.5244 (0.669)	-0.2348 (0.041)	0.8151 (0.064)	0.2061 (0.058)	-0.5731 (0.194)	0
$s = 0.1$	0.0275 (0.009)	10.491 (1.248)	-1.5290 (0.666)	-0.2317 (0.040)	0.8087 (0.064)	0.2069 (0.057)	-0.5658 (0.193)	111
$s = 0.3$	0.0356 (0.008)	9.6492 (1.216)	-1.5361 (0.659)	-0.2239 (0.040)	0.7916 (0.063)	0.2079 (0.057)	-0.5463 (0.191)	418
$s = 0.5$	0.0486 (0.005)	7.5887 (1.130)	-1.5473 (0.633)	-0.2039 (0.038)	0.7351 (0.061)	0.2058 (0.055)	-0.4813 (0.184)	1578
$s = 0.7$	0.0442 (0.003)	4.9575 (0.984)	-1.5211 (0.571)	-0.1749 (0.034)	0.6170 (0.055)	0.1858 (0.049)	-0.3396 (0.166)	5948

Note: The table reports average estimates and average standard errors (in parentheses) from 1000 simulated samples.

7 Conclusions

We show that in 2SLS estimation of linear social network models, measurement errors in the network can have no impact on estimation and inference of structural parameters if the magnitude of measurement errors in the adjacency matrix grows sufficiently slowly with the sample size. Moreover, these results hold even if the measurement errors are correlated with

model errors, covariates, and outcomes. A useful agenda for future work would be to see if similar results can be obtained for more general network models.

Appendix

For a generic matrix A , let $A_{(i)}$, $A_{[k]}$ denote its i -th row and k -th column respectively; and A_{ij} denote its (i, j) -th component, so that $A_{(i)\iota}$ is the sum of the i -th row in A . Let $\Delta_n^* \equiv H_n^* - G_n^*$. With row normalization, we can write Δ_n as

$$H_n - G_n = \text{diag} \left\{ \left(\frac{1}{G_{n,(1)\iota}^*}, \dots, \frac{1}{G_{n,(n)\iota}^*} \right) \right\} \Delta_n^* + \text{diag} \left\{ \left(\frac{1}{H_{n,(1)\iota}^*} - \frac{1}{G_{n,(1)\iota}^*}, \dots, \frac{1}{H_{n,(n)\iota}^*} - \frac{1}{G_{n,(n)\iota}^*} \right) \right\} H_n^*.$$

The following two lemmas are useful for the proofs. (In what follows, we suppress the subscript n to H_n, G_n, H_n^*, G_n^* to simplify notation.)

Lemma A1. *Let a_n, b_n be random vectors in \mathbb{R}^n . Suppose there exist constants $M_1, M_2 < \infty$ such that $\Pr\{\sup_{i \leq n} |a_{n,i}| \leq M_1\} = 1$ and $\Pr\{\sup_{j \leq n} E(|b_{n,j}| | \Delta_n) \leq M_2\} = 1$ for all n . Then $\frac{1}{n} a_n' \Delta_n b_n = O_p(n^{s-1})$ and $\frac{1}{n} a_n' \Delta_n^* b_n = O_p(n^{s-1})$ under Assumption 1.*

Proof of Lemma A1. From the triangle inequality,

$$\begin{aligned} E \left(\sum_i \sum_j |\Delta_{n,ij}| \right) &= E \left(\sum_i \sum_j \left| \frac{1}{G_{(i)\iota}^*} \Delta_{n,ij}^* + \frac{(G_{(i)\iota}^* - H_{(i)\iota}^*) \iota}{(G_{(i)\iota}^*)(H_{(i)\iota}^*)} H_{ij}^* \right| \right) \\ &\leq E \left[\sum_i \sum_j \left(\frac{1}{G_{(i)\iota}^*} |\Delta_{n,ij}^*| + \frac{1}{(G_{(i)\iota}^*)(H_{(i)\iota}^*)} |(G_{(i)\iota}^* - H_{(i)\iota}^*) \iota| \times H_{ij}^* \right) \right] \\ &\leq E \left[\sum_i \left(\frac{1}{G_{(i)\iota}^*} \sum_j |\Delta_{n,ij}^*| + \frac{1}{G_{(i)\iota}^*} \sum_j |\Delta_{n,ij}^*| \right) \right] = O(n^s), \end{aligned}$$

Furthermore,

$$E(|\frac{1}{n} a_n' \Delta_n b_n|) \leq \frac{1}{n} E \left[\sup_{i,j} E(|a_{n,i} b_{n,j}| | \Delta_n) \cdot \left(\sum_i \sum_j |\Delta_{n,ij}| \right) \right] = O(n^{s-1}).$$

Similar arguments can be applied to show $\frac{1}{n} a_n' \Delta_n^* b_n = O_p(n^{s-1})$. \square

Lemma A2. *Under Assumption 2, $\sup_{i \leq n} |V_{iq}| = O(1)$ and $\sup_{i \leq n} V_{iq}^2 = O(1)$ for $q = 1, \dots, K$, and there exists constant $M^* < \infty$ such that $\Pr\{\sup_i E(|y_i| | \Delta_n) \leq M^*\} = 1$ for all n .*

Proof of Lemma A2. Note

$$\sup_{i \leq n} \left([G_{(i)}^2 X_{[q]}]^2 \right) \leq \left(\sup_{i \leq n} \sum_k |G_{ik}| \right)^2 \left(\sup_{k \leq n} \sum_j |G_{kj}| \right)^2 \left(\sup_{j \leq n} x_{jq}^2 \right) = O(1).$$

This implies $\sup_i V_{iq}^2 = O(1)$. By Liapounov's Inequality, $\sup_i V_{iq}^2 = O(1)$ implies $\sup_i |V_{iq}| = O(1)$ for all $q = 1, \dots, K$.

It then follows from reduced form for Y_n that

$$\begin{aligned} \sup_i E(|y_i||\Delta_n) &= \sup_i E \left(\left| \sum_j (S_n^{-1})_{ij} \left(\alpha_0 + x'_j \beta_0 + \sum_k G_{jk} x'_k \gamma_0 + \varepsilon_j \right) \right| \middle| \Delta_n \right) \\ &\leq \sup_i \left[\sum_j (S_n^{-1})_{ij} \right] \times \sup_j E \left(|\alpha_0| + |x'_j \beta_0| + \sum_k |G_{jk}| \times |x'_k \gamma_0| + |\varepsilon_j| \middle| \Delta_n \right). \end{aligned}$$

Hence, there exists some constant $M^* < \infty$ with $\Pr\{\sup_i E(|y_i||\Delta_n) \leq M^*\} = 1$. \square

Proof of Proposition 1. Recall

$$\hat{\theta} - \theta_0 = \left[\frac{\tilde{R}'_n \tilde{V}_n}{n} \left(\frac{\tilde{V}'_n \tilde{V}_n}{n} \right)^{-1} \frac{\tilde{V}'_n \tilde{R}_n}{n} \right]^{-1} \frac{\tilde{R}'_n \tilde{V}_n}{n} \left(\frac{\tilde{V}'_n \tilde{V}_n}{n} \right)^{-1} \frac{\tilde{V}'_n \tilde{\epsilon}_n}{n} \quad (4)$$

where

$$\begin{aligned} \frac{1}{n} \tilde{V}'_n \tilde{R}_n &= \frac{1}{n} V'_n R_n + \frac{1}{n} V'_n (0, \Delta_n Y_n, 0, \Delta_n X_n) \\ &\quad + \frac{1}{n} (0, (G_n \Delta_n + \Delta_n G_n + \Delta_n^2) X_n, 0, \Delta_n X_n)' R_n \\ &\quad + \frac{1}{n} (0, (G_n \Delta_n + \Delta_n G_n + \Delta_n^2) X_n, 0, \Delta_n X_n)' (0, \Delta_n Y_n, 0, \Delta_n X_n). \end{aligned}$$

$$\begin{aligned} \frac{1}{n} \tilde{V}'_n \tilde{V}_n &= \frac{1}{n} V'_n V_n + \frac{1}{n} V'_n (0, (G_n \Delta_n + \Delta_n G_n + \Delta_n^2) X_n, 0, \Delta_n X_n) \\ &\quad + \frac{1}{n} (0, (G_n \Delta_n + \Delta_n G_n + \Delta_n^2) X_n, 0, \Delta_n X_n)' V_n \\ &\quad + \frac{1}{n} (0, (G_n \Delta_n + \Delta_n G_n + \Delta_n^2) X_n, 0, \Delta_n X_n)' (0, (G_n \Delta_n + \Delta_n G_n + \Delta_n^2) X_n, 0, \Delta_n X_n) \end{aligned}$$

$$\begin{aligned} \frac{1}{n} \tilde{V}'_n \tilde{\epsilon}_n &= \frac{1}{n} V'_n \epsilon_n - \frac{1}{n} \lambda_0 V'_n \Delta_n Y_n - \frac{1}{n} V'_n \Delta_n X_n \gamma_0 \\ &\quad + \frac{1}{n} (0, (G_n \Delta_n + \Delta_n G_n + \Delta_n^2) X_n, 0, \Delta_n X_n)' (\epsilon_n - \lambda_0 \Delta_n Y_n - \Delta_n X_n \gamma_0). \end{aligned} \quad (5)$$

Due to Assumption 2 and Lemma A2, $\sup_i V_i V'_i = O(1)$. Lemma A2 also suggests that V_n and $X_n \gamma_0$ satisfy the dominance conditions on the vectors a_n ; Y_n and ϵ_n satisfy the dominance conditions b_n in Lemma A1. The second to the fourth terms on the RHS of (5) can all be expressed as $\frac{1}{n} a'_n \Delta_n b_n$ in Lemma A1, and hence are $O_p(n^{s-1})$. Because $\frac{1}{n} V'_n \epsilon_n = O_p(n^{-1/2})$, it then follows that $\frac{1}{n} \tilde{V}'_n \tilde{\epsilon}_n = O_p(n^{-1/2} \vee n^{s-1})$.

Proof of Proposition 2. As

$$\sqrt{n}(\hat{\theta} - \theta_0) = \left[\frac{R'_n V_n}{n} \left(\frac{V'_n V_n}{n} \right)^{-1} \frac{V'_n R_n}{n} \right]^{-1} \frac{R'_n V_n}{n} \left(\frac{V'_n V_n}{n} \right)^{-1} \frac{V'_n \epsilon_n}{\sqrt{n}} + O_p(n^{s-1/2}),$$

when $s < 1/2$, $\sqrt{n}(\hat{\theta} - \theta_0)$ has the same asymptotic distribution as the 2SLS estimator using true network links.

Consider the asymptotic variance Ω . Let Σ_n be the diagonal matrix of the error variance, i.e., $\Sigma_{ii} = E(\varepsilon_i^2)$. We have $\Omega = A^{-1}BA^{-1}$, where

$$\begin{aligned} A &= p \lim \frac{R'_n V_n}{n} \left(\frac{V'_n V_n}{n} \right)^{-1} \frac{V'_n R_n}{n}; \\ B &= p \lim \frac{R'_n V_n}{n} \left(\frac{V'_n V_n}{n} \right)^{-1} \left(\frac{1}{n} V'_n \Sigma_n V_n \right) \left(\frac{V'_n V_n}{n} \right)^{-1} \frac{V'_n R_n}{n}. \end{aligned}$$

Using Lemma A1, we can show that

$$\hat{A} = A + O_p(n^{s-1})$$

and

$$\hat{B} = B + \frac{R'_n V_n}{n} \left(\frac{V'_n V_n}{n} \right)^{-1} \left(\frac{1}{n} \tilde{V}'_n \hat{\Sigma}_n \tilde{V}_n - \frac{1}{n} V'_n \Sigma_n V_n \right) \left(\frac{V'_n V_n}{n} \right)^{-1} \frac{V'_n R_n}{n} + O_p(n^{s-1}).$$

Then, what left is to show that from the fact that $\frac{1}{n} \tilde{V}'_n \hat{\Sigma}_n \tilde{V}_n - \frac{1}{n} V'_n \Sigma_n V_n$ is $o_p(1)$. As

$$\frac{1}{n} \tilde{V}'_n \hat{\Sigma}_n \tilde{V}_n - \frac{1}{n} V'_n \Sigma_n V_n = \frac{1}{n} V'_n \left(\hat{\Sigma}_n - \Sigma_n \right) V_n + O_p(n^{s-1}),$$

and the first term on the RHS is $O_p(n^{-1/2} \vee n^{s-1})$ because

$$\begin{aligned} \frac{1}{n} V'_n \left(\hat{\Sigma}_n - \Sigma_n \right) V_n &= \frac{1}{n} \sum_{i=1}^n \left((Y_n - \tilde{R}_n \hat{\theta})_{(i)}^2 - E(\varepsilon_i^2) \right) v_i v'_i \\ &= \frac{1}{n} \sum_{i=1}^n v_i v'_i [\varepsilon_i^2 - E(\varepsilon_i^2)] + \frac{1}{n} \sum_{i=1}^n v_i v'_i \left([\tilde{R}_i(\theta_0 - \hat{\theta})]^2 + [(\lambda_0 \Delta_n Y_n + \Delta_n X_n \gamma_0)_{(i)}]^2 \right) \\ &\quad + \frac{2}{n} \sum_{i=1}^n v_i v'_i \tilde{R}_i(\theta_0 - \hat{\theta}) \varepsilon_i - \frac{2}{n} \sum_{i=1}^n v_i v'_i [\tilde{R}_i(\theta_0 - \hat{\theta}) + \varepsilon_i] (\lambda_0 \Delta_n Y_n + \Delta_n X_n \gamma_0)_{(i)} \\ &= O_p(n^{-1/2}) + O_p(\theta_0 - \hat{\theta}) + O_p(n^{s-1}) = O_p(n^{-1/2} \vee n^{s-1}) \end{aligned}$$

Together, we have $\hat{A}^{-1} \hat{B} \hat{A}^{-1} - A^{-1} B A^{-1} = O_p(n^{-1/2} \vee n^{s-1}) = o_p(1)$.

References

Blume, L., W. Brock, S. Durlauf, R. Tayaraman, 2015. Linear social interactions models. *Journal of Political Economy* 123 (2): 444-496.

Boucher, V., Y. Bramoullé, H. Djebbari, B. Fortin, 2014. Do peers affect student achievement? Evidence from Canada using group size variation. *Journal of Applied Econometrics* 29: 91-109.

- Bramoullé, Y, H. Djebbari, B. Fortin, 2009. Identification of peer effects through social networks. *Journal of Econometrics* 150: 41–55.
- Calvó-Armengol, A., E. Patacchini, Y., Zenou, 2009. Peer effects and social networks in education. *Review of Economic Studies* 76: 1239–1267.
- De Paula Á., I. Rasul, P. CL. Souza, 2018. Recovering social networks from panel data: identification, simulations and an application, CeMMAP working papers CWP58/18.
- Griffith, A. 2021. Name your friends, but only five? The importance of censoring in peer effects estimates using social network data. Working paper, The University of Washington.
- Hauser, C., M. Pfaffermayr, G. Tappeiner, J. Walde, 2009. Social capital formation and intra familial correlation: A social panel perspective. *Singapore Economic Review* 54: 473–488.
- Lee, L., 2007. Identification and estimation of econometric models with group interactions, contextual factors and fixed effects. *Journal of Econometrics* 140: 333–374.
- Lee, L., X. Liu, X. Lin, 2010. Specification and estimation of social interaction models with network structures. *Econometrics Journal* 13: 145–176.
- Lewbel, A., X. Qu, X. Tang, 2021 Social Networks with Unobserved Links. Boston College Working Papers in Economics 1004.
- LeSage, J. P., R. K. Pace, 2009. Introduction to spatial econometrics. Taylor Francis/CRC Press, Boca Raton.
- Lin, X., 2010. Identifying peer effects in student academic achievement by spatial autoregressive models with group unobservables. *Journal of Labor Economics* 28: 825–860.
- Lin, X., L. Lee, 2010. GMM estimation of spatial autoregressive models with unknown heteroskedasticity. *Journal of Econometrics* 157, 34-52.
- Lin, X., 2015. Utilizing spatial autoregressive models to identify peer effects among adolescents. *Empirical Economics*, 49(3), 929-960.
- Liu X, E. Patacchini, Y. Zenou, 2014. Endogenous peer effects: local aggregate or local average? *Journal of Economic Behavior and Organization* 103: 39–59.
- Manski, C F., 1993. Identification of Endogenous Social Effects: The Reflection Problem. *Review of Economic Studies* 60 (3): 531–42.
- Patacchini E., Y. Zenou, 2012. Juvenile delinquency and conformism. *Journal of Law, Economics, and Organization* 28: 1–31.
- Patacchini, E., G. Venanzoni, 2014. Peer effects in the demand for housing quality. *Journal of Urban Economics*, 83(0), 6-17.