Online Supplement

"*Identifying peer effects on academic achievement through students' effort*"

Elysée Aristide Houndetoungan and Cristelle Kouame

S.1 Additional Notes for the Proofs

S.1.1 Some Basic Properties

In this section, we state and prove some basic properties used throughout the paper.

P.1 Let $[\mathbf{F}_s,\ \bar{\boldsymbol{\ell}}_s/\sqrt{\bar{n}_s},\ \hat{\boldsymbol{\ell}}_s/\sqrt{\hat{n}_s}]$ be the orthonormal matrix of $\mathbf{J}_s$, where the columns in $\mathbf{F}_s$ are eigenvectors of $\mathbf{J}_s$ corresponding to the eigenvalue one. $\|\mathbf{F}_s\|_2 = 1$, where $\|.\|_2$ is the operator norm induced by the ℓ^2-norm.

Proof. $\|\mathbf{F}_s\|_2 = \max\limits_{\mathbf{u}_s'\mathbf{u}_s=1} \sqrt{(\mathbf{F}_s\mathbf{u}_s)'(\mathbf{F}_s\mathbf{u}_s)} = \max\limits_{\mathbf{u}_s'\mathbf{u}_s=1} \sqrt{\mathbf{u}_s'\mathbf{u}_s}$ because $\mathbf{F}_s'\mathbf{F}_s = \mathbf{I}_{n_s-2}$, the identity matrix of dimension $n_s - 2$. Thus, $\|\mathbf{F}_s\|_2 = 1$. □

P.2 For any $n_s \times n_s$ matrix, $\mathbf{B}_s = [b_{s,ij}]$, $|b_{s,ii}| \leqslant \|\mathbf{B}_s\|_2$.

Proof. Let $\mathbf{u}_s$ be the n_s-vector of zeros except for the i-th element, which is one. Note that $\|\mathbf{u}_s\|_2 = 1$. The i-th entry of $\mathbf{B}_s\mathbf{u}$ of $b_{s.ii}$. As a result, $|b_{s,ii}| \leqslant \sqrt{\sum_{j=1}^{n_s} b_{s,ji}^2} = \sqrt{(\mathbf{B}_s\mathbf{u})'(\mathbf{B}_s\mathbf{u})} \leqslant \|\mathbf{B}_s\|_2$. □

P.3 If $\mathbf{B}_s$ is a symmetric matrix of dimension $n_s \times n_s$, then $\|\mathbf{B}_s\|_2 = \pi_{\max}(\mathbf{B}_s)$, where $\pi_{\max}(.)$ is the largest eigenvalue.

Proof. $\|\mathbf{B}_s\|_2 = \max\limits_{\mathbf{u}_s'\mathbf{u}_s=1} \sqrt{(\mathbf{B}_s\mathbf{u}_s)'(\mathbf{B}_s\mathbf{u}_s)} = \max\limits_{\mathbf{u}_s'\mathbf{u}_s=1} \sqrt{\mathbf{u}_s'\mathbf{B}_s^2\mathbf{u}_s} = \sqrt{\pi_{\max}(\mathbf{B}_s^2)} = \pi_{\max}(\mathbf{B}_s)$. □

P.4 If $\mathbf{B}_s$ is a symmetric matrix of dimension $n_s \times n_s$, then $\pi_{\max}(\mathbf{F}_s'\mathbf{B}_s\mathbf{F}_s) \leqslant \pi_{\max}(\mathbf{B}_s)$.

Proof. $\pi_{\max}(\mathbf{F}_s'\mathbf{B}_s\mathbf{F}_s) = \max\limits_{\mathbf{u}_s'\mathbf{u}_s=1} \mathbf{u}_s'\mathbf{F}_s'\mathbf{B}_s\mathbf{F}_s\mathbf{u}_s = \max\limits_{\mathbf{u}_s'\mathbf{u}_s=1} (\mathbf{F}_s\mathbf{u}_s)'\mathbf{B}_s(\mathbf{F}_s\mathbf{u}_s)$. As $(\mathbf{F}_s\mathbf{u}_s)'(\mathbf{F}_s\mathbf{u}_s) = 1$, then $\max\limits_{\mathbf{u}_s'\mathbf{u}_s=1} (\mathbf{F}_s\mathbf{u}_s)'\mathbf{B}_s(\mathbf{F}_s\mathbf{u}_s) \leqslant \max\limits_{\mathbf{u}_s'\mathbf{u}_s=1} \mathbf{u}_s'\mathbf{B}_s\mathbf{u}_s = \pi_{\max}(\mathbf{B}_s)$. □

P.5 Let $\mathbf{B}_{s,1}$ and $\mathbf{B}_{s,2}$ be $n_s \times n_s$ matrices. If $\mathbf{B}_{s,1}$ and $\mathbf{B}_{s,2}$ are absolutely bounded in row and column sums, then $\mathbf{B}_{s,1}\mathbf{B}_{s,2}$ is absolutely bounded in row and column sums.

Proof. It is sufficient to show that the entries of $\mathbf{B}_{s,1}\mathbf{B}_{s,2}\mathbf{u}_s$ and $\mathbf{u}_s'\mathbf{B}_{s,1}\mathbf{B}_{s,2}$ are absolutely bounded for all n_s-vector $\mathbf{u}_s$ whose entries take -1 or 1. Assume that $\mathbf{B}_{s,1}$ is absolutely bounded in row sum by $C_{b,1}$ and absolutely bounded in the row sum by $R_{b,1}$. Assume also that $\mathbf{B}_{s,2}$ is absolutely bounded in the row sum by $C_{b,2}$ and absolutely bounded in row sum by $R_{b,2}$. We have $\mathbf{B}_{s,2}\mathbf{u}_s \preceq R_{b,2}\mathbf{1}_{n_s}$ and $\mathbf{B}_{s,1}\mathbf{1}_{n_s} \preceq R_{b,1}\mathbf{1}_{n_s}$, where $\preceq$ is the pointwise inequality $\leqslant$ and $\mathbf{1}_{n_s}$

is an n_s-vector of ones. Thus, $\mathbf{B}_{s,1}\mathbf{B}_{s,2}\mathbf{u}_s \preceq R_{b,2}\mathbf{B}_{s,1}\mathbf{1}_{n_s} \preceq R_{b,1}R_{b,2}\mathbf{1}_{n_s}$. Hence, $\mathbf{B}_{s,1}\mathbf{B}_{s,2}$ is bounded in row sum. Analogously, we have $\mathbf{u}_s'\mathbf{B}_{s,1} \preceq C_{b,1}\mathbf{1}_{n_s}'$ and $\mathbf{1}_{n_s}'\mathbf{B}_{s,2} \preceq C_{b,2}\mathbf{1}_{n_s}'$. Thus, $\mathbf{u}_s'\mathbf{B}_{s,1}\mathbf{B}_{s,2} \preceq C_{b,1}\mathbf{1}_{n_s}'\mathbf{B}_{s,2} \preceq C_{b,1}C_{b,2}\mathbf{1}_{n_s}'$. Hence, $\mathbf{B}_{s,1}\mathbf{B}_{s,2}$ is bounded in column sum. □

P.6 If an $n_s \times n_s$ matrix $\mathbf{B}_s$ is absolutely bounded in both row and column sums, then $|\pi_{\max}(\mathbf{B}_s)| < \infty$ and $\|\mathbf{B}_s\|_2 < \infty$.

Proof. $|\pi_{\max}(\mathbf{B}_s)| < \infty$ is a direct implication of the Gershgorin circle theorem.[1]
Besides, $\|\mathbf{B}_s\|_2 = \sqrt{\pi_{\max}(\mathbf{B}_s'\mathbf{B}_s)} < \infty$ because $\mathbf{B}_s'\mathbf{B}_s$ is absolutely bounded in row and column sums by P.5. □

P.7 Let $\mathbf{B}_s = [b_{ij}]$, $\dot{\mathbf{B}}_s = [\dot{b}_{ij}]$ be $n_s \times n_s$ matrices. Let $\mathbf{G} = \operatorname{diag}(\mathbf{G}_1, \dots, \mathbf{G}_S)$, where diag is the block diagonal operator. Let also $\mu_{4\eta} = \mathbb{E}(\eta_{s,i}^4|\mathbf{G}_s, \mathbf{X}_s)$, $\mu_{4\epsilon} = \mathbb{E}(\varepsilon_{s,i}^4|\mathbf{G}_s, \mathbf{X}_s)$, $\mu_{22} = \mathbb{E}(\eta_{s,i}^2\varepsilon_{s,i}^2|\mathbf{G}_s, \mathbf{X}_s)$, $\mu_{31} = \mathbb{E}(\eta_{s,i}^3\varepsilon_{s,i}|\mathbf{G}_s, \mathbf{X}_s)$, and $\mu_{13} = \mathbb{E}(\eta_{s,i}\varepsilon_{s,i}^3|\mathbf{G}_s, \mathbf{X}_s)$. Under Assumptions 3.1 and A.3,

$\mathbb{V}(\boldsymbol{\eta}_s'\mathbf{B}_s\boldsymbol{\eta}_s|\mathbf{G}) = (\mu_{4\eta} - 3\sigma_{0\epsilon}^4)\sum_{i=1}^{n_s} b_{ii}^2 + \sigma_{0\epsilon}^4(\operatorname{Tr}(\mathbf{B}_s\mathbf{B}_s') + \operatorname{Tr}(\mathbf{B}_s^2))$,

$\mathbb{V}(\boldsymbol{\varepsilon}_s'\mathbf{B}_s\boldsymbol{\varepsilon}_s|\mathbf{G}) = (\mu_{4\epsilon} - 3\sigma_{0\epsilon}^4)\sum_{i=1}^{n_s} b_{ii}^2 + \sigma_{0\epsilon}^4(\operatorname{Tr}(\mathbf{B}_s\mathbf{B}_s') + \operatorname{Tr}(\mathbf{B}_s^2))$,

$\mathbb{V}(\boldsymbol{\varepsilon}_s'\mathbf{B}_s\boldsymbol{\eta}_s|\mathbf{G}) = (\mu_{22} - 3\sigma_{0\eta}^2\sigma_{0\epsilon}^2)\sum_{i=1}^{n_s} b_{ii}^2 + (1-\rho^2)\sigma_{0\eta}^2\sigma_{0\epsilon}^2(\operatorname{Tr}(\mathbf{B}_s))^2 + \sigma_{0\eta}^2\sigma_{0\epsilon}^2\operatorname{Tr}(\mathbf{B}_s\mathbf{B}_s') + \rho^2\sigma_{0\eta}^2\sigma_{0\epsilon}^2\operatorname{Tr}(\mathbf{B}_s^2)$,

$\mathbb{C}\mathbf{ov}(\boldsymbol{\eta}_s'\mathbf{B}_s\boldsymbol{\eta}_s, \boldsymbol{\varepsilon}_s'\dot{\mathbf{B}}_s\boldsymbol{\eta}_s|\mathbf{G}) = (\mu_{31} - 3\rho\sigma_{0\eta}^3\sigma_{0\epsilon})\sum_{i=1}^{n_s} b_{ii}\dot{b}_{ii} + \rho\sigma_{0\eta}^3\sigma_{0\epsilon}(\operatorname{Tr}(\mathbf{B}_s\dot{\mathbf{B}}_s') + \operatorname{Tr}(\mathbf{B}_s\dot{\mathbf{B}}_s))$,

$\mathbb{C}\mathbf{ov}(\boldsymbol{\varepsilon}_s'\mathbf{B}_s\boldsymbol{\varepsilon}_s, \boldsymbol{\eta}_s'\dot{\mathbf{B}}_s\boldsymbol{\varepsilon}_s|\mathbf{G}) = (\mu_{13} - 3\rho\sigma_{0\eta}\sigma_{0\epsilon}^3)\sum_{i=1}^{n_s} b_{ii}\dot{b}_{ii} + \rho\sigma_{0\eta}\sigma_{0\epsilon}^3(\operatorname{Tr}(\mathbf{B}_s\dot{\mathbf{B}}_s') + \operatorname{Tr}(\mathbf{B}_s\dot{\mathbf{B}}_s))$,

$\mathbb{C}\mathbf{ov}(\boldsymbol{\eta}_s'\mathbf{B}_s\boldsymbol{\eta}_s, \boldsymbol{\varepsilon}_s'\mathbf{B}_s\boldsymbol{\varepsilon}_s|\mathbf{G}) = (\mu_{22} - 2\rho^2\sigma_{0\eta}^2\sigma_{0\epsilon}^2 - \sigma_{0\eta}^2\sigma_{0\epsilon}^2)\sum_{i=1}^{n_s} b_{ii}\dot{b}_{ii} + \rho^2\sigma_{0\eta}^2\sigma_{0\epsilon}^2(\operatorname{Tr}(\mathbf{B}_s\dot{\mathbf{B}}_s') + \operatorname{Tr}(\mathbf{B}_s\dot{\mathbf{B}}_s))$.

The proof of the lemma is straightforward using the classical definition of variance and covariance.

S.1.2 Identification and Consistent Estimator of $(\sigma_{\epsilon 0}^2,\ \tau_0,\ \rho_0)$

We must show that $\mathbb{V}\left(\hat{\sigma}_\epsilon^2(\tau,\rho)|\mathbf{G}\right) = o_p(1)$.
We have $\hat{\sigma}_\epsilon^2(\tau,\rho) = \displaystyle\sum_{s=1}^{S}\frac{((\mathbf{I}_{n_s} - \lambda_0\mathbf{G}_s)\boldsymbol{\eta}_s + \boldsymbol{\varepsilon}_s)'\mathbf{F}_s\boldsymbol{\Omega}_s^{-1}(\lambda_0,\tau,\rho)\mathbf{F}_s'((\mathbf{I}_{n_s} - \lambda_0\mathbf{G}_s)\boldsymbol{\eta}_s + \boldsymbol{\varepsilon}_s)}{n - 2S}$. Thus,

$$\begin{aligned}\mathbb{V}(\hat{\sigma}_\epsilon^2(\tau,\rho)|\mathbf{G}) = &\frac{1}{(n-2S)^2}\sum_{s=1}^{S}\big(\mathbb{V}(\boldsymbol{\eta}_s'\ddot{\mathbf{M}}_s\boldsymbol{\eta}_s|\mathbf{G}) + 4\mathbb{V}(\boldsymbol{\eta}_s'\dot{\mathbf{M}}_s\boldsymbol{\varepsilon}_s|\mathbf{G}) + \mathbb{V}(\boldsymbol{\varepsilon}_s'\mathbf{M}_s\boldsymbol{\varepsilon}_s|\mathbf{G}) + \\ &4\,\mathbb{C}\mathbf{ov}(\boldsymbol{\eta}_s'\ddot{\mathbf{M}}_s\boldsymbol{\eta}_s, \boldsymbol{\eta}_s'\dot{\mathbf{M}}_s\boldsymbol{\varepsilon}_s|\mathbf{G}) + 2\,\mathbb{C}\mathbf{ov}(\boldsymbol{\eta}_s'\ddot{\mathbf{M}}_s\boldsymbol{\eta}_s, \boldsymbol{\varepsilon}_s'\mathbf{M}_s\boldsymbol{\varepsilon}_s|\mathbf{G}) + \\ &4\,\mathbb{C}\mathbf{ov}(\boldsymbol{\varepsilon}_s'\mathbf{M}_s\boldsymbol{\varepsilon}_s, \boldsymbol{\eta}_s'\dot{\mathbf{M}}_s\boldsymbol{\varepsilon}_s|\mathbf{G})\big),\end{aligned} \tag{S.1}$$

where $\mathbf{M}_s = \mathbf{F}_s\boldsymbol{\Omega}_s^{-1}(\lambda_0,\tau,\rho)\mathbf{F}_s'$, $\dot{\mathbf{M}}_s = (\mathbf{I}_{n_s} - \lambda_0\mathbf{G}_s)'\mathbf{M}_s$, and $\ddot{\mathbf{M}}_s = \dot{\mathbf{M}}_s(\mathbf{I}_{n_s} - \lambda_0\mathbf{G}_s)$.
As $\pi_{\min}(\boldsymbol{\Omega}_s(\lambda_0,\tau,\rho)$ is bounded away from zero (Assumption A.2), we have $|\pi_{\max}(\boldsymbol{\Omega}_s^{-1}(\lambda_0,\tau,\rho)| = O_p(1)$. Thus, $\max_s\|\boldsymbol{\Omega}_s^{-1}(\lambda_0,\tau,\rho)\|_2 = O_p(1)$ by P.3. This implies that $\max_s\|\mathbf{M}_s\|_2 = O_p(1)$, $\max_s\|\dot{\mathbf{M}}_s\|_2 = O_p(1)$, and $\max_s\|\ddot{\mathbf{M}}_s\|_2 = O_p(1)$ because $\|\mathbf{F}_s\|_2 = 1$ and $\|\mathbf{I}_{n_s} - \lambda_0\mathbf{G}_s\|_2 = O_p(1)$ by P.6.

[1] See Horn, R. A. and C. R. Johnson (2012): *Matrix analysis*, Cambridge university press.

We now need to show that the sum over s of each term of the variance (S.1) is $o_p((n-2S)^2)$. By P.2, the trace of any product of matrices chosen among $\mathbf{M}_s$, $\dot{\mathbf{M}}_s$, and $\ddot{\mathbf{M}}_s$ is $O_p(n_s)$ and thus, $o_p((n-2S)^2)$. For example, $|\mathrm{Tr}(\mathbf{M}_s\dot{\mathbf{M}}_s)| \leqslant n_s\|\mathbf{M}_s\dot{\mathbf{M}}_s\|_2 \leqslant n_s\|\mathbf{M}_s\|_2\|\dot{\mathbf{M}}_s\|_2 = O_p(n_s) = o_p((n-2S)^2)$. On the other hand, $\sum_{s=1}^{S}(\mathrm{Tr}(\mathbf{M}_s))^2 = O_p(\sum_{s=1}^{S} n_s^2) = o_p((n-2S)^2)$. Moreover, $\sum_{i=1}^{n_s} m_{ii}^2 \leqslant n_s\|\mathbf{M}_s\|_2^2 = O_p(n_s) = o_p((n-2S)^2)$ by P.2. Analogously, $\sum_{i=1}^{n_s} m_{ii}\dot{m}_{ii} = o_p((n-2S)^2)$. As a result, $\mathbb{V}(\hat{\sigma}_\epsilon^2(\tau,\rho)|\mathbf{G}) = o_p(1)$.

The proof implies, by Chebyshev inequality, that $\hat{\sigma}_\epsilon^2(\tau,\rho) - \mathbb{E}\left(\hat{\sigma}_\epsilon^2(\tau,\rho)|\mathbf{G}_1,\ldots,\mathbf{G}_S\right)$ converges in probability to zero. The convergence is uniform in the space of $(\tau,\ \rho)$ because $\hat{\sigma}_\epsilon^2(\tau,\rho)$ and $\mathbb{E}\left(\hat{\sigma}_\epsilon^2(\tau,\rho)|\mathbf{G}_1,\ldots,\mathbf{G}_S\right)$ can be expressed as a polynomial function in $(\tau,\ \rho)$. Thus, $\frac{1}{n}(L_c(\tau,\rho) - L_c^*(\tau,\rho))$ converges uniformly to zero. This proof also implies that $\operatorname{plim}\hat{\sigma}_\epsilon^2(\tau_0,\rho_0) = \sigma_{\epsilon 0}^2$.

S.1.3 Necessary Conditions for the Identification$(\sigma_{\epsilon 0}^2,\ \tau_0,\ \rho_0)$

As $\lambda_0 \neq 0$ (Condition (i) of Assumption 3.3) and is identified, $\mathbb{E}(\boldsymbol{v}_s\boldsymbol{v}_s'|\mathbf{G}_s)$ implies a unique $(\sigma_{\eta 0},\ \sigma_{\epsilon 0},\ \rho_0)$ if $\mathbf{J}_s$, $\mathbf{J}_s(\mathbf{G}_s+\mathbf{G}_s')\mathbf{J}_s$ and $\mathbf{J}_s\mathbf{G}_s\mathbf{G}_s'\mathbf{J}_s$ are linearly independent. We present a simple subnetwork structure that verifies this condition.

Let $\mathbf{C}_s$ be an arbitrary $n_s \times n_s$ matrix. Unless otherwise stated, we use $\mathbf{C}_{s,ij}$ to denote the $(i,\ j)$-th entry of $\mathbf{C}_s$. Assume that i and j are from the subset of students who have friends in the school s. The (i,j)-th entry of $\mathbf{J}_s\mathbf{C}_s\mathbf{J}_s$ is $\mathbf{C}_{s,ij} - \hat{\mathbf{C}}_{s,\bullet j} - \hat{\mathbf{C}}_{s,i\bullet} + \hat{\mathbf{C}}_{s,\bullet\bullet}$, where $\hat{\mathbf{C}}_{s,\bullet j} = (1/\hat{n}_s)\sum_{k\in\hat{\mathcal{V}}_s}^{n_s}\mathbf{C}_{s,kj}$, $\hat{\mathbf{C}}_{s,i\bullet} = (1/\hat{n}_s)\sum_{l\in\hat{\mathcal{V}}_s}^{n_s}\mathbf{C}_{s,il}$, and $\hat{\mathbf{C}}_{s,\bullet\bullet} = (1/\hat{n}_s^2)\sum_{k,l\in\hat{\mathcal{V}}_s}^{n_s}\mathbf{C}_{s,kl}$.

Let $\tilde{\mathbf{G}}_s = \mathbf{G}_s\mathbf{G}_s'$ and $i_1,\ \ldots,\ i_4$ be four students from $\hat{\mathcal{V}}_s$ who are not directly linked and where only two of them have common friends. Without loss of generality, assume that i_1 and i_3 have common friends. For any $i \in \{i_1,\ i_2\}$ and $j \in \{i_3,\ i_4\}$, $\mathbf{J}_{s,ij} = -1/\hat{n}_s$, $\mathbf{G}_{s,ij} = 0$, and $\mathbf{G}_{s,ij}' = 0$. Moreover, $\tilde{\mathbf{G}}_{s,ij} = 0$ except for the pair $(i_i,\ i_3)$, who have common friends. Let $\mathbf{L}_s = b_1\mathbf{J}_s + b_2\mathbf{J}_s(\mathbf{G}_s + \mathbf{G}_s')\mathbf{J}_s + b_3\mathbf{J}_s\mathbf{G}_s\mathbf{G}_s'\mathbf{J}_s = 0$ for some $b_1,\ b_2,\ b_3 \in \mathbb{R}$. We have $\mathbf{L}_{s,ij} = -b_1/\hat{n}_s - b_2(\mathbf{G}_{s,ij} - \mathbf{G}_{s,\bullet j} - \mathbf{G}_{s,i\bullet} + \mathbf{G}_{s,\bullet\bullet} + \mathbf{G}_{s,ij}' - \mathbf{G}_{s,\bullet j}' - \mathbf{G}_{s,i\bullet}' + \mathbf{G}_{s,\bullet\bullet}') + b_3(\tilde{\mathbf{G}}_{s,ij} - \tilde{\mathbf{G}}_{s,\bullet j} - \tilde{\mathbf{G}}_{s,i\bullet} + \tilde{\mathbf{G}}_{s,\bullet\bullet})$. This implies that $\mathbf{L}_{s,i_1i_3} + \mathbf{L}_{s,i_2i_4} - \mathbf{L}_{s,i_2i_3} - \mathbf{L}_{s,i_1i_4} = b_3\tilde{\mathbf{G}}_{s,i_1i_3}$. Thus, if the combination $\mathbf{L}_s$ is zero, then $b_3 = 0$.

Let $j_1,\ \ldots,\ j_4$ be four students from $\hat{\mathcal{V}}_s$, where only two of them are directly linked (mutually or not), and the others are not directly linked. Without loss of generality, assume that only j_1 to j_3 are linked, i.e., for any $i \in \{j_1,\ j_2\}$ and $j \in \{j_3,\ j_4\}$, $\mathbf{G}_{s,ij} = 0$ and $\mathbf{G}_{s,ij}' = 0$ except for the pairs $(j_1,\ j_3)$ and $(j_3,\ j_1)$. As $b_3 = 0$, we have $\mathbf{L}_{s,j_1j_3} + \mathbf{L}_{s,j_2j_4} - \mathbf{L}_{s,j_2j_3} - \mathbf{L}_{s,j_1j_4} = b_2(\mathbf{G}_{s,j_1j_3} + \mathbf{G}_{s,j_1j_3}')$. Thus if $\mathbf{L}_s$ is zero, then $b_2 = 0$, and it follows that $b_1 = 0$.

As a result, $\mathbf{J}_s$, $\mathbf{J}_s(\mathbf{G}_s + \mathbf{G}_s')\mathbf{J}_s$, and $\mathbf{J}_s\mathbf{G}_s\mathbf{G}_s'\mathbf{J}_s$ are linearly independent if, in some school s, there are four students from $\hat{\mathcal{V}}_s$ who are not directly linked and only two of them have common friends, and if in some school s, there are four students from $\hat{\mathcal{V}}_s$, where only two of them are linked.

We present an example of this condition by adding three nodes to Figure 1 with two additional links (see Figure S.1). There are no links within the nodes i_1, i_4, i_5, and i_6, and only i_5 and i_6 have common a friends (i_7). Besides, only i_5 and i_7 are linked within the nodes i_1, i_2, i_5, and i_7.

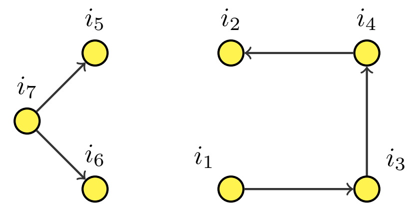

Figure S.1: Illustration of the identification
Note: $\rightarrow$ means that the node on the left side is a friend of the node on the right side.

Many other situations lead to $b_1 = b_2 = b_3 = 0$. In practice, one can easily verify if $\mathbf{J}_s$, $\mathbf{J}_s(\mathbf{G}_s + \mathbf{G}_s')\mathbf{J}_s$ and $\mathbf{J}_s\mathbf{G}_s\mathbf{G}_s'\mathbf{J}_s$ are linearly independent.

S.2 Bayesian Estimation of the Network Formation Model

In the Bayesian approach, we assume that $\mu_{0,s,i}^{in}$ and $\mu_{0,s,i}^{out}$ are random effects following $\mathcal{N}(0,\ \sigma_{in}^2)$ and $\mathcal{N}(0,\ \sigma_{out}^2)$, respectively, with $\mathbb{E}(\mu_{0,s,i}^{in}\mu_{0,s,i}^{out}) = \rho_\mu$. To simulate the posterior distribution of $\mu_{0,s,i}^{in}$ and $\mu_{0,s,i}^{out}$, we use the data augmentation technique.[2]

Let $a_{s,ij}^* = \ddot{\mathbf{x}}_{s,ij}'\ddot{\boldsymbol{\beta}}_0 + \mu_{0,s,i}^{in} + \mu_{0,s,j}^{out} + u_{s,ij}$, such that $a_{s,ij} = 1$ if $a_{s,ij}^* > 0$ and $a_{s,ij} = 0$ otherwise, where $u_{s,ij} \sim \mathcal{N}(0,\ 1)$. Let $\mathbf{a}_s = (a_{s,ij}; i \neq j)'$ and $\mathbf{a}_s^* = \left(a_{s,ij}^*; i \neq j\right)'$. The density function of $\mathbf{a}_s^*$, conditional on $\mathbf{a}_s$, $\ddot{\mathbf{X}}_s = [\ddot{\mathbf{x}}_{s,ij}; i \neq j]'$, $\ddot{\boldsymbol{\beta}}_0$, $\boldsymbol{\mu}_s^{in} = (\mu_{0,s,1}^{in}, \dots, \mu_{0,s,i}^{in})'$, and $\boldsymbol{\mu}_s^{out} = (\mu_{0,s,1}^{out}, \dots, \mu_{0,s,i}^{out})'$ is proportional to

$$\prod_{i\neq j} \left\{\mathrm{I}\left(a_{s,ij}^* \geqslant 0\right)\mathrm{I}\left(a_{s,ij} = 1\right) + \mathrm{I}\left(a_{s,ij}^* < 0\right)\mathrm{I}\left(a_{s,ij} = 0\right)\right\} \exp\left\{-\tfrac{1}{2}(a_{s,ij}^* - \ddot{\mathbf{x}}_{s,ij}'\ddot{\boldsymbol{\beta}}_0 - \mu_{0,s,i}^{in} - \mu_{0,s,j}^{out})^2\right\},$$

where I(.) is the indicator function. This implies that the distribution of $a_{s,ij}^*|\mathbf{a}_s, \ddot{\mathbf{X}}_s, \ddot{\boldsymbol{\beta}}_0, \boldsymbol{\mu}_s^{in}, \boldsymbol{\mu}_s^{out}$ is $\mathcal{N}(\ddot{\mathbf{x}}_{s,ij}'\ddot{\boldsymbol{\beta}}_0 + \mu_{0,s,i}^{in} + \mu_{0,s,j}^{out},\ 1)$, truncated at the left by 0 if $a_{s,ij} = 1$, and at the right by 0 if $a_{s,ij} = 0$. Given that the number of observations in the network formation model is high, we set a flat prior distribution for $\ddot{\boldsymbol{\beta}}_0$, σ_{in}^2, σ_{out}^2, and ρ_μ. Thus,

$$\ddot{\boldsymbol{\beta}}_0|\mathbf{a}_1, \mathbf{a}_1^*, \ddot{\mathbf{X}}_1, \boldsymbol{\mu}_1^{in}, \boldsymbol{\mu}_1^{out}, \dots, \mathbf{a}_S, \mathbf{a}_S^*, \ddot{\mathbf{X}}_S, \boldsymbol{\mu}_S^{in}, \boldsymbol{\mu}_S^{out}, \sim \mathcal{N}\left(\left(\ddot{\mathbf{X}}'\ddot{\mathbf{X}}\right)^{-1}\textstyle\sum_{s=1}^S \ddot{\mathbf{X}}_s'\ddot{\mathbf{a}}_s^*,\ \left(\ddot{\mathbf{X}}'\ddot{\mathbf{X}}\right)^{-1}\right),$$

where $\ddot{\mathbf{X}}'\ddot{\mathbf{X}} = \sum_{s=1}^S \ddot{\mathbf{X}}_s'\ddot{\mathbf{X}}_s$ and $\ddot{\mathbf{a}}_s^* = (a_{s,ij}^* - \mu_{0,s,i}^{in} - \mu_{0,s,j}^{out} : i \neq j)'$. For any i,

$$\mu_{0,s,i}^{in}|\ddot{\boldsymbol{\beta}}_0, \mathbf{a}_s, \mathbf{a}_s^*, \ddot{\mathbf{X}}_s, \boldsymbol{\mu}_{s,-i}^{in}, \boldsymbol{\mu}_s^{out} \sim \mathcal{N}\left(\hat{u}_{s,in},\ \hat{\sigma}_{s,in}^2\right),$$

[2]See Albert, J. H., & Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American statistical Association*, 88(422), 669-679.

where $\hat{u}_{s,in} = \hat{\sigma}^2_{s,in} \sum_{i \neq j} (a^*_{s,ij} - \ddot{\mathbf{x}}'_{s,ij} \ddot{\boldsymbol{\beta}}_0 - \mu^{out}_{0,s,j})$ and $\hat{\sigma}^2_{s,in} = \dfrac{\sigma^2_{in}}{1 + (n_s - 1)\, \sigma^2_{in}}$. Analogously,

$$\mu^{out}_{0,s,i} | \ddot{\boldsymbol{\beta}}_0, \mathbf{a}_s, \mathbf{a}^*_s, \ddot{\mathbf{X}}_s, \boldsymbol{\mu}^{in}, \boldsymbol{\mu}^{out}_{-i} \sim \mathcal{N}\left(\hat{u}_{s,out},\ \hat{\sigma}^2_{s,out}\right),$$

where $\hat{u}_{s,out} = \hat{\sigma}^2_{s,out} \sum_{i \neq j} (a^*_{ji} - \ddot{\mathbf{x}}'_{s,ij} \ddot{\boldsymbol{\beta}}_0 - \mu^{in}_{0,s,j})$, and $\hat{\sigma}^2_{s,out} = \dfrac{\sigma^2_{out}}{1 + (n_s - 1)\, \sigma^2_{out}}$.

For the sake of identification, we normalize $\boldsymbol{\mu}^{in}$ and $\boldsymbol{\mu}^{out}$ to zero mean in each subnetwork for each step in the Gibbs sampling. The means of $\boldsymbol{\mu}^{in}$ and $\boldsymbol{\mu}^{out}$ before this normalization are added to the intercept of the subnetwork for the posterior likelihood not to change.

Finally, let $\boldsymbol{\Sigma}_{\mu,\nu} = \begin{pmatrix} \sigma^2_{in} & \rho_\mu \sigma_{in} \sigma_{out} \\ \rho_\mu \sigma_{in} \sigma_{out} & \sigma^2_{out} \end{pmatrix}$,

$$\boldsymbol{\Sigma}_{\mu,\nu} | \ddot{\boldsymbol{\beta}}_0, \mathbf{a}, \mathbf{a}^*, \ddot{\mathbf{X}}_s, \boldsymbol{\mu}^{in}, \boldsymbol{\mu}^{out} \sim \text{Inverse-Wishart}\left(n, \widehat{\mathbf{V}}_{\boldsymbol{\Sigma}_{\mu,\nu}}\right),$$

where $\widehat{\mathbf{V}}_{\boldsymbol{\Sigma}_{\mu,\nu}} = \sum_{i=1}^{n} (\mu^{in}_{0,s,i}, \mu^{out}_{0,s,i})$.