Impact of Academic Interaction Networks on Grades & Jobs: Evidence from Four-year University Trajectories

Tuan Phan (tphan@hku.hk)
The University of Hong Kong

Dan Ding
Southwestern University of Finance and Economics

Prasanta Bhattacharya
Institute of High Performance Computing (IHPC), A*STAR

Edoardo Airoldi
Temple University

Social Sciences - Article

Keywords: Interaction networks, Academic performance, Peer influence, Job placements

DOI: https://doi.org/10.21203/rs.3.rs-646237/v1

License: ☑️ This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License
Impact of Academic Interaction Networks on Grades & Jobs: Evidence from Four-year University Trajectories

Tuan Q. Phan,\textsuperscript{a,1} Dan Ding,\textsuperscript{b} Prasanta Bhattacharya,\textsuperscript{c} Edoardo M. Airoldi,\textsuperscript{d}
\textsuperscript{a}The University of Hong Kong, Hong Kong
\textsuperscript{b}Southwestern University of Finance and Economics, China
\textsuperscript{c}Institute of High Performance Computing (IHPC), A*STAR, Singapore
\textsuperscript{d}Temple University, Philadelphia, PA 19122, USA

Student interactions play a critical role in education. Interactions and collaborative partnerships formed within universities can have a profound impact on academic and job-related performance. However, little is known about the mechanisms using which these networks develop inside the classroom, generate network effects, and impact academic and employment outcomes. We analyze 1.5 million de-identified in-class interactions, grades, and job outcomes for 6,355 undergraduates through a four-year education program to show that students with larger networks exhibit better academic and employment outcomes, and that high-performing students benefit from stronger network effects, as compared to their mid- and low-performing students. Our results also suggest the existence of a network premium - mid-performers with larger networks can achieve higher average salaries than high-performers with smaller networks. We adapted a stochastic actor-oriented model (SAOM) to model the underlying mechanisms of course selection, network formation, and peer influence among students. Our research sheds light on the importance of interaction networks as drivers of academic and employment performance, as well as the different mechanisms through which high- and low-performing students benefit from them. As education shifts to online, educators and policymakers should invest in the design of digital interactions and communities (e.g. forums) which can generate and sustain healthy networks among students.

\textbf{Keywords:} Interaction networks, Academic performance, Peer influence, Job placements

\textsuperscript{1}To whom correspondence should be addressed; E-mail: tphan@hku.hk.

University education offers students the opportunity to gain skills and form valuable con-
connections. The higher-education experience has a profound impact on an individual’s social network (1,2) and lifelong earnings (3). However, little is known about how academic interactions form inside the classroom, generate network effects, and impact future academic and employment outcomes. Interactions between students can help individuals learn from each other and improve their performance. In contrast, forming new strategic connections can provide access to new resources (2) and help find employment (4).

Understanding the impact of networks on academic performance has been challenging due to a lack of granular and long term data. Previous studies have looked at short-term academic outcomes such as choice of major (5) or academic failure (6). A few other studies have found links between networks and: (i) hiring processes, and success in finding jobs (4,7,8), (ii) women’s leadership success in the work place (3), and (iii) the reduced probability of job loss (9). Prior studies focusing on educational contexts have largely relied on self-reported surveys (6,4), assumed static networks and looked at short-term network evolution (10,6). While self-reported surveys can provide rich multi-relational information (6), they are often limited in time span, and are vulnerable to various sample-selection and response biases (11,12). Other network studies in this area have focused on interactions and non-academic outcomes including students’ consumption behaviors (13,14,15), and tastes and preferences (7). Although longitudinal data on interaction networks are increasingly available (2,16), separating network formation from peer influence using observational data has been statistically challenging due to various endogeneity concerns (17,18). Previous studies have focused on using in vitro approaches or specific conditions to separate the underlying mechanisms (19,2,5). While these approaches can help scholars make statistical inferences under very specific conditions and scenarios, they do not fully represent the natural processes, e.g., in vivo, that govern the overall network effects prevalent among university students.

In this study, we collected 1,574,355 de-identified interactions occurring between students
attending the same course, 234,265 course enrollment records and grades, and first job placements for an entire cohort of 6,355 undergraduate students from the National University of Singapore (NUS) spanning a four-year tenure. Our data on activity and interactions thus provides a proxy of the interactions between students (See Table I in the SI Appendix for detailed data descriptions and Section I for additional information on data context). We first use a set of panel regression analyses to examine the relationship between interaction network size, academic outcomes and first job placements. Next, we deploy a 3-stage variant of the stochastic actor-oriented model (SAOM) to jointly estimate the underlying processes of course selection, and overall network effects\textsuperscript{1}. In this study, we provide empirical evidence to help guide policies on new teaching modes that can help develop and nurture productive student networks. Details on the network and performance effects used in this study have been illustrated in the supplementary materials.

Results

Model-free analyses. Students actively interact and form connections with other students over the course of a four-year university program. Their academic and employment outcomes are affected by their level of network embeddedness. To model differences in degree centrality, we categorized students with more than 10 links ($75^{th}$ percentile) per term as the High Degree group and those with fewer than 2 links ($25^{th}$ percentile) as the Low Degree group. All other students were categorized as the Middle Degree group. For academic performance, we categorized students with a Cumulative Average Point (CAP), the cumulative academic performance of the student, of 4-5 points into the High CAP group, and those with a CAP lower than 3 points into the Low CAP group. All other students were categorized as the Middle CAP group. Fig-

\textsuperscript{1}We use the phrase overall network effects to denote both (i) network change based on homophily, and (ii) behavior change through peer influence.
Figure 1 reports the academic performance, employment status, and starting salary for students in various levels of academic performance and degree centrality of their interaction network. Figure 1 shows that variance exists among students in different CAP groups and degree centrality groups in terms of their academic performance, employment status and starting salary. Detailed explanations and further analyses are in Section 2 in the supplementary material.

In conjunction, interactions serve a dual role of benefiting academic performance and job search. While high-performing students can use their network as a premium to boost success in the job market, mid- and low-performing students can use their networks to compensate for the lack of other type of resources (e.g., academic abilities). Despite variance in the probability of securing jobs, we do not observe such large variances in the magnitude of starting salary across the degree centrality groups as shown in Figure 1 Panel C. However, the differences in starting salary across different levels of degree centrality are statistically significant overall ($p = 0.001$), in the High CAP group ($p = 0.05$) and in the Low CAP group ($p = 0.02$), but not in the Middle CAP group ($p = 0.16$). We estimated a set of regression models quantifying these relationships among degree centrality, CAP, employment status and starting salary while controlling for gender, nationality, faculty and entrance scores. The estimation results are provided in the supplemental material, Table 3. The results of these regression analyses are consistent with our descriptive summaries from Figure 1 and suggest that degree centrality is correlated with employment status upon graduation. In addition, we observed some individual differences in employment related outcomes among the students. Females in our sample were more likely to secure a job, but with a lower starting salary, as compared to their male counterparts. Entrance exam scores were positively linked to overall CAP, but negatively linked to their employment status, as well as starting salaries.

Our model-free analyses provide evidence of the importance of interaction networks and academic performance on employment outcomes. Next, we model the underlying mechanisms
through which networks affect academic performance and employment. Networks can affect employment outcomes directly or through the mediating role of academic performance. Similarly, academic performance might also affect employment outcomes through the mediating role of networks. We performed a set of mediation analyses using Baron and Kenny’s approach to test these competing hypotheses (See Tables 4-7 in the supplemental material for estimation results). Our analyses show that both interaction networks and academic performance partially mediate the impact of each other on employment outcomes.

**Tracking the evolution of interaction networks and academic performance.** We further investigate the dynamics and underlying mechanisms by focusing on a subset of representative students from the business school. Figure 2 illustrates the dynamics of term-wise academic performance, the evolution of homophily, and the clustering of the interaction networks. Figure 2 Panel A shows that, overall, academic performance increases over the course of eight semesters. This is accompanied by an increase in homophily among high GPA students (Figure 2 Panel B). Figure 2 Panel C shows increasing clustering of students over time. Specifically, we observe that (i) the size of the collaborative network increased over time; and (ii) most students evolved from low and middle performers to high performers in terms of academic performance over time. These results indicate the presence of an endogenous relationship between network evolution and changes in students’ academic performance.

The students’ interaction network and academic performance co-evolve by affecting each other. While past work has attempted to disentangle network formation and behavior change (20, 21), our empirical analysis is further complicated by the presence of course selection decisions that affect both network formation as well as course outcomes. In the present study, we extended a SAOM framework with a simple sample-selection stage (27, 20, 22) to develop a three-stage co-evolution model that characterizes the multiplexity of course selection, inter-
action network formation, and academic performance. The conceptual framework is shown in Figure 3. The objective functions and detailed model specification are briefly described in the Methods section, and detailed summaries are provided in Section 3 as part of the supplementary materials. We present estimation results using a MCMC-based method of moments estimator. Considering the computational costs of estimating the SAOM model, we estimated the three-stage SAOM model on the sample of business school students ($N=683$). We provide similar estimates for students from other schools in the supplementary material (Table 10). We further used this model to analyze the heterogeneous effects of network formation and peer influence among students of varying academic abilities.

**Linking peer-effects in interactions network to academic performance.** In Table 1, we present estimates of the peer effect specified in our three stage model. The homophily estimate in Table 1(a) is significantly positive ($5.64$, $p <0.001$), indicating that, on average, students are likely to build connections with peers who have similar academic performance. The peer influence estimate in Table 1(b) is significantly positive ($5.79$, $p <0.001$) suggesting that students are likely to converge in performance to the performance levels of their peers. For details on the estimation procedure and results please see the supplementary material, Section 4.

To evaluate the effectiveness of our proposed model, we compare the model fitness of our three-stage SAOM model against a baseline model that only considers the co-evolution of interaction network and performance changes, without accounting for course selection decisions from Stage (i). The results in Table 1 show that our proposed model outperforms the baseline model on two aspects. First, the focal parameters of interests from our three-stage SAOM model are significant and consistent, i.e., homophily ($5.64$, $p <0.001$) and peer influence ($5.79$, $p <0.001$), whereas the estimates from the baseline model are insignificant, i.e., homophily ($9.85$, $p=0.13$) and peer influence ($5.78$, $p=0.40$). This is unlikely given the abundance of past
work demonstrating the presence of network effects in student networks. Second, we conducted a goodness of fit (GOF) test to assess how well our proposed model specification fits the empirical data. The GOF is measured by computing the differences between the observed values of network measures and the simulated statistics for the same measures at the end of each period (23). To reduce the risk of over-fitting, we used auxiliary statistics based on network measures which were not explicitly fitted in our model estimation. In Table 1, we report the difference in model fit using the Mahalanobis distance (MHD) (24), and show that the three-stage SAOM model (MHD=301.57) has a substantially better model fit than the baseline two-stage model (MHD=21384.20). We further incorporated a latent space model proposed by (25) to check the sensitivity of our proposed three-stage model to any latent homophily related concerns, as described in the supplementary material, Section 7 and Table 13.

Table 1: Estimates for the Baseline and Three-stage Models

| (a) Network Parameters | Baseline model | Three-stage model | Heterogeneity | (b) Performance Parameters | Baseline model | Three-stage model | Heterogeneity |
|------------------------|----------------|-------------------|---------------|-----------------------------|----------------|-------------------|---------------|
| Network rates (period 1-7) (θX) | Included | Included | Included | Performance rates (period 1-7) (θG) | Included | Included | Included |
| Reciprocity (SX1) | 35.73 | 4.29*** | 4.00*** | Performance Trending (linear shape) (SθG1) | 0.23** | 0.18*** | 0.16*** |
| Transitivity (SX2) | 2.79* | 2.78*** | 3.16*** | Performance Trending (quadratic shape) (SθG2) | -0.58*** | -0.63*** | -0.74*** |
| 3-cycles (SX3) | 11.37 | -0.01 | -0.38 | Peer influence (SθG3) | 5.78 | 5.79*** | 8.35*** |
| Female (SX4) | -36.56** | 2.62* | 10.11*** | High GPA on peer influence (SθG4) | | 6.68*** |
| Local (SX5) | 5.22 | 3.04*** | 1.97** | Low GPA on peer influence (SθG5) | | -3.06*** |
| High GPA ego (SX6) | | | | | | | |
| Low GPA ego (SX6) | | | | | | | |
| Prob. of taking course together (Sθ1) | 18.77*** | 12.42*** | | | | | |
| Homophily (SX7) | 9.85 | 5.64*** | 3.26*** | | | | |
| High GPA on homophily (SθG7) | 12.87*** | 13.90*** | | | | | |
| Low GPA on homophily (SθG7) | -7.57*** | -7.57*** | | | | | |
| Goodness of fit (MHD) | 21384.20 | 301.57 | 3888.06 | | 21384.20 | 301.57 | 3888.06 |

Notes. 1. Column Baseline model presents estimates from a baseline model, and column Three-stage model reports estimates from the three-stage model. 2. Signif. codes: *** p < 0.001, ** p < 0.05, * p < 0.1. 3. Smaller MHD indicates a better performed estimation.

To further examine the prevalence of heterogeneity in overall network effects, we conducted a subgroup analysis based on differences in academic performance levels. The results in Table 1 show that the estimate for homophily effect is significant and positive (3.26, p < 0.01), indicating that, on average, students are likely to build ties with students of similar performance. How-
ever, compared to mid-performers, high-performing students are significantly more likely to connect to other high-performing students \((3.26 + 12.87 = 16.13, p < 0.01)\), while low performing students are less likely to connect to other low-performing students \((3.26 - 7.57 = -4.31, p < 0.01)\). This provides further evidence that students with different academic abilities tend to display distinct patterns of social behavior. We also note that the estimate for peer influence is significant and positive \((8.35, p < 0.01)\). However, compared to mid-performers, the peer influence for high performers is stronger \((8.35 + 6.68 = 15.13, p < 0.01)\), whereas the peer influence for low-performing students, while still positive, is considerably weaker \((8.35 - 3.06 = 5.29, p < 0.01)\). We provide additional subgroup analyses using degree level measures in the supplemental material, Table [12].

**Discussion**

We investigated the impact of interaction networks on academic performance and job placements. Using de-identified and longitudinal data from a four-year university program, we modeled the complexity of student in-class interactions through the three stages of course selection, network formation, and peer influence. First, we found that student interactions have an impact on academic performance and job market outcomes. The results highlight the important role of social collaborations in driving both academic performance and job placements. Moreover, the significant variance in job placement outcomes across network centrality levels suggests that for many students, their social embeddedness serve to offer a job market premium, and help partially offset academic performance.

To further unravel the underlying mechanisms, we deployed a three-stage SAOM framework to model the multiplexity of course selection, network dynamics, and academic performance over the course of a four-year academic program. We found that the overall network effects have a significant and positive role in student academic performance and future course
selection decisions, after controlling for selection decisions at both course and individual levels. Moreover, we found significant heterogeneous effects across different academic ability groups. Compared to mid-performing students, high-performing students are more likely to connect to similar high-performing students, while low-performing students are less likely to connect to similar low-performing students. The positive impact of peer influence among high-performing students is stronger than that among both mid-performers and low-performers. However, our study does not capture the modes and channels of communication between students such as information sharing, collaboration or cooperative work.

Our findings suggest the importance of interactions within the academic setting of universities. Student-to-student interactions are a critical part of the university education experience and can have a lasting impact on academic performance and job placements. New modes and technologies for education must continue to enable efficient interactions. In offline education environments, engagement in network formation happens through various signaling mechanisms such as social and extracurricular activities, living arrangements, and other incidental occurrences. These touch-points and signals may be harder to come by in the emerging modes and technologies for online education. This lack of interaction opportunities might have a detrimental effect on academic performance and job placements. With an increasing focus on online learning, policy makers and educators should invest in the design of digital supplements (e.g. forums, virtual conferencing tools) that can create an environment where such social and interaction networks among students are allowed to form and thrive.

Methods

Data information. We collected de-identified and longitudinal data from the National University of Singapore (NUS), a major English-speaking university in Asia. Our dataset contains 234,265 course enrollment records and 1,574,355 digital interactions and activity traces of
6,355 (N=6,355) full-time undergraduate students who entered the university in August 2012 and graduated in May 2016. We obtained anonymized data from their application records, such as gender, nationality, and entrance exam scores etc., enrollment records on their choice of faculty and degree program, a full history of their curriculum, term-by-term course grades, and anonymized logs of student interactions through the university’s official Learning Management System (LMS). The platform, similar to Blackboard or Moodle, supports classroom learning by allowing instructors to post lecture slides, assignments, and teaching material.

**A three-stage SAOM model of network formation and academic performance.** The results in this paper for the three-stage model analyses are estimated using an objective function for student $i$, which is a function of individual interaction ($X$), academic performance ($G$), homophily in course selection ($M$) in Stage 1, and other covariates ($V$) as shown in Eq.1. The details of model specification and estimation are provided in supplementary material, Section 3.

$$U_i(\beta, \delta, X_t, G_t, \hat{M}, V) = \sum_k \beta S^{[X]}_{ki}(X_t, G_t, \hat{M}, V) + \sum_k \delta S^{[G]}_{ki}(X_t, G_t, \hat{M}, V) + \epsilon$$  \hspace{1cm} (1)

To measure the heterogeneous effects across students with differing academic ability, we included two additional dummy variables to represent whether a student ranked higher than the 75th percentile of students in each term (High_GPA), or ranked below the 25th percentile (Low_GPA) of the distribution for academic scores in each term, while holding the middle 50th percentile (Middle_GPA) as a baseline. The final model specification is shown in Eq[2]. The complete list of model variables and associated definitions have been provided in supplementary material, Table[2].
\[ U_i(\beta, \delta, X_t, G_t, \hat{M}, V, \text{High\_GPA}, \text{Low\_GPA}) = \sum_k \beta S_{kt}^X(X_t, G_t, \hat{M}, V, \text{High\_GPA}, \text{Low\_GPA}) + \sum_k \delta S_{kt}^G(X_t, G_t, \hat{M}, V, \text{High\_GPA}, \text{Low\_GPA}) + \epsilon \]

Data Availability. Data cannot be shared publicly because of legal and university restrictions where the research was conducted. In compliance with the Singapore Personal Data Protection Act, data stored on the NUS Institute for Applied Learning Sciences and Educational Technology (ALSET) Data Lake is defined as personal data and cannot be shared publicly without student consent. Data can be accessed and analyzed on the ALSET Data Lake server with approval by the NUS Learning Analytics Committee on ethics, in accordance with NUS data management policies. Researchers who wish to access the data should contact ALSET at NUS (email: alsbox1@nus.edu.sg).

Acknowledgments. We would like to thank the Institute of Applied Learning Sciences and Educational Technology (ALSET) at the National University of Singapore (NUS) and Bernard Tan for providing the data and support for the project. We would also like to thank Wang Gang, Burton Singer, Sinan Aral, Aaron Clauset, Jayson Shi Jia, and Joseph Rilling for their helpful comments on the paper.

References

1. Lewis, K., Kaufman, J., Gonzalez, M., Wimmer, A. & Christakis, N. Tastes, Ties, and Time: A New Social Network Dataset Using Facebook.com. Social Networks 30, 330–342 (2008).
2. Phan, T. Q. & Airoldi, E. M. A natural experiment of social network formation and dynamics. *Proceedings of the National Academy of Sciences* **112**, 6595–6600 (2015). URL [http://www.pnas.org/content/112/21/6595](http://www.pnas.org/content/112/21/6595).

3. Yang, Y., Chawla, N. V. & Uzzi, B. A network’s gender composition and communication pattern predict women’s leadership success. *Proceedings of the National Academy of Sciences* 201721438 (2019). URL [https://www.pnas.org/content/early/2019/01/14/1721438116](https://www.pnas.org/content/early/2019/01/14/1721438116).

4. Munshi, K. Networks in the Modern Economy: Mexican Migrants in the U.S. Labor Market. *The Quarterly Journal of Economics* (2003).

5. Sacerdote, B. Peer Effects with Random Assignment: Results from Dartmouth Roommates. *The Quarterly Journal of Economics* **116**, 681–704 (2001).

6. Stadtfeld, C., Vöros, A., Elmer, T., Boda, Z. & Raabe, I. J. Integration in emerging social networks explains academic failure and success. *Proceedings of the National Academy of Sciences* **116**, 792–797 (2019). URL [https://www.pnas.org/content/116/3/792](https://www.pnas.org/content/116/3/792).

7. Clauset, A., Arbesman, S. & Larremore, D. B. Systematic inequality and hierarchy in faculty hiring networks. *Science advances* **1**, e1400005 (2015).

8. Marmaros, D. & Sacerdote, B. Peer and social networks in job search. *European economic review* **46**, 870–879 (2002).

9. Wu, L. Social network effects on productivity and job security: Evidence from the adoption of a social networking tool. *Information systems research* **24**, 30–51 (2013). Publisher: INFORMS.

10. Kossinets, G. & Watts, D. J. Empirical Analysis of an Evolving Social Network. *Science* **311**, 88–90 (2006).

11. Eagle, N., Pentland, A. S. & Lazer, D. Inferring friendship network structure by using mobile phone data. *Proceedings of the National Academy of Sciences* **106**, 15274–15278 (2009). URL [https://www.pnas.org/content/106/36/15274](https://www.pnas.org/content/106/36/15274) [https://www.pnas.org/content/106/36/15274.full.pdf](https://www.pnas.org/content/106/36/15274.full.pdf).

12. Frensch & Peter, A. Composition during serial learning: A serial position effect. *Journal of Experimental Psychology Learning Memory & Cognition* **20**, 423–442 (1994).

13. Reingen, P. H., Foster, B. L., Brown, J. J. & Seidman, S. B. Brand Congruence in Interpersonal Relations: A Social Network Analysis. *The Journal of Consumer Research* **11**, 771–783 (1984).
14. Reingen, P. H. & Kernan, J. B. Analysis of Referral Networks in Marketing: Methods and Illustration. *Journal of Marketing Research* **23**, 370–378 (1986).

15. Brown, J. J. & Reingen, P. H. Social Ties and Word-of-Mouth Referral Behavior. *The Journal of Consumer Research* **14**, 350–362 (1987).

16. Lazer, D. M. J. et al. Computational social science: Obstacles and opportunities. *Science* **369**, 1060–1062 (2020). URL [https://science.sciencemag.org/content/369/6507/1060](https://science.sciencemag.org/content/369/6507/1060). Publisher: American Association for the Advancement of Science Section: Policy Forum.

17. Manski, C. F. Identification of Endogenous Social Effects: The Reflection Problem. *The Review of Economic Studies* **60**, 531–542 (1993).

18. Shalizi, C. R. & Thomas, A. C. Homophily and Contagion Are Generically Confounded in Observational Social Network Studies. *Sociological Methods & Research* **40**, 211–239 (2011). URL [http://journals.sagepub.com/doi/10.1177/0049124111404820](http://journals.sagepub.com/doi/10.1177/0049124111404820).

19. Aral, S., Muchnik, L. & Sundararajan, A. Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *Proceedings of the National Academy of Sciences* **106**, 21544–21549 (2009). URL [http://www.pnas.org/content/106/51/21544.short](http://www.pnas.org/content/106/51/21544.short).

20. Snijders, T. A., van de Bunt, G. G. & Steglich, C. E. Introduction to stochastic actor-based models for network dynamics. *Social Networks* **32**, 44–60 (2010). URL [http://linkinghub.elsevier.com/retrieve/pii/S0378873309000069](http://linkinghub.elsevier.com/retrieve/pii/S0378873309000069).

21. Snijders, T., Steglich, C. & Schweinberger, M. *Modeling the coevolution of networks and behavior* (na, 2007). URL [http://ppswmm.ppsw.rug.nl/steglich/pdf/SnijdersSteglichSchweinberger.pdf](http://ppswmm.ppsw.rug.nl/steglich/pdf/SnijdersSteglichSchweinberger.pdf).

22. Steglich, C., Snijders, T. A. & West, P. Applying SIENA: An Illustrative Analysis of the Co-evolution of Adolescents’ Friendship Networks, Taste in Music, and Alcohol Consumption. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences* **2**, 48 (2006). URL [http://psycnet.apa.org/journals/med/2/1/48/](http://psycnet.apa.org/journals/med/2/1/48/).

23. Snijders, T. A., Van de Bunt, G. G. & Steglich, C. E. Introduction to stochastic actor-based models for network dynamics. *Social networks* **32**, 44–60 (2010).

24. Hosmer, D. W. & Lemesbow, S. Goodness of fit tests for the multiple logistic regression model. *Communications in statistics-Theory and Methods* **9**, 1043–1069 (1980).

25. Hoff, P. D., Raftery, A. E. & Handcock, M. S. Latent space approaches to social network analysis. *Journal of the american Statistical association* **97**, 1090–1098 (2002).
Supplementary Information

Data and Methods
Supplementary Text
Robustness Checks
Figs. S1 to S2
Tables S1 to S14
References (1-50)

Figure 1: The associations between student academic performance, degree centrality and job placements. (A) Network degree explains the CAP scores. The boxplots represent the CAP scores of all students, students in the High Degree group, students in the Middle Degree and in the Low Degree group. The asterisk in the Panel A indicate there exists a significant difference between all students, High Degree students and Low Degree students. (B) The average overall employment rates across CAP groups and degree groups. The boxplots show the values of $\text{mean} - 2\text{sd}$, $\text{mean} - \text{sd}$, $\text{mean}$, $\text{mean} + \text{sd}$, and $\text{mean} + 2\text{sd}$, respectively. (C) The average amount of annual salary across CAP groups and different degree groups. The boxplots illustrate the values of minimum, first quartile, median, third quartile and maximum of students’ annual salary. We also tested the difference of students’ mean employment status and starting salary within each CAP group using a Kruskal-Wallis test.
Figure 2: The dynamics of student academic performance and social networks for the Business School. (A) The flow of group members between high vs. middle vs. low GPA groups over time. $T$ from 1 to 8 indicates eight semesters and the colors of bars in each $T$ denotes the three GPA groups: green denotes high GPA (4-5) group, yellow denotes the middle GPA (3-3.99) group and red denotes the low GPA (<3) group. The length of bars indicates the number of students in that particular group. Similarly, the color and width of flows indicate the number of students transferred from the source group to the target group. For instance, the green colored flow indicates the mobility of students from a high performer group (i.e. source group) to other groups (i.e. target groups). (B) The proportion of same group neighborhoods over 8 terms is illustrated through the trends of homophily in the three academic groups. Here, the homophily is measured by the number of the same GPA group neighbors divided by the number of total neighbors. The homophily levels for the high, middle, low performing academic groups, and overall are shown in green, yellow, red, and black, respectively. (C) Dynamic processes of social network evolution and academic performance change. Nodes represent students and ties between nodes indicate the interpersonal interactions. The co-evolving processes of social interactions and academic performance suggests that there might be endogenous effects in assessing overall network effects during a four-year university education.
Figure 3: Decision making process spanning course selection, student collaborations, and academic performance. In stage (1), we model course selection in period $t$ as a function of preference similarity with linked students, performance in period $t-1$, and other covariates such as gender and nationality. In stages (2) and (3), we model the interpersonal network and academic performance changes jointly. To estimate peer influence, we control for two types of homophily effects: course-level homophily from stage (1), and individual-level homophily, stemming from attribute similarity and performance similarity.
Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SocialnetworksandeducationaloutcomesNatureSI.pdf