Online Appendix

Title: Beyond the "Old Boys' Network": Social Networks and Job Finding at Community Colleges

Author: Maria Zhu[1]

Abstract: Research and popular discourse on social networks in higher education have focused on elite universities, even though most students do not attend such institutions. This paper sheds novel light on how classmate networks at community colleges influence job finding. Using data from a state community college system linked to matched employer-employee records, I exploit quasi-random variation in class section enrollment within courses to detect network effects. I find students are significantly more likely to get a job at a classmate's firm later than at the firm of a peer enrolled in a different section of the same course.

Disclosure statement: I have no relevant material or financial interests relating to the research in this paper to disclose.

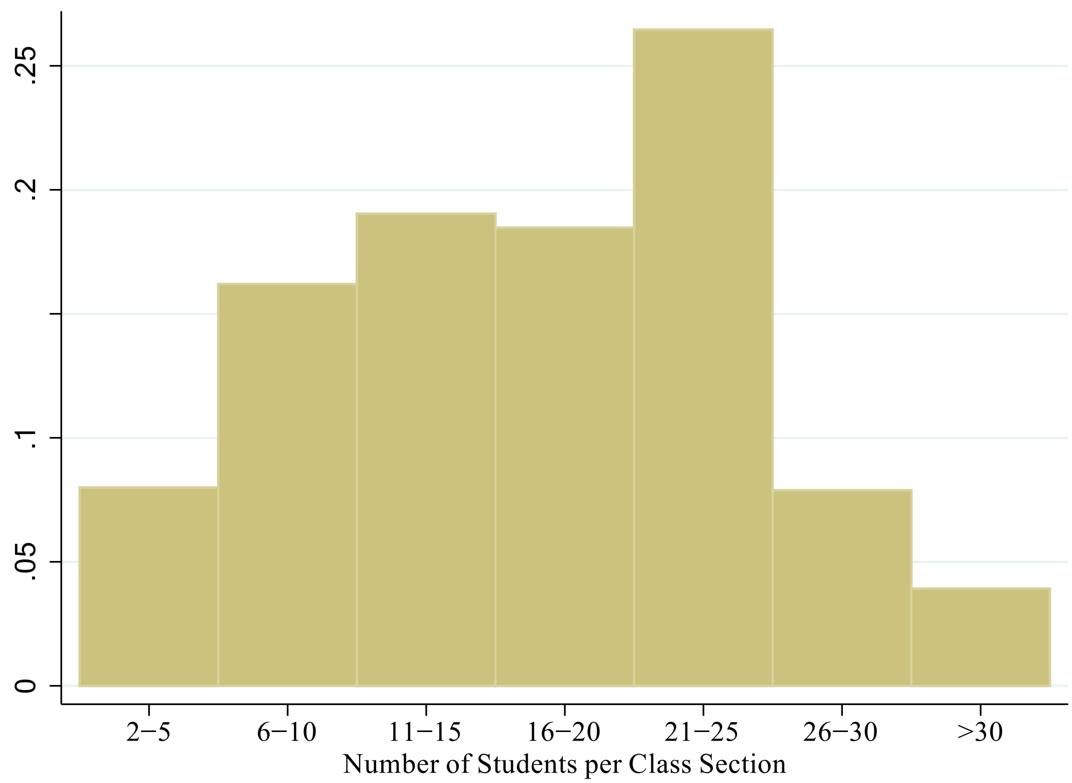
Data availability statement: The research presented in this article uses proprietary data from the Arkansas Research Center, through which researchers must apply for access. The data are available for researchers through an application process, and I commit to providing guidance on how to obtain the data.

JEL codes: I23, I26, J2

[1] Affiliations: Maria Zhu is an assistant professor in the Department of Economics at Syracuse University (Email: mzhu33@syr.edu). I am grateful to my advisors—Peter Arcidiacono, Pat Bayer, Rob Garlick, Joe Hotz, and Arnaud Maurel—for their guidance and encouragement. I thank Gorkem Bostanci, Hugh Cassidy, Liz Cook, David Eil, Alfonso Flores-Lagunes, Paul Goldsmith-Pinkham, Attila Gyetvai, Hugo Jales, Mike Kofoed, Margaux Luflade, Hugh Macartney, Margie McElroy, John Pepper, Michael Ransom, Tyler Ransom, Seth Sanders, Amy Schwartz, David Selover, John Singleton, Juan Carlos Suárez Serrato, and Xiao Yu Wang for helpful comments. This paper also benefitted from seminar participants at Duke, SOLE, AEFP, Young Economists Symposium, SEA, and WEAI. Finally, I thank three anonymous reviewers for their helpful feedback on this paper.

A Additional Tables and Figures

Figure A1: Distribution of Students per Section



Note: Class sections with one student are dropped from the sample.

Table A1: Student Characteristics: Analysis Sample vs. All Students

	Analysis Sample (1)	All Students (2)
Female (%)	63.37	63.78
Part-time (%)	54.30	52.61
Employed (%)	61.42	62.30
In-state (%)	93.90	94.92
Transfer student (%)	24.86	24.43
Age	26.87 (9.07)	26.23 (8.77)
Number of sections	3.09 (1.82)	3.16 (1.70)
N	422,022	944,003

Arkansas two-year college students, 2004-2012. Observations denote student-semester units. Table displays means (standard deviations in parentheses). Age is imputed from year of high school graduation, assuming students are 18 at the end of high school. Column (1) represents observations that appear in the analysis sample for this paper. Column (2) represents observations of all students from Table 1.

Table A2: Class Section Characteristics: Analysis Sample vs. All Classes

	Analysis Sample (1)	All Classes (2)
Class section size	17.37	15.74
	(8.20)	(8.42)
Mean GPA	2.62	2.80
	(0.96)	(0.95)
Female instructor (%)	57	55
Class subject (%)		
Basic skills and remedial	19.2	14.3
Social sciences	14.2	12.4
English	12.7	9.8
Business, management, and marketing	9.4	11.4
Biology	5.6	4.5
Mathematics and statistics	5.1	4.6
Computer and information sciences	4.5	4.2
Health professions	4.1	9.0
Visual and performing arts	4.0	4.2
Humanities	3.2	2.9
Physical sciences	2.2	2.4
Communication, journalism	2.1	1.2
Mechanic and repair	1.9	2.8
Engineering	1.7	2.6
Other	10.1	13.7
Observations	75,833	181,413

Arkansas two-year college class sections, 2004-2012. Units of observations are class sections. Table displays means (standard deviations in parentheses). Sections with one student have been dropped from the sample. Course subjects are categorized using 2010 Classification of Instructor Program (CIP) codes. Column (1) represents class sections that appear in the analysis sample for this paper from Table 2. Column (2) represents observations of all class sections.

Table A3: Effects of Sharing Additional Class Section on Probability of Finding Job at Peer's Firm (Year-by-year Breakdown)

	+1yr	+2yrs	+3yrs	+4yrs	+5yrs	+6yrs
Baseline	0.230	0.259	0.259	0.252	0.236	0.222
Number of Sections	0.0135***	0.0132***	0.0131***	0.0103***	0.0105***	0.0112***
	(0.00166)	(0.00177)	(0.00188)	(0.00205)	(0.00238)	(0.00238)
Course Bundle FE	X	X	X	X	X	X
N	59,192,230	52,294,258	43,748,074	34,621,106	26,337,746	18,624,800

***p<0.01, **p<0.05, *p<0.1. Observations represent matched pairs of students i with peers j. Standard errors are multi-way clustered by student i, student j, and course bundle. Outcome is whether student i gets a job at a firm where peer j is incumbent within three years of their first shared course. Baseline and coefficient estimates are multiplied by 100, so baselines represent percent values and coefficient estimates represent percentage changes.

Table A4: Course Subject Classification

Liberal Arts	Social Sciences
	Basic Skills
	English Language and Literature
	Biological and Biomedical Sciences
	Mathematics and Statistics
	Computer and Information Sciences
	Visual and Performing Arts
	Humanities
	Physical Sciences
	Liberal Arts and sciences
	Multi/Interdisciplinary Studies
Vocational	Business, Management, and Marketing
	Health Professions
	Homeland Security, Law Enforcement, and Service
	Mechanic and Repair Technologies
	Engineering Technologies/Technicians
	Education
	Family and Consumer Sciences
	Parks, Recreation, Leisure, and Fitness
	Communication and Journalism
	Leisure and Recreation Activities
	Precision Production
	Personal and. Culinary Services
	Transportation and Materials Moving
	Personal Awareness and Self-Improvement
	Legal Professions
	Agriculture
	Library Science
	Construction Trades
	Communications Technologies/Technicians
	Natural Resources and Conservation
	Public Administration and Social Services
	Interpersonal and Social Skills
	Military Technologies
	Military Science, Leadership, and Operations
	Science Technologies/Technicians

Classification of course subjects into liberal arts and vocational categories. Course subjects are classified using 2010 Classification of Instructional Program (CIP) codes.

Table A5: School Location Classification

Metropolitan	National Park Community College Northwest Arkansas Community College Pulaski Technical College Mid-South Community College
Micropolitan	East Arkansas Community College Arkansas State University–Mountain Home U. of Arkansas Community College at Batesville Southern Arkansas University Tech North Arkansas College Arkansas State University–Beebe Arkansas Northeastern College College of the Ouachitas
Neither	Rich Mountain Community College U. of Arkansas Community College at Morrilton Cossatot Community College Ozarka College Black River Technical College University of Arkansas Community College at Hope

Schools are classified based on the county in which the school is located. Classifications are based on 2013 definitions of statistical areas from the Office of Management and Budget (OMB).

B Robustness Checks

1. Examining Sorting on Observables

I first assess the degree of correlation on observable characteristics between students and peers in their section, conditioning on overall course composition. While this evaluation does not explicitly provide information on the amount of sorting on unobservable characteristics, the degree of selection on observables is informative of the amount of selection on unobservables (Altonji, Elder, and Taber 2005). To implement this test, for each section in the sample, I randomly select one student. Next, for each observable characteristic of interest in the data, I calculate the mean value of the characteristic of the non-selected students in the section to construct a measure of average section characteristics. I sample only one individual per section to ensure that estimates are not biased by the fact that each individual would significantly affect average group characteristics if multiple individuals per section were sampled.

Table B1 shows correlations along observable characteristics. Column 1 shows unconditional correlations, and Column 2 shows correlations conditional on course fixed effects. To obtain values in Column 2, I first regress both individual and section measures separately on course fixed effects and then report the correlation between the residuals. These findings isolate the portion of the overall correlation to sorting patterns across sections within a course. As expected, results in column 1 show a significant degree of sorting across courses. Reassuringly, Column 2 shows that with the addition course fixed effects, correlation on observable characteristics falls considerably. For example, the correlation between whether an individual is female and the proportion of peers in her section who are female is 0.344 overall, but this value falls to 0.068 with the addition of course fixed effects. On average, the addition of course fixed effects reduces estimated correlations by 76 percent for observable characteristics, and the average correlation between individuals and peers across characteristics is 0.049.

Table B1: Correlation between Individual and Average Characteristics of Peers.

	Unconditional (1)	Conditional on Course (2)
Female	0.344	0.068
Employed	0.331	0.072
In-state	0.282	0.028
Transfer student	0.340	0.065
Age 18-21	0.230	0.066
Age 22-26	0.056	0.021
Age 27+	0.078	0.026

Values shown represent correlations between individual characteristics and corresponding mean group characteristics of other individuals in the same section. Column 1 shows unconditional correlations, and Column 2 shows correlations conditional on course fixed effects.

Although the correlations of characteristics in Table B1 are much smaller after conditioning for course fixed effects, these values are not identical to zero. Thus, I next assess whether remaining sorting into sections along observables after conditioning for course fixed effects would predict a significantly higher probability of an individual getting a job at a firm with an incumbent same-section peer. The intuition behind this exercise is to understand whether differences in student composition on observable characteristics across sections explain any of the network effects measured in the main results. To do so, I extend Equation 1 to allow for heterogeneity along pair-level covariates:

$$F_{ijc} = \rho_c + \beta_1' X_{ij} + (\gamma + \beta_2' X_{ij}) N_{ij} + \varepsilon_{ijc} \tag{2}$$

where X_{ij} represents a vector of pair-level covariates describing i and j for each characteristic listed in Table B1. Covariates of i and j are interacted with one another for each characteristic.[2] In this setting, $\beta_1' X$ measures how the baseline propensity for an individual to start working at a firm with a course peer with whom they share no sections varies based on observable

[2] For example, with regard to gender, X_{ij} includes controls for $Female_i$* $Female_j$, $Female_i$* $Male_j$, and $Male_i$* $Female_j$ (with $Male_i$*$Male_j$ as the omitted category).

characteristics for each attribute of *i* and *j*. I first estimate the coefficients in Equation 2. Next, I use these parameter estimates to test whether the correlation on observable characteristics across sections of a course would lead to a significantly higher predicted probability for a student to work at a firm where a peer with whom they shared a section works, compared to peers from the same course with no shared sections.[3] To do so, I take the parameter estimates $\widehat{\beta}_1$, which were derived using the full sample, and compare the average value of $\widehat{\beta}_1{}'X$ for pairs who were not in the same section with the average value of $\widehat{\beta}_1{}'X$ for pairs who were in the same section. This comparison indicates whether differences in correlation on observable characteristics across sections of a course would lead to a higher predicted propensity for an individual to start working at the same firm as a same-section peer, as opposed to a different-section peer from the same course. If the estimated value of $\widehat{\beta}_1{}'X$ is much higher for same-section pairs, this would suggest that sorting into sections on observables may play a significant role in driving the finding that taking a section with a peer leads to a significant increase in the propensity to get a job at the firm where the peer works.

I estimate Equation 2 separately for each year τ after the pair's first course together, $\tau \in \{1,2,\ldots,6\}$. Table B2 presents predicted propensities for a student to get a job at the firm where their peer works τ years after a course, for pairs not in the same section with pairs who were in the same section. Results indicate that for all years of analysis, the estimated values of $\widehat{\beta}_1$ predict that the propensity to work at a firm where a peer is incumbent is very similar for pairs who shared a section and those who did not. For example, three years after the course, the predicted likelihood that an individual works at a firm where a peer she shared a section with was incumbent at time of hire and the probability of working at a firm where a peer she didn't share a section with was incumbent are both 0.290 percent. Findings in Table B2 provide support that the small remaining amount of sorting on observables within courses do not drive the increased propensity for an individual to get a job at a firm where a peer from the same

[3] I restrict this analysis to pairs who took one course together, which represents over 90 percent of pairs, to isolate predicted propensities due to variation in shared sections for pairs who shared the same number of courses.

section works compared to a firm where a peer from a different section of the same course works.

Table B2: Counterfactual Predicted Propensities

τ	Different Section	Same Section
+1 year	0.241	0.240
+2 years	0.282	0.282
+3 years	0.290	0.290
+4 years	0.290	0.290
+5 years	0.271	0.271
+6 years	0.259	0.259

Numbers in table represent the predicted propensity for an individual i to work at a firm where their peer j is incumbent, τ years after their first shared course. Estimates are reported as percentages.

2. *Estimation Using Subsamples*

For the first subsample robustness check, I run two analyses, one on the 50 percent of schools and one on the 50 percent of course subjects that display the lowest sorting across sections within a course. To do so, for each school (or course subject) in the sample, I analyze the correlation between individual and average characteristics of peers in her section, conditional on course, as in Column (2) of Table B1. I then take the average correlation across all observable characteristics. I select the 50 percent of schools (or course subjects) with the lowest average correlations to use for the restricted subsample.

Next, I re-estimate Equation 1 using these two subsamples and assess whether results finding significant classmate network effects hold. Since these sample restrictions substantially change the composition of the courses analyzed, I do not expect results to be identical to main results in Figure 5 and Table 4. However, if sizable network effects are still found in courses that experience the least amount of sorting into sections on observables, this provides further support to the ability of the research design to isolate the role of social interactions in the job finding process.

Figures B1 and B2 display results of this analysis for the subsample of schools and course subjects with the lowest degrees of sorting based on observables, respectively. Results indicate that for all years of analysis, the estimated effects of taking an additional section with a peer on the propensity for a student to work in a job where their peer was incumbent at time of hire are positive and statistically significant. Moreover, they are similar in magnitude to the estimates from the main sample in Table A3, although estimates in Figures B1 and B2 have wider confidence bands due to decreased sample size. Results are presented in table form with baseline values in Tables B3 and B4. These tables also include aggregate outcome results over the first three years after a pair's first shared course.

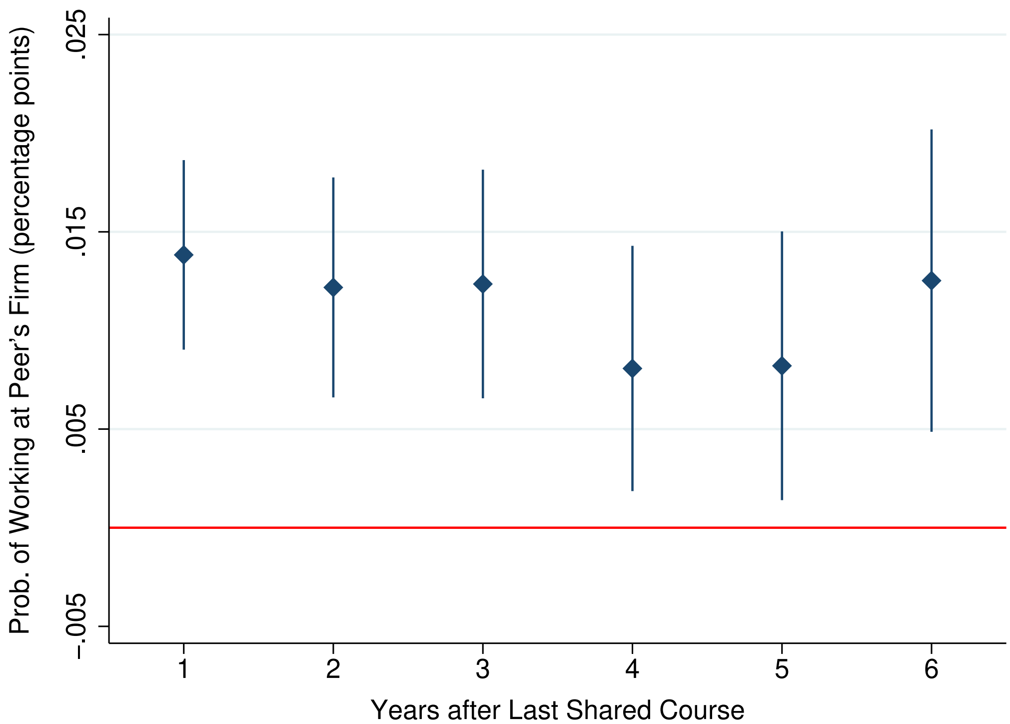
Table B3: Effects of Sharing Additional Class Section, Subsample 50% Schools with Lowest Sorting

	(1) Aggregate	(2) +1yr	(3) +2yrs	(4) +3yrs	(4) +4yrs	(6) +5yrs	(7) +6yrs
Baseline	0.573	0.208	0.242	0.240	0.241	0.232	0.219
Number of sections	0.029***	0.014***	0.013***	0.013***	0.008**	0.009**	0.012***
	(0.005)	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)
Course bundle FE	X	X	X	X	X	X	X
N	21,930,602	28,212,662	25,719,556	21,912,068	17,427,152	13,384,298	9,497,526

*** p<0.01, ** p<0.05, * p<0.1. Observations represent matched pairs of students with peers, *{i,j}*. Standard errors multiway clustered bycourse bundle, student i, and student j. Baseline and coefficient estimates are multiplied by 100, so baselines represent percent values and coefficient estimates represent percentage changes. Outcome variable in Column (1) is whether student starts working at firm where peer is incumbent within three years of their first shared course, and Columns (2)-(7) show year-by-year estimates.

Next, I re-estimate my results after eliminating the restriction of isolating my analysis to sectionswithin courses taught by the same instructor. In the preferred specification, I restrict my analysis

Figure B1: Effects of Sharing Additional Class Section: Subsample of 50% of Schools with Lowest Amount of Sorting on Observables



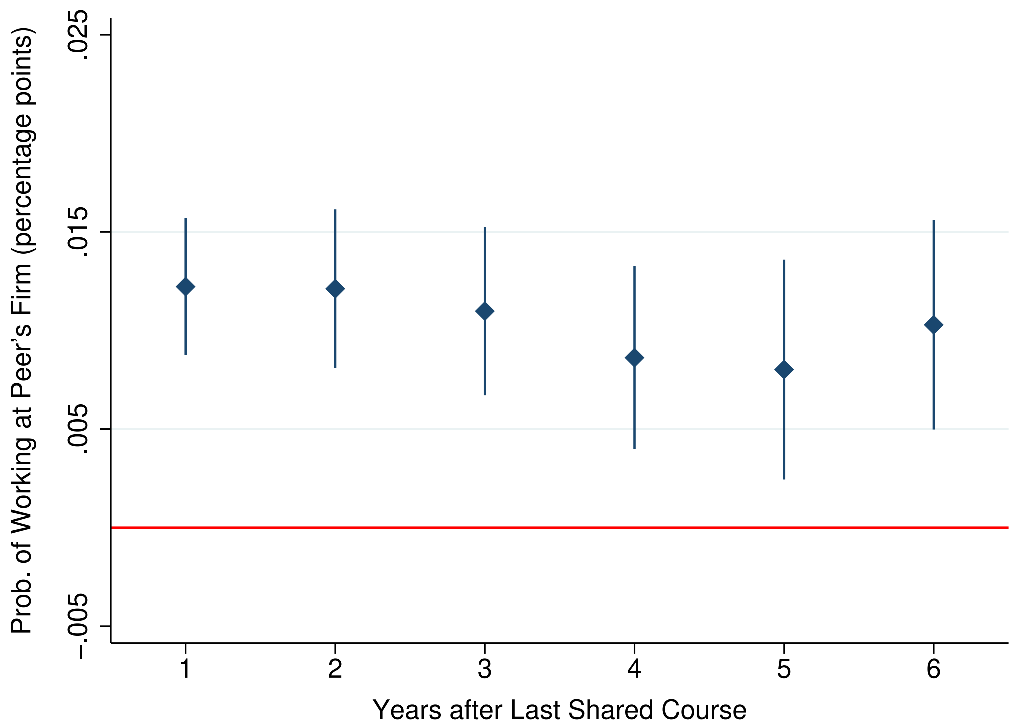
Note: Outcome is whether student *i* is working in a job where student *j* was incumbent at time of hire, *x* years after their first shared course. Coefficient estimates have been multiplied by 100 to reflect percentage changes. Estimates are displayed with 95 percent confidence intervals, and standard errors are multiway clustered by course bundle, student *i*, and student *j*. Results are displayed in table format with information on baseline value and sample sizes in Table B3. Estimates include controls for the attendance status (part-time or full-time) of *i* and *j*.

Table B4: Effects of Sharing Additional Class Section, Subsample 50% Course Subjects withLowest Sorting

	(1) Aggregate	(2) +1yr	(3) +2yrs	(4) +3yrs	(4) +4yrs	(6) +5yrs	(7) +6yrs
Baseline	0.663	0.236	0.277	0.282	0.275	0.257	0.243
Number of sections	0.020***	0.014***	0.013***	0.011***	0.008***	0.008**	0.010***
	(0.004)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)
Course bundle FE	X	X	X	X	X	X	X
N	23,447,710	31,367,690	27,789,966	23,426,624	18,672,868	14,344,494	10,243,348

*** p<0.01, ** p<0.05, * p<0.1. Observations represent matched pairs of students with peers, *{i,j}*. Standard errors multiway clustered bycourse bundle, student *i*, and student *j*. Baseline and coefficient estimates are multiplied by 100, so baselines represent percent values and coefficient estimates represent percentage changes. Outcome variable in Column (1) is whether student starts working at firm where peer is incumbent within three years of their first shared course, and Columns (2)-(7) show year-by-year estimates.

Figure B2: Effects of Sharing Additional Class Section: Subsample of 50% of Course Subjects with Lowest Amount of Sorting on Observables



Note: Outcome is whether student i is working in a job where student j was incumbent at time of hire, x years after their first shared course. Coefficient estimates have been multiplied by 100 to reflect percentage changes. Estimates are displayed with 95 percent confidence intervals, and standard errors are multiway clustered by course bundle, student i, and student j. Results are displayed in table format with information on baseline value and sample sizes in Table B3. Estimates include controls for the attendance status (part-time or full-time) of i and j.

to within-instructor variation in section enrollment due to concerns that students may be selecting into certain instructors within a course on dimensions that are correlated with subsequent job outcomes, or that there may be instructor-specific effects influencing students' job outcomes. However, one limitation of this restriction is that this could exacerbate selection issues induced by scheduling constraints by ruling out students who take the same course at the same time of day since the same instructor cannot be in two places at once. Since I do not observe the time of day or day of week that a section occurs, as a robustness check, I estimate results on a sample without within-instructor restrictions, which will include pairs of students who are enrolled in sections of a course that occur contemporaneously.

Table B5 displays results of this analysis. Results are estimated on a 25 percent random sample of pairs for computational ease since removing within-instructor restrictions significantly increases the sample size for the pair-level analysis. Using this sample, I find that sharing a class with a peer leads to significantly larger effects. In the linear specification, I find that each additional shared section increases the propensity that a student will get a job at her peer's firm within three years of their first shared course by 0.061 percentage points. This represents an 11.1 percent increase from the baseline propensity of getting a job at a peer's firm for this sample of 0.547 percent. Unfortunately, I am unable to separate what portion of these effects are due to students sorting into instructors or instructor-specific effects compared to removal of selection concerns induced by scheduling constraints. However, these results provide some reassurance that allowing for comparisons of students in sections occurring simultaneously does not reduce my estimated effects.

Table B5: Aggregate Effects of Sharing Additional Class Sections on Probability of Working at Peer's Firm (All instructors)

	(1)	(2)
Baseline	0.547	0.547
Number of sections	0.0606***	0.0771***
	(0.00549)	(0.00547)
(Number of sections)^2		-0.00618***
		(0.00136)
Course bundle FE	X	X
N	39,640,828	39,640,828

*** p<0.01, ** p<0.05, * p<0.1. Observations represent matched pairs of students with peers, $\{i,j\}$. Standard errors multiway clustered by course bundle, student i, and student j. Outcome: whetherstudent i gets a job at the firm where peer j works within three years oftheir first shared course. Baseline and coefficient estimates are multiplied by 100, so baselines represent percent values and coefficient estimates represent percentage changes. Sample includes courses for all instructors, not just instructors who taught multiple sections of a course, and observations are a 25 percent random sample of all matched pairs.

The above robustness check assesses one form of sorting that is induced by scheduling constraints. Another potential concern is sorting into instructors that could bias estimates even using a within-instructor approach due to scheduling constraints. I address this concern in the next robustness check. Following the approach implemented in Fairlie, Hoffman, and Oreopoulos (2014), I estimate Equation 1 using the sample of courses in which students do not have a choice over instructors in a semester. This restriction addresses concerns that results are biased by student sorting into certain instructors, which could be a potential concern even when restricting my analysis to comparisons of section enrollment within courses taught by the same instructor. For example, suppose some students are only able to take a certain course in the morning due to scheduling constraints while other students are only able to take the course in the afternoon. Instructor B teaches a section of the course in the morning and Instructor A teaches a section of the course in both the morning and the afternoon. If students who are more like each other (and thus tend to sort into the same firms) tend to sort themselves differentially by instructors conditional on their preferred course timing, this would make it more likely that students in the morning section of Instructor A end up at the same firm as students in the afternoon section of Instructor A, even in the absence of network effects.

Results of this estimation are displayed in Table B6. Reassuringly, I find that sharing a class with a peer has a positive and significant effect on job finding. Specifically, each additional shared section increases the propensity that a student will get a job at her peer's firm within three years of their first shared course by 0.041 percentage points. This represents a 5.0 percent increase from the baseline propensity of getting a job at a peer’s firm for this sample of 0.819 percent. This is actually a larger effect than the main estimation in Table 4 using the full sample, which finds a 3.8 percent increase.

Table B6: Aggregate Effects of Sharing Additional Class Sections on Probability of Working at Peer's Firm (No Instructor Choice)

	(1)	(2)
Baseline	0.819	0.819
Number of sections	0.0406***	0.0709***
	(0.00677)	(0.0105)
(Number of sections)^2		-0.00462***
		(0.00109)
Course bundle FE	X	X
N	6,899,822	6,899,822

*** p<0.01, ** p<0.05, * p<0.1. Observations represent matched pairs of students with peers, {*i, j*}. Standard errors multiway clustered by course bundle, student *i*, and student *j*. Outcome: whetherstudent *i* gets a job at the firm where peer *j* works within three years oftheir first shared course. Baseline and coefficient estimates are multiplied by 100, so baselines represent percent values and coefficientestimates represent percentage changes. Sample is restricted to courses for which students did not have a choice over instructors.

3. *Specification with Individual Fixed Effects*

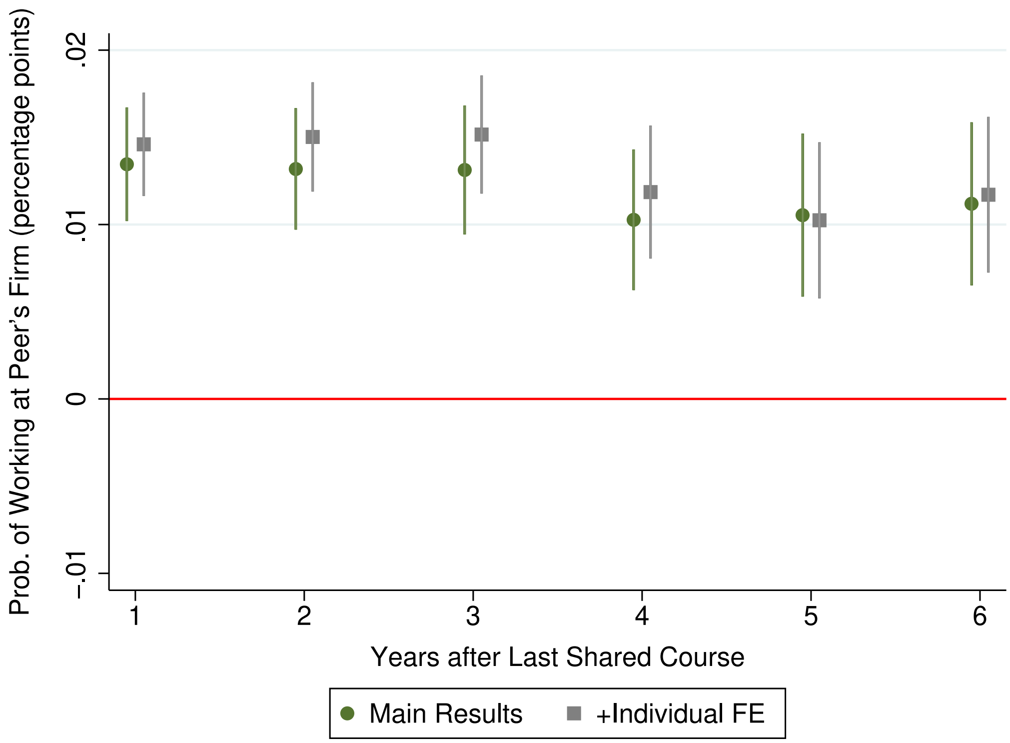
To analyze student sorting unobservable characteristics directly, I extend Equation 1 to include individual fixed effects for each member of the pair:

$$F_{ijc} = \tilde{\rho}_c + \tilde{\gamma} N_{ij} + \lambda_i + \lambda_j + \tilde{\varepsilon}_{ijc} \tag{3}$$

where λ_i and λ_j represent individual fixed effects for *i* and *j*, respectively. Since students appear multiple times in the data, the inclusion of individual fixed effects can be used to test for one type sorting on unobservables. Specifically, this analysis accounts for sorting in situations where (i) certain types of workers are more likely to work with those in the same section for unobserved reasons, and (ii) these workers sort themselves into similar sections within a course. Equation 3 controls for this concern by capturing any inherent differences across individuals to work with peers with fixed effects.

Figure B3 shows estimated coefficients on the number of shared sections for Equation 3, compared to estimated coefficients for the primary specification in Equation 1. Results indicate that estimates including fixed effects produce very similar estimates to the original specification on the effect of taking an additional section together on the propensity for a student to work at a firm with an incumbent peer for any of the years of analysis. For five out of the first six years after the first shared course, the specification with individual fixed effects estimates network effects that are slightly *higher* than estimates from the main specification, although estimates with and without individual fixed effects are not significantly different for any years. All results are reported with 95 percent confidence interval bars, and individual fixed effects results are reported numerically in Table B7.

Figure B3: Effects of Sharing Additional Class Section: Adding Individual Fixed Effects



Note: Outcome is whether student i is working in a job where student j was incumbent at time of hire, x years after their first shared course. Coefficient estimates have been multiplied by 100 to reflect percentage changes. Estimates are displayed with 95 percent confidence intervals, and standard errors are multiway clustered by course bundle, student i, and student j. Estimates include controls for the attendance status (part-time or full-time) of i and j. Results estimated using the augmented specification that includes individual fixed effects, Equation 3, are displayed in in Table B7.

Table B7: Effects of Sharing Additional Class Section, Specification with Individual Fixed Effects

	(1) Aggregate	(2) +1yr	(3) +2yrs	(4) +3yrs	(4) +4yrs	(6) +5yrs	(7) +6yrs
Baseline	0.634	0.230	0.259	0.259	0.252	0.236	0.222
Number of sections	0.029***	0.015***	0.015***	0.015***	0.012***	0.010***	0.012***
	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Course bundle FE	X	X	X	X	X	X	X
Individual FE	X	X	X	X	X	X	X
N	44,611,609	59,944,330	53,061,894	44,576,624	35,423,320	27,099,464	19,348,024

*** p<0.01, ** p<0.05, * p<0.1. Observations represent matched pairs of students with peers, $\{i, j\}$. Standard errors multiway clustered by course bundle, student i, and student j. Baseline and coefficient estimates are multiplied by 100, so baselines represent percent values andcoefficient estimates represent percentage changes. Outcome variable in Column (1) is whether student starts working at firm where peer is incumbent within three years of their first shared course, and Columns (2)-(7) show year-by-year estimates.

4. *Examining Role of Sorting on Pre-Existing Relationships*

This section assesses whether students sort into sections with pre-existing friends in a way that significantly affects the outcome of interest. Since I do not observe directly whether a pair of students share a pre-existing relationship, I proxy for this information by looking at geographic proximity between a pair and whether they worked together previously. I measure residential proximity of a pair by calculating the travel distance between their high schools of attendance, and I measure prior coworker status using an indicator for whether the pair worked at the same place at any time in the three years prior to their first course together. This proxy is created on the intuition that pairs who live closer together or who were prior coworkers are more likely to have established social ties before taking a course together.

To test whether sorting on pre-existing relationships drives outcomes, I first analyze whether students do in fact exhibit a higher propensity to enroll in the same section with peers who they worked with previously or with peers from a closer geographic proximity:

$$N_{ijc} = \varphi_c + \phi_1 Prior\_Cowork_{ij} + \phi_2 Dist_{ij} + \nu_{ijc} \quad (4)$$

where $Prior\_Cowork_{ij}$ is an indicator variable for that takes on a value of one if i and j worked at the same firm in the three years prior to taking their first course together and zero otherwise. The variable $Dist_{ij}$ measures the travel distance in miles between the pair's high schools. The outcome of interest, N_{ijc}, which acts as the independent variable in Equation 1, represents the number of sections i and j share, and φ_c is a course bundle fixed effect. The coefficients ϕ_1 and ϕ_2 measure whether pairs who worked together previously or live closer enroll in more sections together, conditional on the courses in which they enroll.

Table B8 shows estimation results for Equation 4. Column (1) displays unconditional results without course bundle fixed effects, and Column (2) includes course bundle fixed effects. In Column (1), estimates indicate a sizable positive correlation for the propensity of prior coworkers to enroll in a class section together. The coefficient on travel distance is positive, indicating pairs who are further away from each other are more likely to enroll in a section together. The inclusion of course bundle fixed effects reduces the magnitude of coefficient estimates on previous coworkers considerably, and the coefficient on travel distance becomes negative. Results from Column (2) indicate pairs who were prior coworkers take approximately 0.034 more sections together, compared to pairs who did not work together before taking courses. An increase in the travel distance between pairs by one standard deviation decreases the number of sections individuals share by 0.006 sections. These findings suggest students display an increased propensity to enroll in sections with those with whom they were connected to previously, as measured by working together and geographic proximity.

Since students display an increased propensity to enroll in sections with peers they worked with or lived closer to, a concern to identification is that unobservable traits that lead people to sign up for the same section may also affect their propensity to work at the same firm, independent of network effects. In this situation, the coefficient estimate of γ in Equation 1 may overestimate the true effect of

Table B8: Effects of Prior Coworkers and Travel Distance on Number of Class Sections Together

	(1)	(2)
Previous coworker	0.102***	0.034***
	(0.006)	(0.001)
Travel distance	0.004**	-0.006***
	(0.002)	(0.001)
Course FE		X
N	39,432,852	39,432,852

*** p<0.01, ** p<0.05, * p<0.1. Observations represent matched pairs of students with peers, {*i, j*}. Standard errors multiway clustered by course bundle, student *i*, and student *j*. "Travel Distance" captures the distance between the high school attended by student *i* and student *j* is measured using z-scores. "Previous Coworker" is an indicator variable that takes a value of one if i and j worked at the same firm in the three years leading up to their first shared course and zero otherwise. Outcome variable is number of sections pair *i* and *j* share.

taking an additional section with a peer on the propensity of getting a job at the firm where a peer works.[4] To assess this concern, I augment Equation 1 to include controls for prior coworkers and travel distance:

$$F_{ijc} = \breve{\rho}_c + \breve{\gamma} N_{ij} + \phi_1 Prior\_Cowork_{ij} + \phi_2 Dist_{ij} + \breve{\varepsilon}_{ijc} \qquad (5)$$

where F_{ijc} is an indicator variable taking a value of one if i works at a firm where j was working when i was hired, and zero otherwise. The course bundle fixed effect, $\breve{\rho}_c$, controls for sorting into courses, and N_{ij} measures the number of sections i and j take together. I add variables $Prior\_Cowork_{ij}$ and $Dist_{ij}$ to account for the role sorting into sections along preexisting relationships plays in affecting subsequent job finding outcomes. If estimates of $\breve{\gamma}$ do not differ significantly from estimates of γ in Equation 1, this provides reassurance that sorting into sections within a course group based on preexisting relationships as captured by former coworkers and peers in closer geographical proximity

[4] It is possible that friends are both more likely to sign up for the same section and more likely to end up at the same firm, both through network effects. However, it is also possible that individuals tend to sign up for sections with friends, and friends share unobserved characteristics that make it more likely for them to end up at the same firm, even absent any network effects. The latter scenario is the one I am concerned about.

does not drive results.[5]

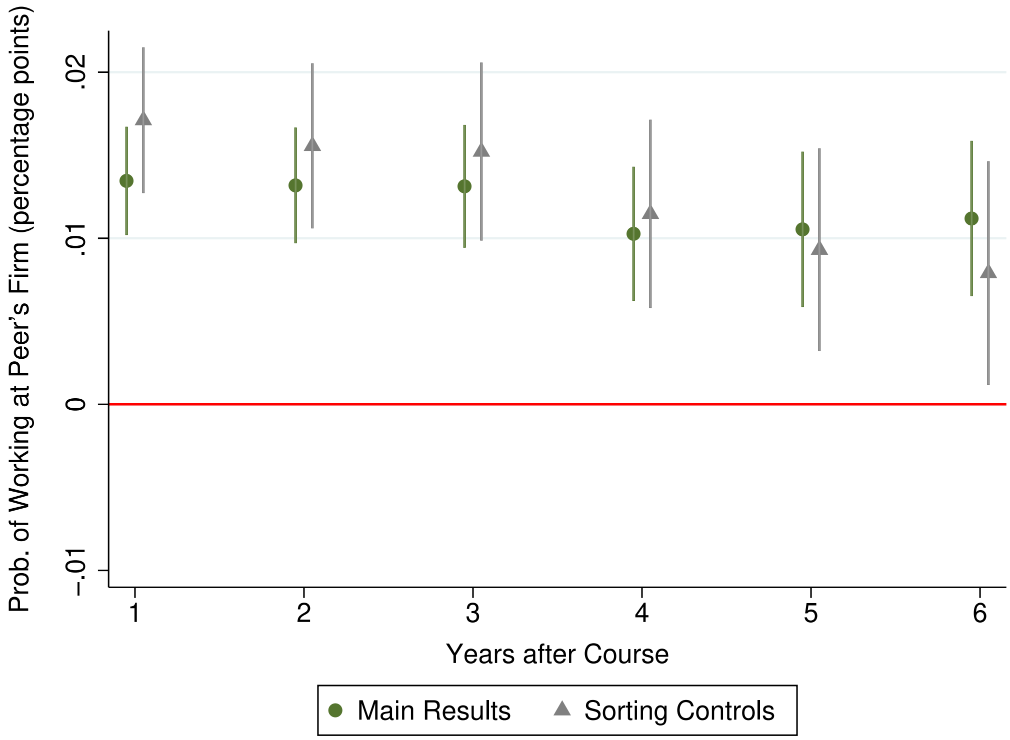
Figure B4 plots estimation results of Equation 5, in addition to original estimation results from Equation 1. All estimates are displayed with 95 percent confidence interval bars. Results indicate that estimates including controls for travel distance (as measured by high school location) and whether a pair were prior coworkers produce very similar estimates to the original specification on the effect of taking an additional section together on the propensity for a student to work at a firm with an incumbent peer for any of the years of analysis. For four out of the first six years after the last course, the specification with prior relationship controls estimates network effects that are slightly *higher* than estimates from the main specification. Estimates are reported numerically in Table B9, indicating results are not being driven by students sorting into sections within a course with peers based on preexisting relationships.[6]

---

[5] Since students typically enroll in college either out of high school or after working for some time, geographic proximity based on high school and former workplace contacts intuitively should capture a significant portion of an individual's relationships. Furthermore, many other places where people form relationships will also be geographically correlated (for example, family, places of worship, and social groups). However, I cannot rule out the possibility that students systematically form preexisting relationships in ways that are not captured by residential proximity or prior work history.

[6] As an additional robustness check, I conduct an event study analysis looking at the evolution of the probability for pairs of students to work together over time for students who are and are not in the same section of a course. This approach allows me to include individual fixed effects, which further addresses concerns about student sorting along preexisting relationships. As expected, I find a sizable jump in the propensity for a student to get a job at a firm where her peer works immediately following a shared section, compared to a student whose peer enrolled in a different section of a course.

Figure B4: Effects of Sharing Additional Class Section: Controlling for Preexisting Relationships



Note: Outcome is whether student i is working in a job where student j was incumbent at time of hire, x years after their first shared course. Coefficient estimates have been multiplied by 100 to reflect percentage changes, and standard errors are clustered by course bundle. Estimates are displayed with 95 percent confidence intervals, and standard errors are multiway clustered by course bundle, student i, and student j. Estimation results from the main specification estimated using Equation 1 are displayed numerically in the appendix in Table A3. Results estimated using the augmented specification that includes sorting controls for proxies for pre-existing relationships, Equation 5, are displayed in Table B9.

Table B9: Effects of Sharing Additional Class Section, Controlling for Pre-existing Relationships

	(1)	(2)	(3)	(4)	(4)	(6)
	+1yr	+2yrs	+3yrs	+4yrs	+5yrs	+6yrs
Baseline	0.230	0.259	0.259	0.252	0.236	0.222
Number of sections	0.0171***	0.0156***	0.0152***	0.0115***	0.00931***	0.00790**
	(0.00223)	(0.00253)	(0.00273)	(0.00289)	(0.00311)	(0.00343)
Course bundle FE	X	X	X	X	X	X
N	31,807,940	28,225,720	23,883,746	19,208,472	14,862,692	10,723,794

*** p<0.01, ** p<0.05, * p<0.1. Observations represent matched pairs of students with peers, $\{i, j\}$. Standard errors multiway clustered by course bundle, student i, and student j. Baseline and coefficient estimates are multiplied by 100, so baselines represent percent values and coefficient estimates represent percentage changes. Outcome variable in Column (1) is whether student starts working at firm where peer is incumbent within three years of their first shared course, and Columns (2)-(7) show year-by-year estimates. All specifications include controls for the travel distance between the high schools that student i and j attended, as well as a control for whether i and j worked at the same firm in the three years leading up to their first shared course.