Online Appendix

"Identifying Peer Effects on Student Academic 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 $\mathbf{B}_s\mathbf{u}$ is $b_{s.ii}$. 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 $\hat{\sigma}_\epsilon^2(\tau,\rho) = \sum_{s=1}^{S} \dfrac{((\mathbf{I}_{n_s} - \lambda_0\mathbf{G}_s)\boldsymbol{\eta}_s + \boldsymbol{\varepsilon}_s)'\mathbf{F}_s\boldsymbol{\Omega}_s^{-1}(\lambda_0,\tau,\rho)\mathbf{F}_s'((\mathbf{I}_{n_s} - \lambda_0\mathbf{G}_s)\boldsymbol{\eta}_s + \boldsymbol{\varepsilon}_s)}{n - 2S}$. 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 $\hat{\sigma}_\epsilon^2(\tau,\rho) - \mathbb{E}\left(\hat{\sigma}_\epsilon^2(\tau,\rho)|\mathbf{G}_1,\dots,\mathbf{G}_S\right)$ converges in probability to zero. The convergence is uniform in the space of $(\tau,\ \rho)$ because $\hat{\sigma}_\epsilon^2(\tau,\rho)$ and $\mathbb{E}\left(\hat{\sigma}_\epsilon^2(\tau,\rho)|\mathbf{G}_1,\dots,\mathbf{G}_S\right)$ 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 of $(\sigma_{\epsilon 0}^2,\ \tau_0,\ \rho_0)$

As $\lambda_0 \neq 0$ (Condition (i) of Assumption 3.2) 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 $i_1,\ \dots,\ i_4$ 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 $j_1,\ \dots,\ j_4$ 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, that is, 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.

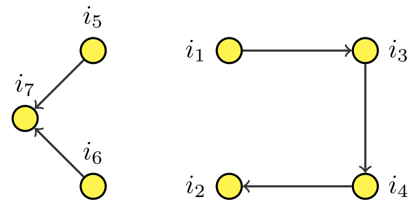

Figure S.1: Illustration of the identification

Note: $\rightarrow$ means that the node on the right side is a friend of the node on the left 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.1.4 Asumptotic Normality in the Case of Endogenous Networks

In the specification controlling for network endogeneity, we replace $\mu_{0,s,i}^{out}$ and $\mu_{0,s,i}^{in}$ with their estimator and replace $\mathbf{h}_s^{\eta}$ and $\mathbf{h}_s^{\epsilon}$ with cubic B-spline approximations. Let ζ be the number of knots in the splines. The knots are points that split the support of $\mu_{0,s,i}^{out}$ and $\mu_{0,s,i}^{in}$ into intervals. The smooth functions are approximated by cubic polynomials on each interval. The case $\zeta = 0$ is equivalent to approximating h_η and h_ϵ by cubic polynomial.

The cubic B-spline approximation of each $\mathbf{h}_s^{\eta}$ and $\mathbf{h}_s^{\epsilon}$ is a linear combination of bases B_k^{in}, $B_{k'}^{out}$, where k and k' take the values $1, \dots, \zeta + 3$, and B_k^{in} and $B_{k'}^{out}$ are piecewise polynomial functions of $\mu_{0,s,i}^{out}$ and $\mu_{0,s,i}^{in}$ respectively (for more details, see Hastie, 2017). We also include a linear combination of $B_k^{in}B_{k'}^{out}$ to account for interaction between $\mu_{0,s,i}^{out}$ and $\mu_{0,s,i}^{in}$. This more flexible approximation is known as a tensor product of the cubic B-splines. Therefore, $\mathbf{h}_s^{\eta}$ and $\mathbf{h}_s^{\epsilon}$ are approximated by combinations of $(\zeta + 3)(\zeta + 5)$ piecewise polynomial functions of $\mu_{0,s,i}^{out}$ and $\mu_{0,s,i}^{in}$. As $\mathbf{h}_s^{\eta}$ multiplies $\mathbf{G}_s$ in Equation (11), the case $\zeta = 10$ leads to plugging 390 new regressors in the initial specification (6).[2]

Let $\dot{\mathbf{X}}_s$ be the matrix of the additional variables (including the new contextual variables). Let also $\hat{\mathbf{R}}_s = [\mathbf{R}_s, \mathbf{J}_s\dot{\mathbf{X}}_s]$ be the new design matrix. We keep the same instrument matrix $\mathbf{J}_s\mathbf{G}_s^2\mathbf{X}_s$ for $\mathbf{J}_s\mathbf{G}_s\mathbf{y}_s$. We define $\hat{\mathbf{Z}}_s = [\mathbf{J}_s\mathbf{G}_s^2\mathbf{X}_s,\ \tilde{\mathbf{X}}_s,\ \mathbf{J}_s\dot{\mathbf{X}}_s]$, $\hat{\mathbf{R}}'\hat{\mathbf{Z}} = \sum_{s=1}^S \hat{\mathbf{R}}_s'\hat{\mathbf{Z}}_s$, $\hat{\mathbf{Z}}'\hat{\mathbf{Z}} = \sum_{s=1}^S \hat{\mathbf{Z}}_s'\hat{\mathbf{Z}}_s$, and $\hat{\mathbf{Z}}'\mathbf{y} = \sum_{s=1}^S \hat{\mathbf{Z}}_s'\mathbf{J}_s\mathbf{y}_s$. Let $\hat{\mathbf{\Gamma}}$ be the estimator of the coefficients associated with $\hat{\mathbf{R}}_s$. We have $\hat{\mathbf{\Gamma}} = ((\hat{\mathbf{R}}'\hat{\mathbf{Z}})(\hat{\mathbf{Z}}'\hat{\mathbf{Z}})^{-1}(\hat{\mathbf{R}}'\hat{\mathbf{Z}})')^{-1}(\hat{\mathbf{R}}'\hat{\mathbf{Z}})(\hat{\mathbf{Z}}'\hat{\mathbf{Z}})^{-1}(\hat{\mathbf{Z}}'\mathbf{y})$.

We also have $\mathbf{h}_s^{\eta} + \mathbf{h}_s^{\epsilon} - \lambda\mathbf{G}_s\mathbf{h}_s^{\eta} = \dot{\mathbf{X}}_s\check{\mathbf{\Gamma}}_0 + \hat{\boldsymbol{\mathcal{E}}}_s$ for some parameter $\check{\mathbf{\Gamma}}_0$, where $\hat{\boldsymbol{\mathcal{E}}}_s$ is an approximation

[2] In the literature on generalized additive models, a variable selection approach is used to eliminate irrelevant explanatory variables among the new regressors. This approach requires a penalty function and tuning parameters that are chosen using cross-validation. This goes beyond the scope of this paper because most cross-validation methods need $\mu_{0,s,i}^{out}$ and $\mu_{0,s,i}^{in}$ to be independent across i, which is not the case here. In our empirical analysis, there is not much difference between the results for $\zeta = 0$ and $\zeta = 10$. Consequently, we do not really need to care about the number of new regressors.

error owing to the B-spline approximation on the one hand and on the other, $\mu^{out}_{0,s,i}$ and $\mu^{in}_{0,s,i}$ being replaced with their estimators. The regulatory assumption we need for the asymptotic normality is $\sum_{s=1}^{S}\hat{\mathbf{Z}}'_s\hat{\boldsymbol{\mathcal{E}}}_s/\sqrt{n} = o_p(1)$. If this holds, then $\sqrt{n}(\hat{\mathbf{\Gamma}} - \mathbf{\Gamma}_0) \xrightarrow{d} \mathcal{N}\big(0, \lim_{n\to\infty} n\,\mathbb{V}(\hat{\mathbf{\Gamma}})\big)$, where $\mathbf{\Gamma}_0 = (\boldsymbol{\psi}'_0,\ \check{\mathbf{\Gamma}}'_0)'$ and $\lim_{n\to\infty} n\,\mathbb{V}(\hat{\mathbf{\Gamma}}) = \sigma^2_{\epsilon 0}\hat{\mathbf{B}}_0^{-1}\hat{\mathbf{D}}_0\hat{\mathbf{B}}_0^{-1}$. The matrices $\hat{\mathbf{B}}_0$ and $\hat{\mathbf{D}}_0$ are defined as the original $\mathbf{B}_0$ and $\mathbf{D}_0$, where $\mathbf{R}_s$ and $\mathbf{Z}_s$ are replaced by $\hat{\mathbf{R}}_s$ and $\hat{\mathbf{Z}}_s$.

S.2 Bayesian Estimation of the Network Formation Model

In the Bayesian approach, we assume that $\mu^{out}_{0,s,i}$ and $\mu^{in}_{0,s,i}$ are random effects following $\mathcal{N}(0,\ \sigma^2_{out})$ and $\mathcal{N}(0,\ \sigma^2_{in})$, respectively, with $\mathbb{E}(\mu^{out}_{0,s,i}\mu^{in}_{0,s,i}) = \rho_\mu$. To simulate the posterior distribution of $\mu^{out}_{0,s,i}$ and $\mu^{in}_{0,s,i}$, we use the data augmentation technique.[3]

Let $a^*_{s,ij} = \ddot{\mathbf{x}}'_{s,ij}\ddot{\boldsymbol{\beta}}_0 + \mu^{out}_{0,s,i} + \mu^{in}_{0,s,j} + u_{s,ij}$, 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 $\mathbf{a}_s$, $\ddot{\mathbf{X}}_s = [\ddot{\mathbf{x}}_{s,ij}; i \neq j]'$, $\ddot{\boldsymbol{\beta}}_0$, $\boldsymbol{\mu}^{out}_s = (\mu^{out}_{0,s,1}, \dots, \mu^{out}_{0,s,i})'$, and $\boldsymbol{\mu}^{in}_s = (\mu^{in}_{0,s,1}, \dots, \mu^{in}_{0,s,i})'$ is proportional to

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

where I(.) is the indicator function. This implies that the distribution of $a^*_{s,ij}|\mathbf{a}_s, \ddot{\mathbf{X}}_s, \ddot{\boldsymbol{\beta}}_0, \boldsymbol{\mu}^{in}_s, \boldsymbol{\mu}^{out}_s$ is $\mathcal{N}(\ddot{\mathbf{x}}'_{s,ij}\ddot{\boldsymbol{\beta}}_0 + \mu^{out}_{0,s,i} + \mu^{in}_{0,s,j},\ 1)$, 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^2_{in}, σ^2_{out}, and ρ_μ. Thus,

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

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

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

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

[3]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.

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

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

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

S.3 Additional Results on the Application

Tables S.1–S.3 present the estimation results after controlling for network endogeneity in our structural model. The unobserved attributes $\mu_{0,s,i}^{out}$ and $\mu_{0,s,i}^{in}$ are estimated using a logit model with fixed effects (see Yan et al., 2019) and a Bayesian random effect model (see OA S.2). We only present the results where the number of knots (No. knots) takes the values 0, 4, 5, and 6. Because the number of plugged variables implied by the cubic B-spline approach is large, we do not show the estimates of the coefficients associated with these variables. However, the line "Endo. Wald prob." in the tables indicates the p-value of the Wald test of the global significance of these variables.

Table S.1: Estimation results controlling for network endogeneity: fixed effect approach with B-spline approximations (full sample)

	Sandard model				Proposed structural model			
	No. knots: 0		No. knots: 10		No. knots: 0		No. knots: 10	
	Coef	Sd Err	Coef	Sd Err	Coef	Sd Err	Coef	Sd Err
Peer Effects	0.578	0.033	0.672	0.036	0.823	0.046	0.818	0.047
Own effects								
Female	0.170	0.006	0.173	0.006	0.169	0.006	0.169	0.006
Age	−0.021	0.003	−0.032	0.003	−0.043	0.003	−0.043	0.003
Hispanic	−0.102	0.010	−0.099	0.010	−0.092	0.010	−0.092	0.010
Race								
Black	−0.141	0.012	−0.123	0.012	−0.109	0.013	−0.107	0.014
Asian	0.215	0.013	0.210	0.013	0.194	0.014	0.194	0.014
Other	−0.023	0.011	−0.029	0.011	−0.031	0.011	−0.031	0.011
Lives with both parents	0.102	0.007	0.097	0.007	0.090	0.007	0.090	0.007
Years in school	0.030	0.003	0.029	0.003	0.024	0.003	0.024	0.003
Member of a club	0.184	0.012	0.151	0.013	0.144	0.013	0.150	0.014
Mother's education								
< High	−0.073	0.009	−0.068	0.009	−0.065	0.009	−0.064	0.009
> High	0.142	0.007	0.142	0.008	0.129	0.008	0.131	0.008
Missing	0.030	0.012	0.028	0.012	0.027	0.012	0.027	0.012
Mother's job								
Professional	0.036	0.009	0.035	0.009	0.031	0.009	0.030	0.009
Other	−0.040	0.007	−0.040	0.008	−0.039	0.008	−0.040	0.008
Missing	−0.077	0.011	−0.075	0.011	−0.071	0.011	−0.072	0.011
Contextual effects								
Female	−0.112	0.012	−0.108	0.012	−0.120	0.014	−0.117	0.014
Age	−0.051	0.003	−0.015	0.004	0.025	0.006	0.026	0.006
Hispanic	0.058	0.017	0.078	0.017	0.082	0.020	0.082	0.020
Race								
Black	0.015	0.016	0.048	0.017	0.072	0.020	0.076	0.020
Asian	−0.065	0.022	−0.087	0.023	−0.123	0.027	−0.121	0.028
Other	−0.040	0.020	−0.026	0.020	−0.002	0.022	−0.003	0.022
Lives with both parents	−0.035	0.016	−0.027	0.016	−0.014	0.018	−0.014	0.018
Years in school	0.017	0.004	0.003	0.005	−0.007	0.006	−0.007	0.006
Member of a club	−0.134	0.027	−0.110	0.027	−0.078	0.030	−0.094	0.031
Mother's education								
< High	−0.035	0.017	−0.008	0.017	0.024	0.019	0.024	0.019
> High	0.011	0.017	−0.012	0.018	−0.025	0.022	−0.024	0.022
Missing	−0.062	0.024	−0.049	0.024	−0.027	0.026	−0.026	0.026
Mother's job								
Professional	−0.053	0.018	−0.044	0.018	−0.034	0.019	−0.033	0.019
Other	−0.088	0.014	−0.059	0.014	−0.025	0.016	−0.026	0.016
Missing	−0.090	0.021	−0.047	0.022	0.007	0.024	0.006	0.024
σ_η^2					0.280		0.279	
σ_ϵ^2	0.505		0.510		0.057		0.059	
ρ					0.527		0.514	
Weak instrument F	157		130		119		116	
Endogeneity Wald prob.	0.000		0.000		0.000		0.000	
Sargan test prob.	0.000		0.020		0.421		0.454	

This table presents the estimation results of the proposed model after controlling for network endogeneity. The functions h_η and h_ϵ are approximated by cubic B-splines, where $\mu_{0,s,i}^{out}$ and $\mu_{0,s,i}^{in}$ are estimated using a logit model with individual fixed effects. The line "Endo. Wald prob." indicates the p-value of the Wald test of the significance of the plugged variables to control for endogeneity.

Table S.2: Estimation results controlling for network endogeneity: fixed effect approach with tensor products of B-spline approximations (full sample)

	Sandard model				Proposed structural model			
	No. knots: 0		No. knots: 10		No. knots: 0		No. knots: 10	
	Coef	Sd Err	Coef	Sd Err	Coef	Sd Err	Coef	Sd Err
Peer Effects	0.586	0.033	0.692	0.037	0.826	0.047	0.834	0.049
Own effects								
Female	0.17	0.006	0.174	0.006	0.169	0.006	0.171	0.006
Age	−0.022	0.003	−0.037	0.003	−0.043	0.003	−0.045	0.004
Hispanic	−0.102	0.01	−0.100	0.010	−0.091	0.010	−0.092	0.010
Race								
Black	−0.139	0.012	−0.113	0.013	−0.106	0.013	−0.102	0.014
Asian	0.213	0.013	0.206	0.013	0.194	0.014	0.192	0.014
Other	−0.024	0.011	−0.029	0.011	−0.031	0.011	−0.029	0.011
Lives with both parents	0.102	0.007	0.095	0.007	0.090	0.007	0.089	0.007
Years in school	0.03	0.003	0.028	0.003	0.024	0.003	0.024	0.003
Member of a club	0.185	0.012	0.154	0.013	0.147	0.013	0.157	0.014
Mother's education								
$<$ High	−0.073	0.009	−0.066	0.009	−0.065	0.009	−0.062	0.009
$>$ High	0.142	0.008	0.141	0.008	0.129	0.008	0.131	0.008
Missing	0.03	0.012	0.028	0.012	0.027	0.012	0.028	0.012
Mother's job								
Professional	0.036	0.009	0.034	0.009	0.031	0.009	0.031	0.009
Other	−0.04	0.007	−0.040	0.008	−0.039	0.008	−0.039	0.008
Missing	−0.076	0.011	−0.076	0.011	−0.071	0.011	−0.071	0.011
Contextual effects								
Female	−0.112	0.012	−0.107	0.012	−0.120	0.014	−0.120	0.014
Age	−0.049	0.003	−0.008	0.004	0.026	0.006	0.028	0.006
Hispanic	0.062	0.017	0.083	0.018	0.082	0.020	0.084	0.021
Race								
Black	0.017	0.016	0.046	0.017	0.071	0.020	0.082	0.020
Asian	−0.063	0.022	−0.095	0.023	−0.126	0.028	−0.127	0.028
Other	−0.037	0.02	−0.022	0.020	−0.002	0.022	−0.001	0.022
Lives with both parents	−0.035	0.016	−0.025	0.016	−0.015	0.018	−0.015	0.018
Years in school	0.017	0.004	0.003	0.005	−0.007	0.006	−0.006	0.006
Member of a club	−0.135	0.027	−0.106	0.027	−0.078	0.031	−0.094	0.032
Mother's education								
$<$ High	−0.032	0.017	−0.002	0.017	0.025	0.019	0.025	0.019
$>$ High	0.008	0.017	−0.016	0.018	−0.025	0.022	−0.030	0.023
Missing	−0.061	0.024	−0.046	0.024	−0.027	0.026	−0.025	0.026
Mother's job								
Professional	−0.053	0.018	−0.040	0.018	−0.034	0.019	−0.033	0.019
Other	−0.085	0.014	−0.050	0.015	−0.025	0.016	−0.022	0.016
Missing	−0.087	0.021	−0.036	0.022	0.008	0.024	0.010	0.024
σ_η^2					0.280		0.277	
σ_ϵ^2	0.506		0.511		0.055		0.056	
ρ					0.544		0.544	
Weak instrument F	115		127		115		110	
Endogeneity Wald prob.	0.000		0.000		0.000		0.000	
Sargan test prob.	0.000		0.123		0.324		0.486	

This table presents the estimation results of the proposed model after controlling for the network endogeneity. The functions h_η and h_ϵ are approximated by tensor products of cubic B-splines, where $\mu_{0,s,i}^{out}$ and $\mu_{0,s,i}^{in}$ are estimated using a logit model with individual fixed effects. The line "Endo. Wald prob." indicates the p-value of the Wald test of the significance of the plugged variables to control for endogeneity.

Table S.3: Estimation results controlling for network endogeneity: Bayesian random effect approach with B-spline approximations (full sample)

	Sandard model				Proposed structural model			
	No. knots: 0		No. knots: 10		No. knots: 0		No. knots: 10	
	Coef	Sd Err	Coef	Sd Err	Coef	Sd Err	Coef	Sd Err
Peer Effects	0.478	0.029	0.478	0.029	0.817	0.047	0.816	0.047
Own effects								
Female	0.178	0.006	0.178	0.006	0.167	0.006	0.167	0.006
Age	−0.016	0.003	−0.016	0.003	−0.044	0.003	−0.044	0.003
Hispanic	−0.101	0.01	−0.101	0.010	−0.091	0.010	−0.091	0.010
Race								
Black	−0.119	0.012	−0.120	0.012	−0.109	0.013	−0.109	0.013
Asian	0.217	0.013	0.217	0.013	0.194	0.014	0.194	0.014
Other	−0.033	0.011	−0.033	0.011	−0.033	0.011	−0.032	0.011
Lives with both parents	0.105	0.007	0.105	0.007	0.090	0.007	0.090	0.007
Years in school	0.031	0.003	0.031	0.003	0.025	0.003	0.025	0.003
Member of a club	0.168	0.012	0.167	0.012	0.152	0.012	0.151	0.012
Mother's education								
< High	−0.072	0.009	−0.072	0.009	−0.065	0.009	−0.064	0.009
> High	0.156	0.007	0.156	0.007	0.130	0.008	0.131	0.008
Missing	0.03	0.012	0.030	0.012	0.025	0.012	0.026	0.012
Mother's job								
Professional	0.036	0.009	0.036	0.009	0.031	0.009	0.031	0.009
Other	−0.044	0.007	−0.044	0.007	−0.040	0.008	−0.040	0.008
Missing	−0.081	0.011	−0.081	0.011	−0.073	0.011	−0.073	0.011
Contextual effects								
Female	−0.102	0.012	−0.101	0.012	−0.118	0.014	−0.117	0.014
Age	−0.072	0.004	−0.072	0.004	0.026	0.006	0.025	0.006
Hispanic	0.044	0.017	0.044	0.017	0.080	0.020	0.081	0.020
Race								
Black	−0.004	0.015	−0.004	0.016	0.071	0.019	0.070	0.019
Asian	−0.033	0.022	−0.033	0.022	−0.124	0.028	−0.122	0.028
Other	−0.045	0.02	−0.046	0.020	−0.004	0.022	−0.003	0.022
Lives with both parents	−0.034	0.016	−0.034	0.016	−0.012	0.018	−0.011	0.018
Years in school	0.029	0.004	0.029	0.004	−0.007	0.006	−0.007	0.006
Member of a club	−0.141	0.028	−0.140	0.028	−0.082	0.029	−0.084	0.029
Mother's education								
< High	−0.048	0.016	−0.049	0.016	0.022	0.019	0.022	0.019
> High	0.033	0.017	0.033	0.017	−0.023	0.022	−0.023	0.022
Missing	−0.063	0.024	−0.064	0.024	−0.026	0.026	−0.026	0.026
Mother's job								
Professional	−0.057	0.018	−0.057	0.018	−0.032	0.019	−0.031	0.019
Other	−0.109	0.014	−0.109	0.014	−0.025	0.016	−0.025	0.016
Missing	−0.117	0.021	−0.117	0.021	0.006	0.024	0.006	0.024
σ_η^2					0.280		0.279	
σ_ϵ^2	0.500		0.501		0.059		0.060	
ρ					0.508		0.506	
Weak instrument F	190		190		115		114	
Endogeneity Wald prob.	0.000		0.000		0.000		0.000	
Sargan test prob.	0.000		0.000		0.540		0.502	

This table presents the estimation results of the proposed model after controlling for the network endogeneity. The functions h_η and h_ϵ are approximated by cubic B-splines, where $\mu_{s,i}^{0,in}$ and $\mu_{0,s,i}^{out}$ are estimated using the Bayesian random effect model. The line "Endo. Wald prob." indicates the p-value of the Wald test of the significance of the plugged variables to control for endogeneity.

Table S.4 presents the estimation results for the data excluding "fully isolated" students and without controlling for network endogeneity. Model 3 is the standard linear-in-means model by approximating student effort by GPA, and Model 4 is based on our approach.

Table S.4: Estimation results without controlling for network endogeneity (sample excluding "fully isolated" students)

	Model 3'		Model 4'	
	Coef	Sd Err	Coef	Sd Err
Peer Effects	0.561	0.030	0.878	0.044
Own effects				
Female	0.182	0.006	0.165	0.007
Age	−0.008	0.004	−0.045	0.004
Hispanic	−0.096	0.011	−0.086	0.011
Race				
Black	−0.113	0.013	−0.102	0.015
Asian	0.199	0.014	0.173	0.015
Other	−0.030	0.011	−0.029	0.012
Lives with both parents	0.098	0.008	0.083	0.008
Years in school	0.032	0.003	0.023	0.003
Member of a club	0.169	0.013	0.150	0.013
Mother's education				
< High	−0.072	0.009	−0.062	0.009
> High	0.146	0.008	0.118	0.008
Missing	0.017	0.013	0.013	0.013
Mother's job				
Professional	0.040	0.009	0.034	0.010
Other	−0.035	0.008	−0.031	0.008
Missing	−0.070	0.012	−0.061	0.012
Contextual effects				
Female	−0.122	0.012	−0.127	0.013
Age	−0.082	0.004	0.028	0.006
Hispanic	0.049	0.017	0.086	0.021
Race				
Black	−0.010	0.017	0.055	0.021
Asian	−0.051	0.022	−0.129	0.028
Other	−0.040	0.020	−0.001	0.022
Lives with both parents	−0.047	0.016	−0.021	0.018
Years in school	0.032	0.004	−0.008	0.006
Member of a club	−0.160	0.028	−0.091	0.029
Mother's education				
< High	−0.043	0.016	0.026	0.019
> High	0.014	0.017	−0.036	0.021
Missing	−0.068	0.024	−0.031	0.026
Mother's job				
Professional	−0.062	0.018	−0.036	0.020
Other	−0.103	0.014	−0.021	0.016
Missing	−0.110	0.021	0.010	0.024
σ_η^2			0.292	
σ_ϵ^2	0.493		0.047	
ρ			0.485	
Weak instrument F	158.76		105.47	
Sargan test prob.	0.000		0.493	

Table S.5 presents the estimation results after controlling for network endogeneity using the data excluding "fully isolated" students.

Table S.5: Estimation results after controlling for network endogeneity (sample excluding "fully isolated" students)

	Sandard model				Proposed structural model			
	No. knots: 0		No. knots: 10		No. knots: 0		No. knots: 10	
	Coef	Sd Err	Coef	Sd Err	Coef	Sd Err	Coef	Sd Err
Peer Effects	0.649	0.036	0.728	0.039	0.846	0.048	0.854	0.050
Own effects								
Female	0.175	0.007	0.176	0.007	0.172	0.007	0.174	0.007
Age	−0.022	0.004	−0.035	0.004	−0.046	0.004	−0.047	0.004
Hispanic	−0.097	0.011	−0.094	0.011	−0.086	0.011	−0.086	0.011
Race								
Black	−0.116	0.014	−0.092	0.014	−0.074	0.015	−0.072	0.016
Asian	0.190	0.014	0.189	0.014	0.171	0.015	0.168	0.015
Other	−0.029	0.011	−0.031	0.011	−0.036	0.012	−0.034	0.012
Lives with both parents	0.093	0.008	0.087	0.008	0.082	0.008	0.081	0.008
Years in school	0.027	0.003	0.024	0.003	0.021	0.003	0.020	0.003
Member of a club	0.195	0.013	0.166	0.014	0.168	0.014	0.173	0.015
Mother's education								
< High	−0.068	0.009	−0.062	0.009	−0.058	0.009	−0.057	0.009
> High	0.135	0.008	0.134	0.008	0.125	0.008	0.128	0.008
Missing	0.015	0.013	0.015	0.013	0.013	0.013	0.014	0.013
Mother's job								
Professional	0.036	0.009	0.034	0.009	0.032	0.010	0.032	0.010
Other	−0.035	0.008	−0.035	0.008	−0.034	0.008	−0.034	0.008
Missing	−0.068	0.012	−0.067	0.012	−0.063	0.012	−0.063	0.012
Contextual effects								
Female	−0.126	0.012	−0.119	0.012	−0.126	0.014	−0.126	0.015
Age	−0.048	0.004	−0.008	0.004	0.031	0.006	0.032	0.006
Hispanic	0.064	0.018	0.078	0.018	0.082	0.021	0.083	0.021
Race								
Black	0.005	0.017	0.030	0.018	0.045	0.021	0.057	0.021
Asian	−0.071	0.023	−0.096	0.024	−0.118	0.028	−0.117	0.029
Other	−0.031	0.020	−0.021	0.020	−0.001	0.022	0.000	0.022
Lives with both parents	−0.039	0.016	−0.028	0.016	−0.017	0.018	−0.018	0.018
Years in school	0.019	0.005	0.005	0.005	−0.005	0.006	−0.005	0.006
Member of a club	−0.144	0.027	−0.115	0.027	−0.086	0.031	−0.102	0.032
Mother's education								
< High	−0.022	0.017	0.002	0.017	0.025	0.019	0.025	0.019
> High	−0.005	0.018	−0.021	0.018	−0.030	0.022	−0.035	0.023
Missing	−0.058	0.024	−0.044	0.024	−0.027	0.026	−0.027	0.026
Mother's job								
Professional	−0.055	0.018	−0.044	0.018	−0.036	0.019	−0.034	0.019
Other	−0.078	0.014	−0.049	0.014	−0.024	0.016	−0.022	0.016
Missing	−0.073	0.021	−0.031	0.022	0.010	0.024	0.011	0.024
σ^2_η					0.289		0.285	
σ^2_ϵ	0.498		0.504		0.058		0.059	
ρ					0.396		0.395	
Weak instrument F	122		108		98		94	
Endogeneity Wald prob.	0.000		0.000		0.000		0.000	
Sargan test prob.	0.000		0.337		0.718		0.820	