Online Appendix

Table of Contents

B Proof of Theorem 1: Additional Material

Lemma 2 (Differentiability). *$\mathbb{E}[\mathbf{m}_m(\boldsymbol{\theta}, \boldsymbol{\rho})]$ is continuously differentiable in $(\boldsymbol{\theta}, \boldsymbol{\rho})$.*

Proof. Since $\mathbf{m}_m(\boldsymbol{\theta}, \boldsymbol{\rho})$ is continuously differentiable in $\boldsymbol{\theta}$ and absolutely integrable, along with its derivative with respect to $\boldsymbol{\theta}$, it follows that $\mathbb{E}[\mathbf{m}_m(\boldsymbol{\theta}, \boldsymbol{\rho})]$ is continuously differentiable in $\boldsymbol{\theta}$ by the Leibniz integral rule. However, $\mathbf{m}_m(\boldsymbol{\theta}, \boldsymbol{\rho})$ is continuously differentiable in $\boldsymbol{\rho}$ only for *almost all* $\boldsymbol{\rho}$. We now show that $\mathbb{E}[\mathbf{m}_m(\boldsymbol{\theta}, \boldsymbol{\rho})]$ is continuously differentiable for all $\boldsymbol{\rho}$.

Consider $\mathbf{B}_m\dot{\mathbf{G}}_m(\boldsymbol{\rho}) = \mathbf{B}_m f(\dot{\mathbf{A}}_m(\boldsymbol{\rho}))$ for some (conformable) matrix $\mathbf{B}_m$. We have:

$$\hat{\mathbb{E}}_m(\mathbf{B}_m\dot{\mathbf{G}}_m(\boldsymbol{\rho})) = \sum_{\hat{\mathbf{A}}_m} \mathbf{B}_m f(\hat{\mathbf{A}}_m) P(\mathbf{A}_m(\boldsymbol{\rho}) = \hat{\mathbf{A}}_m \mid \mathbf{X}_m, \kappa(\mathcal{A}_m)), \tag{10}$$

where the sum is taken over all the possible network configurations $\hat{\mathbf{A}}_m$, and where $P(\mathbf{A}_m(\boldsymbol{\rho}) = \hat{\mathbf{A}}_m|\mathbf{X}_m, \kappa(\mathcal{A}_m)) = \Pi_{ij} P(a_{ij,m}(\boldsymbol{\rho}) = \hat{a}_{ij,m}|\mathbf{X}_m, \kappa(\mathcal{A}_m))$, as defined in Equation (2). Thus $\hat{\mathbb{E}}_m(\mathbf{B}_m\dot{\mathbf{G}}_m(\boldsymbol{\rho}))$ is continuously differentiable in $\boldsymbol{\rho}$ by Assumption 9. By adapting this argument for the appropriate definition of the matrix $\mathbf{B}_m$, and for $\hat{\mathbb{E}}_m$, and $\mathbb{E}_m^{(0)}$, this shows that $\mathbb{E}[\mathbf{m}_{m,rst}(\boldsymbol{\theta}, \boldsymbol{\rho})|\mathbf{X}_m, \kappa(\mathcal{A}_m)]$ is continuously differentiable in $(\boldsymbol{\theta}, \boldsymbol{\rho})$. As a result, $\mathbb{E}[\mathbf{m}_m(\boldsymbol{\theta}, \boldsymbol{\rho})]$ is continuously differentiable in $(\boldsymbol{\theta}, \boldsymbol{\rho})$ by the Leibniz integral rule. □

Lemma 3 (Uniform convergence). *We establish the following results.*

(a) $\mathbb{E}[\bar{\mathbf{m}}_M(\boldsymbol{\theta}, \hat{\boldsymbol{\rho}})] - \mathbb{E}[\bar{\mathbf{m}}_M(\boldsymbol{\theta}, \boldsymbol{\rho}_0)]$ *converges uniformly to* $\mathbf{0}$ *in* $\boldsymbol{\theta}$ *as* $M \to \infty$.

(b) $\bar{\mathbf{m}}_M(\boldsymbol{\theta}, \hat{\boldsymbol{\rho}}) - \mathbb{E}[\bar{\mathbf{m}}_M(\boldsymbol{\theta}, \hat{\boldsymbol{\rho}})]$ *converges uniformly in probability to* $\mathbf{0}$ *in* $\boldsymbol{\theta}$ *as* $M \to \infty$.

(c) $\bar{\mathbf{m}}_M(\boldsymbol{\theta}, \hat{\boldsymbol{\rho}}) - \mathbb{E}[\bar{\mathbf{m}}_M(\boldsymbol{\theta}, \boldsymbol{\rho}_0)]$ *converges uniformly in probability to* $\mathbf{0}$ *in* $\boldsymbol{\theta}$ *as* $M \to \infty$.

Proof. As $\mathbb{E}[\bar{\mathbf{m}}_M(\boldsymbol{\theta}, \boldsymbol{\rho})]$ is nonstochastic and continuously differentiable in $\boldsymbol{\rho}$ (Lemma 2), it follows that $\mathbb{E}[\bar{\mathbf{m}}_M(\boldsymbol{\theta}, \hat{\boldsymbol{\rho}})] - \mathbb{E}[\bar{\mathbf{m}}_M(\boldsymbol{\theta}, \boldsymbol{\rho}_0)]$ converges to $\mathbf{0}$ pointwise for each $\boldsymbol{\theta}$, given that $\text{plim}\, \hat{\boldsymbol{\rho}} = \boldsymbol{\rho}_0$. To establish uniform convergence, we apply Lemma 2.9 in Newey and McFadden (1994). The required conditions are satisfied: (i) the space of $\boldsymbol{\theta}$ is compact (Assumption 6), (ii) $\mathbb{E}[\bar{\mathbf{m}}_M(\boldsymbol{\theta}, \hat{\boldsymbol{\rho}})] - \mathbb{E}[\bar{\mathbf{m}}_M(\boldsymbol{\theta}, \boldsymbol{\rho}_0)]$ converges to $\mathbf{0}$ for all $\boldsymbol{\theta}$, and (iii) the derivative of $\mathbb{E}[\bar{\mathbf{m}}_M(\boldsymbol{\theta}, \hat{\boldsymbol{\rho}})] - \mathbb{E}[\bar{\mathbf{m}}_M(\boldsymbol{\theta}, \boldsymbol{\rho}_0)]$ with respect to $\boldsymbol{\theta}$ is bounded in probability (Assumptions 1 and 7). As a result, $\mathbb{E}[\bar{\mathbf{m}}_M(\boldsymbol{\theta}, \hat{\boldsymbol{\rho}})] - \mathbb{E}[\bar{\mathbf{m}}_M(\boldsymbol{\theta}, \boldsymbol{\rho}_0)]$ uniformly converges to zero in $\boldsymbol{\theta}$. This completes the proof of Statement (a).

For Statement (b), we first establish pointwise convergence by showing that the variance of $\bar{\mathbf{m}}_M(\boldsymbol{\theta}, \hat{\boldsymbol{\rho}})$ vanishes asymptotically. By the law of iterated variances, we have:

$$\begin{aligned}
\mathbb{V}(\bar{\mathbf{m}}_M(\boldsymbol{\theta}, \hat{\boldsymbol{\rho}})) &= \mathbb{V}\left\{\mathbb{E}\big(\bar{\mathbf{m}}_M(\boldsymbol{\theta}, \hat{\boldsymbol{\rho}})|\mathbf{X}, \kappa(\mathcal{A}), \hat{\boldsymbol{\rho}}\big)\right\} + \mathbb{E}\left\{\mathbb{V}\big(\bar{\mathbf{m}}_M(\boldsymbol{\theta}, \hat{\boldsymbol{\rho}})|\mathbf{X}, \kappa(\mathcal{A}), \hat{\boldsymbol{\rho}}\big)\right\}, \\
\lim \mathbb{V}(\bar{\mathbf{m}}_M(\boldsymbol{\theta}, \hat{\boldsymbol{\rho}})) &= \mathbb{V}\left\{\operatorname{plim} \mathbb{E}\big(\bar{\mathbf{m}}_M(\boldsymbol{\theta}, \hat{\boldsymbol{\rho}})|\mathbf{X}, \kappa(\mathcal{A}), \hat{\boldsymbol{\rho}}\big)\right\} + \\
&\quad \mathbb{E}\left\{\operatorname{plim} \mathbb{V}\big(\bar{\mathbf{m}}_M(\boldsymbol{\theta}, \hat{\boldsymbol{\rho}})|\mathbf{X}, \kappa(\mathcal{A}), \hat{\boldsymbol{\rho}}\big)\right\}, \\
\lim \mathbb{V}(\bar{\mathbf{m}}_M(\boldsymbol{\theta}, \hat{\boldsymbol{\rho}})) &= \mathbb{V}\left\{\operatorname{plim} \mathbb{E}\big(\bar{\mathbf{m}}_M(\boldsymbol{\theta}, \hat{\boldsymbol{\rho}})|\mathbf{X}, \kappa(\mathcal{A}), \hat{\boldsymbol{\rho}}\big)\right\} + \\
&\quad \mathbb{E}\left\{\operatorname{plim} \frac{1}{M^2} \sum_{m=1}^{M} \mathbb{V}\big(\mathbf{m}_m(\boldsymbol{\theta}, \hat{\boldsymbol{\rho}})|\mathbf{X}_m, \mathcal{A}_m, \hat{\boldsymbol{\rho}}\big)\right\}. \qquad (11)
\end{aligned}$$

The second equality holds by the dominated convergence theorem.[32] Equation (11) holds by the fact that $\mathbf{m}_m(\boldsymbol{\theta}, \hat{\boldsymbol{\rho}})$ are independent across m, conditional on $\mathbf{X}_m, \mathcal{A}_m, \hat{\boldsymbol{\rho}}$.

From Equation (11), it is thus sufficient to show that $\operatorname{plim} \mathbb{E}(\bar{\mathbf{m}}_M(\boldsymbol{\theta}, \hat{\boldsymbol{\rho}})|\mathbf{X}, \kappa(\mathcal{A}), \hat{\boldsymbol{\rho}})$ is nonstochastic and that $\operatorname{plim} \dfrac{1}{M^2} \sum_{m=1}^{M} \mathbb{V}(\mathbf{m}_m(\boldsymbol{\theta}, \hat{\boldsymbol{\rho}})|\mathbf{X}_m, \mathcal{A}_m, \hat{\boldsymbol{\rho}}) = \mathbf{0}$. First, the variance of $\mathbb{E}[\bar{\mathbf{m}}_M(\boldsymbol{\theta}, \boldsymbol{\rho})|\mathbf{X}_m, \kappa(\mathcal{A}_m)]$ vanishes asymptotically by being the variance of an average of independent elements. Therefore, it converges in $\mathcal{L}^2$ and, thus, in probability to its expectation, $\mathbb{E}[\bar{\mathbf{m}}_M(\boldsymbol{\theta}, \boldsymbol{\rho})]$, which is nonstochastic. Consequently, $\operatorname{plim} \mathbb{E}(\bar{\mathbf{m}}_M(\boldsymbol{\theta}, \hat{\boldsymbol{\rho}})|\mathbf{X}, \kappa(\mathcal{A}), \hat{\boldsymbol{\rho}})$ is also nonstochastic and, thus, $\mathbb{V}\{\operatorname{plim} \mathbb{E}(\bar{\mathbf{m}}_M(\boldsymbol{\theta}, \hat{\boldsymbol{\rho}})|\mathbf{X}, \kappa(\mathcal{A}), \hat{\boldsymbol{\rho}})\} = \mathbf{0}$.

[32] Specifically, $\mathbb{E}(\bar{\mathbf{m}}_M(\boldsymbol{\theta}, \hat{\boldsymbol{\rho}}) \mid \mathbf{X}, \kappa(\mathcal{A}), \hat{\boldsymbol{\rho}})$ and $\mathbb{V}(\bar{\mathbf{m}}_M(\boldsymbol{\theta}, \hat{\boldsymbol{\rho}}) \mid \mathbf{X}, \kappa(\mathcal{A}), \hat{\boldsymbol{\rho}})$ are bounded by Assumptions 1 and 7, as well as by the fact that the conditional variance of $\boldsymbol{\varepsilon}_m$ is uniformly bounded (Assumption 8). Consequently, we can interchange the expectation and probability limit operators.

Second, $\mathbb{V}(\mathbf{m}_m(\boldsymbol{\theta}, \hat{\boldsymbol{\rho}})|\mathbf{X}_m, \mathcal{A}_m, \hat{\boldsymbol{\rho}}) < \infty$ because $\mathbf{m}_m(\boldsymbol{\theta}, \hat{\boldsymbol{\rho}})$ is linear in $\boldsymbol{\varepsilon}_m$ which has a bounded variance (Assumption 8). Thus, $\text{plim}\,\frac{1}{M^2}\sum_{m=1}^{M}\mathbb{V}(\mathbf{m}_m(\boldsymbol{\theta}, \hat{\boldsymbol{\rho}})|\mathbf{X}_m, \mathcal{A}_m, \hat{\boldsymbol{\rho}}) = 0$ and $\mathbb{E}\{\text{plim}\,\frac{1}{M^2}\sum_{m=1}^{M}\mathbb{V}(\mathbf{m}_m(\boldsymbol{\theta}, \hat{\boldsymbol{\rho}})|\mathbf{X}_m, \mathcal{A}_m, \hat{\boldsymbol{\rho}})\} = \mathbf{0}$.

Consequently, $\bar{\mathbf{m}}_M(\boldsymbol{\theta}, \hat{\boldsymbol{\rho}}) - \mathbb{E}[\bar{\mathbf{m}}_M(\boldsymbol{\theta}, \boldsymbol{\rho}_0)]$ converges in probability to $\mathbf{0}$. The convergence is uniform in $\boldsymbol{\theta}$ from Lemma 2.9 of Newey and McFadden (1994) as in Statements (a). This completes the proof of Statement (b) and, consequently, the lemma, given that Statements (a) and (b) together imply Statement (c). □

Lemma 4. *The stochastic equicontinuity condition CD1 is verified.*

Proof. As $\boldsymbol{\theta}_0$ in CD1 is fixed, we ignore it in our notations and define $\tilde{\mathbf{m}}_m(\boldsymbol{\rho}) = \mathbf{m}_m(\boldsymbol{\theta}_0, \boldsymbol{\rho})$. We follow Andrews (1994) and define

$$\nu_M(\boldsymbol{\rho}) = \frac{1}{\sqrt{M}}\sum_m [\tilde{\mathbf{m}}_m(\boldsymbol{\rho}) - \mathbb{E}(\tilde{\mathbf{m}}_m(\boldsymbol{\rho}))],$$

so that conditions CD1 is equivalent to $\nu_M(\hat{\boldsymbol{\rho}}) - \nu_M(\boldsymbol{\rho}_0) = o_p(1)$. Consider the following pseudo-metric, for any dimension k of the moment function

$$d_k(\boldsymbol{\rho}_1, \boldsymbol{\rho}_2) = \sup_m(\mathbb{E}[\tilde{\mathbf{m}}_{m,[k]}(\boldsymbol{\rho}_1) - \tilde{\mathbf{m}}_{m,[k]}(\boldsymbol{\rho}_2)]^2)^{1/2}.$$

We say that the process ν_M is *stochastically equicontinuous* if, $\forall \epsilon > 0$, $\exists \delta > 0$ such that

$$\text{plim} \sup_{d_k(\boldsymbol{\rho}_1, \boldsymbol{\rho}_2) < \delta} |\nu_{M,[k]}(\boldsymbol{\rho}_1) - \nu_{M,[k]}(\boldsymbol{\rho}_2)| < \epsilon,$$

for each dimension $[k]$. To see that stochastic equicontinuity implies CD1, note that, for any $\epsilon > 0$:

$$\begin{aligned}
& \lim P(|\nu_{M,[k]}(\hat{\boldsymbol{\rho}}) - \nu_{M,[k]}(\boldsymbol{\rho}_0)| > \epsilon) \\
\leq\; & \lim P(|\nu_{M,[k]}(\hat{\boldsymbol{\rho}}) - \nu_{M,[k]}(\boldsymbol{\rho}_0)| > \epsilon, d_k(\hat{\boldsymbol{\rho}}, \boldsymbol{\rho}_0) \leq \delta) + \lim P(d_k(\hat{\boldsymbol{\rho}}, \boldsymbol{\rho}_0) > \delta)
\end{aligned}$$

$$\leq \lim P\left(\sup_{d(\boldsymbol{\rho}_1,\boldsymbol{\rho}_2)<\delta} |\nu_{M,[k]}(\hat{\boldsymbol{\rho}}) - \nu_{M,[k]}(\boldsymbol{\rho}_0)| > \epsilon\right)$$

The last inequality holds because $\lim P(d_k(\hat{\boldsymbol{\rho}}, \boldsymbol{\rho}_0) > \delta) = 0$ by the consistency of $\hat{\boldsymbol{\rho}}$. Stochastic equicontinuity implies that δ can be chosen so that $\lim P\Big(\sup_{d(\boldsymbol{\rho}_1,\boldsymbol{\rho}_2)<\delta} |\nu_{M,[k]}(\hat{\boldsymbol{\rho}}) - \nu_{M,[k]}(\boldsymbol{\rho}_0)| > \epsilon\Big)$ is as small as desired. Thus, $\lim P(|\nu_{M,[k]}(\hat{\boldsymbol{\rho}}) - \nu_{M,[k]}(\boldsymbol{\rho}_0)| > \epsilon)$ can also be made arbitrarily small, that is, $\nu_M(\hat{\boldsymbol{\rho}}) - \nu_M(\boldsymbol{\rho}_0) = o_p(1)$, which corresponds to our condition CD1. It is thus sufficient to show that ν_M is stochastically equicontinuous.

Following Andrews (1994), Section 5, we say that $\tilde{\mathbf{m}}_m$ is Type $IV(p = 2)$ if the parameter space is bounded (Assumption 6) and

$$\sup_m \left(\mathbb{E}\Big(\sup_{\boldsymbol{\rho}_1:\|\boldsymbol{\rho}_1-\boldsymbol{\rho}_2\|<\delta} (\tilde{\mathbf{m}}_{m,[k]}(\boldsymbol{\rho}_1) - \tilde{\mathbf{m}}_{m,[k]}(\boldsymbol{\rho}_2))^2\Big)\right)^{1/2} \leq C\delta^{\psi}, \tag{12}$$

for all $\boldsymbol{\rho}_2$ and all $\delta > 0$ in a neighborhood of 0, for some finite positive constants C and ψ, and for all dimensions $[k]$.

We can express $\tilde{\mathbf{m}}_{m,[k]}(\boldsymbol{\rho})$ as a linear function of $\boldsymbol{\varepsilon}_m$ (e.g., see Equation (6) in Appendix A.1). Thus, $\tilde{\mathbf{m}}_{m,[k]}(\boldsymbol{\rho}_1) - \tilde{\mathbf{m}}_{m,[k]}(\boldsymbol{\rho}_2) = u_{1,m,[k]}(\boldsymbol{\rho}_1, \boldsymbol{\rho}_2) + \mathbf{u}_{2,m,[k]}(\boldsymbol{\rho}_1, \boldsymbol{\rho}_2)\boldsymbol{\varepsilon}_m$ for some scalar $u_{1,m,[k]}(\boldsymbol{\rho}_1, \boldsymbol{\rho}_2)$ and row vector $\mathbf{u}_{2,m,[k]}(\boldsymbol{\rho}_1, \boldsymbol{\rho}_2)$. Additionally, by Assumptions 1 and 7, $|u_{1,m,[k]}(\boldsymbol{\rho}_1, \boldsymbol{\rho}_2)|$ and $\|\mathbf{u}_{2,m,[k]}(\boldsymbol{\rho}_1, \boldsymbol{\rho}_2)\|$ are uniformly bounded by some $\bar{u}_{1,[k]}$ and $\bar{u}_{2,[k]}$, respectively, where $\bar{u}_{1,[k]}$ and $\bar{u}_{2,[k]}$ do not depend on $\boldsymbol{\rho}_1$ and $\boldsymbol{\rho}_2$. Therefore, $(\tilde{\mathbf{m}}_{m,[k]}(\boldsymbol{\rho}_1) - \tilde{\mathbf{m}}_{m,[k]}(\boldsymbol{\rho}_2))^2$ is dominated by $(\bar{u}_{1,[k]} + \bar{u}_{2,[k]}\|\boldsymbol{\varepsilon}_m\|)^2$, for any sub-multiplicative norm $\|\cdot\|$.

Since $(\bar{u}_1 + \bar{u}_2\|\boldsymbol{\varepsilon}_m\|)^2$ is integrable (see Assumption 8), we can apply the dominated convergence theorem and interchange the expectation and the second supremum symbol in (12). A sufficient condition for (12) is thus:

$$\sup_m \Big(\sup_{\boldsymbol{\rho}_1:\|\boldsymbol{\rho}_1-\boldsymbol{\rho}_2\|<\delta} \mathbb{E}((\tilde{\mathbf{m}}_{m,[k]}(\boldsymbol{\rho}_1) - \tilde{\mathbf{m}}_{m,[k]}(\boldsymbol{\rho}_2))^2)\Big)^{1/2} \leq C\delta^{\psi}. \tag{13}$$

Now, note that $\mathbb{E}(\tilde{\mathbf{m}}_{m,[k]}(\boldsymbol{\rho}_1) - \tilde{\mathbf{m}}_{m,[k]}(\boldsymbol{\rho}_2))^2$ is continuously differentiable in $\boldsymbol{\rho}_1$ with bounded derivatives following the argument in Lemma 2. See in particular Equation (10).[33]

Then, by the Mean Value Theorem, we have $\mathbb{E}(\tilde{\mathbf{m}}_{m,[k]}(\boldsymbol{\rho}_1) - \tilde{\mathbf{m}}_{m,[k]}(\boldsymbol{\rho}_2))^2 = \mathcal{D}(\boldsymbol{\rho}^+, \boldsymbol{\rho}_2)(\boldsymbol{\rho}_1 - \boldsymbol{\rho}_2)$, where $\mathcal{D}(\boldsymbol{\rho}^+, \boldsymbol{\rho}_2)$ is the derivative of $\mathbb{E}(\tilde{\mathbf{m}}_{m,[k]}(\boldsymbol{\rho}_1) - \tilde{\mathbf{m}}_{m,[k]}(\boldsymbol{\rho}_2))^2$ with respect to $\boldsymbol{\rho}_1$ at some $\boldsymbol{\rho}^+$ lying between $\boldsymbol{\rho}_1$ and $\boldsymbol{\rho}_2$. Thus, $\tilde{\mathbf{m}}_m$ is of Type IV with $p = 2$, $\psi = 1/2$, and C as the upper bound of $\mathcal{D}(\boldsymbol{\rho}^+, \boldsymbol{\rho}_2)$. As a result, Condition (13), and thus Condition (12), hold.

By Theorem 4 in Andrews (1994), stochastic equicontinuity CD1 holds if Ossianders' condition (his condition D) holds, $\lim \frac{1}{M}\sum_m \mathbb{E}\sup_{\boldsymbol{\rho}} |\tilde{\mathbf{m}}_m|^{2+\eta} < \infty$ for some $\eta > 0$ (his condition B), and if groups m are independent (Assumption 1) implied by his condition C). By Theorem 5 in Andrews (1994), Ossianders' condition holds if $\tilde{\mathbf{m}}_m$ is Type $IV(p = 2)$, which we just shown. His condition B is verified because $\mathbb{E}(\|\boldsymbol{\varepsilon}_m\|^{2+\eta})$ is bounded for some $\eta > 0$ (Assumption 8). Thus, as above, we can employ the dominated convergence theorem and interchange the expectation and the second supremum. Since $\mathbb{E}|\tilde{\mathbf{m}}_m|^{2+\eta}$ is bounded, then condition B follows. $\square$

B.1 Identification

In this section, we show that Assumption 11 can be expressed as an identification condition on a concentrated objective function.

As $\mathbf{W}_0$ is positive definite (Assumption 10), Assumption 11 is equivalent to stating that $Q_0(\boldsymbol{\theta})$ has a unique minimizer. Since $Q_0(\boldsymbol{\theta})$ depends on the true value $\boldsymbol{\rho}_0$ and not its estimator, all simulated networks in this section are drawn from the true network distribution.

[33]The derivative is bounded because the linking probabilities have bounded derivatives (Assumption 9), and $\mathbf{X}_m$, $\mathbb{E}(\|\boldsymbol{\varepsilon}_m\|^2)$, and the entries of the network matrices are bounded (Assumptions 1, 7, and 8).

We therefore omit $\boldsymbol{\rho}_0$ from the notation for simplicity.

We define:

$$\mathbf{B}_m(\alpha) = \frac{1}{RST}\sum_{rst}\dot{\mathbf{Z}}_m^{(r)\prime}(\mathbf{I}_m - \alpha\ddot{\mathbf{G}}_m^{(s)})(\mathbf{I}_m - \alpha\dddot{\mathbf{G}}_m^{(t)})^{-1}\dddot{\mathbf{V}}_m^{(t)},$$

$$\mathbf{D}_m(\alpha) = \frac{1}{RS}\sum_{rs}\dot{\mathbf{Z}}_m^{(r)\prime}(\mathbf{I}_m - \alpha\ddot{\mathbf{G}}_m^{(s)}).$$

We have $\bar{\mathbf{m}}_M(\boldsymbol{\theta}, \boldsymbol{\rho}_0) = \frac{1}{M}\sum_m[\mathbf{D}_m(\alpha)\mathbf{y}_m - \mathbf{B}_m(\alpha)\tilde{\boldsymbol{\theta}}]$. The first-order condition of the minimization of $Q_0(\boldsymbol{\theta})$ with respect to $\tilde{\boldsymbol{\theta}}$ is:

$$\left[\lim\frac{1}{M}\sum_m\mathbb{E}\left[\mathbf{B}_m(\alpha)\right]\tilde{\boldsymbol{\theta}}\right]'\mathbf{W}_0\left[\lim\frac{1}{M}\sum_m\left(\mathbb{E}\left[\mathbf{D}_m(\alpha)\mathbf{y}_m\right] - \mathbb{E}\left[\mathbf{B}_m(\alpha)\right]\tilde{\boldsymbol{\theta}}\right)\right] = 0.$$

A sufficient condition for the last equation to have a unique solution in $\tilde{\boldsymbol{\theta}}$ is that $\bar{\mathbf{B}}_0(\alpha) := \lim\frac{1}{M}\sum_m\mathbb{E}\left[\mathbf{B}_m(\alpha)\right]$ is a full rank matrix for all α. Under this condition, the solution $\tilde{\boldsymbol{\theta}}$ can be expressed as:

$$\hat{\tilde{\boldsymbol{\theta}}}(\alpha) = [\bar{\mathbf{B}}_0'(\alpha)\mathbf{W}_0\bar{\mathbf{B}}_0]^{-1}\bar{\mathbf{B}}_0'(\alpha)\mathbf{W}_0\bar{\mathbf{F}}_0(\alpha),$$

where $\bar{\mathbf{F}}_0(\alpha) = \operatorname{plim}\frac{1}{M}\sum_m\mathbf{D}_m(\alpha)\mathbf{y}_m$. By replacing $\mathbf{y}_m = (\mathbf{I}_m - \alpha_0\mathbf{G}_m)^{-1}(\mathbf{V}_m\tilde{\boldsymbol{\theta}}_0 + \boldsymbol{\varepsilon}_m)$ in the expression of $\bar{\mathbf{F}}_0(\alpha)$, we obtain:

$$\bar{\mathbf{F}}_0(\alpha) = \lim\frac{1}{M}\sum_m\mathbb{E}\left[\mathbf{F}_m(\alpha)\right], \quad \text{where}$$

$$\mathbf{F}_m(\alpha) = \frac{1}{RS}\sum_{rs}\mathbb{E}\left(\dot{\mathbf{Z}}_m^{(r)\prime}(\mathbf{I}_m - \alpha\ddot{\mathbf{G}}_m^{(s)})(\mathbf{I}_m - \alpha_0\mathbf{G}_m)^{-1}\mathbf{V}_m\tilde{\boldsymbol{\theta}}_0\right).$$

Since $\bar{\mathbf{F}}_0(\alpha_0) = \bar{\mathbf{B}}_0(\alpha_0)\tilde{\boldsymbol{\theta}}_0$, it follows that $\hat{\tilde{\boldsymbol{\theta}}}(\alpha_0) = \tilde{\boldsymbol{\theta}}_0$, which means that $\tilde{\boldsymbol{\theta}}_0$ is identified if α_0 is identified; the underlying condition being that $\bar{\mathbf{B}}_0(\alpha)$ is full rank.

By replacing the solution $\hat{\tilde{\boldsymbol{\theta}}}(\alpha)$ in the objective function, we can concentrate $Q_0(\boldsymbol{\theta})$ around α as $Q_0^c(\alpha) = \bar{\mathbf{Q}}^c(\alpha)'\mathbf{W}_0\bar{\mathbf{Q}}^c(\alpha)$, Where

$$\bar{\mathbf{Q}}^c(\alpha) = \bar{\mathbf{F}}_0(\alpha) - \bar{\mathbf{B}}_0(\alpha)[\bar{\mathbf{B}}_0'(\alpha)\mathbf{W}_0\bar{\mathbf{B}}_0]^{-1}\bar{\mathbf{B}}_0'(\alpha)\mathbf{W}_0\bar{\mathbf{F}}_0(\alpha).$$

Let $\mathbf{W}_0^{1/2}$ be the positive definite square root of $\mathbf{W}_0$. We have $\mathbf{W}_0^{1/2}\bar{\mathbf{Q}}^c(\alpha) = [\mathbf{I}_w - \mathbf{P}_{\mathbf{B}}(\alpha)]\mathbf{W}_0^{1/2}\bar{\mathbf{F}}_0(\alpha)$, where $\mathbf{P}_{\mathbf{B}}(\alpha) := \mathbf{W}_0^{1/2}\bar{\mathbf{B}}_0(\alpha)[\bar{\mathbf{B}}_0'(\alpha)\mathbf{W}_0\bar{\mathbf{B}}_0]^{-1}\bar{\mathbf{B}}_0'(\alpha)\mathbf{W}_0^{1/2}$ is a projection matrix onto the space of the column of $\mathbf{W}_0^{1/2}\bar{\mathbf{B}}_0(\alpha)$ and $\mathbf{I}_w$ if the identity matrix of the same dimension as $\mathbf{W}_0$. The concentrated objective function can be written as:

$$Q_0^c(\alpha) = [\mathbf{W}_0^{1/2}\bar{\mathbf{F}}_0(\alpha)]'[\mathbf{I}_w - \mathbf{P}_{\mathbf{B}}(\alpha)]\mathbf{W}_0^{1/2}\bar{\mathbf{F}}_0(\alpha).$$

For Assumption 11 to hold, the equation $Q_0^c(\alpha) = 0$ must not have multiple solutions. Unfortunately, simplifying this condition is challenging due to the nonlinearity of $Q_0^c(\alpha)$. A similar issue arises with the maximum likelihood estimator even when the network is fully observed. In this case, the identification condition also leads to a nonlinear equation in the peer effect parameter (see Lee, 2004, Assumption 9).

Nevertheless, it is straightforward to sketch the empirical counterpart of $Q_0^c(\alpha)$ since it is a function of a single variable. In numerous simulation exercises, we observe that $Q_0^c(\alpha)$ is strictly convex, even when all entries of $\mathbf{A}_m$ are simulated from an estimated distribution (without setting some entries to observed data). This evidence is encouraging and suggests that the solution to $Q_0^c(\alpha) = 0$ is likely unique.

B.2 Asymptotic variance estimation

Estimating the asymptotic variance of $\sqrt{M}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ can be challenging, as it requires computing the derivative of the expected moment function to estimate $\boldsymbol{\Gamma}_0 = \text{plim}\,\dfrac{\partial \bar{\mathbf{m}}_M^*(\boldsymbol{\theta}_0, \boldsymbol{\rho}^+)}{\partial \boldsymbol{\rho}'}$ (see Appendix A.2). We now present a simple method for estimating this asymptotic variance by adapting Houndetoungan and Maoude (2024).

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

$$\left(\frac{\partial \bar{\mathbf{m}}_M(\hat{\boldsymbol{\theta}}, \boldsymbol{\rho})}{\partial \boldsymbol{\theta}'}\right)' \mathbf{W}_M[\bar{\mathbf{m}}_M(\hat{\boldsymbol{\theta}}, \boldsymbol{\rho})] = \mathbf{0}.$$

By applying the mean value theorem to $\bar{\mathbf{m}}_M(\boldsymbol{\theta}, \boldsymbol{\rho})$, we have

$$\sqrt{M}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = -\left[\mathbf{H}_M(\hat{\boldsymbol{\theta}})'\mathbf{W}_M\mathbf{H}_M(\boldsymbol{\theta}^*)\right]^{-1}\mathbf{H}_M(\hat{\boldsymbol{\theta}})'\mathbf{W}_M\frac{1}{\sqrt{M}}\sum_{m=1}^{M}\mathbf{m}_m(\boldsymbol{\theta}_0, \hat{\boldsymbol{\rho}}),$$

where $\mathbf{H}_M(\boldsymbol{\theta}) = \dfrac{\partial \bar{\mathbf{m}}_M(\boldsymbol{\theta}, \hat{\boldsymbol{\rho}})}{\partial \boldsymbol{\theta}'}$ and $\boldsymbol{\theta}^*$ is some point between $\hat{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}_0$.

Let $\boldsymbol{\Omega}_M = \mathbb{V}(\frac{1}{\sqrt{M}}\sum_{m=1}^{M}\mathbf{m}_m(\boldsymbol{\theta}_0, \hat{\boldsymbol{\rho}}))$. We assume the following:

Assumption 13. $\operatorname{plim}\boldsymbol{\Omega}_M = \boldsymbol{\Omega}_0$ *and* $\operatorname{plim}\mathbf{H}_M(\boldsymbol{\theta}_0) = \mathbf{H}_0$ *exist and are finite matrices.*

Under this assumption, the asymptotic variance of $\sqrt{M}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ is:

$$\mathbb{V}_0(\sqrt{M}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)) = (\mathbf{H}_0'\mathbf{W}_0\mathbf{H}_0)^{-1}\mathbf{H}_0'\mathbf{W}_0\boldsymbol{\Omega}_0\mathbf{W}_0\mathbf{H}_0(\mathbf{H}_0'\mathbf{W}_0\mathbf{H}_0)^{-1}.$$

The expression for the asymptotic variance is similar to the variance of the standard GMM estimator. The key difference is that $\boldsymbol{\Omega}_0$, which is the asymptotic variance of $\dfrac{1}{\sqrt{M}}\sum_{m=1}^{M}\mathbf{m}_m(\boldsymbol{\theta}_0, \hat{\boldsymbol{\rho}})$, accounts for the uncertainty in $\boldsymbol{\varepsilon}_m$, the first-stage estimator $\hat{\boldsymbol{\rho}}$, and the finite number of simulated networks from the estimated network distribution.

To estimate $\mathbb{V}_0(\sqrt{M}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0))$, one can replace $\mathbf{H}_0$ and $\mathbf{W}_0$ with their usual estimators. Specifically, $\mathbf{H}_0$ can be estimated by $\mathbf{H}_M(\hat{\boldsymbol{\theta}})$ and $\mathbf{W}_0$ can be estimated by $\mathbf{W}_M$. Let $\mathcal{S}$ be the set of simulated networks from the network distribution and the true network. To estimate $\boldsymbol{\Omega}_0$, we rely on the following decomposition.

$$\boldsymbol{\Omega}_M = \mathbb{V}\left(\frac{1}{\sqrt{M}}\sum_{m=1}^{M}\mathbf{m}_m(\boldsymbol{\theta}_0, \hat{\boldsymbol{\rho}})\right),$$

$$\boldsymbol{\Omega}_M = \mathbb{E}\left\{\mathbb{V}\left(\frac{1}{\sqrt{M}}\sum_{m=1}^{M}\mathbf{m}_m(\boldsymbol{\theta}_0, \hat{\boldsymbol{\rho}})|\mathbf{X}_m, \mathcal{S}\right)\right\} + \mathbb{V}\left\{\mathbb{E}\left(\frac{1}{\sqrt{M}}\sum_{m=1}^{M}\mathbf{m}_m(\boldsymbol{\theta}_0, \hat{\boldsymbol{\rho}})|\mathbf{X}_m, \mathcal{S}\right)\right\},$$

$$\mathbf{\Omega}_M = \mathbb{E}\left(\frac{1}{M}\sum_{m=1}^{M}\mathbf{V}_m\right) + \mathbb{V}\left(\frac{1}{\sqrt{M}}\sum_{m=1}^{M}\mathcal{E}_m\right).$$

where $\mathbf{V}_m = \mathbb{V}\left(\mathbf{m}_m(\boldsymbol{\theta}_0, \hat{\boldsymbol{\rho}})|\mathbf{X}_m, \hat{\boldsymbol{\rho}}, \kappa(\mathcal{A}_m)\right)$ and $\mathcal{E}_m = \mathbb{E}\left(\mathbf{m}_m(\boldsymbol{\theta}_0, \hat{\boldsymbol{\rho}})|\mathbf{X}_m, \hat{\boldsymbol{\rho}}, \kappa(\mathcal{A}_m)\right)$.

Note that both $\mathbf{V}_m$ and $\mathcal{E}_m$ can be easily computed and estimated. They represent the conditional variance and the conditional expectation of the moment function, given $\mathcal{S}$.

Let $\mathbf{V}_M = \frac{1}{M}\sum_{m=1}^{M}\mathbf{V}_m$ and $\mathcal{E}_M = \frac{1}{\sqrt{M}}\sum_{m=1}^{M}\mathcal{E}_m$. It follows that:

$$\mathbf{\Omega}_0 = \operatorname{plim}\mathbf{V}_M + \lim\mathbb{V}\left(\mathcal{E}_M\right).$$

The first term is due to the error term of the model $\boldsymbol{\varepsilon}_m$, whereas the second term reflects the uncertainty associated with the estimation of $\hat{\boldsymbol{\rho}}$ and the simulated networks. In practice, $\operatorname{plim}\mathbf{V}_M$ can be estimated by the average of the conditional variance of the moment function without accounting for the uncertainty in the simulated network. To estimate $\lim\mathbb{V}\left(\mathcal{E}_M\right)$, we repeatedly generate many $\mathcal{S}$, each associated with a new $\boldsymbol{\rho}$ simulated from the estimator of the distribution of $\hat{\boldsymbol{\rho}}$. For each $\mathcal{S}$, we compute the associated $\mathcal{E}_M$. Finally, the sample variance of the generated values of $\mathcal{E}_M$ serves as an estimator of $\lim\mathbb{V}\left(\mathcal{E}_M\right)$.[34]

C Additional Results

C.1 Simple estimators

When the network is fully observed, the moment function of the standard instrumental variables approach is linear in parameters (Bramoullé et al., 2009). Consequently, the estimator can be computed without requiring numerical optimization and identification conditions can be easily tested. Our SGMM estimator does not exhibit such simplicity when $\mathbf{G}_m\mathbf{y}_m$ is not

[34] Our R package offers tools to compute this variance.

observed. In this section, we discuss other straightforward estimators that result from a linear moment function. We first discuss the case where $\mathbf{G}_m\mathbf{X}_m$ and $\mathbf{G}_m\mathbf{y}_m$ are observed.

Proposition 1. *[Conditions] Suppose that $\mathbf{G}_m\mathbf{X}_m$ and $\mathbf{G}_m\mathbf{y}_m$ are observed. Let $\boldsymbol{H}_m$ be a matrix such that (1) at least one column of $\mathbf{H}_m^k\mathbf{X}_m$ is (strongly) correlated with $\mathbf{G}_m\mathbf{y}_m$, conditional on $[\mathbf{1}_m, \mathbf{X}_m, \mathbf{G}_m\mathbf{X}_m]$ for $k \geq 2$, and (2) $\mathbb{E}[\boldsymbol{\varepsilon}_m|\mathbf{X}_m, \mathbf{A}_m, \mathbf{H}_m] = \mathbf{0}$ for all m. Finally, define the matrix $\mathbf{Z}_m = [\mathbf{1}_m, \mathbf{X}_m, \mathbf{G}_m\mathbf{X}_m, \mathbf{H}_m^2\mathbf{X}_m, \mathbf{H}_m^3\mathbf{X}_m...]$.*

[Results] Then, under classical assumptions (e.g., Cameron and Trivedi (2005), Proposition 6.1), the (linear) GMM estimator based on the moment function $\frac{1}{M}\sum_m \mathbf{Z}_m'\boldsymbol{\varepsilon}_m$ is consistent and asymptotically normally distributed with the usual asymptotic variance-covariance matrix.

Condition (1) is the relevancy condition, whereas condition (2) is the exogeneity condition.[35] Although Proposition 1 holds for any matrix $\mathbf{H}_m$ such that conditions (1) and (2) hold, the most sensible example in our context is when $\mathbf{H}_m$ is constructed using a draw from $\hat{P}(\mathbf{A}_m|\hat{\boldsymbol{\rho}}, \mathbf{X}_m, \kappa(\mathcal{A}_m))$.

Importantly, the moment conditions remain valid even when the researcher uses a *misspecified* estimator of the distribution $P(\mathbf{A}_m|\mathbf{X}_m, \kappa(\mathcal{A}_m))$, as long as the specification error on $P(\mathbf{A}_m|\mathbf{X}_m, \kappa(\mathcal{A}_m))$ does not induce a correlation with $\boldsymbol{\varepsilon}_m$.[36] This could be of great practical importance, especially if the estimation of $\hat{P}(\mathbf{A}_m|\hat{\boldsymbol{\rho}}, \mathbf{X}_m, \kappa(\mathcal{A}_m))$ suffers from a small sample bias.

Second, we present a simple, but asymptotically biased, linear GMM estimator when

[35] Although (for simplicity) in Proposition 1, we use the entire matrix $\mathbf{X}_m$ to generate the instruments $\mathbf{H}_m\mathbf{X}_m$, in practice, one should avoid including instruments (i.e., columns of $\mathbf{H}_m\mathbf{X}_m$) that are weakly correlated with $\mathbf{G}_m\mathbf{y}_m$.

[36] We would like to thank Chih-Sheng Hsieh and Arthur Lewbel for discussions on this important point.

$\mathbf{G}_m\mathbf{X}_m$ is observed and $\mathbf{G}_m\mathbf{y}_m$ is not. The presentation of such an estimator is useful because simulations show that the asymptotic bias turns out to be negligible in many cases, especially for moderate group sizes. Moreover, the estimator is computationally attractive because the estimator can be written in a closed form.

Proposition 2. *[Conditions] Assume that $\mathbf{G}_m\mathbf{X}_m$ is observed. Let $\ddot{\mathbf{S}}_m = [\mathbf{1}_m, \mathbf{X}_m, \mathbf{G}_m\mathbf{X}_m, \ddot{\mathbf{G}}_m\mathbf{X}_m, \ddot{\mathbf{G}}_m\mathbf{y}_m]$ and $\dot{\mathbf{Z}}_m = [\mathbf{1}_m, \mathbf{X}_m, \mathbf{G}_m\mathbf{X}_m, \ddot{\mathbf{G}}_m\mathbf{X}_m, \dot{\mathbf{G}}_m^2\mathbf{X}_m, \dot{\mathbf{G}}_m^3\mathbf{X}_m, ...]$. We consider the moment function $\frac{1}{M}\sum_{m=1}^{M}\dot{\mathbf{Z}}_m(\mathbf{y}_m - \ddot{\mathbf{S}}_m\ddot{\boldsymbol{\theta}})$ and $\check{\boldsymbol{\theta}}$ be the associated GMM estimator of $\ddot{\boldsymbol{\theta}}$. We define the sensitivity matrix $\mathbf{R} = \left(\frac{\sum_m \ddot{\mathbf{S}}_m'\dot{\mathbf{Z}}_m}{M}\mathbf{W}_M\frac{\sum_m \dot{\mathbf{Z}}_m'\ddot{\mathbf{S}}_m}{M}\right)^{-1}\frac{\sum_m \ddot{\mathbf{S}}_m'\dot{\mathbf{Z}}_m}{M}\mathbf{W}_M$.*

[Result] Under classical assumptions (see proof), the asymptotic bias of $\hat{\boldsymbol{\theta}}$ is given by $\alpha_0 \operatorname{plim}\frac{\mathbf{R}\sum_m \dot{\mathbf{Z}}_m'(\mathbf{G}_m - \ddot{\mathbf{G}}_m)\mathbf{y}_m}{M}$. Moreover, letting $\mathbf{W}_M$ be an identity matrix minimizes the asymptotic bias in the sense of minimizing the Frobenius norm of $\mathbf{R}$.

Although there are no obvious ways to obtain a consistent estimate of the asymptotic bias (because $\mathbf{y}_m$ is a function of $\mathbf{G}_m$ and α_0), simulations show that the bias is very small in practice.

The intuition behind Proposition 2 comes from the literature on error-in-variable models with repeated observations (e.g., Bound et al. (2001)). The instrumental variable uses two independent draws from the (estimated) distribution of the true network. One draw is used to proxy the unobserved variable (i.e., $\mathbf{G}_m\mathbf{y}_m$), whereas the other is used to proxy the instrument (i.e., $\mathbf{G}_m\mathbf{X}_m$). This approach greatly reduces the bias compared with a situation in which only one draw would be used.

The argument in Proposition 2 is very similar to the one in Andrews et al. (2017), although here perturbation with respect to the true model is not *local*.[37] We also show that we expect the identity matrix weight to minimize the asymptotic bias. Our result therefore provides a theoretical justification for the simulations in Onishi and Otsu (2021) who show that using the identity matrix to weight the moments greatly reduces the bias in the context studied by Andrews et al. (2017).

C.1.1 Proof of Proposition 2

Part 1: Asymptotic bias

Let $\ddot{\boldsymbol{\theta}}_0$ be the true value the parameter when regressors are defined as $\ddot{\mathbf{S}}_m = [\mathbf{1}_m, \mathbf{X}_m, \mathbf{G}_m\mathbf{X}_m, \ddot{\mathbf{G}}_m\mathbf{X}_m, \ddot{\mathbf{G}}_m\mathbf{y}_m]$; that is, the true coefficient vector associated with $[\mathbf{1}_m, \mathbf{X}_m\mathbf{G}_m\mathbf{X}_m, \ddot{\mathbf{G}}_m\mathbf{y}_m]$ is $\boldsymbol{\theta}_0$ while the true coefficient vector associated with $\ddot{\mathbf{G}}_m\mathbf{X}_m$ is zero.

We now use matrix notation as the sample level to avoid summations over m and the index m. For example, $\dot{\mathbf{Z}}'\ddot{\mathbf{S}} = \sum_m \dot{\mathbf{Z}}_m'\ddot{\mathbf{S}}_m$. The Linear GMM estimator can be written as

$$\check{\boldsymbol{\theta}} = \left(\frac{\ddot{\mathbf{S}}'\dot{\mathbf{Z}}}{M}\mathbf{W}_M\frac{\dot{\mathbf{Z}}'\ddot{\mathbf{S}}}{M}\right)^{-1}\frac{\ddot{\mathbf{S}}'\dot{\mathbf{Z}}}{M}\mathbf{W}_M\dot{\mathbf{Z}}'\left(\frac{\ddot{\mathbf{S}}\ddot{\boldsymbol{\theta}}_0 + \boldsymbol{\eta} + \boldsymbol{\varepsilon}}{M}\right)$$

where $\boldsymbol{\eta} = \alpha_0(\mathbf{G} - \ddot{\mathbf{G}})\mathbf{y}$ is due to the approximation of $\mathbf{G}\mathbf{y}$ by $\ddot{\mathbf{G}}\mathbf{y}$ in $\ddot{\mathbf{S}}$. Therefore $\check{\boldsymbol{\theta}} = \ddot{\boldsymbol{\theta}}_0 + \mathbf{R}\left(\dfrac{\dot{\mathbf{Z}}'\boldsymbol{\eta} + \dot{\mathbf{Z}}'\boldsymbol{\varepsilon}}{M}\right)$ and the asymptotic bias of $\check{\boldsymbol{\theta}}$ is $\text{plim}(\check{\boldsymbol{\theta}} - \ddot{\boldsymbol{\theta}}_0) = \alpha_0 \,\text{plim}\, \dfrac{\mathbf{R}\dot{\mathbf{Z}}'(\mathbf{G} - \ddot{\mathbf{G}})\mathbf{y}}{M}$.

Part 2: Choice of W (we omit the index M for simplicity)

Let $\boldsymbol{\Delta} = \mathbf{G} - \ddot{\mathbf{G}}$, $\mathbf{K} = \dot{\mathbf{Z}}'\boldsymbol{\Delta}\mathbf{G}^2/M$, if $\boldsymbol{\gamma} = \mathbf{0}$, and $\mathbf{K} = \dot{\mathbf{Z}}'\boldsymbol{\Delta}/M$ otherwise. Consider $\|\mathbf{RK}\|_F = \sqrt{trace(\mathbf{K}'\mathbf{R}'\mathbf{R}\mathbf{K})} = \sqrt{trace(\mathbf{K}\mathbf{K}'\mathbf{R}'\mathbf{R})}$. We have

$$(1/M^2)\mathbf{R}\mathbf{R}' = [\ddot{\mathbf{S}}'\dot{\mathbf{Z}}\mathbf{W}\dot{\mathbf{Z}}'\ddot{\mathbf{S}}]^{-1}\ddot{\mathbf{S}}'\dot{\mathbf{Z}}\mathbf{W}\mathbf{W}\dot{\mathbf{Z}}'\ddot{\mathbf{S}}[\ddot{\mathbf{S}}'\dot{\mathbf{Z}}\mathbf{W}\dot{\mathbf{Z}}'\ddot{\mathbf{S}}]^{-1}.$$

[37] See page 1562 in Andrews et al. (2017).

Let $\mathbf{W} = \mathbf{C}'\mathbf{C}$ and let $\mathbf{B} = \ddot{\mathbf{S}}'\dot{\mathbf{Z}}\mathbf{C}'$. We have

$$(1/M^2)\mathbf{R}\mathbf{R}' = (\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}'\mathbf{C}\mathbf{C}'\mathbf{B}(\mathbf{B}'\mathbf{B})^{-1}.$$

Now, define $\mathbf{J}' = (\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}'\mathbf{C} - (\mathbf{B}'(\mathbf{C}')^{-1}\mathbf{C}^{-1}\mathbf{B})^{-1}\mathbf{B}'(\mathbf{C}')^{-1}$. We have

$$(1/M^2)\mathbf{R}\mathbf{R}' = \mathbf{J}'\mathbf{J} + (\mathbf{B}'(\mathbf{C}')^{-1}\mathbf{C}^{-1}\mathbf{B})^{-1} = \mathbf{J}'\mathbf{J} + (\ddot{\mathbf{S}}'\dot{\mathbf{Z}}\dot{\mathbf{Z}}'\ddot{\mathbf{S}})^{-1}.$$

Therefore, we have

$$(1/M^2)\|\mathbf{R}\|_F = \sqrt{trace(\mathbf{J}'\mathbf{J} + (\ddot{\mathbf{S}}'\dot{\mathbf{Z}}\dot{\mathbf{Z}}'\ddot{\mathbf{S}})^{-1})} = \sqrt{trace(\mathbf{J}'\mathbf{J}) + trace((\ddot{\mathbf{S}}'\dot{\mathbf{Z}}\dot{\mathbf{Z}}'\ddot{\mathbf{S}})^{-1})}.$$

When $\mathbf{W} = \mathbf{I}$, we have that $\mathbf{J} = \mathbf{0}$ and the Frobenius norm of $\mathbf{R}$ is given by $M^2\sqrt{trace((\ddot{\mathbf{S}}'\dot{\mathbf{Z}}\dot{\mathbf{Z}}'\ddot{\mathbf{S}})^{-1})}$.

C.2 Corollaries to Theorem 1

Corollary 1. *Assume that $\mathbf{G}_m\mathbf{X}_m$ is observed but that $\mathbf{G}_m\mathbf{y}_m$ is not observed. Let $\dot{\mathbf{Z}}_m^{(r)} = [\mathbf{1}_m, \mathbf{X}_m, \mathbf{G}_m\mathbf{X}_m, \dot{\mathbf{G}}_m^{(r)}\mathbf{G}\mathbf{X}, (\dot{\mathbf{G}}_m^{(r)})^2\mathbf{G}_m\mathbf{X}_m, ...]$, $\ddot{\mathbf{Z}}_m^{(r,s)} = [\mathbf{1}_m, \mathbf{X}_m, \ddot{\mathbf{G}}_m^{(s)}\mathbf{X}_m, \dot{\mathbf{G}}_m^{(r)}\ddot{\mathbf{G}}_m^{(s)}\mathbf{X}_m, (\dot{\mathbf{G}}_m^{(r)})^2\ddot{\mathbf{G}}_m^{(s)}\mathbf{X}_m, ...]$, $\mathbf{V}_m = [\mathbf{1}_m, \mathbf{X}_m, \mathbf{G}_m\mathbf{X}_m]$, and $\ddot{\mathbf{V}}_m^{(s)} = [\mathbf{1}_m, \mathbf{X}_m, \ddot{\mathbf{G}}_m^{(s)}\mathbf{X}_m]$. Then, the results from Theorem 1 hold for the following (simulated) moment function:*

$$\bar{\mathbf{m}}_M(\boldsymbol{\theta}) = \frac{1}{M}\sum_m \frac{1}{R}\sum_r \dot{\mathbf{Z}}_m^{(r)\prime}(\mathbf{I}_m - \alpha\dot{\mathbf{G}}_m^{(r)})\mathbf{y}_m - \frac{1}{RS}\sum_{r,s}(\dot{\mathbf{Z}}_m^{(r)\prime}\mathbf{V}_m - \ddot{\mathbf{Z}}_m^{(r,s)\prime}\ddot{\mathbf{V}}_m^{(s)})\tilde{\boldsymbol{\theta}}$$
$$-\frac{1}{RS}\sum_{r,s}\ddot{\mathbf{Z}}_m^{(r,s)\prime}(\mathbf{I}_m - \alpha\dot{\mathbf{G}}_m^{(r)})(\mathbf{I}_m - \alpha\ddot{\mathbf{G}}_m^{(s)})^{-1}\ddot{\mathbf{V}}_m^{(s)}\tilde{\boldsymbol{\theta}} \qquad (14)$$

under the same conditions.

Corollary 2. *Assume that $\mathbf{G}_m\mathbf{y}_m$ is observed but that $\mathbf{G}_m\mathbf{X}_m$ is not observed. Let $\dot{\mathbf{Z}}_m^{(r)} = [\mathbf{1}_m, \mathbf{X}_m, \dot{\mathbf{G}}_m^{(r)}\mathbf{X}_m, (\dot{\mathbf{G}}_m^{(r)})^2\mathbf{X}_m, ...]$, and $\ddot{\mathbf{V}}_m^{(s)} = [\mathbf{1}_m, \mathbf{X}_m, \ddot{\mathbf{G}}_m^{(s)}\mathbf{X}_m]$. Then, the results from The-*

orem 1 hold for the following (simulated) moment function:

$$\frac{1}{R}\sum_r \dot{\mathbf{Z}}_m^{(r)\prime}(\mathbf{I}_m - \alpha\mathbf{G}_m)\mathbf{y}_m - \frac{1}{RS}\sum_{r,s}\dot{\mathbf{Z}}_m^{(r)\prime}\ddot{\mathbf{V}}_m^{(s)}\tilde{\boldsymbol{\theta}} \tag{15}$$

under the same conditions.

D Specific network formation models

D.1 Graham (2017)

Graham (2017) presents a network formation process for *undirected* networks. As mentioned, all of our results hold for undirected networks with the appropriate notation changes. The network formation process in Graham (2017) is as follows:

$$P(a_{ij,m} = 1|\mathbf{X}_m) = \frac{\exp\{\mathbf{w}_{ij,m}\boldsymbol{\rho} + \nu_{i.m} + \nu_{j,m}\}}{1 + \exp\{\mathbf{w}_{ij,m}\boldsymbol{\rho} + \nu_{i,m} + \nu_{j,m}\}}$$

for all pairs $ij : i < j$.

Graham (2017) assumes (his Assumption 3) that the researchers observes a random sample of pairs $(a_{ij,m}, \mathbf{w}_{ij,m})$. While the individual level fixed effect ν_i cannot be consistently estimated when the group sizes are bounded (see our Assumption 1), Graham (2017) shows that the degree distribution (i.e., the number of links individuals have) is a sufficient statistic for $\boldsymbol{\nu}$, the vector of $\nu_{i,m}$'s. He presents an estimator (the "Tetrad Logit") that allows for the consistent estimation of $\boldsymbol{\rho}$ conditional on the degree sequence. Specifically, he shows that:

$$P(\mathbf{A}_m|\mathbf{X}_m, \bar{\mathbf{a}}) = \frac{\exp\{\sum_{ij:i<j} a_{ij}\mathbf{w}_{ij}\boldsymbol{\rho}\}}{\sum_{\mathbf{B}:\bar{\mathbf{b}}=\bar{\mathbf{a}}}\exp\{\sum_{ij:i<j} b_{ij}\mathbf{w}_{ij}\boldsymbol{\rho}\}}, \tag{16}$$

where $\bar{\mathbf{a}}$ and $\bar{\mathbf{b}}$ are the degree sequences of the adjacency matrices $\mathbf{A}$ and $\mathbf{B}$. Note that the denominator in Equation (16) sums only over the network structures that have the same degree sequence as $\mathbf{A}$. Note also that Equation (16) does not depend on $\boldsymbol{\nu}$.

Thus, the estimation of (16) is possible if one observes (i) a random sample of pairs, and (ii) the number of links for each individual belonging to a sampled pair (see Theorem 1 in Graham (2017)).[38]

D.2 Boucher and Mourifié (2017)

Boucher and Mourifié (2017) present a network formation model for *undirected* networks. Their network formation model, which is a special case of an exponential random graph model (ERGM) is as follows:

$$P(a_{ij,m}=1|\mathbf{X}_m)=\frac{\exp\{\mathbf{w}_{ij,m}\tilde{\boldsymbol{\rho}}+(n_{i,m}+n_{j,m})\rho_1+\psi(d(\tilde{\mathbf{x}}_{i,m},\tilde{\mathbf{x}}_{j,m}))\rho_2\}}{1+\exp\{\mathbf{w}_{ij,m}\tilde{\boldsymbol{\rho}}+(n_{i,m}+n_{j,m})\rho_1+\psi(d(\tilde{\mathbf{x}}_{i,m},\tilde{\mathbf{x}}_{j,m}))\rho_2\}}$$

for all pairs $ij : i < j$, where $n_{i,m}$ and $n_{j,m}$ represent the number of links that i and j have. Importantly, $\tilde{\mathbf{x}}_{i,m}$ and $\tilde{\mathbf{x}}_{j,m}$ are non-stochastic "positions" of i and j on an underlying Euclidean space (e.g., geographical distance), and $d(\tilde{\mathbf{x}}_{i,m},\tilde{\mathbf{x}}_{j,m})$ is a distance (and ψ is some increasing function, see below). Under the restriction that $d(\tilde{\mathbf{x}}_{i,m},\tilde{\mathbf{x}}_{j,m})\geq d_0>0$ for all i and j and that $\rho_2<\underline{\rho}<0$, Boucher and Mourifié (2017) show that $\boldsymbol{\rho}=[\tilde{\boldsymbol{\rho}},\rho_1,\rho_2]$ is consistently estimated by a simple pseudo-logistic regression.

While they do not consider bounded groups, their setup is compatible with our framework. To do so, however, we need to adapt our framework and ensure that individuals (and groups) are also drawn on some non-stochastic Euclidean space (e.g., geographical location). Then, their estimator is valid if we assume that (1) groups are drawn in a way that the distance (on the non-stochastic space) between each pair of individuals within the group is bounded

[38] The theorem also requires weak conditions on the asymptotic degree sequence, see Assumption 4 in Graham (2017).

below, that (2) any pairs of individuals from two different groups are located at a distance (again, on the non-stochastic space) greater than $\bar{d} > d_0$, and that (3) $\psi(d) = d$ when $d \leq \bar{d}$, while $\psi(d) = \infty$ if $d > \bar{d}$. Essentially, this ensures that no link can be created between individuals of different groups.

Then, as for Graham (2017), the estimation of the network formation process in Boucher and Mourifié (2017) can be done if the researcher observes (i) a random sample of pairs, and (ii) the number of links for each individual belonging to a sampled pair.

D.3 Other ERGM

Exponential Random Graph models (ERGM) are such that:[39]

$$P(\mathbf{A}_m|\mathbf{X}_m) = \frac{\exp\{\mathbf{q}(\mathbf{A}_m, \mathbf{X}_m)\boldsymbol{\rho}\}}{\sum_{\mathbf{B}_m} \exp\{\mathbf{q}(\mathbf{B}_m, \mathbf{X}_m)\boldsymbol{\rho}\}},$$

where $\mathbf{q}$ is a known function and the sum in the denominator is over all the possible network structures $\mathbf{B}_m$ for the group m. Microfoundations for ERGM can be found in Mele (2017) and Hsieh et al. (2020). Since ERGM are from the exponential family, $\mathbf{q}(\mathbf{A}_m, \mathbf{X}_m)$ are the sufficient statistics for $\boldsymbol{\rho}$. This means that consistent estimation of $\boldsymbol{\rho}$ requires consistent estimation of these sufficient statistics. This in turns implies that the sampling process that generates $\mathcal{A}_m$ must allow for this. We give two simple examples below.

D.3.1 Reciprocal links

This simplest possible ERGM is such that:

$$P(\mathbf{A}_m|\mathbf{X}_m) \propto \exp\{\sum_{ij}(a_{ij}w_{ij}\tilde{\boldsymbol{\rho}} + \rho_1 a_{ij}a_{ji})\},$$

[39]We focus on cases for which the term inside the exponential is linear in $\boldsymbol{\rho}$ for simplicity.

where $\boldsymbol{\rho} = [\rho_1,\ \tilde{\boldsymbol{\rho}}]$. When $\rho_1 = 0$, the model reduces to the baseline model in Equation (2). Here $\rho_1 > 0$ implies that reciprocal links (when i is linked to j and j is linked to i) are more likely than what would be expected from a model with conditionally independent links.

Estimation of such a model requires sampling *pairs of individuals*, and observing their linking status without error, in order to recover the sufficient statistics: the fraction of links (given a set of observed pair characteristics $\mathbf{w}_{ij}$), and the fraction of reciprocal links. For this simple example, the estimation via maximum likelihood is straightforward and our results follow.

D.3.2 Transitive triads

A typical feature of the data that is hard to replicate using the model in Equation (2) is the fraction of transitive triads. That is, if i is linked to j and j is linked to k, then the probability that i and k are linked is higher than what would be predicted by a model with conditionally independent links. Consider the following ERGM:

$$P(\mathbf{A}_m|\mathbf{X}_m) \propto \exp\{\sum_{ij}(a_{ij}w_{ij}\tilde{\boldsymbol{\rho}} + \rho_1 a_{ij}a_{ji} + \rho_2 \sum_k a_{ij}a_{jk}a_{ki})\},$$

which now includes the number of directed triangles (cycles of length 3). If $\rho_2 > 0$, then network configurations in which i is linked to j, j is linked to k, and k is linked to i are more likely (everything else equal).

Here, the sufficient statistics are: the fraction of links (given a set of observed pair characteristics $\mathbf{w}_{ij}$), the fraction of reciprocal links, and the fraction of closed directed triangles. This already requires sampling triads of individuals, which substantially complicates the sampling design. We are not aware of any such application. We also note that even with the consistent estimation of sufficient statistics, the estimation of $\boldsymbol{\rho}$ is not straightforward and computationally intensive. These considerations are left for future research.

E Full Simulation Results

Tables E.1–E.4 report simulation results. We also report simulations for cases where $\mathbf{G}_m\mathbf{y}_m$ and/or $\mathbf{G}_m\mathbf{X}_m$ is observed, including cases that account for unobserved group heterogeneity. The estimator still performs well in these settings. Precision improves significantly when $\mathbf{G}_m\mathbf{y}_m$ is observed, more so than when $\mathbf{G}_m\mathbf{X}_m$ is observed. This occurs because $\mathbf{G}_m\mathbf{y}_m$ is a nonlinear function of the true network $\mathbf{G}_m$ and peer effect coefficient α_0. Therefore, its approximation is more challenging than that of $\mathbf{G}_m\mathbf{X}_m$, which is exogenous.

Table E.1: Full simulation results under missing links without group fixed effects

Proportion of missing links	0%		25%		50%		75%	
Statistic	Mean	Std	Mean	Std	Mean	Std	Mean	Std
	Classical IV: $\mathbf{Gy}$ observed and $\mathbf{GX}$ unobserved; Instruments: $\mathbf{GX}^2$							
$\alpha = 0.538$	0.537	0.008	0.531	0.014	0.525	0.023	0.518	0.055
$c = 3.806$	3.807	0.132	4.378	0.168	4.885	0.267	5.336	0.638
$\beta_1 = -0.072$	−0.072	0.009	−0.089	0.011	−0.106	0.015	−0.124	0.030
$\beta_2 = 0.133$	0.133	0.027	0.136	0.030	0.141	0.030	0.143	0.033
$\gamma_1 = 0.086$	0.086	0.005	0.063	0.005	0.046	0.006	0.033	0.010
$\gamma_2 = -0.003$	−0.003	0.037	−0.009	0.036	−0.013	0.040	−0.013	0.052
	Classical IV: $\mathbf{Gy}$ and $\mathbf{GX}$ unobserved; Instruments: $\mathbf{GX}^2$							
$\alpha = 0.538$	0.537	0.008	0.442	0.014	0.362	0.021	0.293	0.043
$c = 3.806$	3.807	0.132	6.598	0.325	8.931	0.412	10.789	0.474
$\beta_1 = -0.072$	−0.072	0.009	−0.176	0.021	−0.273	0.028	−0.358	0.033
$\beta_2 = 0.133$	0.133	0.027	0.151	0.058	0.168	0.072	0.186	0.084
$\gamma_1 = 0.086$	0.086	0.005	0.030	0.008	−0.006	0.011	−0.027	0.022
$\gamma_2 = -0.003$	−0.003	0.037	−0.028	0.045	−0.046	0.054	−0.050	0.078
	SGMM: $\mathbf{Gy}$ and $\mathbf{GX}$ observed; $T = 100$							
$\alpha = 0.538$	0.537	0.008	0.537	0.012	0.538	0.015	0.539	0.021
$c = 3.806$	3.807	0.132	3.811	0.133	3.801	0.136	3.805	0.150
$\beta_1 = -0.072$	−0.072	0.009	−0.073	0.009	−0.072	0.009	−0.072	0.010
$\beta_2 = 0.133$	0.133	0.027	0.132	0.028	0.134	0.026	0.132	0.026
$\gamma_1 = 0.086$	0.086	0.005	0.086	0.006	0.086	0.007	0.086	0.010
$\gamma_2 = -0.003$	−0.003	0.037	−0.003	0.036	−0.003	0.037	−0.003	0.038
	SGMM: $\mathbf{Gy}$ observed and $\mathbf{GX}$ unobserved; $S = T = 100$							
$\alpha = 0.538$	0.537	0.008	0.538	0.012	0.538	0.019	0.540	0.033
$c = 3.806$	3.807	0.132	3.812	0.150	3.806	0.178	3.802	0.225
$\beta_1 = -0.072$	−0.072	0.009	−0.073	0.010	−0.072	0.011	−0.072	0.014
$\beta_2 = 0.133$	0.133	0.027	0.132	0.030	0.134	0.030	0.132	0.032
$\gamma_1 = 0.086$	0.086	0.005	0.086	0.006	0.086	0.010	0.085	0.016
$\gamma_2 = -0.003$	−0.003	0.037	−0.003	0.042	−0.003	0.056	−0.004	0.081
	SGMM: $\mathbf{Gy}$ unobserved and $\mathbf{GX}$ observed; $S = T = 100$							
$\alpha = 0.538$	0.537	0.008	0.538	0.015	0.538	0.027	0.540	0.064
$c = 3.806$	3.807	0.132	3.817	0.263	3.819	0.350	3.803	0.492
$\beta_1 = -0.072$	−0.072	0.009	−0.073	0.016	−0.073	0.021	−0.072	0.030
$\beta_2 = 0.133$	0.133	0.027	0.132	0.047	0.133	0.057	0.135	0.066
$\gamma_1 = 0.086$	0.086	0.005	0.086	0.009	0.086	0.014	0.085	0.031
$\gamma_2 = -0.003$	−0.003	0.037	−0.002	0.050	−0.005	0.076	−0.004	0.136
	SGMM: $\mathbf{Gy}$ and $\mathbf{GX}$ unobserved; $R = S = T = 100$							
$\alpha = 0.538$	0.537	0.008	0.538	0.016	0.539	0.029	0.542	0.073
$c = 3.806$	3.807	0.132	3.816	0.314	3.821	0.423	3.794	0.580
$\beta_1 = -0.072$	−0.072	0.009	−0.073	0.019	−0.073	0.026	−0.071	0.036
$\beta_2 = 0.133$	0.133	0.027	0.132	0.055	0.133	0.069	0.136	0.079
$\gamma_1 = 0.086$	0.086	0.005	0.086	0.009	0.086	0.016	0.084	0.035
$\gamma_2 = -0.003$	−0.003	0.037	−0.002	0.052	−0.005	0.083	−0.005	0.153

Note: We perform 1,000 simulations. 'Std' denotes the standard deviation.

Table E.2: Full simulation results under missing links with group fixed effects

Proportion of missing links	0%		25%		50%		75%	
Statistic	Mean	Std	Mean	Std	Mean	Std	Mean	Std
	Classical IV: $\mathbf{Gy}$ observed and $\mathbf{GX}$ unobserved; Instruments: $\mathbf{GX}^2$							
$\alpha = 0.538$	0.538	0.009	0.531	0.015	0.523	0.027	0.522	0.067
$\beta_1 = -0.072$	-0.072	0.009	-0.090	0.011	-0.107	0.016	-0.123	0.035
$\beta_2 = 0.133$	0.133	0.028	0.136	0.030	0.139	0.032	0.143	0.033
$\gamma_1 = 0.086$	0.086	0.005	0.063	0.005	0.046	0.006	0.032	0.011
$\gamma_2 = -0.003$	-0.001	0.038	-0.009	0.039	-0.012	0.042	-0.012	0.052
	Classical IV: $\mathbf{Gy}$ and $\mathbf{GX}$ unobserved; Instruments: $\mathbf{GX}^2$							
$\alpha = 0.538$	0.538	0.009	0.432	0.016	0.343	0.025	0.272	0.049
$\beta_1 = -0.072$	-0.072	0.009	-0.178	0.022	-0.273	0.029	-0.353	0.033
$\beta_2 = 0.133$	0.133	0.028	0.149	0.058	0.165	0.074	0.182	0.081
$\gamma_1 = 0.086$	0.086	0.005	0.033	0.009	0.000	0.013	-0.019	0.024
$\gamma_2 = -0.003$	-0.001	0.038	-0.028	0.051	-0.044	0.061	-0.045	0.081
	SGMM: $\mathbf{Gy}$ and $\mathbf{GX}$ observed; $T = 100$							
$\alpha = 0.538$	0.538	0.009	0.537	0.013	0.538	0.017	0.536	0.027
$\beta_1 = -0.072$	-0.072	0.009	-0.073	0.009	-0.072	0.010	-0.073	0.011
$\beta_2 = 0.133$	0.133	0.028	0.133	0.027	0.133	0.027	0.133	0.027
$\gamma_1 = 0.086$	0.086	0.005	0.086	0.006	0.086	0.008	0.087	0.012
$\gamma_2 = -0.003$	-0.001	0.038	-0.002	0.039	-0.004	0.041	-0.002	0.040
	SGMM: $\mathbf{Gy}$ observed and $\mathbf{GX}$ unobserved; $S = T = 100$							
$\alpha = 0.538$	0.538	0.009	0.537	0.013	0.538	0.022	0.539	0.042
$\beta_1 = -0.072$	-0.072	0.009	-0.073	0.010	-0.072	0.012	-0.072	0.016
$\beta_2 = 0.133$	0.133	0.028	0.133	0.030	0.133	0.032	0.133	0.032
$\gamma_1 = 0.086$	0.086	0.005	0.086	0.007	0.086	0.011	0.085	0.019
$\gamma_2 = -0.003$	-0.001	0.038	-0.003	0.046	-0.004	0.059	0.002	0.085
	SGMM: $\mathbf{Gy}$ unobserved and $\mathbf{GX}$ observed; $S = T = 100$							
$\alpha = 0.538$	0.538	0.009	0.537	0.018	0.538	0.031	0.542	0.076
$\beta_1 = -0.072$	-0.072	0.009	-0.074	0.017	-0.075	0.022	-0.073	0.031
$\beta_2 = 0.133$	0.133	0.028	0.133	0.047	0.133	0.059	0.136	0.066
$\gamma_1 = 0.086$	0.086	0.005	0.085	0.010	0.084	0.016	0.081	0.036
$\gamma_2 = -0.003$	-0.001	0.038	-0.003	0.057	-0.007	0.082	0.001	0.140
	SGMM: $\mathbf{Gy}$ and $\mathbf{GX}$ unobserved; $R = S = T = 100$							
$\alpha = 0.538$	0.538	0.009	0.537	0.018	0.538	0.033	0.539	0.084
$\beta_1 = -0.072$	-0.072	0.009	-0.074	0.020	-0.075	0.027	-0.074	0.037
$\beta_2 = 0.133$	0.133	0.028	0.133	0.055	0.133	0.071	0.137	0.078
$\gamma_1 = 0.086$	0.086	0.005	0.085	0.011	0.084	0.017	0.082	0.039
$\gamma_2 = -0.003$	-0.001	0.038	-0.003	0.060	-0.007	0.087	0.003	0.153

Note: We perform 1,000 simulations. 'Std' denotes the standard deviation.

Table E.3: Full simulation results under misclassified links without group fixed effects

False pos. rate	0%		0%		15%		15%	
False neg. rate	15%		30%		0%		15%	
Statistic	Mean	Std	Mean	Std	Mean	Std	Mean	Std
	Classical IV: $\mathbf{Gy}$ observed and $\mathbf{GX}$ unobserved; Instruments: $\mathbf{GX}^2$							
$\alpha = 0.538$	0.534	0.011	0.529	0.015	0.611	0.112	0.612	0.143
$c = 3.806$	4.154	0.151	4.489	0.190	3.904	1.534	3.981	2.186
$\beta_1 = -0.072$	−0.082	0.010	−0.093	0.012	−0.087	0.068	−0.083	0.097
$\beta_2 = 0.133$	0.134	0.028	0.138	0.029	0.136	0.033	0.136	0.036
$\gamma_1 = 0.086$	0.072	0.005	0.060	0.005	0.044	0.020	0.032	0.023
$\gamma_2 = -0.003$	−0.008	0.038	−0.011	0.038	−0.010	0.075	−0.009	0.080
	Classical IV: $\mathbf{Gy}$ and $\mathbf{GX}$ unobserved; Instruments: $\mathbf{GX}^2$							
$\alpha = 0.538$	0.479	0.012	0.424	0.015	0.366	0.168	0.275	0.171
$c = 3.806$	5.538	0.276	7.100	0.359	8.870	1.563	9.876	1.502
$\beta_1 = -0.072$	−0.135	0.018	−0.196	0.024	−0.421	0.034	−0.425	0.036
$\beta_2 = 0.133$	0.143	0.048	0.158	0.060	0.191	0.087	0.191	0.088
$\gamma_1 = 0.086$	0.049	0.007	0.022	0.009	0.064	0.049	0.036	0.050
$\gamma_2 = -0.003$	−0.022	0.043	−0.035	0.048	0.022	0.194	0.026	0.197
	SGMM: $\mathbf{Gy}$ and $\mathbf{GX}$ observed; $T = 100$							
$\alpha = 0.538$	0.537	0.011	0.538	0.013	0.538	0.020	0.537	0.023
$c = 3.806$	3.809	0.133	3.808	0.139	3.802	0.151	3.814	0.156
$\beta_1 = -0.072$	−0.073	0.009	−0.072	0.010	−0.072	0.010	−0.073	0.011
$\beta_2 = 0.133$	0.132	0.027	0.133	0.026	0.132	0.027	0.133	0.025
$\gamma_1 = 0.086$	0.086	0.006	0.086	0.006	0.086	0.009	0.086	0.010
$\gamma_2 = -0.003$	−0.003	0.037	−0.004	0.037	−0.003	0.038	−0.004	0.038
	SGMM: $\mathbf{Gy}$ observed and $\mathbf{GX}$ unobserved; $S = T = 100$							
$\alpha = 0.538$	0.538	0.011	0.537	0.014	0.539	0.037	0.540	0.047
$c = 3.806$	3.806	0.145	3.806	0.165	3.805	0.243	3.805	0.273
$\beta_1 = -0.072$	−0.072	0.010	−0.072	0.011	−0.072	0.015	−0.072	0.017
$\beta_2 = 0.133$	0.132	0.029	0.134	0.029	0.133	0.034	0.133	0.034
$\gamma_1 = 0.086$	0.086	0.006	0.087	0.007	0.085	0.018	0.086	0.023
$\gamma_2 = -0.003$	−0.004	0.042	−0.004	0.046	−0.006	0.086	−0.007	0.105
	SGMM: $\mathbf{Gy}$ unobserved and $\mathbf{GX}$ observed; $S = T = 100$							
$\alpha = 0.538$	0.538	0.012	0.537	0.018	0.539	0.077	0.536	0.103
$c = 3.806$	3.803	0.232	3.799	0.292	3.812	0.501	3.812	0.622
$\beta_1 = -0.072$	−0.072	0.015	−0.072	0.018	−0.072	0.032	−0.073	0.038
$\beta_2 = 0.133$	0.132	0.041	0.135	0.048	0.135	0.067	0.134	0.070
$\gamma_1 = 0.086$	0.086	0.007	0.087	0.010	0.085	0.039	0.088	0.050
$\gamma_2 = -0.003$	−0.005	0.047	−0.004	0.055	−0.005	0.170	−0.010	0.209
	SGMM: $\mathbf{Gy}$ and $\mathbf{GX}$ unobserved; $R = S = T = 100$							
$\alpha = 0.538$	0.538	0.013	0.537	0.018	0.543	0.090	0.539	0.124
$c = 3.806$	3.800	0.273	3.794	0.350	3.802	0.596	3.792	0.753
$\beta_1 = -0.072$	−0.072	0.017	−0.072	0.022	−0.071	0.038	−0.072	0.047
$\beta_2 = 0.133$	0.132	0.048	0.135	0.058	0.135	0.081	0.134	0.085
$\gamma_1 = 0.086$	0.086	0.008	0.087	0.010	0.083	0.045	0.087	0.061
$\gamma_2 = -0.003$	−0.005	0.047	−0.004	0.057	−0.008	0.205	−0.014	0.251

Note: We perform 1,000 simulations. 'Std' denotes the standard deviation. 'False pos. rate' refers to the proportion of false positives among actual negatives, which include true negatives and false positives. 'False neg. rate' refers to the proportion of false negatives among actual positives, which include true positives and false negatives. A positive indicates a friendship, while a negative indicates a non-friendship.

Table E.4: Full simulation results under misclassified links with group fixed effects

False pos. rate	0%		0%		15%		15%	
False neg. rate	15%		30%		0%		15%	
Statistic	Mean	Std	Mean	Std	Mean	Std	Mean	Std
	Classical IV: $\mathbf{Gy}$ observed and $\mathbf{GX}$ unobserved; Instruments: $\mathbf{GX}^2$							
$\alpha = 0.538$	0.532	0.020	0.525	0.026	0.655	0.172	0.645	0.217
$\beta_1 = -0.072$	−0.082	0.010	−0.092	0.011	−0.093	0.054	−0.097	0.131
$\beta_2 = 0.133$	0.135	0.028	0.138	0.029	0.138	0.033	0.135	0.036
$\gamma_1 = 0.086$	0.072	0.004	0.059	0.005	0.042	0.018	0.033	0.023
$\gamma_2 = -0.003$	−0.007	0.039	−0.010	0.039	−0.009	0.083	−0.006	0.095
	Classical IV: $\mathbf{Gy}$ and $\mathbf{GX}$ unobserved; Instruments: $\mathbf{GX}^2$							
$\alpha = 0.538$	0.472	0.019	0.412	0.024	0.380	0.156	0.289	0.165
$\beta_1 = -0.072$	−0.111	0.013	−0.147	0.016	−0.280	0.023	−0.282	0.025
$\beta_2 = 0.133$	0.140	0.037	0.148	0.043	0.173	0.058	0.170	0.061
$\gamma_1 = 0.086$	0.061	0.006	0.041	0.006	0.064	0.030	0.045	0.033
$\gamma_2 = -0.003$	−0.015	0.042	−0.021	0.045	0.011	0.145	0.009	0.146
	SGMM: $\mathbf{Gy}$ and $\mathbf{GX}$ observed; $T = 100$							
$\alpha = 0.538$	0.538	0.019	0.537	0.023	0.535	0.042	0.539	0.051
$\beta_1 = -0.072$	−0.072	0.010	−0.072	0.010	−0.072	0.011	−0.072	0.011
$\beta_2 = 0.133$	0.133	0.026	0.133	0.027	0.133	0.026	0.131	0.027
$\gamma_1 = 0.086$	0.086	0.005	0.086	0.006	0.087	0.010	0.086	0.011
$\gamma_2 = -0.003$	−0.004	0.039	−0.003	0.039	−0.001	0.039	−0.005	0.040
	SGMM: $\mathbf{Gy}$ observed and $\mathbf{GX}$ unobserved; $S = T = 100$							
$\alpha = 0.538$	0.538	0.019	0.537	0.024	0.533	0.075	0.532	0.109
$\beta_1 = -0.072$	−0.072	0.010	−0.072	0.011	−0.073	0.016	−0.073	0.019
$\beta_2 = 0.133$	0.133	0.028	0.134	0.029	0.134	0.033	0.131	0.035
$\gamma_1 = 0.086$	0.086	0.005	0.086	0.006	0.087	0.018	0.088	0.026
$\gamma_2 = -0.003$	−0.004	0.044	−0.004	0.048	−0.002	0.100	0.001	0.115
	SGMM: $\mathbf{Gy}$ unobserved and $\mathbf{GX}$ observed; $S = T = 100$							
$\alpha = 0.538$	0.538	0.020	0.536	0.027	0.544	0.105	0.552	0.142
$\beta_1 = -0.072$	−0.072	0.011	−0.072	0.013	−0.072	0.022	−0.071	0.026
$\beta_2 = 0.133$	0.133	0.031	0.134	0.035	0.134	0.043	0.131	0.046
$\gamma_1 = 0.086$	0.086	0.006	0.086	0.007	0.084	0.025	0.083	0.034
$\gamma_2 = -0.003$	−0.005	0.046	−0.003	0.052	−0.002	0.132	−0.003	0.157
	SGMM: $\mathbf{Gy}$ and $\mathbf{GX}$ unobserved; $R = S = T = 100$							
$\alpha = 0.538$	0.538	0.021	0.536	0.028	0.540	0.126	0.537	0.178
$\beta_1 = -0.072$	−0.072	0.013	−0.072	0.015	−0.072	0.027	−0.073	0.033
$\beta_2 = 0.133$	0.133	0.036	0.134	0.042	0.135	0.055	0.131	0.059
$\gamma_1 = 0.086$	0.086	0.006	0.086	0.008	0.085	0.030	0.087	0.042
$\gamma_2 = -0.003$	−0.004	0.047	−0.003	0.054	−0.003	0.164	−0.002	0.198

Note: We perform 1,000 simulations. 'Std' denotes the standard deviation. 'False pos. rate' refers to the proportion of false positives among actual negatives, which include true negatives and false positives. 'False neg. rate' refers to the proportion of false negatives among actual positives, which include true positives and false negatives. A positive indicates a friendship, while a negative indicates a non-friendship.

F Bayesian estimator

F.1 Posterior Distributions for Algorithm 1.

To compute the posterior distributions, we set prior distributions on $\tilde{\alpha}$, $\mathbf{\Lambda}$, and σ^2, where $\tilde{\alpha} = \log(\frac{\alpha}{1-\alpha})$ and $\mathbf{\Lambda} = [\boldsymbol{\beta}, \boldsymbol{\gamma}]$. In Algorithm 1, we therefore sample $\tilde{\alpha}$ and compute α, such that $\alpha = \frac{\exp(\tilde{\alpha})}{1+\exp(\tilde{\alpha})}$. Using this functional form for computing α ensures that $\alpha \in (0, 1)$. The prior distributions are set as follows:

$$\tilde{\alpha} \sim \mathcal{N}(\mu_{\tilde{\alpha}}, \sigma^2_{\tilde{\alpha}}),$$

$$\mathbf{\Lambda}|\sigma^2 \sim \mathcal{N}(\boldsymbol{\mu}_{\mathbf{\Lambda}}, \sigma^2 \mathbf{\Sigma}_{\mathbf{\Lambda}}),$$

$$\sigma^2 \sim IG(\frac{a}{2}, \frac{b}{2}).$$

For the simulations and estimations in this paper, we set $\mu_{\tilde{\alpha}} = -1$, $\sigma^{-2}_{\tilde{\alpha}} = 2$, $\boldsymbol{\mu}_{\mathbf{\Lambda}} = \mathbf{0}$, ${\mathbf{\Sigma}_{\mathbf{\Lambda}}}^{-1} = \frac{1}{100}\boldsymbol{I}_K$, $a = 4$, and $b = 4$, where $\boldsymbol{I}_K$ is the identity matrix of dimension K and $K = \dim(\mathbf{\Lambda})$.

Following Algorithm 1, α is updated at each iteration t of the MCMC by drawing $\tilde{\alpha}^*$ from the proposal $\mathcal{N}(\tilde{\alpha}_{t-1}, \xi_t)$, where the jumping scale ξ_t is also updated at each t following Atchadé and Rosenthal (2005) for an acceptance rate of a^* targeted at 0.44. As the proposal is symmetrical, $\alpha^* = \frac{\exp(\tilde{\alpha}^*)}{1+\exp(\tilde{\alpha}^*)}$ is accepted with the probability

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

The parameters $\mathbf{\Lambda}_t = [\boldsymbol{\beta}_t, \boldsymbol{\gamma}_t]$ and σ_t^2 are drawn from their posterior conditional distributions, given as follows:

$$\begin{aligned} \mathbf{\Lambda}_t | \boldsymbol{y}, \boldsymbol{A}_t, \alpha_t, \sigma_{t-1}^2 &\sim \mathcal{N}(\hat{\boldsymbol{\mu}}_{\mathbf{\Lambda}_t}, \sigma_{t-1}^2 \hat{\mathbf{\Sigma}}_{\mathbf{\Lambda}_t}), \\ \sigma_t^2 | \boldsymbol{y}, \boldsymbol{A}_t, \boldsymbol{\theta}_t &\sim IG\left(\frac{\hat{a}_t}{2}, \frac{\hat{b}_t}{2}\right), \end{aligned}$$

where,

$$\begin{aligned} \hat{\mathbf{\Sigma}}_{\mathbf{\Lambda}_t}^{-1} &= \mathbf{V}_t'\mathbf{V}_t + \mathbf{\Sigma}_{\mathbf{\Lambda}}^{-1}, \\ \hat{\boldsymbol{\mu}}_{\mathbf{\Lambda}_t} &= \hat{\mathbf{\Sigma}}_{\mathbf{\Lambda}_t}\left(\mathbf{V}_t'(\boldsymbol{y} - \alpha_t \boldsymbol{G}_t \boldsymbol{y}) + \mathbf{\Sigma}_{\mathbf{\Lambda}}^{-1}\boldsymbol{\mu}_{\mathbf{\Lambda}}\right), \\ \hat{a}_t &= a + N, \\ \hat{b}_t &= b + (\mathbf{\Lambda}_t - \boldsymbol{\mu}_{\mathbf{\Lambda}})'\mathbf{\Sigma}_{\mathbf{\Lambda}}^{-1}(\mathbf{\Lambda}_t - \boldsymbol{\mu}_{\mathbf{\Lambda}}) + (\boldsymbol{y} - \alpha_t \boldsymbol{G}_t \boldsymbol{y} - \mathbf{V}_t \mathbf{\Lambda}_t)'(\boldsymbol{y} - \alpha_t \boldsymbol{G}_t \boldsymbol{y} - \mathbf{V}_t \mathbf{\Lambda}_t), \\ \mathbf{V}_t &= [\mathbf{1},\ \boldsymbol{X},\ \boldsymbol{G}_t \boldsymbol{X}]. \end{aligned}$$

F.2 Network Sampling

This section explains how we sample the network in Algorithm 1 using Gibbs sampling. As discussed above, a natural solution is to update only one entry of the adjacency matrix at every step t of the MCMC. The entry (i, j) is updated according to its conditional posterior distribution. For each entry, however, we need to compute $\mathcal{P}(\mathbf{y}|0, \mathbf{A}_{-ij})$ and $\mathcal{P}(\mathbf{y}|1, \mathbf{A}_{-ij})$, which are the respective likelihoods of replacing a_{ij} by 0 or by 1. The likelihood computation requires the determinant of $(\mathbf{I} - \alpha\mathbf{G})$, which has a complexity $O(N^3)$ where N is the dimension of $\mathbf{G}$. This implies that we must compute $2N(N-1)$ times $\det(\mathbf{I} - \alpha\mathbf{G})$ to update the adjacency matrix at each step of the MCMC. As $\boldsymbol{G}$ is row-normalized, alternating any off-diagonal entry (i, j) in $\boldsymbol{A}$ between 0 and 1 perturbs all off-diagonal entries of the row i in

$(\mathbf{I} - \alpha\mathbf{G})$. We show that $\boldsymbol{A}_{ij}$ and $\det(\mathbf{I} - \alpha\mathbf{G})$ can be updated by computing a determinant of an auxiliary matrix that requires only updating two entries.

Assume that we want to update the entry (i, j). Let h be the function defined in $\mathbb{N}$ such that $\forall\ x \in \mathbb{N}^*$, $h(x) = x$, and $h(0) = 1$. Let $\mathbf{L}$ be an $\mathrm{N} \times \mathrm{N}$ diagonal matrix, where $\mathbf{L}_{ii} = h(n_i)$, and n_i stands for the degree of i, while $\mathbf{L}_{kk} = 1$ for all $k \neq i$, and $\mathbf{W}$ is the matrix $\boldsymbol{G}$ where the row i of $\mathbf{W}$ is replaced by the row i of $\boldsymbol{A}$. Then, as the determinant is linear in each row, we can obtain $\mathbf{I} - \alpha\mathbf{G}$ by dividing the row i of $\mathbf{L} - \alpha\mathbf{W}$ by $h(n_i)$. We get:

$$\det(\mathbf{I} - \alpha\mathbf{G}) = \frac{1}{h(n_i)}\det(\mathbf{L} - \alpha\mathbf{W}).$$

When a_{ij} changes (from 0 to 1, or 1 to 0), note that only the entries (i, i) and (i, j) change in $\mathbf{L} - \alpha\mathbf{W}$. Two cases can be distinguished.

- If $a_{ij} = 0$ before the update, then the new degree of i will be $n_i + 1$. Thus, the entry (i, i) in $\mathbf{L} - \alpha\mathbf{W}$ will change from $h(n_i)$ to $h(n_i + 1)$ (as the diagonal of $\mathbf{W}$ equals 0), and the entry (i, j) will change from 0 to $-\alpha$. The new determinant is therefore given by

  $$\det(\mathbf{I} - \alpha\mathbf{G}^*) = \frac{1}{h(n_i + 1)}\det(\mathbf{L}^* - \alpha\mathbf{W}^*),$$

  where $\mathbf{G}^*$, $\mathbf{L}^*$, and $\alpha\mathbf{W}^*$ are the new matrices once a_{ij} has been updated.

- If $a_{ij} = 1$ before the update, then the new degree of k will be $n_i - 1$. Thus, the entry (i, i) in $\mathbf{L} - \alpha\mathbf{W}$ will change from $h(n_i)$ to $h(n_i - 1)$, and the entry (i, j) will change from $-\alpha$ to 0. The new determinant is therefore given by

  $$\det(\mathbf{I} - \alpha\mathbf{G}^*) = \frac{1}{h(n_i - 1)}\det(\mathbf{L}^* - \alpha\mathbf{W}^*).$$

Then, to update $\det(\mathbf{L} - \alpha\mathbf{W})$ when only the entries (i, i) and (i, j) change, we adapt the

Lemma 1 in Hsieh et al. (2019) as follows:

Proposition 3. *Let* $\mathbf{e}_i$ *be the i'th unit basis vector in* $\mathbb{R}^N$. *Let* $\mathbf{M}$ *denote an* $N \times N$ *matrix and* $\mathbf{B}_{ij}(\mathbf{Q}, \epsilon)$ *an* $N \times N$ *matrix as a function of an* $N \times N$ *matrix* $\mathbf{Q}$ *and a real value* ϵ, *such that*

$$\mathbf{B}_{ij}(\mathbf{Q}, \epsilon) = \frac{\mathbf{Q}\mathbf{e}_i\mathbf{e}_j'\mathbf{Q}}{1 + \epsilon\mathbf{e}_j'\mathbf{Q}\mathbf{e}_i}. \tag{17}$$

Adding a perturbation ϵ_1 *in the (i, i)th position and a perturbation* ϵ_2 *in the (i, j)th position to the matrix* $\mathbf{M}$ *can be written as* $\tilde{\mathbf{M}} = \mathbf{M} + \epsilon_1\mathbf{e}_i\mathbf{e}_i' + \epsilon_2\mathbf{e}_i\mathbf{e}_j'$.

1. *The inverse of the perturbed matrix can be written as*

$$\tilde{\mathbf{M}}^{-1} = \mathbf{M}^{-1} - \epsilon_1\mathbf{B}_{ii}(\mathbf{M}^{-1}, \epsilon_1) - \epsilon_2\mathbf{B}_{ij}\left(\mathbf{M}^{-1} - \epsilon_1\mathbf{B}_{ii}(\mathbf{M}^{-1}, \epsilon_1), \epsilon_2\right).$$

2. *The determinant of the perturbed matrix can be written as*

$$det\left(\tilde{\mathbf{M}}\right) = \left(1 + \epsilon_2\mathbf{e}_j'\left(\mathbf{M}^{-1} - \epsilon_1\mathbf{B}_{ii}(\mathbf{M}^{-1}, \epsilon_1)\mathbf{e}_i\right)\right)(1 + \epsilon_1\mathbf{e}_i'\mathbf{M}^{-1}\mathbf{e}_i)\,det\,(\mathbf{M}).$$

Proof. 1. By the Sherman–Morrison formula (Mele, 2017), we have

$$\left(\mathbf{M} + \epsilon\mathbf{e}_i\mathbf{e}_j'\right)^{-1} = \mathbf{M}^{-1} - \epsilon\frac{\mathbf{M}^{-1}\mathbf{e}_i\mathbf{e}_j'\mathbf{M}^{-1}}{1 + \epsilon\mathbf{e}_j'\mathbf{M}^{-1}\mathbf{e}_i} = \mathbf{M}^{-1} - \epsilon\mathbf{B}_{ij}(\mathbf{M}, \epsilon).$$

Thus,

$$\tilde{\mathbf{M}}^{-1} = \left((\mathbf{M} + \epsilon_1\mathbf{e}_i\mathbf{e}_i') + \epsilon_2\mathbf{e}_i\mathbf{e}_j'\right)^{-1},$$

$$\tilde{\mathbf{M}}^{-1} = (\mathbf{M} + \epsilon_1\mathbf{e}_i\mathbf{e}_i')^{-1} - \epsilon_2\mathbf{B}_{ij}((\mathbf{M} + \epsilon_1\mathbf{e}_i\mathbf{e}_i')^{-1}, \epsilon_2),$$

$$\tilde{\mathbf{M}}^{-1} = \mathbf{M}^{-1} - \epsilon_1\mathbf{B}_{ii}(\mathbf{M}^{-1}, \epsilon_1) - \epsilon_2\mathbf{B}_{ij}\left(\mathbf{M}^{-1} - \epsilon_1\mathbf{B}_{ii}(\mathbf{M}^{-1}, \epsilon_1), \epsilon_2\right).$$

2. By the matrix determinant lemma (Johnson and Horn, 1985), we have

$$\det\left(\mathbf{M} + \epsilon\mathbf{e}_i\mathbf{e}_j'\right) = (1 + \epsilon\mathbf{e}_j'\mathbf{M}^{-1}\mathbf{e}_i)\det(\mathbf{M}).$$

It follows that

$$\begin{aligned}
\det\left(\tilde{\mathbf{M}}\right) &= \det\left((\mathbf{M}+\epsilon_1\mathbf{e}_i\mathbf{e}_i')+\epsilon_2\mathbf{e}_i\mathbf{e}_j'\right), \\
\det\left(\tilde{\mathbf{M}}\right) &= (1+\epsilon_2\mathbf{e}_j'(\mathbf{M}+\epsilon_1\mathbf{e}_i\mathbf{e}_i')^{-1}\mathbf{e}_i)\det\left(\mathbf{M}+\epsilon_1\mathbf{e}_i\mathbf{e}_i'\right), \\
\det\left(\tilde{\mathbf{M}}\right) &= \left(1+\epsilon_2\mathbf{e}_j'\left(\mathbf{M}^{-1}-\epsilon_1\mathbf{B}_{ii}(\mathbf{M}^{-1},\epsilon_1)\mathbf{e}_i\right)\right)(1+\epsilon_1\mathbf{e}_i'\mathbf{M}^{-1}\mathbf{e}_i)\det\left(\mathbf{M}\right).
\end{aligned}$$

□

The method proposed above becomes computationally intensive when many entries must be updated simultaneously. We also propose an alternative method that allows updating the block for entries in $\boldsymbol{A}$. Let $\mathbf{D}=(\mathbf{I}-\alpha\mathbf{G})$; we can write

$$\det(\mathbf{D})=\sum_{j=1}^{N}(-1)^{i+j}\mathbf{D}_{\mathrm{ij}}\delta_{\mathrm{ij}}, \tag{18}$$

where i denotes any row of $\mathbf{D}$ and δ_{ij} is the minor[40] associated with the entry (i,j). The minors of row i do not depend on the values of entries in row i. To update any block in row i, we therefore compute the N minors associated with i and use this minor within the row. We can then update many entries simultaneously without increasing the number of times that we compute $\det(\mathbf{D})$.

One possibility is to update multiple links simultaneously by randomly choosing the number of entries to consider and their position in the row. As suggested by Chib and Ramamurthy (2010), this method would help the Gibbs sampling to converge more quickly. We can summarize how we update the row i as follows:

1. Compute the N minors $\delta_{i1},\ldots,\delta_{in}$.

2. Let $\Omega_{\boldsymbol{G}}$ be the entries to update in the row i, and $n_{\boldsymbol{G}}=|\Omega_{\boldsymbol{G}}|$ the number of entries in

[40]The determinant of the submatrix of $\mathbf{M}$ by removing row i and column j.

$\Omega_{\boldsymbol{G}}$.

(a) Choose r, the size of the block to update, as a random integer number such that $1 \leq r \leq n_{\boldsymbol{G}}$. In practice, we choose $r \leq \min(5, n_{\boldsymbol{G}})$ because the number of possibilities of links to consider grows exponentially with r.

(b) Choose the r random entries from $\Omega_{\boldsymbol{G}}$. These entries define the block to update.

(c) Compute the posterior probabilities of all possibilities of links inside the block and update the block (there are 2^r possibilities). Use the minors calculated at 1 and the formula (18) to quickly compute $det(\mathbf{D})$.

(d) Remove the r drawn positions from $\Omega_{\boldsymbol{G}}$ and let $n_{\boldsymbol{G}} = n_{\boldsymbol{G}} - r$. Replicate 2a, 2b, and 2c until $n_{\boldsymbol{G}} = 0$.

F.3 How to build prior distributions

The two following examples discuss how to construct prior distributions depending on whether the first stage is estimated by a classical or Bayesian estimator.

Example 4 (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 5 (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 (5). Performing such a sequential Bayesian updating approach comes with a well-known numerical issue.*[41]

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 $O(N_m)$. 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.[42]

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

[42]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 accompanying our R package.

G Application

G.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 errors, 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.

G.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 males and $s = f$ for females. 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 $\ell(\mathcal{S}_{im}) + \ell(\mathcal{S}_{if})$ 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 $\boldsymbol{\rho}$.

G.3 Tables

Table G.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.

Table G.2: Empirical results (Bayesian method)

	Model 1		Model 2		Model 3	
Statistic	Mean	Std	Mean	Std	Mean	Std
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" 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 G.3: Empirical results (SGMM Method)

Statistic	Model 1 Mean	Model 1 Std	Model 2 Mean	Model 2 Std	Model 3 Mean	Model 3 Std
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" 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 G.4: Empirical results (Standard IV estimator based on the sample of students without missing network data)

Statistic	Mean	Std
Peer effects	0.408	(0.243)
Own effect		
Female	0.178	(0.075)
Hispanic	−0.068	(0.099)
Race (white)		
Black	−0.170	(0.084)
Asian	−0.239	(0.212)
Other	0.101	(0.131)
Mother's education (High)		
$<$high	−0.041	(0.104)
$>$high	0.230	(0.089)
missing	0.171	(0.122)
Mother's job (Stay-at-home)		
Professional	0.060	(0.121)
Other	−0.098	(0.092)
missing	−0.116	(0.116)
Age	−0.011	(0.028)
Contextual effects		
Female	−0.063	(0.153)
Hispanic	0.421	(0.246)
Race (white)		
Black	0.010	(0.169)
Asian	−0.242	(0.909)
Other	−0.204	(0.304)
Mother's education (High)		
$<$high	−0.292	(0.257)
$>$high	−0.076	(0.196)
missing	0.077	(0.295)
Mother's job (Stay-at-home)		
Professional	0.168	(0.247)
Other	−0.141	(0.191)
missing	−0.126	(0.283)
Age	−0.065	(0.05)

Note: These results are based on the standard IV estimation considering the sample of students without missing network data and who are not affected by the censoring issue ($N = 561$). Column "Mean" indicates the estimates, and Column "Std" indicates the posterior standard deviations in parentheses. 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.

G.4 Key Player

Figure G.1 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 G.2 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.

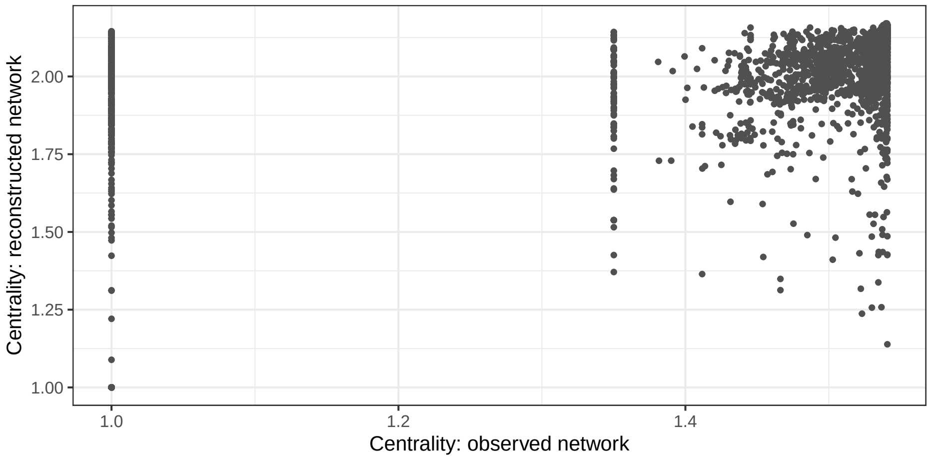


Figure G.1: Centrality

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.

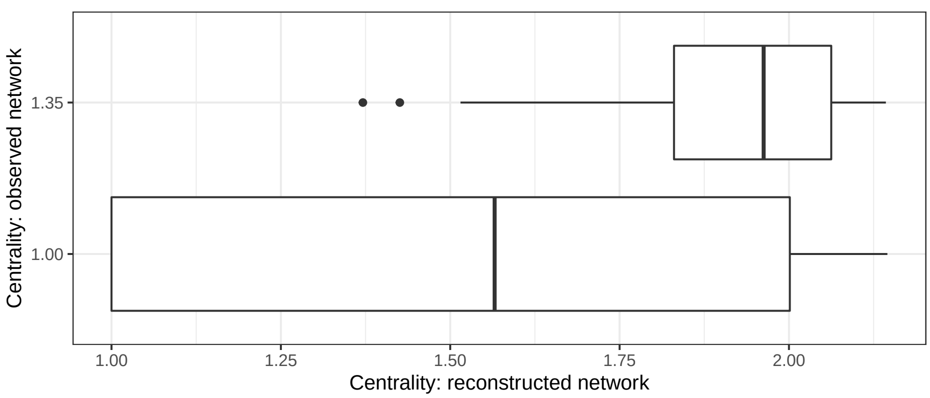


Figure G.2: Centrality

Note: See note of Figure G.1.

Figure G.3: MCMC Simulations – Peer Effect Model

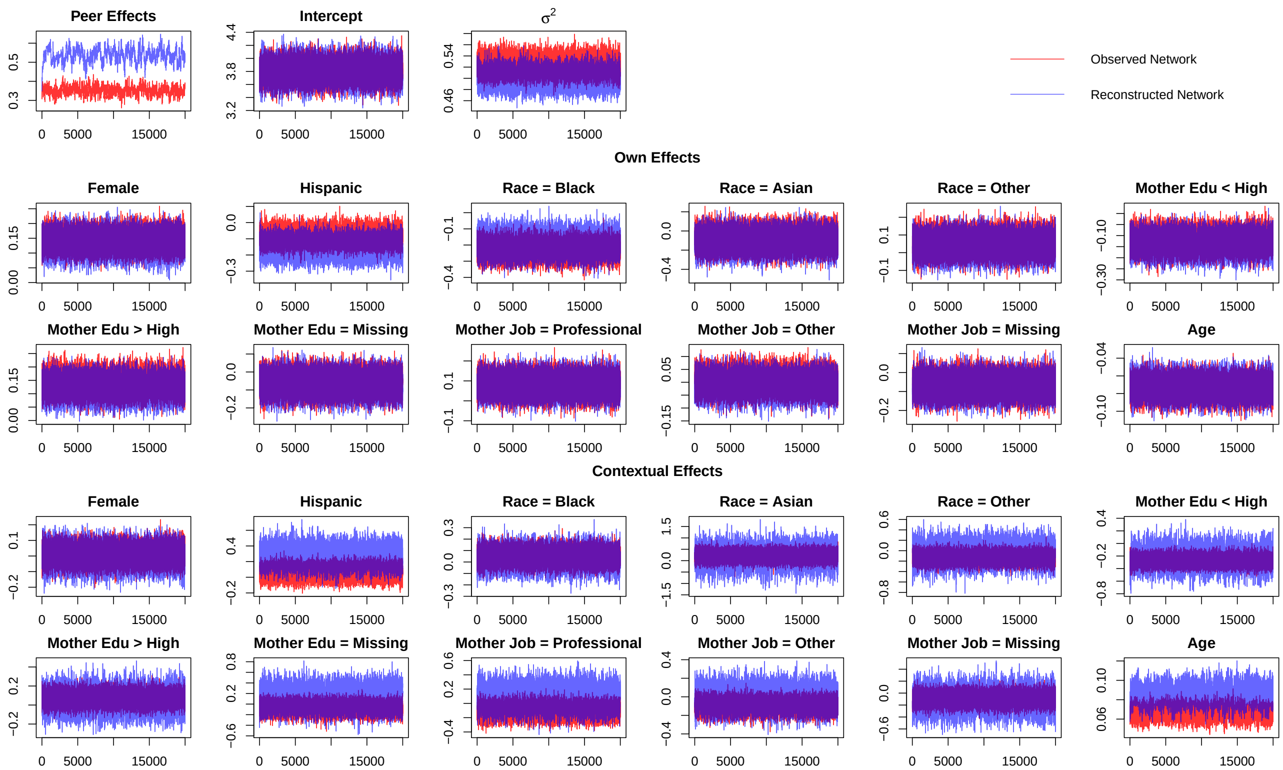
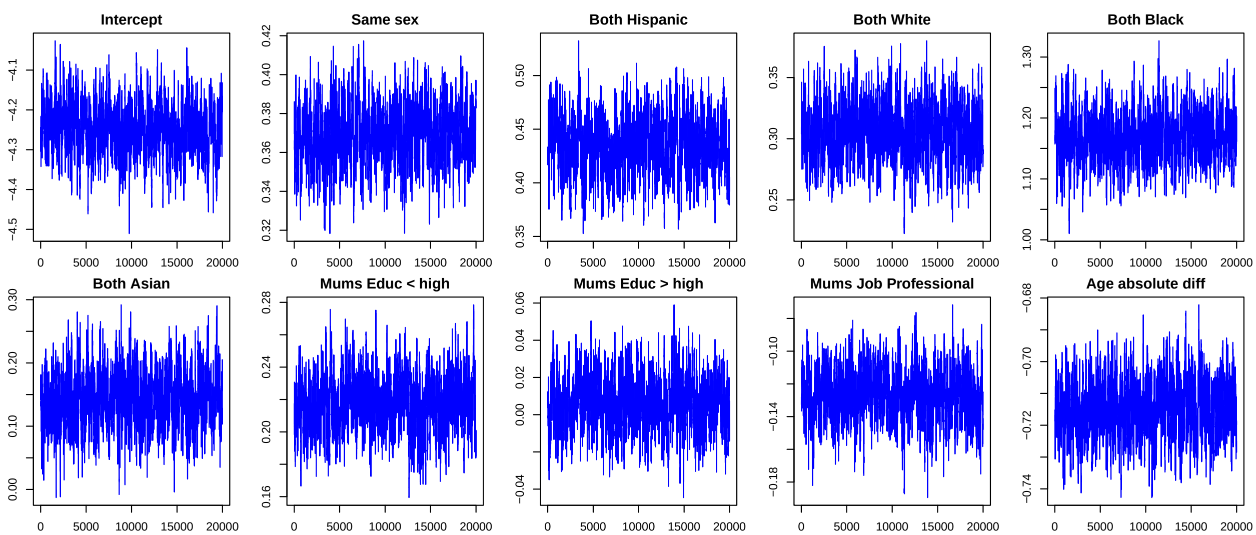


Figure G.4: MCMC Simulations – Network Formation Model



H Aggregated Relational Data

Example 6 (Aggregated Relational Data – *Assumption 5 does not hold*)**.** *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:

$$P(a_{ij,m} = 1) = \frac{\exp\{\nu_i + \nu_j + \zeta \mathbf{z}_i'\mathbf{z}_j\}}{1 + \exp\{\nu_i + \nu_j + \zeta \mathbf{z}_i'\mathbf{z}_j\}}. \tag{19}$$

Here, $\boldsymbol{\rho} = [\{\nu_i, \mathbf{z}_i\}_i, \zeta]$ is not observed by the econometrician. The 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)). This model differs from the ones presented in Examples 1–3 in two fundamental ways.

First, ARD does not provide information on any specific links; [43] *therefore, one could disregard the ARD information and define the predicted distribution estimator of the true network as:*

$$\hat{P}(a_{ij,m} = 1|\hat{\boldsymbol{\rho}}, \mathbf{X}_m) = \frac{\exp\{\hat{\nu}_i + \hat{\nu}_j + \hat{\zeta} \hat{\mathbf{z}}_i'\hat{\mathbf{z}}_j\}}{1 + \exp\{\hat{\nu}_i + \hat{\nu}_j + \hat{\zeta} \hat{\mathbf{z}}_i'\hat{\mathbf{z}}_j\}},$$

where $\hat{\nu}_i$, $\hat{\mathbf{z}}_i$, and $\hat{\zeta}$ are the estimators (e.g., posterior means) of ν_i, $\mathbf{z}_i$, and ζ, respectively.

Second, (and perhaps more importantly) consistent estimation of $\boldsymbol{\rho}$ is only possible as the group size N_m goes to infinity (Breza et al., 2023), which contradicts our Assumption

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

1. Thus, Assumption 5 does not hold. In the online Appendix **??**, *we show that our SGMM estimator (see Section 3) still performs well in finite samples for groups of moderate sizes. In Section 4, we present a Bayesian estimator that allows for inference in finite samples.*

This section provides details about ARD simulation and model estimation using a MCMC method. We simulate the network for a population of 5000 individuals divided into $m = 20$ groups of $n = 250$ individuals. Within each group, the probability of a link is

$$P(a_{ij} = 1) \propto \exp\{\nu_i + \nu_j + \zeta \mathbf{z}_i' \mathbf{z}_j\}. \tag{20}$$

As there is no connection between the groups, the networks are simulated and estimated independently. We first present how we simulate the data following the model (22).

H.1 ARD Simulation

The parameters are defined as follows: $\zeta = 1.5$, $\nu_i \sim \mathcal{N}(-1.25, 0.37)$, and the $\mathbf{z}_i$ are distributed uniformly according to a von Mises–Fisher distribution. We use a hypersphere of dimension 3. We set the same values for the parameter for the 20 groups. We generate the probabilities of links in each network following Breza et al. (2020).

$$P(a_{ij} = 1 | \nu_i, \nu_j, \zeta, \mathbf{z}_i, \mathbf{z}_j) = \frac{\exp\{\nu_i + \nu_j + \zeta \mathbf{z}_i' \mathbf{z}_j\} \sum_{i=1}^N d_i}{\sum_{ij} \exp\{\nu_i + \nu_j + \zeta \mathbf{z}_i' \mathbf{z}_j\}}, \tag{21}$$

where d_i is the degree defined by $d_i \approx \dfrac{C_p(0)}{C_p(\zeta)} \exp(\nu_i) \displaystyle\sum_{i=1}^N \exp(\nu_i)$, and the function $C_p(.)$ is the normalization constant in the von Mises–Fisher distribution density function. After computing the probability of a link for any pair in the population, we sample the entries of the adjacency matrix using a Bernoulli distribution with probability (21).

To generate the ARD, we require the "traits" (e.g., cities) for each individual. We set

$K = 12$ traits on the hypersphere. Their location $\mathbf{v}_k$ is distributed uniformly according to the von Mises-Fisher distribution. The individuals having the trait k are assumed to be generated by a von Mises–Fisher distribution with the location parameter $\mathbf{v}_k$ and the intensity parameter $\eta_k \sim |\mathcal{N}(4, 1)|$, $k = 1, \ldots, 12$.

We attribute traits to individuals given their spherical coordinates. We first define N_k, the number of individuals having the trait k:

$$N_k = \left\lfloor r_k \frac{\sum_{i=1}^{N} f_{\mathcal{M}}(\mathbf{z}_i | \mathbf{v}_k, \eta_k)}{\max_i f_{\mathcal{M}}(\mathbf{z}_i | \mathbf{v}_k, \eta_k)} \right\rfloor,$$

where $\lfloor x \rfloor$ represents the greatest integer less than or equal to x, r_k is a random number uniformly distributed over $(0.8; 0.95)$, and $f_{\mathcal{M}}(\mathbf{z}_i | \mathbf{v}_k, \eta_k)$ is the von Mises–Fisher distribution density function evaluated at $\mathbf{z}_i$ with the location parameter $\mathbf{v}_k$ and the intensity parameter η_k.

The intuition behind this definition for N_k is that when many $\mathbf{z}_i$ are close to $\mathbf{v}_k$, many individuals should have the trait k.

We can finally attribute trait k to individual i by sampling a Bernoulli distribution with the probability f_{ik} given by

$$f_{ik} = N_k \frac{f_{\mathcal{M}}(\mathbf{z}_i | \mathbf{v}_k, \eta_k)}{\sum_{i=1}^{N} f_{\mathcal{M}}(\mathbf{z}_i | \mathbf{v}_k, \eta_k)}.$$

The probability of having a trait depends on the proximity of the individuals to the trait's location on the hypersphere.

H.2 Model Estimation

In practice, we only have the ARD and the traits of each individual. McCormick and Zheng (2015) propose an MCMC approach to infer the parameters in the model (20).

However, the spherical coordinates and the degrees in this model are not identified. The authors solve this issue by fixing some $\mathbf{v}_k$ and use the fixed positions to rotate the latent surface back to a common orientation at each iteration of the MCMC using a Procrustes transformation. In addition, the total size of a subset b_k is constrained in the MCMC.

As discussed by McCormick and Zheng (2015), the number of $\mathbf{v}_k$ and b_k to be set as fixed depends on the dimensions of the hypersphere. In our simulations, $\mathbf{v}_1$, $\mathbf{v}_2$, ..., $\mathbf{v}_5$ are set as fixed to rotate back the latent space. When simulating the data, we let $\mathbf{v}_1 = (1, 0, 0)$, $\mathbf{v}_2 = (0, 1, 0)$, and $\mathbf{v}_3 = (0, 0, 1)$. This ensures that the fixed positions on the hypersphere are spaced, as suggested by the authors, to use as much of the space as possible, maximizing the distance between the estimated positions. We also constrain b_3 to its true value. The results do not change when we constrain a larger set of b_k

Following Breza et al. (2020), we estimate the link probabilities using the parameters' posterior distributions. The gregariousness parameters are computed from the degrees d_i and the parameter ζ using the following equation:

$$\nu_i = \log(d_i) - \log\left(\sum_{i=1}^{N} d_i\right) + \frac{1}{2}\log\left(\frac{C_p(\zeta)}{C_p(0)}\right).$$

H.3 Finite Sample Performance Using ARD

In this section, we study the small sample performance of the estimator presented in Section 3 when the researcher only has access to ARD (as in Example 6). First, we simulate network data using the model proposed by Breza et al. (2020) and simulate outcomes using the linear-in-means model (1) conditional on the simulated networks. Second, we estimate the network formation model using the Bayesian estimator proposed by Breza et al. (2020) (yielding $\hat{\boldsymbol{\rho}}_B$)

and using the classical estimator proposed by Alidaee et al. (2020) (yielding $\hat{\boldsymbol{\rho}}_A$). Third, we estimate the linear-in-means model using the estimators presented in Proposition 2 and Theorem 1 based on $\hat{\boldsymbol{\rho}}_A$ and $\hat{\boldsymbol{\rho}}_B$.

Recall that

$$P(a_{ij} = 1) \propto \exp\{\nu_i + \nu_j + \zeta \mathbf{z}_i' \mathbf{z}_j\}, \tag{22}$$

where ν_i, ν_j, ζ, $\mathbf{z}_i$, and $\mathbf{z}_j$ are not observed by the econometrician but follow parametric distributions. We refer the interested reader to McCormick and Zheng (2015), Breza et al. (2020), and Breza et al. (2023) for a formal discussion of the model, including its identification and consistent estimation.

To study the finite sample performance of our instrumental strategy in this context, we simulate 20 groups, each having 250 individuals. Within each subpopulation, we simulate the ARD responses and a series of observable characteristics. The details of the Monte Carlo simulations can be found below in the Online Appendix H.

Importantly, the model in (22) is based on a single population framework. Thus, the network formation model must be estimated separately for each of the 20 groups. With only 250 individuals in each group, we therefore expect significant small-sample bias.

We contrast the estimator proposed by Breza et al. (2020) with that of Alidaee et al. (2020). Whereas Breza et al. (2020) present a parametric Bayesian estimator, Alidaee et al. (2020) propose a (nonparametric) penalized regression based on a low-rank assumption. One main advantage of the estimator proposed in Alidaee et al. (2020) is that it allows for a wider class of model and ensures that the estimation is fast and easily implementable.[44]

[44] The authors developed user-friendly packages in R and Python. See Alidaee et al. (2020) for links and details.

Note, however, that their method only yields a consistent estimator of $\hat{P}(\mathbf{A})$ if the true network is effectively low rank.

Very intuitively, the low-rank assumption implies that linking probabilities were generated from a small number of parameters. Importantly, the model (22) is not necessarily low rank; for example, if the individuals' latent positions (i.e., the $\mathbf{z}_i$'s) are uniformly distributed, then the model may not be low rank and the method proposed by Alidaee et al. (2020) would perform poorly. If, however, individuals' latent positions are located around a few focal points, then the model might be low-rank because knowledge of these focal points may have high predictive power.

We compare the performance of both estimators as we vary the concentration parameter (that is, κ; see below in the Online Appendix H for details). This has the effect of changing the *effective rank* of the linking probabilities: increasing κ decreases the effective rank.[45] We therefore expect the estimator proposed by Alidaee et al. (2020) to perform better for larger values of κ.

The results are presented in Tables H.1 and H.2. Table H.1 presents the results for the special case where $\mathbf{GX}$ are observed in the data. The table displays the performance of our simulated GMM (see Corollary 1) when the network formation model is estimated by Breza et al. (2020) and Alidaee et al. (2020).

When $\kappa = 0$, the network formation is not low rank. This disproportionately affects the estimator of Alidaee et al. (2020). When $\kappa = 15$, the estimators proposed by Breza et al. (2020) and Alidaee et al. (2020) perform similarly.

[45] We refer the interested reader to Alidaee et al. (2020) for a formal discussion of the effective rank and its importance for their estimator.

We now turn to the more general case where **GX** are not observed. Table H.2 presents the performance of our SGMM estimator (Theorem 1) when the network formation process is estimated using the estimators proposed by Breza et al. (2020) and Alidaee et al. (2020) and when we assume that the researcher knows the true distribution of the network.

We see that the performance of our estimator is strongly affected by the quality of the first-stage network formation estimator. When based on either the estimator proposed by Breza et al. (2020) or Alidaee et al. (2020), for $\kappa = 0$ or $\kappa = 15$, our SGMM estimator performs poorly.

The poor performance of our SGMM estimator in a context where both **Gy** and **GX** are unobserved was anticipated. This occurs for two main reasons. *First*, the consistency of the network formation estimator in Breza et al. (2023) holds as the size of each subpopulation goes to infinity, whereas the consistency of our estimator holds as the number of (bounded) subpopulations goes to infinity. This should affect the performance of our estimator, when based on *estimated* network formation models but not when based on the true distribution of the network.

Second, as discussed in Example 6, ARD provides very little information about the realized network structure in the data (as opposed to censoring issues, for example; see Example 2). Then, if the true distribution is vague in the sense that most predicted probabilities are away from 0 or 1, we expect the estimation to be imprecise. This is what happens when $\kappa = 15$, where our estimation based on the true distribution of the network is very imprecise in a context where the network affects the outcome through both **Gy** and **GX**.

In the next section, we present a likelihood-based estimator, which uses more information on the data-generating process of the outcome to improve the precision of the estimation.

Table H.1: Simulation results with ARD and observed $\boldsymbol{GX}$

Parameter	Breza et al.		Alidaee et al.	
	Mean	Std	Mean	Std
	SMM, $\kappa = 0$, $N = 250$, $M = 20$			
$\alpha = 0.4$	0.392	(0.01)	0.492	(0.057)
$\beta_1 = 1$	1.001	(0.004)	1.002	(0.009)
$\beta_2 = 1.5$	1.500	(0.007)	1.496	(0.016)
$\gamma_1 = 5$	5.013	(0.034)	3.884	(0.295)
$\gamma_2 = -3$	-2.993	(0.052)	-4.048	(0.354)
	SMM, $\kappa = 15$, $N = 250$, $M = 20$			
$\alpha = 0.4$	0.400	(0.009)	0.428	(0.009)
$\beta_1 = 1$	1.000	(0.004)	0.999	(0.004)
$\beta_2 = 1.5$	1.500	(0.008)	1.499	(0.008)
$\gamma_1 = 5$	4.996	(0.034)	4.677	(0.034)
$\gamma_2 = -3$	-3.005	(0.055)	-3.387	(0.055)

Note: In each subnetwork, the spherical coordinates of individuals are generated from a von Mises–Fisher distribution with a location parameter $(1, 0, 0)$ and intensity parameter κ. Predicted probabilities are computed using the mean of the posterior distribution. We chose the weight associated with the nuclear norm penalty to minimize the RMSE through cross-validation. This value of $\lambda = 600$ is smaller than the recommended value in Alidaee et al. (2020). Instruments are build using only second-degree peers, i.e., $\mathbf{G^2X}$.

Table H.2: Simulation results with ARD and unobserved $\boldsymbol{GX}$

Parameter	Breza et al.		Alidaee et al.		True distribution	
	Mean	Std	Mean	Std	Mean	Std
	SMM, $\kappa = 0$, $N = 250$, $M = 20$					
$\alpha = 0.4$	0.717	(0.463)	0.700	(0.268)	0.400	(0.056)
$\beta_1 = 1$	0.988	(0.022)	0.995	(0.017)	1.000	(0.015)
$\beta_2 = 1.5$	1.505	(0.03)	1.503	(0.029)	1.501	(0.021)
$\gamma_1 = 5$	1.778	(4.473)	1.512	(2.37)	4.991	(0.455)
$\gamma_2 = -3$	-2.205	(1.24)	-0.405	(0.955)	-3.005	(0.287)
	SMM, $\kappa = 15$, $N = 250$, $M = 20$					
$\alpha = 0.4$	0.603	(0.069)	0.870	(0.202)	0.434	(0.394)
$\beta_1 = 1$	0.989	(0.014)	0.984	(0.015)	0.998	(0.021)
$\beta_2 = 1.5$	1.504	(0.029)	1.509	(0.029)	1.501	(0.023)
$\gamma_1 = 5$	2.866	(0.566)	0.246	(1.973)	4.638	(3.887)
$\gamma_2 = -3$	-2.458	(0.379)	-1.539	(0.602)	-2.913	(1.037)

Note: see Table H.1.

References (Online Appendix)

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.

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.

Breza, E., A. G. Chandrasekhar, S. Lubold, T. H. McCormick, and M. Pan (2023): "Consistently estimating network statistics using aggregated relational data," *Proceedings of the National Academy of Sciences*, 120, e2207185120.

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

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

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

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

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.

Houndetoungan, A. and A. H. Maoude (2024): "Inference for Two-Stage Extremum Estimators," *arXiv preprint arXiv:2402.05030*.

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.

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

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

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.

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

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.

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