Estimating Peer Effects on Career Choice: A Spatial Multinomial Logit Approach

Bolun Li [*], Robin Sickles [†], and Jenny Williams[‡§]

Dated: January 20, 2019

Abstract

Peers and friends are among the most influential social forces affecting adolescent behavior. In this paper we investigate peer effects on post-high school career decisions and on school choice. We define peers as students who are in the same classes and social clubs and measure peer effects as spatial dependence among them. Utilizing recent development in spatial econometrics, we formalize a spatial multinomial choice model in which individuals are spatially dependent in their preferences. We estimate the model with data from the Texas Higher Education Opportunity Project. We do find that individuals are positively correlated in their career and college preferences and examine how such dependencies impact decisions directly and indirectly as peer effects are allowed to reverberate through the social network in which students reside.

JEL Classification: C31, C35
Keywords: spatial models, peer effect, random utility, multinomial logit models

[*]Department of Economics, Rice University. Email: bolun.li@rice.edu
[†]Department of Economics, Rice University. Email: rsickles@rice.edu
[‡]Department of Economics, University of Melbourne, Email: jenny.williams@unimelb.edu.au
[§]This research uses data from THEOP, a project directed by Marta Tienda, at Princeton University, in collaboration with Teresa A. Sullivan, formerly at the University of Texas at Austin. This research was supported by grants from the Ford, Mellon and Hewlett Foundations and NSF (GRANT # SES-0350990). We gratefully acknowledge institutional support from Princeton University's Office of Population Research (NICHD Grant # R24 H0047879). Special acknowledgment is due to Dawn Koffman, for preparing and documenting data for public use, and to Sunny X. Niu, who has been a stalwart collaborator and co-investigator.

1 Introduction

Peers and friends are among the most influential social forces affecting adolescent behavior. In this paper, we investigate peer effects on post-high school decisions. There is a substantial body of literature that studies peer influences on educational achievement (Ding & Lehrer, 2007; Lin, 2010; Fletcher & Tienda, 2009; Neidell & Waldfogel , 2010; Zimmerman, 2003). Yet little is known about how peers influence post school decisions, such as whether to attend college or not, and career choices. The few studies that investigate peer effects on college attendance typically measure peer effects as the proportion of high-school friends who intend to go to college (Alvarado & Turley, 2012; Arcidiacono & Nicholson, 2005; Fletcher, 2010; Lyle, 2007). However, these studies are silent on how the preferences of friends are formally linked and what mechanism best explains these links.

Utilizing developments in spatial econometrics (Anselin, 2002; Calabrese & Elkink, 2014; Chakir & Parent, 2009; Smirnov, 2010; Smirnov & Egan, 2012), we define peer effects as spatial dependence among individuals. In particular, we define peers as students who are in the same classes and social clubs. We develop and estimate a spatial multinomial logit choice model in which individuals' latent utilities are linked through a spatial weight matrix. In doing so, we are able to account for interdependency among individuals, and hence their related preferences and choices. Our model is estimated using detailed information on high school seniors and their post-school choices from the Texas Higher Education Opportunity Project (THEOP).

The remainder of the paper is organized as follows. In Section 2, we introduce a spatial multinomial choice model that allows interdependence among individual preferences. Section 3 discusses details regarding our estimation strategy, including the calculation of indirect and indirect effects. In Section 4, we describe the data that we apply to our model. The estimation results are presented in Section 5, leading to our conclusions in Section 6.

2 Model

In this section, we outline our approach to measuring peer effects. We first define and construct the peer effects among high school seniors. Next, we discuss the spatial multinomial logit choice model which we use to model the interdependencies in peer choices.

2.1 Measuring Peer Effects

Peer effects are understood as common classes and social clubs which each high school senior i shares with her peers in high school s. We are interested in whether individuals decide to attend college or select a profession, given that they are linked in space. We estimate these different sets of decisions in two separate analyses.[1]

To measure peer effects, we adapt a spatial econometrics approach. Let N^s denote the set of high school seniors in high school s and $|N^s| \equiv n^s$ be the total number of seniors in high school s. We define a symmetric square matrix A^s whose entity a^s_{ij} is the total number of classes and social clubs student i and j have in common in high school s. The dimension of A^s is the $n^s \times n^s$. We construct the spatial weight matrix W^s that measures the peer effect in high school s in the following way: w^s_{ij} is an element of W^s defined as

$$w^s_{ij} = \begin{cases} a^s_{ij}, & \text{if} \quad i \neq j \\ 0, & \text{if} \quad i = j. \end{cases}$$

Notice that the matrix W^s has zero diagonal elements, as we do not allow for an individual to spatially affect herself. In addition, W^s is symmetric, i.e., $w^s_{ij} = w^s_{ji}$, implying the peer effect between any two individuals is symmetric.[2] The cardinal value of a^s_{ij} implies the ordinal closeness among friends.

[1] We assume that W is exogenous to the model and chosen by the researcher. However, W is possible to be endogenous if the friendship formation among individuals is taken into consideration (Kelejian & Piras, 2014; Qu & Lee, 2015). We do not model network formation in this paper.

[2] In what follows, the high school index superscript s is suppressed to ease notational complexity.

2.2 Spatial Random Utility

Let C_i denote the choice set of individual i. We assume that everyone faces the same choice set, i.e., $C_i = C$, and specify spatial dependence as a relationship among individual preferences linked in space via a linear utility specification. The latent utility of individual i choosing alternative j is thus given by:

$$u_{ij} = \rho \sum_{k=1}^{n} w_{ik} u_{kj} + v_j(\beta) + \varepsilon_{ij}.$$

The deterministic utility component of each alternative is denoted by $v_j(\beta)$. The parameters of interest are ρ and β. Idiosyncratic shocks ε_{ij} are introduced to form the random utility structure. The model includes a spatial lag vector $\rho \sum_{k=1}^{n} w_{ik} u_{kj}$ which represents the linear combination of values of the latent dependent variable vector from neighboring observations. Because we allow for endogenous interaction and feedback effects, in the language of of LeSage (2014), the model considered in this paper falls into the category of a global spillover specification.

In addition, we assume ε_{ij} follows a Type I Extreme Value (TIEV) distribution. Let y_{ij} be an observed decision. The decision rule for individual i is:

$$y_{ij} = \begin{cases} 1, & \text{if } u_{ij} \geq u_{il}, \forall j, l \in C \\ 0, & \text{otherwise ,} \end{cases} \tag{1}$$

and thus we allow for one and only one alternative to be chosen. By stacking the equations, we can write latent preferences as:

$$u_j = \rho W u_j + v_j(\beta) + \varepsilon_j, \tag{2}$$

where $u_j = (u_{1j}, u_{2j}, ..., u_{nj})'$ and $\varepsilon_j = (\varepsilon_{1j}, \varepsilon_{2j}, ..., \varepsilon_{nj})'$. Equation (1) and (2) together with the distributional assumption on the error terms fully describe the basic spatial discrete choice model.[3]

3 Estimation

This section describes the statistical model derived from the utility structures in Section 2.2. We proceed in the following steps. First, we collect terms involving u in Equation (2) and premultiply by $Z \equiv (I - \rho V)^{-1}$, where V is the row normalized matrix of W. Row standardization is required because for larger schools with more clubs and classes, the possibility that the same level of engagement with peers could be distorted among different schools simply because the students had more options. One could also impose column standardization to achieve the same goal. Reduced form preferences thus become:

$$u_j = Z(\rho)v_j(\beta) + Z(\rho)\varepsilon_j. \tag{3}$$

Recalling that $Z \equiv (I - \rho V)^{-1}$ and assuming convergence,[4] we have

$$\begin{aligned} Z(\rho) &= \sum_{k=0}^{\infty}(\rho V)^k \\ &= D(\rho) + (Z(\rho) - D(\rho)), \end{aligned}$$

where $D(\rho)$ is the $n \times n$ matrix with the diagonal elements of $Z(\rho)$, which we refer to as the "private" or "direct" effect, while $Z(\rho) - D(\rho)$ is the so-called "social" or indirect effect. Equation (3) then can be written as:

$$u_j = Z(\rho)v_j(\beta) + D(\rho)\varepsilon_j + (Z(\rho) - D(\rho))\varepsilon_j.$$

[3] A richer model also could include the spatial impact δW on deterministic utility components $v_j(\beta)$. However, we do not consider it in this paper given the substantial complexity and nonlinearity of the current model.

[4] Convergence is ensured by the identification conditions of the spatial discrete choice model. See the detailed discussion in Smirnov (2010).

We next make a behavioral assumption that individuals ignore social shocks $(Z(\rho) - D(\rho))\varepsilon$ when making decisions.[5] That is:

$$u_j = Z(\rho)v_j(\beta) + D(\rho)\varepsilon_j. \tag{4}$$

Under the distributional assumption that ε follows the Type I Extreme Value distribution, the conditional probability of individual i choosing alternative j has the closed form expression:

$$p_{ij} = \frac{\exp(\sum_{k=1}^{n} z_{ik}(\rho)v_{kj}(\beta)/d_{ii}(\rho))}{\sum_{j\in J}\exp(\sum_{k=1}^{n} z_{ik}(\rho)v_{kl}(\beta)/d_{ii}(\rho))}.$$

Together with Equation (1), the log-likelihood is:

$$LL(\beta, \rho \mid y, X, W) = \sum_{i=1}^{n}\sum_{j=1}^{J} y_{ij}\ln(p_{ij}). \tag{5}$$

The pseudo maximum likelihood estimates (PMLE) are the values of parameters (ρ, β) that maximize LL in equation (5). Several remarks are worth mentioning. Equation (1) and (4) comprise an auxiliary form of the true model comprised by equations (2) and (1). Both models have the same observed deterministic components of individual random utilities. However, error terms in the auxiliary model are assumed to be independent. Since some of the information about effects of individual interdependence is not present in the auxiliary model, parameter estimates are not necessarily asymptotically efficient.

Another way to estimate spatial multinomial choice models is to integrate out the error dependency structure. Baltagi, LeSage, & Pace (2016) provide a detailed discussion of the methodology and computation feasibility of this method. Although conceptually straightforward, numerical integration introduces a computational burden. Moreover, the parameters

[5] This assumption is violated however, when a student has strong preference over a certain school because she lacks peers at that school. We are reminded of an anecdote relayed to us by Kelley Pace about a student he knew who had applied to a particular college because she did NOT know anyone there.

of interest in this paper, spatial effect ρ and utility parameters β, can be easily recovered without numerical integration.

Because we are interested in how peers influence choices, we define choice sets $C_{college} \equiv$ {attending college, not attending college}, if she faces a binary choice, and $C_{career} \equiv$ {attending college, working, serving military, staying at home}, if she faces multiple choices.[6] Let P denote the total number of individual characteristics. The deterministic utility component is specified as:

$$v_j(\beta_j) = x_i'\beta_j + \delta_j \tag{6}$$

where $x_i = (x_{i1}, x_{i2}, ..., x_{iP})$ denotes individual characteristics and $\beta_j \equiv (\beta_{j1}, \beta_{j2}, ..., \beta_{jP})$ and $\delta_j \equiv (\delta_1, \delta_2, ..., \delta_J)$ are parameter of interest. The marginal utility of each characteristic is alternative specific and thus is indexed by j. Substituting Equation (6) into Equation (4) completes the model.

3.1 Direct & Indirect Impacts

Because the impact of changes in an explanatory variable differs over all observations, this section provides a measure of these varying impacts. The average total impact for coefficient β_{jp} is the average of all derivatives of the latent utility u_{ij} with respect to x_{ip} for any i. The average direct impact for coefficient β_{jp} is the average of all own derivatives. Naturally, the average indirect impact is defined as the difference between average total impact and average direct impact. Formally, the average total effect (TE), average direct effect (DE)

[6] In the case where an individual is working and attending school, her choice is defined as school if she is working part time. Staying at home is served as base and hence parameters β_1 and δ_1 are normalized to 0 for the sake of identification.

and average indirect effect (IE) are defined as:

$$\text{TE}(\beta_{jp}) = \frac{1}{N} l' \big[(I - \rho V)^{-1} \beta_{jp} \big] l \tag{7}$$

$$\text{DE}(\beta_{jp}) = \frac{1}{N} tr\big((I - \rho V)^{-1} \beta_{jp} \big) \tag{8}$$

$$\text{IE}(\beta_{jp}) = \text{TE}(\beta_{jp}) - \text{DE}(\beta_{jp}) \tag{9}$$

where l is a $N \times 1$ vector of one's. It is worth noting that DE (β_{jp}) equals β_{jp} because $\frac{1}{N} tr\big((I - \rho V)^{-1} \beta_{jp}\big) = \frac{1}{N} tr\big((I - \rho V)^{-1}\big) \beta_{jp} = \beta_{jp}$.[7]

4 Data

We estimate our model using data from the Texas Higher Education Opportunity Project (THEOP) (Tienda & Sullivan, 2011).[8] The THEOP is accessible via the Princeton University Office of Population Research (OPR) data archive and requires confidentiality protocols that are rather extensive. It is a multi-year research evaluation study initiated by the Ford Foundation and undertaken by Princeton University's Office of Population Research. The centerpiece of THEOP is a two-cohort longitudinal survey of sophomores and seniors who were enrolled in Texas public high schools as of spring, 2002. Our estimation sample is drawn from the senior cohort only, which is a sample of 13,803 high school seniors attending 96 Texas public high school. Students were randomly selected and surveyed during their last semester in high school and data were collected through a self-administered survey. Our

[7] We have also conducted a check on our methods by conducting a small Monte Carlo simulation. We proceeded in two steps. We allowed the total number of possible choices to be four and the number of different individual characteristics associated with each choice to be 5. We drew 5000 identically and independently distributed idiosyncratic shocks from the TEIV distribution and specified a symmetric and zero diagonal spatial weight matrix. After substituting the true parameter values, covariates, spatial weights, and idiosyncratic shocks into the utility function, we solved for the optimal decision for each observation and, given optimal decisions, individual characteristics and the spatial weight matrix, we used the estimation procedure just outlined to estimate model primitives with the data generated in the previous step. Simulation results based on 500 replications are available on request. Estimated parameters are close to the true parameter values and we are confident that our estimation algorithms are correct.

[8] Papers based on the public and restricted use data are posted on the THEOP Web site at http://theop.princeton.edu/publications/

estimation sample comprises 3125 seniors who responded to the self-administered survey.

Of relevance to this study, the THEOP provides a unique measure of peer effects. The survey asks each student what classes and social clubs she belongs to. We define peers as students who share common classes and social clubs. The THEOP also collects data on students' post high school activities in the second wave, which is used to construct the outcome variables for our analysis. We construct two outcomes. The first is a binary indicator for attending college. This variable is coded as one for all respondents who report attending college, either part or full time, in the follow-up survey. The second outcome is a multinomial measure for post high school career choice. This variable is coded as 1 if the respondent's primary activity is attending college, 2 if the respondent's primary activity is working, 3 if the respondent joins the military, and 4 if they are not attending college, not working and not in the military. Table 1 shows the sample distributions of the outcomes we study. 81% of our sample attend college either part or full time, while 19% report that they do not attend college at all. In terms of their primary activity, 70% of respondents report their primary activity is attending college, 15% report that their primary activity is working, 5% report being int he military, and 10% are not attending college, are not working, and are not in the military (are at home).

In addition, THEOP contains high school characteristics and individual demographics. Table 2 presents the sample descriptive statistics used in estimation. Figures **??** and **??** present undirected networks of peers in for two different Texas high schools. It is clear from these two examples, which are representative of the other Texas high school peer patterns, that such network linkages are pronounced and that there is variation in these networks across the different schools in our sample.

5 Estimation Results

5.1 Binary Models for College Attendance

Table 3 and Table 4 present maximum likelihood estimates of the coefficients of the binary choice logit model, and the spatial binary logit model for attending college. The impact of changes in explanatory variable for the logit model (both average partial effects (APE) and partial effects evaluated at sample averages (PEA)) are also presented in Table 3, and direct and indirect effects of explanatory variables for the spatial logit model are presented in Table 5.

Overall, the first point to note is that the spatial model finds significant evidence of peer effects, as evidenced by the positive and significant coefficient estimate on the peer effect term, ρ. Nonetheless, the coefficient estimates from the binary logit and the spatial binary logit model show general agreement in terms of the characterisitcs of the individual and their parents that are significant determinants of seniors' decision to go to college. In particular, the probability of attending college is lower for females, for seniors who are older, and who perform worse academically as measured by their percentile class rank, and for those whose family rents rather then owns their home. As expected, the probability of attending college is increasing in parental education, although only the education of the father (and not the mother) is significant in the spatial model that accounts for peer effects.

School characteristics are also important determinants of student choices, however, the estimated effects differ when peer effects are accounted for. For example, after accounting for peer effects, the likelihood that an individual attends college is increasing in the college attendance rate of the high school they attend, and decreasing in the dropout rate of the high school they attend whereas,in the binary logit model that does not account for peer effects, the percent attending college or percent dropping out of high school do not impact on individuals decision to attend college. Similarly, high school level variables related to the availability of advanced placement (AP) courses, the percent of students taking AP courses and the percent passing AP course are significantly and positively related to the individual attending college after accounting for peer effects, whereas only whether AP courses are offered is significantly related to attending college in the simple logit model. Finally, after accounting for peer effects, the probability of attending college is decreasing in the distance to a 4 year college and and the distance to a private college, whereas the estimates from the simple logit imply the opposite – that the probability of attending college is increasing in distance to four year and private colleges.

In terms of magnitudes of impacts of individual, family and school characteristics, the spatial model that accounts for peer effects implies much larger effects than the simple logit model, as a comparison of the APE (or PEA) in Table 3 with the direct and indirect effects (the sum of which produce the total partial effect) reported in Table 5 reveals. For example, the simple logit model estimates imply a partial effect of father's education on college attendance of 2 percentage points whereas the spatial model estimates imply an total effect of 14 percentage points, comprised of a 10 percentage point direct effect and a 4 percentage point indirect effect. Similarly, the logit model estimates imply a 4 percentage point reduction in the probability of attending college for seniors whose family live in a rented accommodation compared to those whose family own their home, whereas the coefficient estimates from the spatial model imply a total reduction in the probability of attending college of 60 percentage points, comprised of a 43 percentage point reduction due to the direct effect and a 17 percentage point reduction due to the indirect effect.

5.2 Mulinomial Models for Post High School Choice

Table 6 and Table 9 provide maximum likelihood coefficient estimates of the model of post high school choice, where senior choose to either attend college, to work, to join the military, or to stay home (where stay at home is the base category). Table 7 and Table 8 report partial effects of the explanatory variables for the Multinomial Logit Model and Table 10 reports direct and indirect partial effects from the spatial multinomial model.

As with the binary spatial model for college attendance, the multinomial spatial model provides evidence of the importance of high school peer effects in post school decisions.

Specifically, the coefficient on the peer effect term, ρ, is statistically significant and positive. In terms of the impact of individual, family and school characteristics, the point estimates are quite similar for the simple multinomial and spatial multinomial models, although there are differences in the precision with which the coefficients are estimated. For example, in the spatial multinomial model, all else being equal, females are statistically significantly more likely than males to attend college, to work, and to join the military, whereas the coefficients estimates are significant for working and joining the military only in the simple multinomial model.

Irrespective of the model estimated, however, very few of the individual, family, or school characteristics predict joining the military after high school. And while individual academic performance as measured by percentile rank in class, parents education, and having a family who rents their home are important determinants of attending college in both multinomial and spatial multinomial models, school characteristics have differing effects. In the spatial multinmial model, for example, the percentage of students from the high school attending college increases the likelihood of the individual attending college, while greater distances to 4 and 2 year colleges reduce it. In contrast, estimates from the simple multinomial model indicate that the probability of attending college is higher for students attending high schools a greater distance from a four year college, and lower for students from high schools with a higher proportion of low income students. Similarly, the estimates from the spatial multinomial logit indicate that the probability of choosing to work following high school is statistically significantly lower amongst individuals who attended high schools further away from 4 year colleges and with fewer nonwhites enrollments, whereas the simple multinomial logit estimates indicate a statistically significantly *higher* probability of working for students from high schools further away from 4 year colleges and with fewer nonwhites enrollments.

In terms of the magnitude of effects, Table 10 reports direct and indirect effects for the spatial model that accounts for peer effects and shows that the direct effects are typically around twice the magnitude of the indirect effects.

6 Conclusion

Our study has developed a spatial model of peer decision making based on the random utility paradigm and has applied this econometric model to a unique data set of High School Seniors in Texas, We have developed both binary and multinomial logit versions of the random utility model utilizing parametric specifications based on standard treatments used in this literature. We also have considered a spatially autoregressive generalization of the classical random utility model to account for peer effects within Texas high schools. Allowing interdependence among individual preferences requires us to develop a new estimation strategy and interpretation of marginal effects that include the calculation of direct and indirect effects. Our estimation results clearly point to the importance of peer effects in this age group of young people and that such peer effects, if ignored, may distort our understanding of the determinants of the decisions we model, both in terms of magnitudes and in terms of directions of effects.

References

Alvarado, S. E., & Turley, R. N. L. (2012). College-bound friends and college application choices: Heterogeneous effects for Latino and white students. *Social Science Research*, 41(6), 1451-1468.

Anselin, L. (2002). Under the hood issues in the specification and interpretation of spatial regression models. *Agricultural Economics*, 27(3), 247-267.

Arcidiacono, P., & Nicholson, S. (2005). Peer effects in medical school. *Journal of Public Economics*, 89(2), 327-350.

Baltagi, B. H., LeSage, J. P., & Pace, R. K. (Eds.). (2016). *Spatial econometrics: qualitative and limited dependent variables.* Emerald Group Publishing.

Calabrese, R., & Elkink, J. A. (2014). Estimators of binary spatial autoregressive models: A Monte Carlo study. *Journal of Regional Science*, 54(4), 664-687.

Chakir, R., & Parent, O.(2009). Determinants of land use changes: A spatial multinomial probit approach. *Papers in Regional Science*, 88(2), 327-344.

Ding, W., & Lehrer, S. F. (2007). Do peers affect student achievement in China's secondary schools?. *The Review of Economics and Statistics*, 89(2), 300-312.

Fletcher, J. M. (2010). Peer Influences on College Choices: New Evidence from Texas.

Fletcher, J. M., & Tienda, M. (2009). High school classmates and college success. *Sociology of Education*, 82(4), 287-314.

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

LeSage, J. (2014). What regional scientists need to know about spatial econometrics.

Lin, X. (2010). Identifying peer effects in student academic achievement by spatial autoregressive models with group unobservables. *Journal of Labor Economics*, 28(4), 825-860

Lyle, D. S. (2007). Estimating and interpreting peer and role model effects from randomly assigned social groups at West Point. *The Review of Economics and Statistics,* 89(2), 289-299.

Neidell, M., & Waldfogel, J. (2010). Cognitive and noncognitive peer effects in early education. *The Review of Economics and Statistics*, 92(3), 562-576.

Qu, X., & Lee, L. F. (2015). Estimating a spatial autoregressive model with an endogenous spatial weight matrix. *Journal of Econometrics*, 184(2), 209-232.

Smirnov, O. A. (2010). Modeling spatial discrete choice. *Regional Science and Urban Economics*, 40(5), 292-298.

Smirnov, O. A., & Egan, K. J. (2012). Spatial random utility model with an application to recreation demand. *Economic Modeling*, 29(1), 72-78.

Tienda, M., & Sullivan, T. (2011). Texas Higher Education Opportunity Project. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2011-06-02. https://doi.org/10.3886/ICPSR29841.v1

Zimmerman, D. J. (2003). Peer effects in academic outcomes: Evidence from a natural experiment. *The Review of Economics and Statistics*, 85(1), 9-23.

Table 1: Occupational Choice

	College	Work	Military	Home
Number of Observation	2189	463	146	327
(Proportion)	(70%)	(15%)	(5%)	(10%)

	College	Not College
Number of Observation	2523	602
(Proportion)	(81%)	(19%)

[a] Total number of observations is 3125

Table 2: Descriptive Statistics

	Mean	Std. Dev.
Gender	0.45	0.49
Age	18.42	0.59
House owner	0.14	0.35
Father education	5.33	1.98
Mother education	5.18	1.85
Rank	39.29	48.98
HS College percentage	75.83	19.29
HS dropout rate	1.55	1.05
Low income	31.43	22.78
AP course offered	13.56	5.53
AP taking percentage	13.19	6.66
AP passing percentage	52.02	23.60
Total Enrollment	2393.42	1176.66
Distance 4yr college	10.13	8.89
Distance 2yr college	12.94	15.06
Distance private college	51.63	67.61
HS mean algebra score	1411.54	272.44
Nonwhite	1386.26	1192.10

[a] HS stands for high school.
[b] Gender equals to 1 if female.
[c] The variable House owner is an indicator that equals to 1 if a household owns.
[d] The variable Father/Mother education: 1-Illiterate/Semi-literate, 2-Elementary school, 3-Middle school, 4-High school, 5-some college, 6-2 year college, 7-4 year college, 8-Master degree, 9-professional school
[e] The variable Rank is class ranking of high school students indicated in percentile. The higher the rank, the lower the percentage.
[f] The variable HS college percentage is the percentage of students plan to attend college in each high school.
[g] The variable Low income is the percentage of low income families in each high school.
[h] Nonwhite is the total number of nonwhite students in each high school.
[i] Total number of observation is 3125

Table 3: Binary Logit Model

y = 1 if attending college

	β	APE	PEA
Gender	-0.37***	-0.05***	-0.05***
Age	-0.14*	-0.02**	-0.02**
House owner	-0.31**	-0.04***	-0.04***
Father education	0.14***	0.02***	0.02***
Mother education	0.23***	0.03***	0.03***
Rank	-0.01***	-1.36e-3***	-1.33e-3***
HS college percentage	0.01	2.63e-4	2.58e-4
HS dropout rate	-0.03	-4.62e-3	-4.53e-3
Low income	-0.02***	-2.01e-3***	-1.97e-3***
AP course offered	0.03*	3.81e-3	3.74e-3
AP taking percentage	-1.05e-3	-1.42e-4	-1.38e-4
AP passing percentage	6.39e-4	8.60e-5	8.44e-5
Total Enrollment	-8.12e-5	-1.09e-5	-1.07e-5
Distance to 4yr college	0.02***	2.56e-3***	2.51e-3***
Distance to 2yr college	-0.01***	-1.75e-3***	-1.72e-3***
Distance to private college	2.62e-3**	3.53e-4***	3.46e-4***
HS mean algebra	1.51e-4	2.03e-5	1.99e-5
Nonwhite	1.19e-4	1.61e-5	1.58e-5
Constant	2.89**	-	-

[a] HS stands for high school.
[b] Gender equals to 1 if female.
[c] The variable House owner is an indicator that equals to 1 if a household rents.
[d] The variable Father/Mother education: 1-Illiterate/Semi-literate, 2-Elementary school, 3-Middle school, 4-High school, 5-some college, 6-2 year college, 7-4 year college, 8-Master degree, 9-professional school
[e] The variable Rank is class ranking of high school students indicated in percentile. The lower the percentage, the higher this person is ranked and therefore the coefficient is negative.
[f] The variable HS college percentage is the percentage of students plan to attend college in each high school.
[g] The variable Low income is the percentage of low income families in each high school.
[h] Nonwhite is the total number of nonwhite students in each high school.

Table 4: Spatial Binary Logit Model: Point Estimates

y = 1 if attending college	
Peer effect ρ	0.29**
	β
Gender	-0.19***
Age	-0.14*
House owner	-0.43**
Father education	0.10***
Mother education	-0.13
Rank	-0.07***
HS college percentage	0.01**
HS dropout rate	-0.04**
Low income	-0.03*
AP course offered	0.02*
AP taking percentage	0.02*
AP passing percentage	3.30e-3*
Total Enrollment	2.80e-4
Distance to 4yr college	-0.03***
Distance to 2yr college	9.80e-3
Distance to private college	-4.21e-3**
HS mean algebra	1.92e-4
Nonwhite	-3.00e-4
Constant	2.74**

[a] HS stands for high school.
[b] Gender equals to 1 if female.
[c] The variable House owner is an indicator that equals to 1 if a household rents.
[d] The variable Father/Mother education: 1-Illiterate/Semi-literate, 2-Elementary school, 3-Middle school, 4-High school, 5-some college, 6-2 year college, 7-4 year college, 8-Master degree, 9-professional school
[e] The variable Rank is class ranking of high school students indicated in percentile. The lower the percentage, the higher this person is ranked and therefore the coefficient is negative.
[f] The variable HS college percentage is the percentage of students plan to attend college in each high school.
[g] The variable Low income is the percentage of low income families in each high school.
[h] Nonwhite is the total number of nonwhite students in each high school.

Table 5: Spatial Binary Logit Model: Direct & Indirect Effect

	DE	IE
Gender	-0.19***	-0.08***
Age	-0.14*	-0.06**
House owner	-0.43**	-0.17**
Father education	0.10***	0.04**
Mother education	-0.13	-0.05
Rank	-0.07***	-0.03***
HS college percentage	0.01**	4.05e-3**
HS dropout rate	-0.04**	-0.02*
Low income	-0.03*	-0.01*
AP course offered	0.02*	8.11e-3*
AP taking percentage	0.02*	8.01e-3
AP passing percentage	3.30e-3*	1.34e-3*
Total Enrollment	2.80e-4	1.13e-4
Distance to 4yr college	-0.03***	-0.01
Distance to 2yr college	9.80e-3	3.97e-3*
Distance to private college	-4.21e-3**	-1.71e-3**
HS mean algebra	1.92e-4	7.71e-5
Nonwhite	-3.00e-4	1.21e-4*
Constant	2.74**	1.11**

[a] HS stands for high school.
[b] Gender equals to 1 if female.
[c] The variable House owner is an indicator that equals to 1 if a household rents.
[d] The variable Father/Mother education: 1-Illiterate/Semi-literate, 2-Elementary school, 3-Middle school, 4-High school, 5-some college, 6-2 year college, 7-4 year college, 8-Master degree, 9-professional school
[e] The variable Rank is class ranking of high school students indicated in percentile. The lower the percentage, the higher this person is ranked and therefore the coefficient is negative.
[f] The variable HS college percentage is the percentage of students plan to attend college in each high school.
[g] The variable Low income is the percentage of low income families in each high school.
[h] Nonwhite is the total number of nonwhite students in each high school.

Table 6: Multinomial Logit Model

y = 1 if attending college
y = 2 if working
y = 3 if serving in military
y = 4 if staying at home (base)

	College β_1	Work β_2	Military β_3
Gender	0.07	0.69***	1.86***
Age	-0.03	0.24***	-0.17
House owner	-0.31**	-0.32**	0.03
Father education	0.14***	-0.05	0.09
Mother education	0.20***	0.06	0.04
Rank	-1.35e-3***	-1.49e-3	-1.9e-3*
HS college percentage	4.22e-3	7.94e-3**	2.37e-3
HS dropout rate	-0.02	-0.16**	-0.03
Low income	-0.01*	-5.94e-3	-7.40e-4
AP course offered	0.02	3.26e-4	-0.04
AP taking percentage	-8.51e-3	-5.55e-3	-.02
AP passing percentage	2.56e-3	-5.58e-3	3.51e-4
Total Enrollment	-5.50e-5	1.99e-4	7.65e-5
Distance to 4yr college	0.02***	0.03**	0.02
Distance to 2yr college	-0.02***	-3.09e-3	-0.02*
Distance to private college	1.21e-3	-3.32e-3**	-1.762e-3
HS mean algebra	-4.98e-4	-5.15e-4	-5.72e-4
Nonwhite	1.14e-4	-1.46e-4	1.52e-4
Constant	2.35	-3.24*	2.13

[a] HS stands for high school.
[b] Gender equals to 1 if female.
[c] The variable House owner is an indicator that equals to 1 if a household rents.
[d] The variable Father/Mother education: 1-Illiterate/Semi-literate, 2-Elementary school, 3-Middle school, 4-High school, 5-some college, 6-2 year college, 7-4 year college, 8-Master degree, 9-professional school
[e] The variable Rank is class ranking of high school students indicated in percentile. The lower the percentage, the higher this person is ranked and therefore the coefficient is negative.
[f] The variable HS college percentage is the percentage of students plan to attend college in each high school.
[g] The variable Low income is the percentage of low income families in each high school.
[h] Nonwhite is the total number of nonwhite students in each high school.

Table 7: Multinomial Logit Model: Average Partial Effect

VARIABLES	(1) Gender	(2) Age	(3) House owner	(4) Father education	(5) Mother education	(6) Rank	(7) HS college percentage	(8) HS dropout rate	(9) Low income	(10) AP course offered	(11) AP taking percentage	(12) AP Passing percentage	(13) Total enrollment	(14) Distance to 4 yr college	(15) Distance 2yr college	(16) Distance to private college	(17) HS mean algebra	(18) Nonwhite
y = 1 if College	-0.102***	-0.0239*	-0.0278	0.0277***	0.0306***	-0.00227***	-2.16e-05	0.0117	-0.00177**	0.00415	-0.000417	0.000908	-2.61e-05	0.000894	-0.00228***	0.000575***	-2.80e-05	3.00e-05
	(0.0157)	(0.0130)	(0.0214)	(0.00523)	(0.00544)	(0.000320)	(0.000458)	(0.00839)	(0.000863)	(0.00298)	(0.00133)	(0.000584)	(2.00e-05)	(0.00100)	(0.000610)	(0.000184)	(3.60e-05)	(2.12e-05)
y = 2 if Work	0.0592***	0.0330***	-0.00955	-0.0199***	-0.0117***	0.00107***	0.000534	-0.0167**	0.000473	-0.00108	0.000322	-0.000962**	2.93e-05*	0.00102	0.00126**	-0.000490***	-1.04e-05	-2.91e-05*
	(0.0122)	(0.0102)	(0.0176)	(0.00428)	(0.00444)	(0.000180)	(0.000373)	(0.00708)	(0.000713)	(0.00246)	(0.00104)	(0.000477)	(1.58e-05)	(0.000777)	(0.000491)	(0.000154)	(2.69e-05)	(1.67e-05)
y = 3 if Military	0.0711***	-0.00814	0.0114	0.000408	-0.00478*	0.000306***	-8.71e-05	0.000739	0.000379	-0.00228	-0.000673	0.000129	-3.46e-06	-3.85e-05	-0.000155	-7.97e-05	-5.84e-06	4.53e-06
	(0.0101)	(0.00627)	(0.00988)	(0.00249)	(0.00265)	(9.47e-05)	(0.000209)	(0.00373)	(0.000396)	(0.00139)	(0.000682)	(0.000259)	(9.99e-06)	(0.000483)	(0.000313)	(8.16e-05)	(1.73e-05)	(1.07e-05)
y = 4 if Home	-0.0280***	-0.00101	0.0260*	-0.00825**	-0.0141***	0.000889***	-0.000426	0.00419	0.000922	-0.000793	0.000767	-7.48e-05	2.61e-07	-0.00188**	0.00118***	-6.03e-06	4.42e-05	-5.40e-06
	(0.0105)	(0.00872)	(0.0137)	(0.00369)	(0.00391)	(0.000130)	(0.000303)	(0.00555)	(0.000567)	(0.00204)	(0.000929)	(0.000393)	(1.54e-05)	(0.000746)	(0.000388)	(0.000117)	(3.16e-05)	(1.59e-05)
Observations	3,125	3,125	3,125	3,125	3,125	3,125	3,125	3,125	3,125	3,125	3,125	3,125	3,125	3,125	3,125	3,125	3,125	3,125

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 8: Multinomial Logit Model: Partial Effect at Average

VARIABLES	(1) Gender	(2) Age	(3) House owner	(4) Father education	(5) Mother education	(6) Rank	(7) HS college percentage	(8) HS dropout rate	(9) Low income	(10) AP course offered	(11) AP taking percentage	(12) AP Passing percentage	(13) Total enrollment	(14) Distance to 4 yr college	(15) Distance 2yr college	(16) Distance to private college	(17) HS mean algebra	(18) Nonwhite
y = 1 if College	-0.102***	-0.0275**	-0.0274	0.0300***	0.0321***	-0.00239***	-5.68e-05	0.0132	-0.00184**	0.00414	-0.000530	0.00101	-2.91e-05	0.000887	-0.00247***	0.000616***	-2.95e-05	3.32e-05
	(0.0164)	(0.0137)	(0.0227)	(0.00556)	(0.00582)	(0.000350)	(0.000485)	(0.00890)	(0.000918)	(0.00316)	(0.00140)	(0.000619)	(2.10e-05)	(0.00105)	(0.000647)	(0.000196)	(3.77e-05)	(2.23e-05)
y = 2 if Work	0.0669***	0.0329***	-0.00637	-0.0210***	-0.0137***	0.00121***	0.000504	-0.0167**	0.000604	-0.00144	0.000314	-0.000979**	2.97e-05*	0.000894	0.00137***	-0.000512***	-7.77e-06	-2.98e-05*
	(0.0125)	(0.0103)	(0.0179)	(0.00428)	(0.00449)	(0.000203)	(0.000377)	(0.00714)	(0.000723)	(0.00249)	(0.00105)	(0.000481)	(1.59e-05)	(0.000784)	(0.000499)	(0.000155)	(2.68e-05)	(1.68e-05)
y = 3 if Military	0.0538***	-0.00543	0.00880	-0.000263	-0.00409**	0.000268***	-6.26e-05	0.000295	0.000311	-0.00174*	-0.000481	7.53e-05	-1.98e-06	-4.52e-05	-6.70e-05	-6.97e-05	-3.70e-06	2.69e-06
	(0.00577)	(0.00466)	(0.00733)	(0.00187)	(0.00197)	(7.71e-05)	(0.000155)	(0.00277)	(0.000295)	(0.00103)	(0.000505)	(0.000193)	(7.40e-06)	(0.000359)	(0.000234)	(6.09e-05)	(1.28e-05)	(7.89e-06)
y = 4 if Home	-0.0189*	3.55e-05	0.0250*	-0.00872**	-0.0143***	0.000914***	-0.000384	0.00324	0.000922*	-0.000958	0.000696	-0.000107	1.39e-06	-0.00174**	0.00116***	-3.41e-05	4.10e-05	-6.18e-06
	(0.0102)	(0.00821)	(0.0128)	(0.00345)	(0.00362)	(0.000139)	(0.000284)	(0.00520)	(0.000533)	(0.00191)	(0.000870)	(0.000369)	(1.44e-05)	(0.000692)	(0.000362)	(0.000110)	(2.93e-05)	(1.49e-05)
Observations	3,125	3,125	3,125	3,125	3,125	3,125	3,125	3,125	3,125	3,125	3,125	3,125	3,125	3,125	3,125	3,125	3,125	3,125

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 9: Spatial Multinomial Logit Model: Point Estimates

y = 1 if attending college
y = 2 if working
y = 3 if serving in military
y = 4 if staying at home (base)

Peer effect ρ		0.36^{***}	

	College β_1	Work β_2	Military β_3
Gender	0.08^{*}	0.64^{*}	2.98^{*}
Age	-0.03	0.27^{**}	-0.16^{*}
House owner	-0.31^{**}	-0.31	0.04
Father education	0.15^{*}	-0.20^{*}	0.10
Mother education	0.24^{***}	0.07	0.03
Rank	-0.01^{*}	-1.40e-3	-2.8e-3
HS college percentage	3.8e-4**	9.51e-4	3.5e-3
HS dropout rate	-0.02	0.14^{**}	-0.02
Low income	-0.07	-6.72e-4	9.9e-3*
AP course offered	0.03	-7.50e-4	-0.04
AP taking percentage	-3.51e-3	-3.61e-3	-0.02
AP passing percentage	1.00e-3	-3.5e-4	3.1e-4
Total Enrollment	3.2e-3	7.89e-4	8.7e-4
Distance to 4yr college	-0.01^{***}	-0.04^{*}	4.6e-3
Distance to 2yr college	-0.015^{***}	1.00e-3	-0.01
Distance to private college	7.11e-4	-4.11e-4**	-3.1e-3
HS mean algebra	-5.24e-4	-3.20e-5**	-5.01e-4
Nonwhite	-3.9e-5	1.1e-3*	-8.7e-5
Constant	1.77^{*}	-2.73^{*}	1.34^{*}

[a] HS stands for high school.
[b] Gender equals to 1 if female.
[c] The variable House owner is an indicator that equals to 1 if a household rents.
[d] The variable Father/Mother education: 1-Illiterate/Semi-literate, 2-Elementary school, 3-Middle school, 4-High school, 5-some college, 6-2 year college, 7-4 year college, 8-Master degree, 9-professional school
[e] The variable Rank is class ranking of high school students indicated in percentile. The lower the percentage, the higher this person is ranked and therefore the coefficient is negative.
[f] The variable HS college percentage is the percentage of students plan to attend college in each high school.
[g] The variable Low income is the percentage of low income families in each high school.
[h] Nonwhite is the total number of nonwhite students in each high school.

Table 10: Spatial Multinomial Logit Model: Direct & Indirect Effect

	College		Work		Military	
	DE	IE	DE	IE	DE	IE
Gender	0.08*	0.04*	0.64*	0.36**	2.98*	1.66*
Age	-0.03	-0.02*	0.27**	0.15**	-0.16*	-0.08*
House owner	-0.31**	-0.17*	-0.31	-0.17	0.04	0.02
Father education	0.15*	0.08**	-0.20*	-0.11*	0.10	0.06
Mother education	0.24***	0.13***	0.07	0.04	0.03	0.02
Rank	-0.01*	-5.57e-3*	-1.40e-3	-7.8e-4	-2.8e-3	-1.56e-3
HS college percentage	3.8e-4**	2.23e-4*	9.51e-4	5.3e-4	3.5e-3	1.95e-3
HS dropout rate	-0.02	-0.01	0.14**	0.08*	-0.02	-0.01
Low income	-0.07	-0.04*	-6.72e-4	-3.73e-4*	9.9e-3*	5.52e-3*
AP course offered	0.03	0.02	-7.50e-4	-4.18e-4	-0.04	-0.02
AP taking percentage	-3.51e-3	1.67e-3	-3.61e-3	-2.01e-3	-0.02	-0.01
AP passing percentage	1.00e-3	5.58e-4*	-3.5e-4	-1.95e-4	3.1e-4	1.73e-4
Total Enrollment	3.2e-3	1.79e-4	7.89e-4	4.40e-4	8.7e-4	4.86e-4
Distance to 4yr college	-0.01***	-5.58e-4**	-0.04*	-0.02**	4.6e-3	2.57e-3
Distance to 2yr college	-0.015***	-8.37e-3**	1.00e-3	5.58e-4	-0.01	-5.58e-3
Distance to private college	7.11e-4	3.96e-4	-4.11e-4**	-2.29e-4*	-3.1e-3	-1.73e-4
HS mean algebra	-5.24e-4	-2.92e-4	-3.20e-5**	1.78e-5	-5.01e-4	-2.79e-4
Nonwhite	-3.9e-5	-2.18e-5	1.1e-3*	-6.13e-4*	-8.7e-5	-4.85e-5
Constant	1.77*	0.98*	-2.73*	-1.52*	1.34*	0.75*

[a] HS stands for high school.
[b] Gender equals to 1 if female.
[c] The variable House owner is an indicator that equals to 1 if a household rents.
[d] The variable Father/Mother education: 1-Illiterate/Semi-literate, 2-Elementary school, 3-Middle school, 4-High school, 5-some college, 6-2 year college, 7-4 year college, 8-Master degree, 9-professional school
[e] The variable Rank is class ranking of high school students indicated in percentile. The lower the percentage, the higher this person is ranked and therefore the coefficient is negative.
[f] The variable HS college percentage is the percentage of students plan to attend college in each high school.
[g] The variable Low income is the percentage of low income families in each high school.
[h] Nonwhite is the total number of nonwhite students in each high school.

Figure 1
Example of an undirected Network of Peers-School A

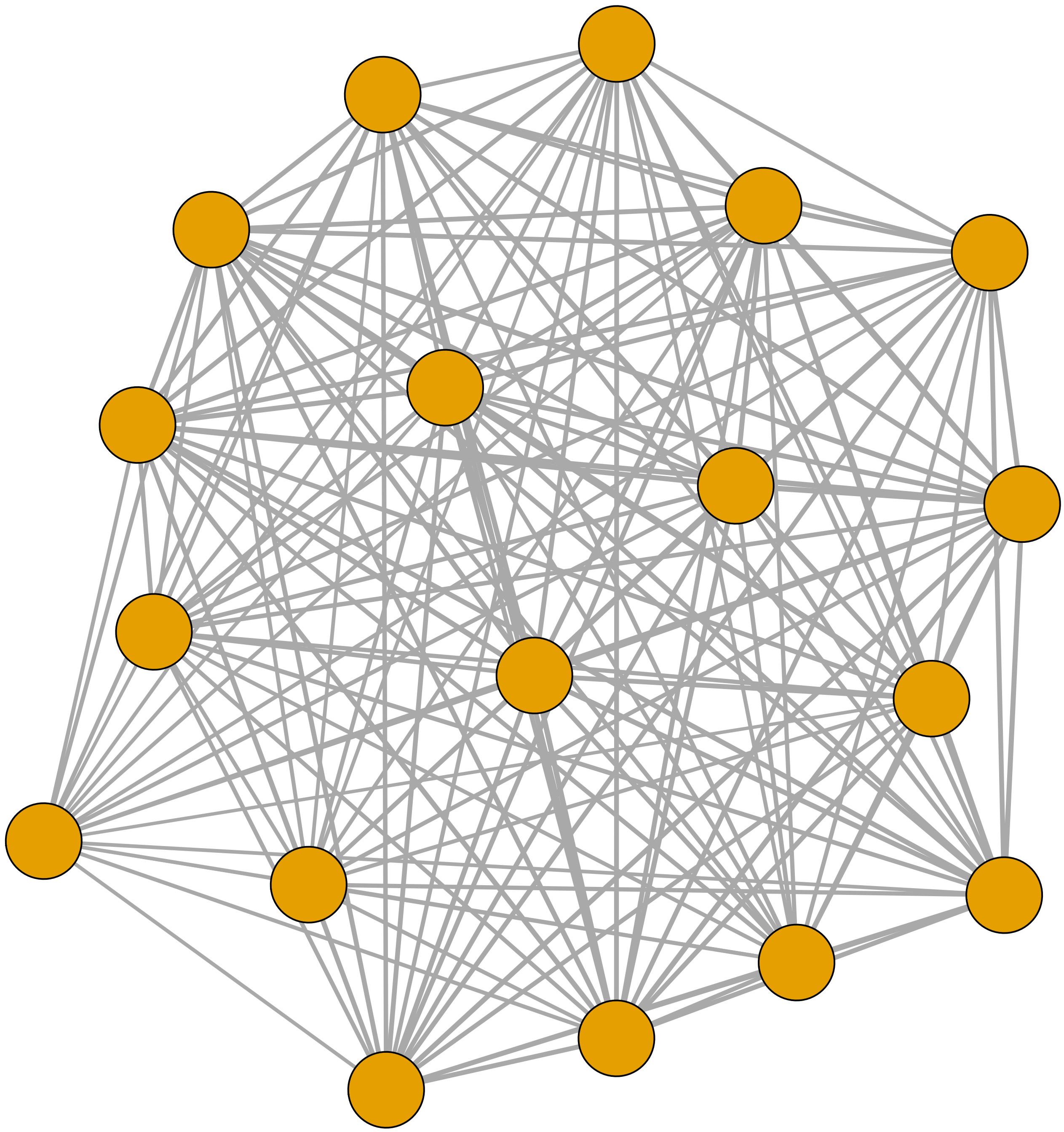

Figure 2
Example of an undirected Network of Peers-School B