

Construction of Weighted Course Co-Enrollment Network

XunFei Li¹ and Renzhe Yu²

¹ University of California, Irvine, University of California, Irvine, Irvine, 92697, USA

² University of California, Irvine, University of California, Irvine, Irvine, 92697, USA

Abstract

The increasing availability of digitized campus administrative data provides researchers with the opportunity to systematically quantify how co-presence in classes shapes individual students' educational outcomes. Social network analysis is appropriate for this purpose through the construction of course co-enrollment network and network-based statistical models. This study intends to explore different ways to construct the course co-enrollment network and evaluate their capacity to capture meaningful student connections through courses. We specifically compare a simple unweighted co-enrollment network and a weighted network based on course characteristics along two dimensions: the relationship between network indices and students' academic performance, and the degree to which students with stronger weighted ties with each other experience more peer influence on individual performance than peers who are less connected through course co-enrollment.

Keywords

Social network analysis, course co-enrollment, network autocorrelation model, peer effect

1. Introduction

Course-taking experience is a critical part of undergraduate students' college life. Exposure to peers who take the same course might significantly impact individual academic achievement. The demographic composition (in regard to gender, ethnicity, etc.) of classmates shapes the socio-cultural contexts of students' academic experience, and the direct (such as group work) and indirect (such as presentations) interactions with peer students exert intangible influence on individual outcomes from time to time (Eckles & Stradley, 2012).

With the availability of campus administrative data, researchers are able to evaluate this important peer influence at scale. Among a few different methodological traditions, social network analysis (SNA) is appropriate for this purpose because it is one of the most used methods to study relational data and can explicitly model how students are connected through the course co-enrollment network as well as the effects of network properties on individual-level outcomes.

Studies applying SNA to course co-enrollment networks have found that network statistics such as degree and density contribute to explaining students' educational outcomes (Fincham et al., 2018; Israel et al., 2020; Weeden & Cornwell, 2020). However, the network edge in most of these studies is defined as a binary indicator of whether two students enroll in the course or not. This is a rather coarse proxy for peer exposure because the strength of connections between students in different courses largely varies with different course types, delivery formats, meeting schedules, among other factors. Given that

Proceedings of the NetSciLA21 workshop, April 12, 2021

EMAIL: xunfeil@uci.edu (A. 1); renzhey@uci.edu (A. 2)

ORCID: 0000-0002-2780-4493 (A. 1); 0000-0002-2375-3537



© 2020 Copyright for this paper by its authors.

Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

the relationship between network statistics and node-level outcomes is affected by how the network is constructed, using an overly simplified construction of course co-enrollment network might mask the actual effect of enrolling in the same course. To date, little effort has been put into examining alternative ways of constructing this network, and this study intends to investigate what network construction(s) best captures class-based peer influence. Specifically, by comparing different weighting strategies that leverage different course-level information, we aim to identify the construction approach that can best predict student achievement from network characteristics. The findings of this inquiry will inspire both researchers and practitioners to get deeper insights into students' college experience from administrative data which is largely standard and usable across different institutions.

2. Related work

2.1. Social Network Analysis and Course Co-Enrollment Networks

Social network analysis (SNA) has been used in studying educational contexts for a long time (Biancani & McFarland, 2013). Traditionally, students' friendship and residence-based networks have gained much attention for examining how significant others' preference and selection affect focal students' educational performance and behavior. At the micro-level, SNA has also been applied to students' posts in online discussion forums to understand how students interact with each other through discourse in individual classrooms (Fincham et al., 2018). While these networks capture different aspects of peer influence in college experience, they are either very context-specific (e.g., course design contexts for discussion forum networks) or require extensive data collection effort from researchers. These characteristics limit the scalability of such analyses.

As various campus-wide data become digitally available from the administrative end, some other aspects of peer influence become measurable on a larger scale and at low cost. A prominent example is course transcript data which can be used to construct course co-enrollment networks. Course co-enrollment captures the most important academic relations between students and their fellow students, but only a handful of studies in the field of higher education have examined how the structure of co-enrollment network relates to students' connections and behavior (Fincham et al., 2018; Weeden & Cornwell, 2020). As a cost of scalability, student-by-course enrollment records can only capture between-course variations in peer exposure and miss out variations in granular peer interaction within a class. Accordingly, the main challenge of constructing a course co-enrollment network is how to understand and model peer exposure and peer influence in relation to course contexts in a more accurate manner.

2.2. Approaches to Network Construction

Different network constructions represent researchers' understanding of the relation(s) being modeled. In friendship networks, the existence of a tie depends on students' self-reports of their best friend(s), which assumes that perceived intimacy between friends has a significant effect on individual students. Ties could also be constructed based on students' direct interactions. The discussion forum network, for example, usually defines a tie as a student's response to another student's post. Networks in the context of small groups such as study groups or orientation groups are also based on the assumption that students affect each other through direct interaction. Another type of network is co-presence networks which define ties as students' physical presence in the same space during the same time, such as networks constructed based on campus network data, course co-enrollment, and campus activity participation (Eckles & Stradley, 2012; Nguyen et al., 2020).

In the case of course co-enrollment network, it can either be constructed as a two-mode course-student network or be projected as two one-mode networks separately (student-student and course-course network). The network structure and tie definition could also be affected by the time span, node inclusion criteria and other research-specific concerns (Gardner et al., 2018; Israel et al., 2020; Weeden

& Cornwell, 2020). Weeden and Cornwell (2020) construct a two-mode course co-enrollment network with a single term's transcript data at Cornell University. Undergraduate, graduate and professional master students are connected to each other if they are in the same class at all in that term, and all the ties are treated equally. Israel and colleagues (2020) project a one-mode course network and a one-mode student network from the full two-mode co-enrollment network, which is based on one single cohort of students' course-taking data over six years. A student forms a tie with another student if they ever enrolled in the same class within six years after they enrolled, and the edge is weighted by the total number of co-enrolled courses. Gardner et al. (2018) use ten years of undergraduate course-taking records to build the network and further specify different edges through link attributes, which change according to the characteristics of co-enrolled peers.

2.3. Link Network Statistics to Students Educational Outcomes

Researchers have applied SNA to explore how network-level features and node-level indices could help understand the connection between students' social relations and their educational outcomes as well as how such relations form and evolve in contexts. Network-level indices such as density, betweenness centralization, clustering coefficient, and two-mode bi-component structure are used to examine overall how students are connected to each other, and how certain classes or students play critical linking roles (Israel et al., 2020; K. A. Weeden & Cornwell, 2020). Node level indices such as degree and demographic and academic features of peers in the network are examined in relation to students' educational outcomes such as retention rate (Eckles & Stradley, 2012), STEM preference (Raabe et al., 2019), and GPA based performance (Gardner et al., 2018). As discussed in Section 2.2, the specific network construction approach would affect the estimated relationship between network statistics and individual outcomes of interest (Fincham et al., 2018). However, previous studies on course co-enrollment networks did not further investigate this perspective.

3. Research questions

This study investigates different ways to construct course co-enrollment networks with course-level information from university administrative data. We specifically focus on weighting network ties by different pieces of course information such as course type, class size, and meeting schedule. The assumption is that co-enrolling in a course means different levels and effects of peer exposure in different course contexts. For example, students may have more in-depth connections in small seminars than in large lectures, in classes with more frequent meeting schedules than in courses with fewer opportunities to meet. This course-relevant information would affect the strength of students' connection through the course co-enrollment network.

RQ 1:

What are the different ways of constructing co-enrollment networks weighted by course information from campus administrative data?

To further validate which construction approach more effectively captures students' connection in different course contexts, we employ two modeling perspectives. We first examine the predictive power of local network statistics on individual outcomes in each network construction. The assumption is that a stronger predictive relationship would indicate a more valid network construction.

RQ 2:

Is the relationship between network indices and students' academic performance in an unweighted baseline co-enrollment network different from that in a weighted network?

We also examine how individual students' academic performance correlates with each other in each network construction through network autocorrelation models. We assume that in a valid co-enrollment network, peers with heavier weights on their connections have stronger correlations in their performance.

RQ 3:

How does the autocorrelation model fit on a weighted co-enrollment network compared to an unweighted baseline network?

4. Methods and proposed analyses

4.1. Data

The data used in constructing the course co-enrollment network come from the administrative data from a large four-year public university in the United States. The administrative data includes student-level courses-taking records and grades, and the course-level information for full-time undergraduate students across multiple years. This context carries good representative value for research on co-enrollment networks for a few reasons. First, the large public university includes a variety of majors and schools that are commonly in place at other institutions. Second, students come from very different family backgrounds including those that are traditionally underrepresented. Third, courses at the university have a variety of class size and delivery format, providing sufficient variations in course contexts and the corresponding network constructions. In this study we restrict our analysis to the data from 2015 to 2020 in order to follow the complete college experience of students from the 2015 and 2016 cohort. We only include course enrollment records for students who completed a course and got a valid grade.

4.2. Course Co-Enrollment Network Construction

4.2.1. Baseline Network

The course co-enrollment network is constructed as a one-mode network that each node represents one student (Zhou et al., 2007). Students have ties with other students if they enrolled and completed the same class. The network is an $m \times m$ matrix that m equal to the total number of students in that term excluding students who were only in courses with only one student or students who failed all classes.

Each cell in the matrix presents the weight of the tie of row m student and column n student. If they went to and completed the same class then their cell would be filled with 1 instead of 0. If row m student and column n student enrolled and completed more than one class, the cell would be filled with the total overlapping courses they had.

4.2.2. Weighted Ties

In the baseline network, the existence of ties between two students solely depends on whether they completed the same courses together, but in reality not all ties are equal. Considering the differences in course contexts, we further add the edge weight based on the combination of different aspects of course-level information. The specific course features we use include:

- Course types, including lecture, seminar, lab, and discussion. Different types correspond to different edge weights in the co-enrollment network based on the chance of interaction they generally offer to students. The order from the most to the least weighted course type is seminar, discussion, lab, and lecture;
- Course schedule (meeting times). Courses that meet more often correspond to larger edge weight than courses with fewer meetings (Srinivasan et al., 2006);
- Class size. Smaller courses lead to larger edge weight because the chance of interaction between students there is higher than in larger classes;
- Courses level (upper-division vs. lower division). Upper-division courses are weighted heavier than lower-division courses since they generally expect more engagement from students.

4.3. Network Autocorrelation Model

The network autocorrelation model enables us to analyze the social influence process among people in an interdependent network (Leenders, 2002). In the autocorrelation model, ego's endogenous outcome variable is not only affected by the ego's own covariates but also affected by other alters in the same network with the ego. The strength of alters' influence is determined by the weight matrix in the autocorrelation model.

In this study, students' term GPA would be the endogenous outcome variable, and the covariates include students' cumulative GPA before the term and demographic characteristics (gender, race, first-generation college student status, low-income status). In the baseline network, the weight matrix is defined as described in Section 4.2.1; in the weighted network, the weight matrix is further computed from the weighted ties following Section 4.2.2. By comparing the model fit on these different network constructions, we can evaluate if incorporating more course information could capture more accurate strength of students' influence to each other in the course co-enrollment.

5. Discussion

This proposed study is contextualized in a specific usage of campus administrative data: understanding students' connection and peer influence through course co-enrollment. We focus on finding the optimal approach to constructing co-enrollment networks from both student transcripts and course-level metadata, largely because these administrative records only reflect co-presence and the actual peer exposure and influence needs to be inferred. While the two analytical perspectives we take (network statistics in relation to individual outcome; network autocorrelation model) aim at evaluating the different network constructions, the results in turn could provide insights into how college students' academic connection with each other varies with course characteristics. For policymakers, this is informative for them to better tailor academic and curricular policies to the goal of promoting student success.

6. Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant No.153500 and the Andrew W. Mellon Foundation under Grant No.1806-05902.

7. References

- [1] Biancani S, McFarland DA. Social networks research in higher education. In *Higher education: Handbook of theory and research 2013* (pp. 151-215). Springer, Dordrecht. doi: https://doi.org/10.1007/978-94-007-5836-0_4
- [2] Eckles, James E., and Eric G. Stradley. "A Social Network Analysis of Student Retention Using Archival Data." *Social Psychology of Education*, vol. 15, no. 2, 2012, doi:10.1007/s11218-011-9173-z.
- [3] Fincham, Ed, et al. "From Social Ties to Network Processes: Do Tie Definitions Matter?" *Journal of Learning Analytics*, vol. 5, no. 2, 2018, doi:10.18608/jla.2018.52.2.
- [4] Gardner, Josh, et al. "Learn From Your (Markov) Neighbour: Co-Enrollment, Assortativity, and Grade Prediction in Undergraduate Courses." *Journal of Learning Analytics*, vol. 5, no. 3, Society for Learning Analytics Research, Dec. 2018, pp. 42–59, doi:10.18608/jla.2018.53.4.
- [5] Israel, Uriah, et al. "Campus Connections: Student and Course Networks in Higher Education." *Innovative Higher Education*, vol. 45, no. 2, Innovative Higher Education, 2020, pp. 135–51, doi:10.1007/s10755-019-09497-3.
- [6] Leenders, Roger Th A. J. "Modeling Social Influence through Network Autocorrelation: Constructing the Weight Matrix." *Social Networks*, 2002, doi:10.1016/S0378-8733(01)00049-1.
- [7] Nguyen, Quan, et al. "Exploring Homophily in Demographics and Academic Performance Using Spatial-Temporal Student Networks." *Proceedings of The 13th International Conference on Educational Data Mining (EDM 2020)*, 2020, pp. 194–201.
- [8] Raabe, Isabel J., et al. "The Social Pipeline: How Friend Influence and Peer Exposure Widen the STEM Gender Gap." *Sociology of Education*, 2019, doi:10.1177/0038040718824095.
- [9] Srinivasan, Vikram, et al. "Analysis and Implications of Student Contact Patterns Derived from Campus Schedules." *Proceedings of the 12th Annual International Conference on Mobile*

Computing and Networking (MobiCom '06), ACM Press, 2006, pp. 86–97, doi:10.1145/1161089.1161100.

- [10] Weeden, Kim A., and Benjamin Cornwell. “The Small-World Network of College Classes: Implications for Epidemic Spread on a University Campus.” *Sociological Science*, vol. 7, 2020, pp. 222–41, doi:10.15195/V7.A9.
- [11] Zhou, Tao, et al. “Bipartite Network Projection and Personal Recommendation.” *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, vol. 76, no. 4, 2007, doi:10.1103/PhysRevE.76.046115.