UC Irvine
UC Irvine Electronic Theses and Dissertations

Title
Ranking and Sorting: An Examination of Positional and Relational Dynamics in Higher Education

Permalink
https://escholarship.org/uc/item/98z4b69v

Author
Jeffrey, Wesley

Publication Date
2024

Copyright Information
This work is made available under the terms of a Creative Commons Attribution License, available at https://creativecommons.org/licenses/by/4.0/

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA,
IRVINE

Ranking and Sorting:
An Examination of Positional and Relational Dynamics in Higher Education

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Sociology

by

Wesley Jeffrey

Dissertation Committee:
Professor David R. Schaefer, Chair
Professor Andrew Penner
Professor Richard Arum

2024

Chapter 1 © Nature Publishing Group
Chapter 2 © Nature Publishing Group
All other materials © 2024 Wesley Jeffrey

DEDICATION

To my parents who continue to teach me.

To Fátima and Camila whose love and support made this possible.

To my family and community who continue to fuel my passion.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to express appreciation to my committee chair, Professor David Schaefer, who introduced me to social networks and helped to strengthen my quantitative skills. His in-depth and timely feedback greatly improved the research findings presented here.

I would also like to thank my committee members, Professor Andrew Penner and Professor Richard Arum, whose work inspired me to pursue research related to stratification and learning in higher education.

Chapter 1 of this dissertation is a reprint of the material as it appears in Jeffrey, Wesley, and Benjamin G. Gibbs. 2024. “Unpacking the Gap: Socioeconomic Background and the Stratification of College Applications in the United States.” *Research in Higher Education* doi:10.1007/s11162-024-09780-z, used with the permission of Nature Publishing Group.

Chapter 2 of this dissertation is a reprint of the material as it appears in Jeffrey, Wesley, David R. Schaefer, Di Xu, Peter McPartlan, and Sabrina Solanki. 2022. “STEM Learning Communities Promote Friendships but Risk Academic Segmentation.” *Scientific Reports* 12(12442):1–9, used with the permission of Nature Publishing Group.

VITA

Wesley Jeffrey

2016	B.S. in Sociology, Brigham Young University
2018	M.S. in Sociology, Brigham Young University
2021	M.A. in Social Science, University of California, Irvine
2024	Ph.D. in Sociology, University of California, Irvine

FIELD OF STUDY

Higher Education, Stratification/Inequality/Mobility, Networks, Organizations

PUBLICATIONS

Jeffrey, Wesley, and Benjamin G. Gibbs. 2024. "Unpacking the Gap: Socioeconomic Background and the Stratification of College Applications in the United States." *Research in Higher Education* doi: 10.1007/s11162-024-09780-z.

Jeffrey, Wesley, David R. Schaefer, Di Xu, Peter McPartlan, and Sabrina Solanki. 2022. "STEM Learning Communities Promote Friendships but Risk Academic Segmentation." *Scientific Reports* 12(12442):1–9.

ABSTRACT OF THE DISSERTATION

Ranking and Sorting:

An Examination of Positional and Relational Dynamics in Higher Education

by

Wesley Jeffrey

Doctor of Philosophy in Sociology

University of California, Irvine, 2024

Professor David R. Schaefer, Chair

Sociologists of education have a longstanding interest in studying the relationship between schooling and inequality in society. While we know education matters for who gets ahead, we still know relatively less about the processes and mechanisms behind this relationship. In my dissertation, I focus on higher education as a key site where stratification plays out and examine various ways in which individuals are ranked and sorted relative to one another. Conceptually, I extend insight into the *positional* (i.e., hierarchical) and *relational* (i.e., network) dynamics of postsecondary schooling. I draw upon sociological theories relevant to the social stratification, organizations, and social networks literatures. Through both regression-based and social network analyses, my dissertation adds to our understanding of differential sorting into and through the first year of college. In so doing, I shed additional light on *how* and *why* students end up at distinct institutions, as well as the causes and consequences of the social networks that emerge once they arrive on campus.

INTRODUCTION

> “Differentiating humans means distinguishing them perceptively and shaping them physically, categorizing them lingually and sorting them classificatorily, but also subjecting them to unequal evaluative treatment, segregating them spatially, and separating them permanently on an institutional basis.”
>
> -Stefan Hirschauer

Sociologists of education have a longstanding interest in studying the relationship between schooling and inequality in society (Jencks 1972). A key insight from this line of work is that education is a principal mechanism linking social origin to occupational destination (Blau and Duncan 1967). Although sociologists have drawn much attention to the potential role of education in stratifying populations, we still know relatively less about *how* and *why* this occurs.

To shed light on possible processes and mechanisms relevant for the intergenerational transmission of advantage, my dissertation focuses on higher education as a key site where stratification is likely to occur. In turn, here I study the transition to college, as this period represents a critical juncture when disparities are likely to emerge (Jeffrey 2020). Theoretically, I aim to analyze *positional* (i.e., hierarchical) and *relational* (i.e., network) dimensions of postsecondary schooling, and their consequences for inequality and learning in higher education.

First, an educational credential can be viewed as a positional good in that its value stems in part from its relative scarcity in the population (Shavit and Park 2016). However, educational credentials are also positional in that different degrees carry unequal earnings potential. For example, research has documented differential returns by institutional selectivity (Chetty, Deming, and Friedman 2023), major or field of study (Wolniak et al. 2008), and academic performance in college (Rumberger and Thomas 2000). Thus, each of these forms of horizontal stratification in higher education (Gerber and Cheung 2008) points to ways in which students may be ranked relative to one another. Importantly, these dimensions of ranking take place across and within institutions.

Second, a relational view of education draws our attention to the networks in which students are embedded (Kolluri and Tierney 2020) and how social connections may impact student outcomes (Felten and Lambert 2020). Sociologists understand that friendships matter in school settings and that these relationships may be especially valuable for traditional-aged college students who are making the transition to adulthood (Epstein and Karweit 1983). College students spend a significant amount of time with their friends, and as such, they represent an important form of peer social capital (McCabe 2016). Consequently, better understanding the nature of who is connected to whom and how friends may influence one another over time seems critically important.

In Chapter 1, I combine insight from Boudon's (1974) inequality of educational opportunity framework with the status attainment perspective to study the gap in college application selectivity between higher- and lower-SES students. Drawing upon data from the *High School Longitudinal Study of 2009*, I investigate the contribution of theoretical factors to the class-based gap in the selectivity of college applications. Through a series of Heckman selection models, as well as a Blinder-Oaxaca decomposition analysis, I study *how* and *why* students differ in their application behavior by SES.

In Chapter 2, I draw upon focus theory to study how organizational practices impact the friendship networks of entering students. Namely, utilizing a novel methodological approach, I combine a quasi-experimental design with social network analysis to examine how placement into a freshman learning community shapes the *size*, *strength*, *structure*, and *composition* of student friendship networks. I use two freshman biological sciences cohorts at a large, selective public university to study these relationships.

In Chapter 3, I extend the concept of curricular differentiation to higher education to study how routine curricular practices shape the friendships that arise. I draw upon complete longitudinal friendship network data across two freshman cohorts to examine *how* and *to what extent* allocating students to curricular groups and grading their performance in class shapes 1) processes of friend selection at the dyadic level and 2) friendship clustering at the network level. I utilize a set of stochastic actor-oriented models, as well as two follow-up analyses, to understand the nature and magnitude of these effects.

Overall, my dissertation project entails an examination of various ways in which individuals are ranked and sorted in higher education. Conceptually, I build upon prior stratification research to reveal different positional and relational mechanisms within postsecondary schooling that may have implications for intergenerational mobility in the US context. Methodologically, my dissertation includes analyses that span the macro, meso, and micro levels. Through both large-scale, nationally representative data as well as smaller-scale, in-depth longitudinal social network data, this project illuminates several processes of stratification occurring at different levels of analysis. It also draws attention to where further research is needed. For example, findings presented here point to the need for more insight into the link between social networks and learning in higher education.

CHAPTER 1

Unpacking the Gap:

Socioeconomic Background and the Stratification of College Applications in the United States

ABSTRACT

While a substantial body of work has shown that higher-SES students tend to apply to more selective colleges than their lower-SES counterparts, we know relatively less about *why* students differ in their application behavior. In this study, we draw upon a sociological approach to educational stratification to unpack the SES-based gap in college application selectivity. Using data from the *High School Longitudinal Study of 2009,* we examine the contribution of theoretical factors to the class-based gap in the selectivity of college applications. Namely, from the rational action model we estimate the contribution of *performance differentials* and *choice differentials*, while from the status attainment model we look at the *level* and *type* of *educational expectations* as well as the *number of applications* submitted. Through a series of Heckman selection models, as well as a Blinder-Oaxaca decomposition analysis, we can explain 85% of the gap in college application selectivity between students in the top and bottom SES quintiles. In turn, we estimate that 60% of this explained portion is due to rational action mechanisms such as grades and test scores while 35% is due to status attainment mechanisms. Finally, we reveal that SES moderates the relationship between type of expectations and application selectivity. We find that the payoff to higher expectations (in terms of selective applications) disproportionately accrues to higher-SES students.

Keywords: college application selectivity, socioeconomic status, rational action, status attainment, educational expectations

INTRODUCTION

Sorting into higher education entails a complex set of actions of both students and colleges (Grodsky and Jackson 2009). Even so, one step that is especially influential in shaping where students ultimately enroll is the application process (Bowen, Chingos, and McPherson 2009; Holzman, Klasik, and Baker 2020). Mirroring enrollment gaps by social background (Alon 2009; Chetty et al. 2020; Reardon, Baker, and Klasik 2012), a growing number of studies have documented application disparities across socioeconomic groups (An 2010; Holland 2014; Mullen and Goyette 2019). Namely, high-SES students are significantly more likely to apply to selective institutions compared to their low-SES counterparts (Hoxby and Avery 2012; Mullen and Goyette 2019; Radford 2013). While it is well documented that these gaps exist, we know less about *why* students vary in their application behavior.

Comparing two well-developed sociological theories linking social origins to educational stratification, we undertake a quantitative investigation of how and why high- and low-SES students differ in their college application behavior. First, applying the rational action model developed within Boudon's (1974) framework of inequality of educational opportunity (IEO), we analyze how *performance differentials* (or "primary effects") and *choice differentials* (or "secondary effects") vary by social class and what role they play in explaining SES-based gaps in college applications. Second, building upon the status attainment model (Sewell, Haller, and Portes 1969), we examine how unequal *educational expectations* contribute to class-based disparities in where students apply to college. Importantly, we measure educational expectations in terms of not only the level of education students plan to obtain (i.e., how far) (Mullen and Goyette 2019), but also the type of college students plan to attend (i.e., where) (Gerber and Cheung 2008). Finally, we include the *number of applications* submitted as a strategy differentially employed by students of varying socioeconomic backgrounds to improve their chances of admission to selective colleges (Radford 2013).

We utilize data from the most recent nationally representative sample of high school students in the United States—the *High School Longitudinal Study of 2009* (HSLS:09)—to examine: 1) how performance differentials, choice differentials, unequal educational expectations, and the number of applications submitted contribute to the class-based gap in college application selectivity, and 2) whether the link between expectations and applications differs by socioeconomic background. While past work in this area has tended to focus on only the highest-performing students (Hoxby and Avery 2012; Lor 2023; Radford 2013), we take a broader approach and examine the application behavior of all high school graduates. Moreover, in this analysis, we do not impose restrictions in terms of which students we deem a "match" for a particular type of institution (Mullen and Goyette 2019; Roderick, Coca, Nagaoka 2011; Roksa and Deutschlander 2018; Smith, Pender, and Howell 2013). For example, because colleges and universities generally rely upon a host of factors when making admissions offers—not just academic metrics—we prefer an open approach to college applications rather than restrict our analysis to an overly narrow set of measures.

To preview our results, we can explain 85% of the gap in college application selectivity between students in the top and bottom SES quintiles. We estimate that 60% of this explained portion is due to rational action mechanisms (e.g., GPA, standardized test scores, etc.), while 35% is due to status attainment mechanisms (i.e., educational expectations and number of applications). In turn, we reveal a significant interaction between SES and type of expectations on application selectivity. For instance, we estimate that average-performing, low-SES students with the highest expectations have a 31% predicted probability of applying to selective colleges, compared to 46% among their high-SES counterparts.[1] Overall, we believe this study provides the most comprehensive quantitative investigation to date into the drivers of the SES-based gap in college application selectivity.

BACKGROUND

Over the past few decades, a growing literature has drawn attention to SES-based gaps in where students apply to college. Prior work analyzing nationally representative datasets of students in the U.S. has consistently shown differential rates of application to selective colleges by socioeconomic background. For example, using data from the *National Education Longitudinal Study of 1988*, Cabrera et al. (2000a) highlight significant gaps in the probability of applying to four-year colleges by SES quartile. Likewise, drawing upon the *Education Longitudinal Study of 2002*, several studies have shown that higher-SES students are more likely to apply to selective colleges compared to their lower-SES counterparts (An 2010; Mullen and Goyette 2019; Roksa and Deutschlander 2018). Finally, recent work using the *High School Longitudinal Study of 2009*, has documented significant gaps in the selectivity of college applications between students in the top and bottom income quartiles (Holzman, Klasik, and Baker 2020). While these and other studies have provided strong evidence that SES shapes stratified college applications, we know relatively less about why. Comparing two sociological models of educational stratification, we aim to unpack the various mechanisms that give rise to these unequal patterns.

Rational Action Model of Stratified Applications

[1] All other factors were held to their mean values (see Figure 1.3).

One potential sociological framework for understanding the SES-based gap in college applications is the rational action model (Grodsky and Jackson 2009) of educational inequality developed most in-depth by Boudon (1974). According to this perspective, social origins shape unequal educational outcomes via performance differentials (or primary effects) and choice differentials (or secondary effects). Whereas the former mechanism draws attention to the role of the achievement gap in linking socioeconomic background with educational stratification, the latter mechanism looks at variation net of (or conditioning on) performance disparities (Jackson 2013). While we know that both mechanisms likely matter for where students apply, relatively little work has estimated the relative contribution of each mechanism to stratified college applications by class background.

First, according to primary effects of the rational action model, students from varying socioeconomic backgrounds may sort into distinct pools of applicants by aligning the selectivity of their applications with their performance levels—as measured by grades and standardized test scores. This would occur if students only applied to schools that seemed to "match" their high school academic credentials (Hoxby and Avery 2012; Mullen and Goyette 2019). In general, past research has shown that academic achievement measures are associated with selective applications (An 2010) and that gaps in academic qualifications help to explain stratified college destinations by family income (Bastedo and Jaquette 2011; Holzman, Klasik, and Baker 2020). However, it remains unclear *how much* performance differentials contribute to the SES-based gap in college applications.

Second, according to secondary effects of the rational action model, there are reasons to suspect that even after accounting for performance disparities, students may still apply to different types of colleges based on their class origins. This could arise for a variety of reasons.

For example, students from varying social backgrounds may have differential access to information about college and ways to finance their education (Hoxby and Avery 2012; Hoxby and Turner 2013; McDonough 1997; Robinson and Roksa 2016). They may also face differing constraints in terms of the factors they consider most important when making their postsecondary decisions (Hossler, Schmit, and Vesper 1999; Mullen 2011; Perna 2006; Roksa and Deutschlander 2018). Indeed, qualitative research of high-performing students has uncovered various ways in which socioeconomic background continues to shape the types of colleges students apply to, even among this very select group (Lor 2023; Radford 2013).

Status Attainment Model of Stratified Applications

A second sociological framework for understanding the SES-based gap in college applications is the status attainment model (Grodsky and Jackson 2009). According to this perspective, differential expectations are a critical mechanism linking social origins with stratified educational outcomes. Indeed, pioneering work among sociologists within the status attainment tradition demonstrated the role of expectations in linking socioeconomic background with unequal educational destinations (Haller and Portes 1973; Sewell, Haller, and Portes 1969). In turn, recent work has revealed that differential educational expectations help to explain SES-based gaps in where students apply to college (Mullen and Goyette 2019).

However, until now most research has been limited to measures of educational expectations in terms of *how far* students plan to go in school (i.e., level). We argue that in the current era of increasing access, differentiation, and competition within higher education (Alon 2009; Bastedo and Jaquette 2011; Mullen 2011; Taylor and Cantwell 2019), student expectations for *where* they will attend college (i.e., type) also matters. Namely, a substantial body of work provides reasons to suspect that more advantaged students may develop selective expectations for certain kinds of educational credentials that can facilitate their access to high-status positions in society (Goyette 2008; Lucas 2001; Mullen 2011; Reay, David, and Ball 2005). These heightened expectations could arise due to more frequent and active discussions surrounding college for high-SES students that take place at home, among peers, and with institutional agents such as high school counselors or private college consultants (Cabrera and La Nasa 2000b, McDonough 1997, Perna and Titus 2005; Roksa and Deutschlander 2018). The culmination of varying family and schooling environments may lead high-SES students to develop a sense of "entitlement" for a particular type of collegiate education (Ford and Thompson 2016; McDonough 1997:9; Roderick, Coca, Nagaoka 2011).

Additionally, due to increasing competition within higher education, we also examine the extent to which the number of applications students submit contributes to the SES-based gap in college application selectivity. For instance, students may submit more applications as a strategy to improve the likelihood that they are accepted to at least one selective college (Roderick, Coca, Nagaoka 2011). Not surprisingly, advantaged students are better positioned to handle the material costs of submitting applications when fee waivers are not available or widely known (Hoxby and Turner 2013). Indeed, past work has shown that higher-SES students submit more college applications, on average (Mullen and Goyette 2019; Radford 2013), and that this practice helps to explain class-based disparities in where students apply (Mullen and Goyette 2019; Roksa and Deutschlander 2018).

In sum, both the rational action model and the status attainment model highlight ways in which class background may shape stratified college applications. While the rational action model draws attention to the costs and benefits of applying to different schools, the status attainment model focuses on the socialized and taken-for-granted aspects of the college application process.

Differential "Returns" to College Plans

Finally, scholars of social stratification argue that differential return processes can also contribute to inequality between groups (Persell, Catsambis, and Cookson Jr. 1992). For example, in our case, high-SES students may not only benefit from the types of colleges they plan to attend but also from greater return on their expectations. This could occur, for instance, if the association between the type of college students expected to attend, and the selectivity of their applications grew as SES increased.

There are a couple reasons to suspect that SES may moderate the relationship between expectations and college applications in this way. First, high-SES students may be more likely to enact their college plans due to greater familiarity with the concrete steps necessary to apply to top colleges (Morgan 2018). Second, even among those who plan to attend a selective college, barriers during the actual application process may differentially impact students from varying class backgrounds. For example, recent work shows that the complexity of the college application process—especially the essay portion—may unequally lead low-SES students to start but not finish their application submissions (Odle and Magouirk 2023).

DATA AND METHODS

For this analysis, we draw upon the first follow-up wave (carried out in 2012) and the 2013 update of the High School Longitudinal Study of 2009 (HSLS:09). The data are a nationally representative sample of 9th graders from more than 900 high schools (public and private) collected by the National Center for Education Statistics (NCES) in 2009. As stated by NCES, the purpose of the data collection is to monitor the transition of a national sample of adolescents from their high school experiences through their postsecondary years.

The data were collected during the spring of their junior year, the spring of their senior year, and three years after high school graduation. One of the innovative features of the HSLS is an enhanced focus on the dynamics of educational decision-making, especially as it relates to college choice factors (Ingels et al. 2013). Thus, the HSLS is an ideal dataset for understanding how and why SES-based disparities arise among students during the application stage of the college-going process.

The analytic sample for this study is restricted to students who have acquired a high school degree or equivalent, since we are interested in looking at college application decisions. Missing values were addressed using multiple imputation with chained equations (MI) in Stata with ten imputed datasets. We use the dependent variable in the imputation equations, but all analyses are estimated using only non-missing values of the dependent variable (Von Hippel 2007).

Measurement

College Application Selectivity

The dependent variable for this study is the selectivity of college applications as indicated by the *highest institutional selectivity* among top choice schools applied to or registered at (up to three available through the 2013 HSLS update data file).[2] Institutional selectivity is measured using the 2011-2012 admissions rate (including open admissions) collected by the NCES

[2] Ideally, we would have information on every school the student applied to, but HSLS is limited to information on up to three schools the student "most seriously considered". This is less problematic, however, given that students in the sample applied to less than three schools, on average.

Integrated Postsecondary Education Data System (IPEDS). To ease interpretation, we reverse code the measure so that a higher value corresponds to greater selectivity (i.e., higher rejection rate). We acknowledge that much of the prior work in this area has drawn upon the *Barron's Competitiveness Index* measure to examine institutional selectivity (An 2010; Brewer, Eide, and Ehrenberg 1999; Holzman, Klasik, and Baker 2020; Roksa and Deutschlander 2018). Supplementary analyses utilizing this measure of selectivity produce strikingly similar results (see S1.1). Ultimately, we utilize the continuous IPEDS measure over the categorical Barron's measure.

Socioeconomic Status (SES)

The main independent variable is a measure of the student's socioeconomic background as indicated by their SES composite score. In the HSLS:09 dataset, this variable is a combined index of parental education level, parental occupational prestige score, and family income (see Ingels et al. 2013 for details). While the main analyses draw upon the continuous measure of SES, we also utilize SES quintiles to compare gaps in the outcome between those in the top 20% and bottom 20%.[3]

Performance Differentials

To estimate how differential academic performance contributes to the SES-based gap in college applications, we draw upon several high school academic metrics. We include overall *11th-grade GPA*, since this most accurately represents the period surrounding a students' expectations during the spring of their junior year and will be what the students ultimately use to apply to colleges the following fall of their senior year. We also include a dichotomous indicator

[3] Supplementary analyses utilizing deciles or quartiles produce similar results (see S1.2, S1.3).

of *AP coursework* (yes=1; no=0)—which is a measure of whether the student has taken any AP classes as of spring of their junior year—since this has been shown to matter for enrollment at top colleges (Espenshade and Radford 2009). Additionally, we account for the students' *standardized test score* since research shows this metric may influence how students gauge their own competitiveness for college admissions (Meyer 1970). This measure indicates 1) the composite SAT score for students who took the test, 2) the converted equivalent using a respondents' ACT score, or 3) the predicted equivalent using the standardized theta score (T-score) gathered as part of the HSLS-administered 11^{th}-grade math assessment.

Choice Differentials

To estimate how choice differentials contribute to the SES-based gap in college applications, we include a host of variables that measure access to informational resources as well as college considerations. In terms of informational resources we include: *attendance of a program at, or taken a tour of, a college campus* (yes=1; no=0); *searching for college options through the internet or through reading college guides* (yes=1; no=0); *talking with a high school counselor* (yes=1; no=0); and *talking about options with a counselor hired to prepare for college admission* (yes=1; no=0). We also include a measure of *taking a course to prepare for a college admission exam* (yes=1; no=0).

In terms of college considerations, we account for the importance of several factors in the decision-making process. Distance measures whether *being close to home* is an important consideration (very important=1; somewhat important/not at all important=0). Differential perceptions of cost are measured in terms of the importance of *cost of attendance* (very important=1; somewhat important/not at all important=0). *Academic quality/reputation* measures the importance of institutional prestige (very important=1; somewhat important/not at all important=0), and *family/friend recommendation* (very important=1; somewhat important/not at all important=0) as well as *family legacy* (very important=1; somewhat important/not at all important=0) capture the familial influence component of the college choice decision. We also include the importance of whether the *degree program* of interest is offered at the school (very important=1; somewhat important/not at all important=0), the importance of *graduate school placement* (very important=1; somewhat important/not at all important=0), *job placement* (very important=1; somewhat important/not at all important=0), the importance of the opportunity to *play school sports* (very important=1; somewhat important/not at all important=0) and the perception of *campus social life* / *school spirit* (very important=1; somewhat important/not at all important=0).

Educational Expectations

We measure educational expectations in terms of both level and type. *Educational expectations (level)* is the conventional measure of expectations and indicates how far a student plans to go in school (less than high school=1; high school=2; some college=3; associate's or AA degree=4; bachelor's or BA degree=5; graduate or professional degree=6).[4] *Educational expectations (type)* indicates the kind of college students plan to attend after high school based on 2011-2012 IPEDS admissions data. To create this measure, we use the student questionnaire from the first follow-up wave in 2012, when students were largely in their junior year. In a section of the interview about future plans and preparations, students were asked "What [school that provides occupational training/ 2-year college/ 4-year college/school or college] are you most likely to attend? (Please type in the full name. Do not use abbreviations)." Thus, this

[4] About 9% of students responded "don't know". We coded these students as missing and estimated a value for them during the imputation process.

question captures a students' college plans unconditioned by whether they actually applied or were admitted (Niu and Tienda 2008). The responses were coded in the same way we coded the selectivity of college applications.[5]

Number of Applications

We include a measure of the *number of applications* submitted, since this likely varies by SES (Radford 2013) and has been shown to matter for the types of colleges students apply to (Mullen and Goyette 2019; Roderick, Coca, Nagaoka 2011).

Control Variables

We account for other covariates at the individual level that could also influence the selectivity of college applications. Specifically, we include measures of *race/ethnicity*—(non-Hispanic white=reference group) with indicators for Hispanic or Latinx, Black/African American, Asian/Asian American, and multiracial/other—and *gender* (female=1; male=0).

At the school level, there are a host of factors that could also influence the selectivity of college applications among students from diverse backgrounds. Namely, the types of high schools that low- and high-SES students attend likely differ in terms of their ability to promote selective college applications (Roderick, Coca, and Nagaoka 2011). Since we know students sort into different high schools based on their family income, we include aspects of the schooling context that past research has shown to differentially promote the transition to college for students from varying socioeconomic backgrounds (Turley 2009). Specifically, we include measures of *school control* (public=0; Catholic=1; other private=2), *school type* (regular=0;

[5] Around 27% of students responded "don't know". We coded these students as missing and estimated a value for them during the imputation process. Supplementary analyses utilizing listwise deletion (see S1.4) as well as analyses of those who responded "don't know" (see S1.5) show similar patterns as the main results.

charter school=1; special program school=2; other including career/ technical/ vocational/ and alternative=3), *school urbanicity* (city=0; suburb=1; town=2; rural=3), and *geographic region* (New England=0; Middle Atlantic=1; East North Central=2; West North Central=3; South Atlantic=4; East South Central=5; West South Central=6; Mountain=7; Pacific=8). Lastly, we account for *high school size* using a measure of the total enrollment of students in grades 9-12, and the *percent low-income* which indicates the percent of the student body receiving free or reduced-price lunch.

Analytic Strategy

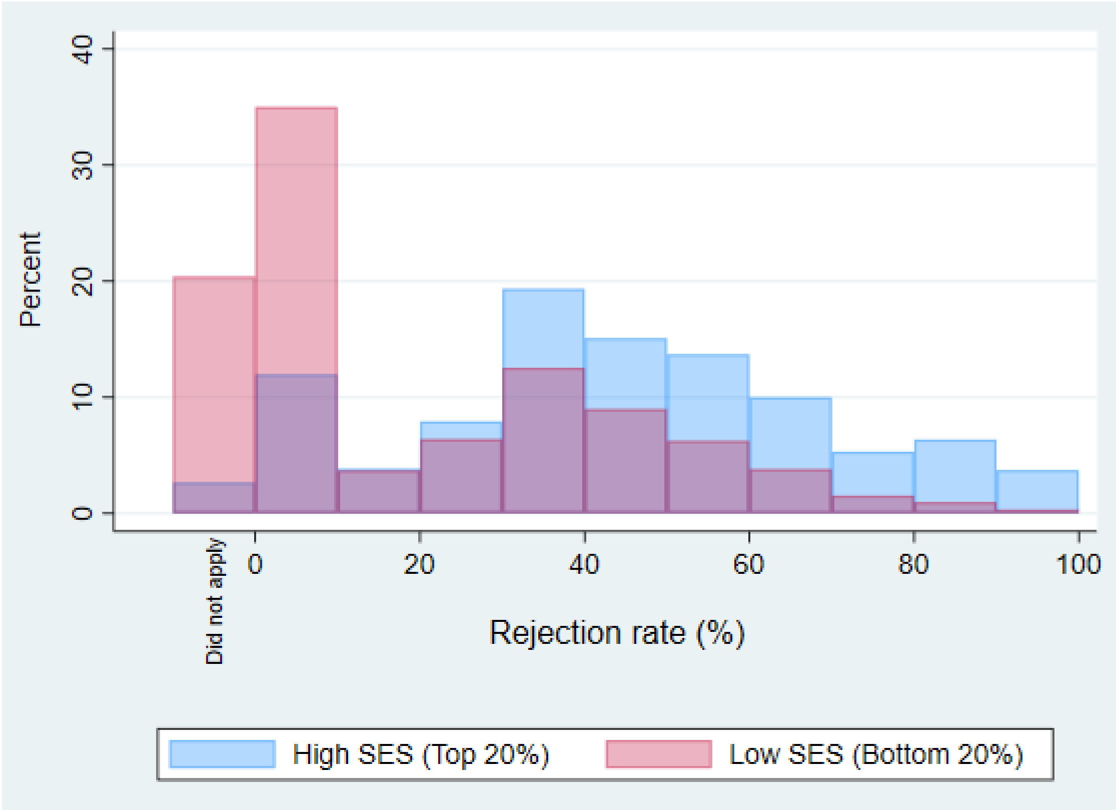
To examine how and why high- and low-SES students differ in their college application behavior, we draw upon a series of Heckman regression models with selection to estimate the selectivity of college applications using measures of performance differentials, choice differentials, unequal educational expectations, and the number of applications submitted. In general, the class of Heckman models are a common strategy for dealing with sample selection bias when observation of the outcome is not missing at random (Heckman 1979; Stolzenberg and Relles 1997). Thus, Heckman selection models are appropriate for purposes of our study since, in our case, we expect systematic differences in who applies to college (i.e., censoring bias). For example, we know SES predicts whether a student will apply to any college (Odle and Magouirk 2023). We utilize the *heckman* command in Stata to implement a two-step model that first estimates a probit equation to predict the probability of applying to college. The second equation then fits a regression model of the highest selectivity level of college applications, conditional on applying. For this analysis, we draw upon the same set of predictors in both the selection and outcome equations, except we omit the number of applications in the selection equation, since inclusion of this variable leads to model degeneracy.

To compare the relative contribution of the rational action model with the status attainment model, we perform a Blinder-Oaxaca decomposition analysis of the application gap between students from the top and bottom SES quintiles. The Blinder-Oaxaca decomposition approach is commonly used to study mean outcome differentials between two groups (Blinder 1973; Oaxaca 1973). We utilize the *oaxaca* command in Stata (Jann 2008) to carry out the decomposition analysis. Doing so allows us to divide the gap in college applications into three components: 1) the part that is due to group differences in the predictors (or "endowments effect"), 2) the part that is due to differences in the coefficients including the intercept (or "coefficients effect"), and 3) an interaction term that accounts for the simultaneous interplay between the first two components (Jann 2008). For the estimation parameters, we draw upon the same Heckman selection model specification as the full Model 5. Consequently, this part of the study will provide an estimate of the contribution of each theoretical factor to the gap in college applications by socioeconomic background. Doing so allows us to make comparisons of the relative contribution of each theoretical mechanism to the SES-based gap in college applications.

Model 1 provides a baseline estimate of inequality in college applications by student SES without any covariates. Model 2 adds the individual- and school-level control variables. Model 3 includes the measures from the rational action model and Model 4 includes the measures from the status attainment model. Model 5 is the full model that includes all measures analyzed in this study. Finally, to test whether the link between expectations and applications differs by socioeconomic background, we run an additional analysis (Model 6) that includes an interaction term between SES and college expectations (type).

RESULTS

Figure 1.1. Distribution of College Application Selectivity by Top and Bottom SES Quintiles



NOTE: Estimates are limited to those with a high school degree or equivalent.

SOURCES: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSLS), 2013.

U.S. Department of Education, National Center for Education Statistics, Integrated Postsecondary Education Data System (IPEDS), 2011-2012.

Figure 1.1 provides a visual representation of the gap in college application selectivity by student SES. Namely, the two histograms displayed highlight the distinct distributions in college applications between students from the top and bottom SES quintiles. We can see that whereas high-SES students tend to apply to more selective colleges (i.e., higher rejection rate or lower admissions rate), low-SES students tend to either apply to less selective colleges or none at all. To unpack why we see these unequal patterns in college applications by socioeconomic background, we first undertake a bivariate analysis before turning to the multivariate regression models.

Table 1.1. Bivariate Analysis of Key Measures between Top and Bottom SES Quintiles

Measure	Top 20% Mean/ prop.	Bottom 20% Mean/ prop.	P-value
Dependent variable			
Application selectivity (most selective based on IPEDS data)	43.95	23.34	***
Applied (yes=1)	.97	.78	***
Main independent variable			
SES composite score (-2 to 2)	1.03	-1.05	***
Rational action model			
Performance differentials			
11th-grade GPA	3.26	2.62	***
AP coursework (yes=1)	.60	.29	***
Standardized test score (SAT or equivalent)	1121	876	***
Choice differentials			
Informational resources			
Attended program at or taken tour of college campus (yes=1)	.68	.45	***
Searched for college options (yes=1)	.92	.80	***
Talked w/ high school counselor (yes=1)	.70	.61	***
Talked w/ college admission's counselor (yes=1)	.13	.13	
Took preparatory course for college admission exam (yes=1)	.60	.41	***
College considerations			
Being close to home (very important=1)	.15	.31	***
Cost of attendance (very important=1)	.54	.72	***
Academic quality/reputation (very important=1)	.85	.71	***
Family/friend recommendations (very important=1)	.21	.23	*
Family legacy (very important=1)	.07	.09	**
Degree program (very important=1)	.78	.72	***
Graduate school placement (very important=1)	.57	.60	**
Job placement (very important=1)	.70	.76	***
Play school sports (very important=1)	.18	.23	***
Campus social life / school spirit (very important=1)	.59	.58	
Status attainment model			
Educational expectations (*level* ranging from 1 <HS to 6 MA+)	5.45	4.51	***
Educational expectations (*type* based on IPEDS data)	40.43	29.13	***
Number of applications submitted	4.02	2.05	***

†p<.10; *p<.05; **p<.01;***p<.001

NOTE: Estimates are limited to those with a high school degree or equivalent.

SOURCES: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSLS), 2012, 2013.

U.S. Department of Education, National Center for Education Statistics, Integrated Postsecondary Education Data System (IPEDS), 2011-2012.

Like the histograms, Table 1.1 highlights significant differences in college applications between students from the top and bottom SES quintiles. As expected, we see that high-SES students not only are more likely to apply to any college, but when they do, they tend to apply to more selective colleges compared to their low-SES counterparts. In turn, looking at the key measures shown, we find that most factors differ significantly between high- and low-SES students. For example, in terms of performance differentials, we see that high-SES students tend to have significantly higher GPAs and test scores than low-SES students ($p<.001$). Likewise, we see that high- and low-SES students vary in several important ways in terms of access to informational resources and the factors they consider when making their college choice decision. Finally, we see that high- and low-SES students differ significantly in terms of how far they plan to go in school ($p<.001$), the type of college they plan to attend ($p<.001$), and the number of applications they submit during the college application phase ($p<.001$). To understand how these differences may matter for inequality in college applications by SES, we turn to a series of Heckman selection models.

Table 1.2. Heckman Regression Model of College Application Selectivity (N=15,130)

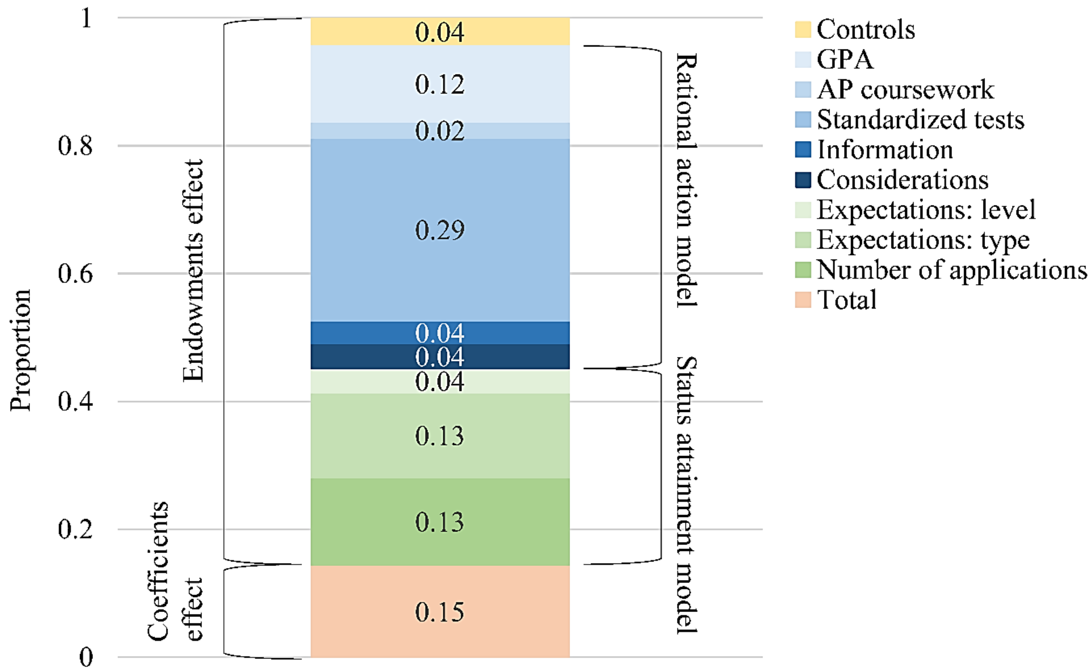
Variable	Model 1	Model 2	Model 3	Model 4	Model 5
SES	23.20***	10.55***	2.47***	3.90***	1.06***
	(4.40)	(.624)	(.324)	(.328)	(.279)
Constant	11.92	25.37	-51.78	-16.69	-42.41
Inverse Mills Ratio (lambda)	88.36**	23.42***	17.08***	8.58**	5.27*
Controls		X	X	X	X
Performance differentials			X		X
Choice differentials			X		X
Educational expectations				X	X
Number of applications				X	X

†p<.10; *p<.05; **p<.01;***p<.001
NOTES: Standard errors in parentheses.
Sample size rounded to the nearest 10.
SOURCES: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSLS), 2012, 2013.

U.S. Department of Education, National Center for Education Statistics, Integrated Postsecondary Education Data System (IPEDS), 2011-2012.

Results of the Heckman selection models in Table 1.2 provide three main takeaways. First, across models we find that the *inverse Mills ratio* is significant, thus providing statistical support for use of the selection model over standard OLS. Second, whereas the baseline Model 1 indicates that a one-unit increase in the SES composite score is associated with an increase in the application selectivity (or rejection rate) by 23.20 ($p<.001$), in the full Model 5 this relationship drops to just 1.06 ($p<.001$). Third, from Models 3 and 4 it seems that the rational action model has greater explanatory power compared to the status attainment model based on the relative size of the SES coefficient across models. Namely, the SES coefficient decreases more in Model 3 compared to Model 4. However, to get a more precise estimate of how much performance differentials, choice differentials, educational expectations, and the number of applications, contribute to the SES-based gap in college applications, we turn to the Blinder-Oaxaca decomposition analysis.

Figure 1.2. Relative Contribution to Gaps in College Application Selectivity between Top and Bottom SES Quintiles from Blinder-Oaxaca Decomposition Analysis



NOTE: Estimates are limited to those with a high school degree or equivalent and are conditional on those applying to college.

SOURCES: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSLS), 2012, 2013.

U.S. Department of Education, National Center for Education Statistics, Integrated Postsecondary Education Data System (IPEDS), 2011-2012.

From the results of the Blinder-Oaxaca decomposition analysis shown in Figure 1.2 (also see S1.3), we get a deeper understanding of how the underlying mechanisms contribute to the SES-based gap in college applications. First, we see that around 85% of the application gap is due to group differences in the predictors (i.e., “endowments effect”), and 15% of the gap is due to differences in the coefficients (i.e., “coefficients effect”). Although not shown, the “interaction effect” overall did not significantly contribute to the gap in applications (p>.10). Second, when we compare the contribution of the two sociological models in terms of the endowments effect, we see that factors from the rational action model contribute relatively more to the college application gap. Namely, 60% of the endowments effect is due to rational action mechanisms, whereas 35% is due to status attainment mechanisms. Finally, while several factors contribute significantly in terms of the coefficients effect, one that is particularly relevant is the type of educational expectations. To further examine this, we test the interaction between SES and educational expectations.

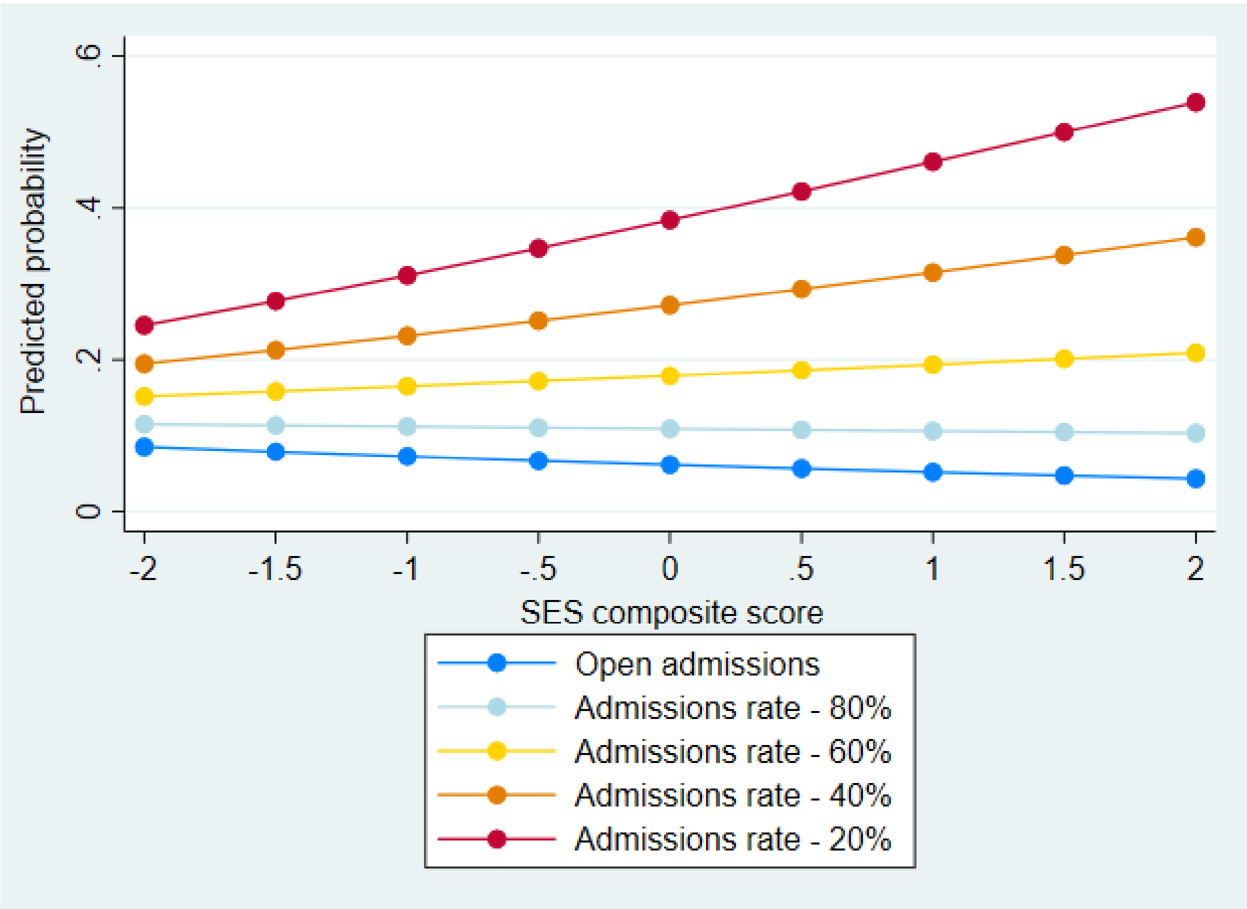
Table 1.3. Heckman Regression Model Testing Interaction Effect of SES and Educational Expectations on College Application Selectivity (N=15,130)

Variable	Model 6
SES	-0.17
	(.524)
Educational expectations (type)	0.25***
	(.010)
SES X Educational expectations (type)	0.03**
	(.012)
Constant	-40.77
Inverse Mills Ratio (lambda)	3.82†
Controls	X
Performance differentials	X
Choice differentials	X
Educational expectations	X
Number of applications	X

†p<.10; *p<.05; **p<.01;***p<.001
NOTES: Standard errors in parentheses.
Sample size rounded to the nearest 10.
SOURCES: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSLS), 2012, 2013.
U.S. Department of Education, National Center for Education Statistics, Integrated Postsecondary Education Data System (IPEDS), 2011-2012.

Model 6 in Table 1.3 reveals a significant interaction between SES and educational expectations (type) (0.03; p<.01). Thus, we find evidence of differential returns to educational expectations by SES in both the Blinder-Oaxaca decomposition analysis as well as this final model. To better grasp how the lower- and higher-order terms combine to shape selective college applications, we utilize Stata's *margins* command to produce a predicted plot of the focal relationships. First, we generate a dichotomous measure of "selective colleges" based on whether the institution accepts less than half its applicants (yes=1; no=0), which is roughly equivalent to Barron's "most competitive" and "highly competitive" categories (see S1.4). Next, we run a *heckprobit* selection model with the interaction terms to estimate the likelihood of applying to a selective college. Finally, we use *margins* to estimate the predicted probability of applying to selective colleges at varying levels of SES and educational expectations, while holding all other factors at their mean values.

Figure 1.3. Likelihood of Selective Applications by SES and College Expectations (Type)



NOTES: Estimates are limited to those with a high school degree or equivalent and are conditional on those applying to college.

All other factors held at their mean values.

SOURCES: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSLS), 2012, 2013.

In general, Figure 1.3 highlights that across SES as expectations increase (blue → red), so does the predicted probability of applying to selective colleges. For example, low-SES students who expect to attend an institution with an 80% acceptance rate (i.e., light blue dot at -1) have an 11% predicted probability of applying to a selective college compared to 23% among those who expect to attend a college with a 40% acceptance rate (i.e., orange dot at -1), controlling for all other factors. We also observe patterns related to the interaction effect. For instance, among high-SES students with the same expectations as the low-SES students just discussed, their predicted probabilities of applying to a selective college are 11% (i.e., light blue dot at 1) and 32% (i.e., orange dot at 1), respectively. Thus, we see that among average-performing high school students in the U.S., high expectations do not translate into selective applications equally for those from high- and low-SES backgrounds. In other words, the returns to high expectations, in terms of selective applications, seem to pay off most for those from more advantaged backgrounds even when accounting for a host of factors that are known to matter for college admissions.

DISCUSSION

Although most high school graduates in the U.S. make the transition to some type of college, gaps in where students apply are evident by socioeconomic background (An 2010; Bowen, Chingos, and McPherson 2009; Holzman, Klasik, and Baker 2020). While a substantial body of work has shown that higher-SES students tend to apply to more selective colleges than their lower-SES counterparts (Hoxby and Avery 2012; Mullen and Goyette 2019; Radford 2013), we know relatively less about why students differ in their application behavior. In this study, we draw upon a sociological approach to compare the rational action model with the status attainment model of educational stratification. Utilizing data from the *High School Longitudinal Study of 2009,* and a series of Heckman selection models, we find that mechanisms related to the rational action model contribute relatively more to the SES-based application gap compared to the status attainment model, although both are important. We also reveal a significant interaction effect between SES and the type of educational expectations.

This study thus adds to our understanding of the processes that lead to unequal sorting by SES during the college application phase (Holzman, Klasik, and Baker 2020). Although past work has helped to uncover some of the factors related to the SES-based gap in college applications (Cabrera et al. 2000a; Mullen and Goyette 2019; Roksa and Deutschlander 2018), to our knowledge, this study provides the most comprehensive analysis to date of the underlying mechanisms that contribute to the observed disparity. First, from the Blinder-Oaxaca decomposition analysis, we find that 85% of the gap in college applications between those in the top and bottom SES quintiles is due to the endowments effect, or differences in the predictors, while 15% is due to differences in the coefficients, or the portion left unexplained. Thus, we can explain most of the SES-based gap in college application selectivity through the factors modeled in this study. Second, we estimate that 60% of the endowments effect is due to rational action mechanisms, while 35% is due to status attainment mechanisms. Consequently, although prior work has tended to focus on either the rational action mechanisms or the status attainment mechanisms, we show that both are important for fully understanding the SES-based gap in college applications.

In terms of the rational action model, this analysis confirms the importance of performance differentials in shaping unequal applications by SES (Holzman, Klasik, and Baker 2020). The Blinder-Oaxaca decomposition analysis shown in Figure 1.2 indicates that 43% of the SES-based selectivity gap in college applications is due to performance differentials, with 29% just from standardized tests. In contrast to prior work that has focused exclusively on high-performing students (Hoxby and Avery 2012; Lor 2023; Radford 2013), our study indicates that choice differentials do not contribute much to the gap in college applications. Specifically, we estimate that only about 8% of the application gap between top and bottom SES quintiles is due directly from differences in access to information and college considerations. Overall, then, it seems that in the case of SES-based disparities in college applications, secondary effects play a relatively minor role compared to primary effects (Jackson 2013). We suspect, however, that some of the secondary effects may operate indirectly through its association with educational expectations.

In terms of the status attainment model, this study highlights the importance of educational expectations. Whereas prior work has tended to focus on level of expectations (Mullen and Goyette 2019), we show that type of expectations seems to matter more. Namely, results shown in Figure 1.2 indicate that only 4% of the SES-based gap in college applications is due to the level of expectations, while 13% is due to the type of expectations. It is important to note, however, that this analysis focuses on college application selectivity rather than application to college. Since our modeling approach conditions on application to any college, it is likely that the level of expectations (i.e., how far) matters more for predicting whether or not a student applied at all (see selection equation section in Appendix B). In turn, as shown in Table 1.3, we see that SES moderates the relationship between type of expectations and the selectivity of college applications. Figure 1.3 highlights that the payoff to higher expectations in terms of selective applications disproportionately accrues to higher-SES students.

This finding thus provides insight on an additional source of advantage for high-SES students during the college application process. For example, average-performing, low-SES students with the highest expectations have a 31% predicted probability of applying to selective colleges, while average-performing, high-SES students with the same expectations have a 46%

predicted probability—holding all other factors at their mean values.[6] This provides evidence that high-SES students are more likely to enact their plans and apply to selective schools regardless of their measured performance. Consequently, these results indicate that equalizing access to information or college considerations will not necessarily lead to equal application behavior among students from differing class backgrounds. Even among students who expect to attend the same type of selective college—and thus must already *know* about these schools and *plan* to attend—still exhibit differential application behavior by socioeconomic background. Future research is needed to better understand when and why this disconnect occurs at the application stage. It may be that some low-SES students do campus visits in the summer after their junior year where they have a negative experience that deters them from applying to selective colleges (Radford 2013).

From a policy standpoint, our results have implications for those aiming to increase the share of low-SES students in the pool of applicants at selective colleges. Specifically, our results reinforce the need to address the SES-based performance gap in high school to bring about greater equity during the transition to college. While past work has tended to focus on the admissions side of the equation, we draw attention to the application side as well. Our analysis shows that students sort into different pools of applicants based on their grades and test scores. In other words, we observe that students largely align their college applications with their own performance metrics.[7] This may arise due to student awareness of the relevant components and academic thresholds specified by a given institution in the admissions process. For example, students may decide where to apply in part based on how competitive they feel they would be for

[6] High SES was estimated at a score of 1 whereas low SES corresponds to a score of -1. Similar gaps are evident using alternative thresholds such as a 2.5 GPA (29% vs 44%) or 3.5 GPA (34% vs 49%).
[7] Interestingly, supplementary analysis looking at differential returns to academic performance shows a marginally significant interaction between SES and GPA on college application selectivity (see S1.6).

admission to a given college. However, because many institutions claim to base admissions decisions on a host of academic as well as non-academic factors, there is likely a larger pool of missed talent among low-SES students than previously discussed (Hoxby and Avery 2012). For instance, recent work has shown that, regardless of academic qualifications, applying to “reach” schools increases the likelihood of enrolling at a more selective institution (Mullen and Goyette 2019).

This study has some important limitations. First, because our data was collected prior to the pandemic, it is not entirely clear how the underlying relationships may have shifted since then. For example, we know that since the pandemic, many colleges and universities have switched to test-optional or test-blind admissions. With these changes schools likely place greater emphasis on grades, and as a result, students may align their applications with their grades instead of test scores. If this were the case, it is unlikely that we would see much difference from the patterns observed here since grades and test scores are moderately correlated. Second, we need a better understanding of why low-SES students even with high expectations do not apply to selective colleges at equal rates to their high-SES counterparts. Our work offers more support for research on the application process itself, and how to decrease (or eliminate) the barriers that students face as they apply to college (Odle and Magouirk 2023).[8] Finally, while our data and analysis has allowed us to undertake a broad examination of the factors that drive differential applications to selective colleges, we acknowledge that our decomposition approach models predictors at one point in time. Future data collection and research would benefit from greater attention to the dynamic interplay of the underlying factors as they emerge over time—in other words, investigating the longitudinal process behind these patterns.

---

[8] Low-SES students with selective expectations may represent an ideal group for targeted interventions related to direct admissions in higher education (Odle and Delaney 2023).

CHAPTER 1

SUPPLEMENTARY MATERIALS

Supplementary Tables and Figures

S1.1. Heckman Regression Model of College Application Selectivity using Barron's Competitiveness Index (N=15,130)

Variable	Model 1	Model 2	Model 3	Model 4	Model 5
SES	1.620***	0.717***	0.188***	0.311***	0.111***
	(.275)	(.037)	(.019)	(.020)	(.017)
Constant	1.90	3.02	-2.49	-0.46	-2.47
Inverse Mills Ratio (lambda)	5.58**	1.14***	0.99***	0.45**	0.45***
Controls		X	X	X	X
Performance differentials			X		X
Choice differentials			X		X
Educational expectations				X	X
Number of applications				X	X

†p<.10; *p<.05; **p<.01;***p<.001
NOTES: Standard errors in parentheses.
Sample size rounded to the nearest 10.
SOURCES: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSLS), 2012, 2013.
Barron's Profiles of American Colleges 2011

S1.2. Results of the Blinder-Oaxaca Decomposition Analysis of the Gap in College Application Selectivity between Students in the Top and Bottom SES Deciles

Component	Coef.	Std. Err.	Sig.
Outcome differential			
Low SES (*prediction*)	19.978	(2.48)	***
High SES (*prediction*)	47.455	(.682)	***
Difference	-27.477	(2.58)	***
Endowments effect			
Controls	-1.778	(1.12)	
GPA	-2.855	(.673)	***
AP coursework	-0.778	(.354)	*
Standardized tests	-8.985	(.973)	***
Information	-1.056	(.369)	**
Considerations	-1.324	(.412)	**
Expectations: level	-1.550	(.727)	*
Expectations: type	-3.867	(.409)	***
Number of applications	-3.903	(.345)	***
Total	-26.095	(1.73)	***
Coefficients effect			
Controls	0.485	(3.14)	
GPA	1.422	(6.30)	
AP coursework	2.472	(1.22)	*
Standardized tests	-14.647	(7.03)	*
Information	2.769	(3.03)	
Considerations	3.588	(2.58)	
Expectations: level	-2.774	(7.07)	
Expectations: type	-5.480	(1.90)	**
Number of applications	11.054	(1.73)	***
Constant	-1.237	(13.62)	
Total	-2.347	(2.24)	
Interaction			
Controls	-0.371	(1.89)	
GPA	-0.283	(1.25)	
AP coursework	-1.300	(.646)	*
Standardized tests	3.679	(1.77)	*
Information	-0.228	(.630)	

Considerations	2.275	(.701)	**
Expectations: level	0.449	(1.14)	
Expectations: type	1.593	(.571)	**
Number of applications	-4.849	(.801)	***
Total	0.966	(2.97)	

†p<.10; *p<.05; **p<.01;***p<.001

NOTE: Controls include: *race/ethnicity, gender, school control, school type, school urbanicity, geographic region, high school size, and percent free/reduced price lunch*. Information includes: *attended program at or taken tour of college campus, searched for college options, talked w/ high school counselor, talked w/ college admission's counselor, took preparatory course for college admission exam*. Considerations include: *being close to home, cost of attendance, academic quality/reputation, family/friend recommendations, family legacy, degree program, graduate school placement, job placement, play school sports, campus social life/school spirit*.

SOURCES: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSLS), 2012, 2013.

U.S. Department of Education, National Center for Education Statistics, Integrated Postsecondary Education Data System (IPEDS), 2011-2012.

S1.3. Results of the Blinder-Oaxaca Decomposition Analysis of the Gap in College Application Selectivity between Students in the Top and Bottom SES Quartiles

Component	Coef.	Std. Err.	Sig.
Outcome differential			
Low SES (*prediction*)	21.636	(1.87)	***
High SES (*prediction*)	42.318	(.483)	***
Difference	-20.682	(1.93)	***
Endowments effect			
Controls	-0.907	(.533)	†
GPA	-2.249	(.358)	***
AP coursework	-0.618	(.172)	***
Standardized tests	-6.413	(.472)	***
Information	-0.629	(.170)	***
Considerations	-0.736	(.192)	***
Expectations: level	-0.794	(.309)	*
Expectations: type	-2.811	(.219)	***
Number of applications	-3.018	(.201)	***
Total	-18.176	(.896)	***
Coefficients effect			
Controls	-1.729	(2.13)	
GPA	2.682	(4.10)	
AP coursework	1.588	(.757)	*
Standardized tests	-13.077	(4.08)	**
Information	1.606	(1.83)	
Considerations	2.547	(1.57)	
Expectations: level	3.874	(4.78)	
Expectations: type	-3.519	(1.12)	**
Number of applications	4.644	(.842)	***
Constant	-1.873	(9.90)	
Total	-3.257	(1.21)	**
Interaction			
Controls	-0.118	(.818)	
GPA	-0.416	(.636)	
AP coursework	-0.702	(.337)	*
Standardized tests	2.517	(.787)	**
Information	-0.143	(.278)	

Considerations	0.847	(.311)	**
Expectations: level	-0.468	(.578)	
Expectations: type	0.839	(.271)	**
Number of applications	-1.605	(.303)	***
Total	0.751	(1.56)	

†p<.10; *p<.05; **p<.01;***p<.001

NOTE: Controls include: *race/ethnicity, gender, school control, school type, school urbanicity, geographic region, high school size, and percent free/reduced price lunch*. Information includes: *attended program at or taken tour of college campus, searched for college options, talked w/ high school counselor, talked w/ college admission's counselor, took preparatory course for college admission exam*. Considerations include: *being close to home, cost of attendance, academic quality/reputation, family/friend recommendations, family legacy, degree program, graduate school placement, job placement, play school sports, campus social life/school spirit*.

SOURCES: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSLS), 2012, 2013.
U.S. Department of Education, National Center for Education Statistics, Integrated Postsecondary Education Data System (IPEDS), 2011-2012.

S1.4. Regression Model of College Application Selectivity using Listwise Deletion (N=6,230)

Variable	Model 1	Model 2	Model 3	Model 4	Model 5
SES	13.957***	11.657***	3.746***	5.309***	2.315***
	(.615)	(.678)	(.648)	(.593)	(.591)
Constant	27.78	28.51	-60.13	-31.62	-63.45
Controls		X	X	X	X
Performance differentials			X		X
Choice differentials			X		X
Educational expectations				X	X
Number of applications				X	X

†p<.10; *p<.05; **p<.01;***p<.001

NOTES: Standard errors in parentheses.

Sample size rounded to the nearest 10.

SOURCES: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSLS), 2012, 2013.

U.S. Department of Education, National Center for Education Statistics, Integrated Postsecondary Education Data System (IPEDS), 2011-2012.

S1.5. Regression Model of College Application Selectivity among "Don't Know" College Plans (N=2,200)

Variable	Model 1	Model 2	Model 3	Model 4	Model 5
SES	19.151***	14.429***	2.783*	5.307***	1.215
	(1.154)	(1.278)	(1.246)	(1.125)	(1.141)
Constant	25.41	15.23	-96.98	-52.67	-90.57
Controls		X	X	X	X
Performance differentials			X		X
Choice differentials			X		X
Educational expectations				X	X
Number of applications				X	X

†p<.10; *p<.05; **p<.01;***p<.001

NOTES: Standard errors in parentheses.

Sample size rounded to the nearest 10.

SOURCES: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSLS), 2012, 2013.

U.S. Department of Education, National Center for Education Statistics, Integrated Postsecondary Education Data System (IPEDS), 2011-2012.

S1.6. Heckman Regression Model Testing Interaction Effect of SES and GPA on College Application Selectivity (N=15,130)

Variable	Model 7
SES	-0.950
	(1.152)
GPA	4.670***
	(.411)
SES X GPA	.640†
	(.359)
Constant	-40.68
Inverse Mills Ratio (lambda)	3.52
Controls	X
Performance differentials	X
Choice differentials	X
Educational expectations	X
Number of applications	X

†p<.10; *p<.05; **p<.01;***p<.001

NOTES: Standard errors in parentheses.
Sample size rounded to the nearest 10.

SOURCES: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSLS), 2012, 2013.

U.S. Department of Education, National Center for Education Statistics, Integrated Postsecondary Education Data System (IPEDS), 2011-2012.

CHAPTER 2

STEM Learning Communities Promote Friendships but Risk Academic Segmentation

ABSTRACT

Universities are increasingly using learning communities (LCs) to promote the academic and social integration of entering students, especially within STEM majors. Examining the causal effect of LCs on student networks is necessary to understand the nature and scope of their impact. This study combines a regression discontinuity design with social network analysis to estimate the effect of a simple LC design on the *size*, *strength*, *structure*, and *composition* of friendship networks among students within the same biological sciences freshman cohort. Results of the quasi-experimental analysis indicate that LC participants acquired one additional friend in the major and increased their share of friends in the LC by 54 percentage-points. Exponential random-graph models that test mediation and alternative friendship mechanisms provide support for the theoretical argument that the LC promoted friendship development by structuring opportunities for interaction through block-registration into courses. Thus, this study shows that even simple LCs can shape the development of friendships through relatively low-cost administrative means. The increased access to resources and support facilitated by the LC is likely beneficial for participating students. However, there is a potential downside when eligibility for participation is determined using academic metrics that separate the student population into distinct classroom environments.

Keywords: learning community, STEM, social networks, higher education

INTRODUCTION

STEM higher education suffers from issues of attrition and academic disparities that threaten to undermine the adequate supply of skilled workers to keep up with societal demand (Xie, Fang, and Shauman 2015). Nearly half of bachelor's degree-seeking students leave STEM fields (Chen 2013), and those who persist and perform best disproportionately come from advantaged backgrounds (Huang, Taddese, and Walter 2000; Hurtado, Eagan, and Chang 2010). Finding ways to help all students thrive in STEM environments is a major goal of the science community (NSF 2014; James and Singer 2016).

One proposed avenue to promote student persistence and academic success is through Learning Communities (LCs) (Andrade 2007; Dagley et al. 2016; Maton et al. 2012; Smith et al. 2004). In 2019, an estimated 13% of first-year students and 22% of seniors reported having participated in some form of LC (NSSE 2015). In general, institutions construct LCs by grouping students through some combination of shared courses, a residential living component, active learning strategies, and informal activity with the goal of promoting student academic and social integration (Otto et al. 2015; Tinto 1987; Tinto 2003). Without the aid of structured interventions such as LCs (Boda et al. 2020), students must make connections and find support to navigate the new college and STEM environment largely on their own. Yet, while LCs have been shown to promote performance and persistence in STEM, their direct impact on social integration in terms of student friendships remains unclear (Andrade 2007; Dagley et al. 2016; Hotchkiss, Moore, and Pitts 2006; Whalen and Shelley 2010; Xu et al. 2018).

Indeed, despite intuition regarding how LCs may guide the development of friendships, there is surprisingly little evidence establishing a causal relationship. Correlational studies have linked LC participation to positive relational outcomes, such as increased socializing (Jaffee et al. 2008), heightened exposure and network formation (Tinto and Goodsell 1994), and social support (Domizi 2008). However, because universities often make LC participation voluntary, confounds between the types of students who opt into LCs and student outcomes are inevitable, thereby precluding causal inferences (Andrade 2007). Thus, the effect of LCs on friendship has yet to receive the rigorous causal evaluation needed to demonstrate their promise.

In this study, we extend prior work on college LCs by exploring the causal impact of LCs on friendship networks in a unique setting where students are assigned to participate in a LC using a strict SAT math score cutoff. This enables us to utilize a regression discontinuity design (RDD) that can credibly support causal inferences because assignment to treatment creates a scenario that is "as good as randomized" for individuals proximate to the threshold (Jacob et al. 2012; Lee and Lemieux 2010). Accordingly, results from this study advance our understanding of the link between organizational practices and relational outcomes generally, and specifically its importance for helping students build meaningful connections with peers in STEM.

BACKGROUND

While social integration is important throughout college, it is particularly critical during the freshman year when students transition into college (Hays and Oxley 1986; Tinto 1987). Advocates of LCs point out that social integration can improve motivation (Freeman, Anderman, and Jensen 2007; Zumbrunn et al. 2014) and provide access to resources and information needed to succeed academically (Hasan and Bagde 2013; Stadtfeld et al. 2019). Although social integration has multiple dimensions (Kraemer 1997), encompassing faculty, staff, and peer interactions (Nora 1993; Tinto 1987), we focus on the friendship networks that first-year students develop within their major. Friendships are crucial in college (Felten and Lambert 2020; McCabe

2016) as they represent some of the strongest influences on students' attitudes, values, and behaviors (Hallinan 1981). Within the STEM context, friendships are a key factor promoting persistence (Thomas 2000; Turetsky et al. 2020) and academic success (Dokuka, Valeeva, and Yudkevich 2020; Stadtfeld et al. 2019).

Evidence from the K-12 context provides reason to suspect that LCs promote friendship by acting as "foci" to structure student interaction patterns (Epstein and Karweit 1983). According to focus theory, foci are any "social, psychological, legal, or physical entity around which joint activities are organized" and which "actively bring people together or passively constrain them to interact" (Feld 1981:pp.1016, pp.1018). By enhancing proximity and promoting regular interaction opportunities, foci are a powerful force behind the development of positive sentiments, relationships, and their change throughout the life course (Mollenhorst, Volker, and Flap 2014; Rivera, Soderstrom, and Uzzi 2010; Thomas 2019). Within secondary schools, research has demonstrated the importance of tracking (Kubitschek and Hallinan 1998) and clustered sets of courses (Frank, Muller, and Mueller 2013) for understanding processes of friendship formation among students. Likewise, some work within higher education has also highlighted the association between shared classes (Kossinets and Watts 2009) and majors (Wimmer and Lewis 2010) and the relationships that arise. Thus, the active manipulation of which students attend class together—at the core of the LC model—will likely shape which friendships emerge.

Based upon insight from focus theory, we expect the LC to concentrate friendships among students assigned to the same courses. We also expect the community cultivated by the LC to lead to more friendships and friendships that are stronger and more group-based than outside the LC. However, an often-overlooked side effect of these processes is that some potential friendships will be inadvertently discouraged. Students placed in the same classroom are primed for friendship while those placed in different classrooms face a structural barrier (Hallinan and Sørensen 1985; Kubitschek and Hallinan 1998; Leszczensky and Pink 2015). Hence, the LC may create divisions within the student body, which can exacerbate inequality (DiMaggio and Garip 2012; Johnson 2019). Our analysis considers multiple friendship network outcomes—size, strength, structure, and composition (see Materials and Methods section for details)—as a way to evaluate the intended goal of social integration, while being cognizant of such unintended consequences (Jaffee et al. 2008).

Our results indicate that participating in the LC led to an additional friend in the major, although this effect was only marginally significant. In addition, LC participation led to a 54 percentage-point increase in students' share of friends in the LC. We did not find evidence that participating in the LC altered the strength or structure of students' friendship networks. Follow-up mediation analyses substantiate the theoretical expectation that increased opportunity for interaction brought about through the LC's block-registration into classes is the main mechanism responsible for the observed differences in friendship network outcomes.

METHODS

Data and Setting

Data come from two sequential cohorts of first-time entering biological sciences freshmen at a large, selective, public R1 university in the Western United States. The case under study represents a diverse environment in terms of race/ethnicity, socioeconomic background, and gender. Namely, the major cohort across years is predominantly female, with around half of students considered first-generation, and about 30-40% classified as underrepresented minorities

(URM) in terms of racial/ethnic status. During the final week of Fall term, electronic surveys were sent to the entire freshman cohort (LC participants and non-participants) to collect data on friendship ties within the major and various aspects of student background to serve as controls (>93% response rate). Information on LC participation and additional student demographic data was provided by the university. The study design and procedures were reviewed and approved by the Institutional Review Board of the University of California, Irvine.

Learning Community Design

For each cohort, the department implemented a simple LC program by block-registering participating students into the same introductory biology and chemistry courses (see Supplementary Materials Section 1). Eligibility for placement into the program was determined using a strict SAT math cutoff score because prior institutional research had identified this metric as a strong predictor of performance and persistence in the major. Each year the freshman cohort consisted of around 1000 students, and approximately 300 students (or one-third) below the cutoff were assigned to participate in the LC. Along with being placed into the same biology and chemistry courses, all participating students took an additional seminar together that met weekly for one hour. Students were split into groups of about 30 students for these weekly meetings that were designed to promote study skills, career advice, and help with navigating the academic environment. This LC design is relatively easy, low-cost, and the predominant model on large campuses, compared to more extensive LCs utilized in smaller settings (Smith et al. 2004).

Friendship Network Measures

We draw upon four basic egocentric measures of students' friendship networks (Perry, Pescosolido, and Borgatti 2018). Network size refers to the number of friends with whom a focal student is connected and is measured using *total degree*, where we do not differentiate who named who as a friend (the focal student or the peer) (Wasserman and Faust 1994). Tie strength reflects the idea that relationships vary along dimensions such as closeness, intensity, and meaningfulness and is measured through the *count of mutual ties*—whereby both students acknowledge the relationship by naming each other as friends (Hartup and Stevens 1997; Hasan and Bagde 2013; Vaquera and Kao 2008). Network structure recognizes that students not only have friends, but that those friends may be connected to one another. We use *density* as our measure of network structure, calculated as the number of observed ties among a focal student's friends divided by the number of potential ties (Wasserman and Faust 1994). Finally, network composition refers to the characteristics of people in one's network (e.g., how homogenous one's friends are). We use the *proportion of friends in the learning community* as our measure of network composition since, given the design of the LC, we expect opportunities and subsequent friendships with LC participants to vary greatly depending upon whether a student belongs to the LC.

Analytic Strategy

Using survey and administrative data from two consecutive first-year cohorts, we tested the LC effect on friendship in two steps. First, we estimate the causal effect of LC participation through a RDD that effectively compares friendship outcomes among students whose SAT math scores placed them just above versus just below the LC threshold. Second, we estimated a series of social network models that test whether the impact of LC participation on friendship was mediated by LC organizational factors, versus alternative mechanisms that may have coincided with the assignment of students to courses and sections (i.e., potential confounds). This second step replicates the findings of the RD analysis and offers insight to *how* the LC had its observed effects.

Regression Discontinuity Approach

The regression discontinuity (RD) approach has been widely used in social science as a compelling quasi-experimental design to estimate program impacts when eligibility to a treatment is determined by whether an individual's score exceeds a designated threshold or cut-point (Jacob et al. 2012). This creates a situation that approximates a "local randomization" (Lee and Lemieux 2010), where the major premise is that within a specified bandwidth around the cutoff, individuals would not be expected to differ significantly from one another, other than eligibility to program participation. In the case of this study, RD is warranted because the program uses a specific cutoff score to determine each student's eligibility to participate in the LC; freshmen with SAT math scores below a cutoff of 600 were assigned to participate in the LC. If we assume the underlying relationship between SAT math score and friendship network measures follows a continuous relationship, and nothing other than the LC participation varies discontinuously at the cutoff, then we may attribute any observed discontinuity in friendship network outcomes at the cutoff to LC participation.

To deal with issues of noncompliance where a small proportion of students below the cutoff were exempted from participating in the LC (see Supplementary Materials Section 3), we use a fuzzy RD design. Specifically, we use learning community eligibility as an instrumental variable for actual participation in the first-year program with a two-stage least squares strategy (Imbens and Lemieux 2008). Namely, we derive estimates of the "local average treatment effects" (Imbens and Angrist 1994) (or LATE) through a pooled local polynomial regression within a bandwidth of ±70 points. For all models measuring the causal impact of the intervention on the four network outcomes, we draw upon the following equations:

$$\text{Enroll}_i = \gamma_0 + \gamma_1(\text{Below}_i) + \gamma_2(\text{SAT Math Distance}_i) + \gamma_3(\text{SAT Math Distance}_i * \text{SAT Math Distance}_i) + \gamma_4(\text{Below}_i * \text{SAT Math Distance}_i) + X_i + \mu_i \quad (1)$$

$$Y_i = \delta_0 + \delta_1(\widehat{Enroll}_i) + \delta_2(\text{SAT Math Distance}_i) + \delta_3(\text{SAT Math Distance}_i * \text{SAT Math Distance}_i) + \delta_4(\text{Below}_i * \text{SAT Math Distance}_i) + X_i + \varepsilon_i \quad (2)$$

Equation (1) represents the first stage of the regression, where we predict LC enrollment as a function of eligibility for placement. *Below*$_i$ is a binary variable indicating whether the student was assigned to the LC based on SAT math score eligibility; *SAT Math Distance*$_i$ is the difference between the student's math SAT score and the cutoff threshold (i.e., 600); *SAT Math Distance*$_i$ * *SAT Math Distance*$_i$ is a quadratic term that allows for nonlinear relationships between the running variable and the outcome; *Below*$_i$ * *SAT Math Distance*$_i$ is an interaction term that allows different slopes above and below the threshold; X_i is a vector of individual-level covariates as outlined above. Equation (2) represents the second stage of the regression, where we use the predicted probability of enrollment to estimate the local average treatment effect as indicated by the δ_1 coefficient. We estimate the impact of the LC on each network outcome separately using the *ivregress* command in STATA version 16.1 (https://www.stata.com).

Social Network Analysis

We used an ERGM (Robins et al. 2007) to estimate the factors that promoted friendships between students at the end of their first quarter on campus. The ERGM considers all possible directed dyads among the sample of students, where an $i \rightarrow j$ friendship was modeled separately from a $j \rightarrow i$ friendship. The model estimates the probability of observing a given network conditioned on the set of effects present in the model. We use two types of effects: nodal covariates represent student characteristics (e.g., LC participation, gender) and dyadic covariates represent similarity (i.e., homophily) or co-presence of students (e.g., in the LC, classes). Specific effects are listed in Supplementary Materials Section 4. Estimated coefficients are interpretable as the log-odds of observing a friendship in a given dyad conditional on the rest of the network. For a given effect, exponentiating the estimated coefficient indicates how a one-unit change affects the odds of a tie, assuming all other model effects remain constant. We estimated a separate ERGM for each first-year student cohort using the *statnet* package in R version 4.1.0 (https://www.r-project.org) (Handcock et al. 2008).

RESULTS

Descriptive Evidence

Figure 2.1 presents the friendship networks and distributions of network outcomes for the full set of first-year students (see Materials and Methods section for details). Descriptively, we find that LC students were more socially integrated, with significantly more friends and a greater share of friends in the LC compared to non-participants across years (panels *c-d*, *i-j*), but more mutual ties (panels *e-f*) and more dense networks (panels *g-h*) in only one of the years (see Supplementary Materials S2.1). Additionally, in examining the odds of having *no friends* (i.e., being an "isolate") in the major, LC participants were 50% less likely to be an isolate, compared to non-participants ($p<.01$; Supplementary Materials S2.2). The sociograms in panels *a-b* make clear the network segmentation based on LC status, which is stronger in Year 2 (see Supplementary Materials Section 1).

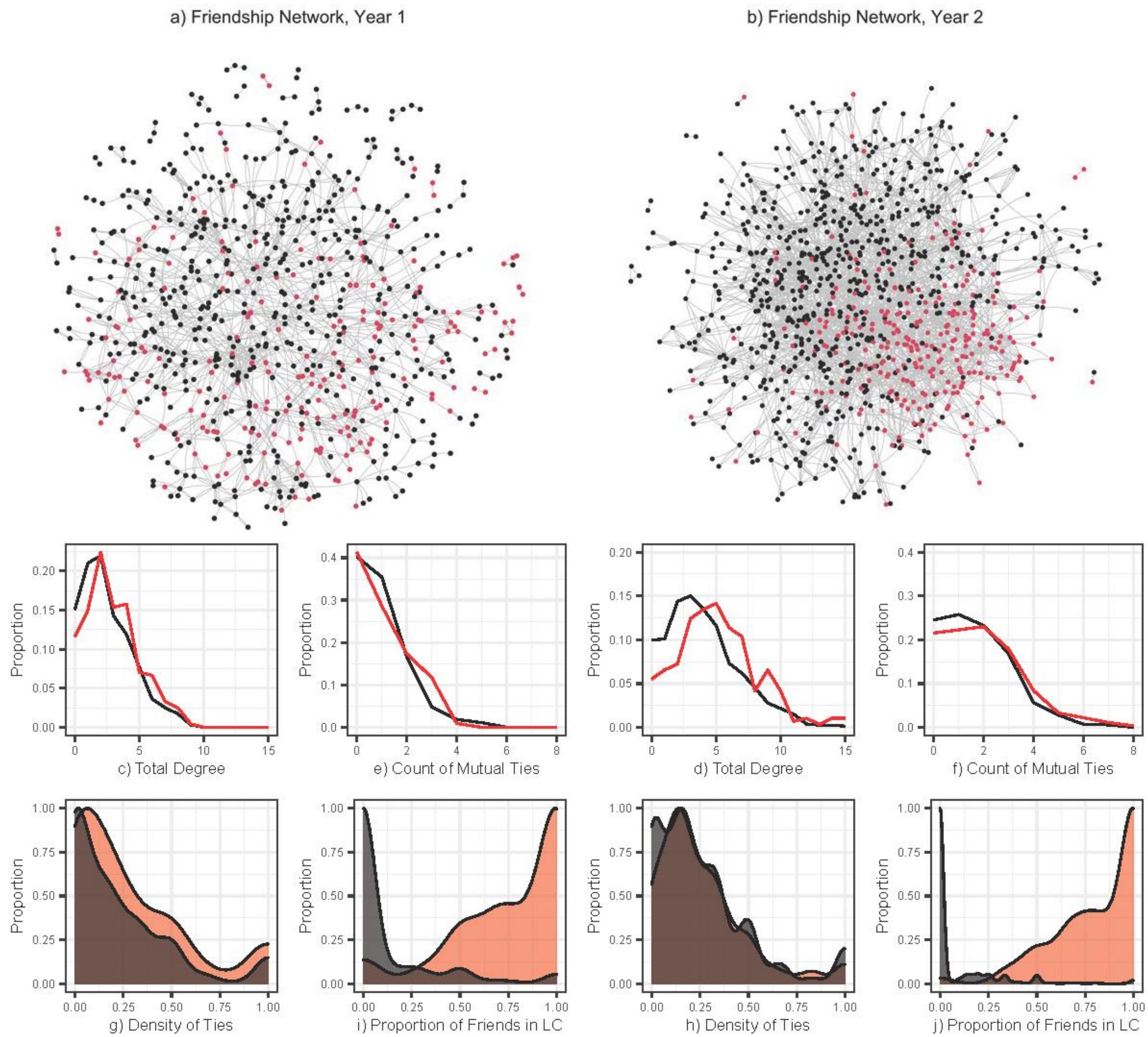


Figure 2.1. The biological sciences friendship networks and distributions of network outcomes by year. (*a*) and (*b*) omit isolates and highlight segmentation of LC from non-LC students, which is stronger in Year 2, likely due to programmatic changes (see Supplementary Materials Section 1). The Year 2 network is also more densely connected, with an average outdegree of 2.93, vs 1.72 the prior year, which we attribute to differences in the survey instrument (see Supplementary Materials Section 1). (*c*-*j*) reflect the distribution of each of our network outcomes in each year. Node, line, and density plot colors indicate LC students (red) and non-LC students (black) in each panel.

Impact of LC on Friendship Network Outcomes

While the descriptive statistics presented above provide support for the positive association between LC participation and friendship development, it is unclear whether observed differences are due to the impact of the LC or baseline differences between LC participants and non-participants. Figure 2 visualizes the discontinuity of the four network outcomes at the SAT cutoff, where quadratic prediction lines are fitted within a bandwidth of ±70 points around the threshold. Overall, we find visual evidence for a discontinuity in network size and network composition at the cutoff, but no discontinuity in tie strength or network structure. These patterns are supported by statistical estimates of the local average treatment effect (LATE) based on pooled local polynomial regressions (see Supplementary Materials S2.3): LC participation led to an additional friend in the major ($p<.10$) as well as a 54 percentage-point increase, on average, in the share of friends in the first-year program ($p<.001$). No significant effects were observed for the count of mutual ties ($p>.10$) or network density ($p>.10$).

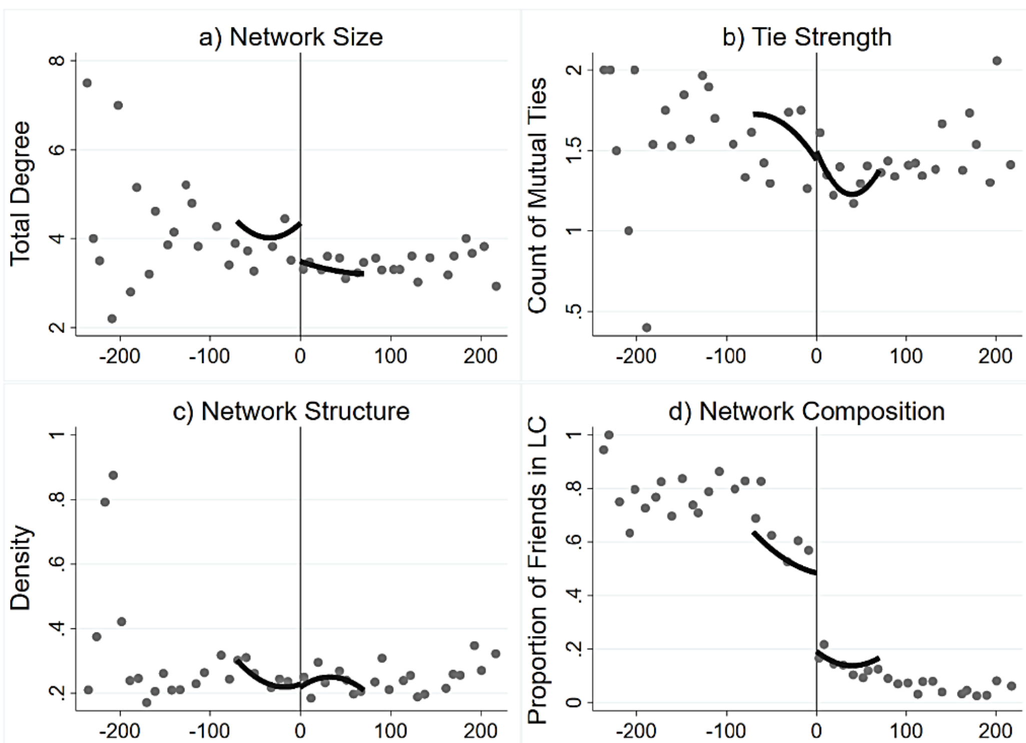


Figure 2.2. RD plots of friendship network outcomes at SAT math score cutoff. In (*a-d*), RD plots are generated on the pooled sample across years, using the *rdplot* command in STATA with bin size selected using the default esmv method (see Calonico et al. 2015 for details). While average outcomes within each bin are plotted on the entire sample, predicted quadratic lines are only fitted to those within ±70 points around the eligibility threshold.

Fitted lines adjust for mass points in the data and control for the following: race/ethnicity, gender, first-generation student status, low-income status, high school GPA, Fall term cumulative GPA, survey completion status, and year.

To aid interpretation of our RD results, we use the LATE estimates to calculate predicted network outcomes for LC participants compared to non-LC participants. As shown in Supplementary Materials S2.4, LC participants are expected to average 4.25 fellow first-year majors as friends compared to 3.25 for their non-LC counterparts. In addition, the LC affected whom students befriend: LC participants are predicted to have almost 70% of their friends in the LC, whereas their similar non-LC peers are predicted to have less than 20% of their friends in the LC program. Together, these results demonstrate that the LC had friendship network *size* and *segmentation effects* for students around the cutoff.

Given we found a marginally significant effect of LC participation on network size, in the next section, we specifically test the hypothesized mechanism through which the LC shaped friendship volume: namely, heightened opportunity to interact brought about through block-registration. Because the RD approach assumes that no other meaningful differences exist that could explain the gap at the cutoff, the follow-up network analyses provide additional validation by explicitly modeling alternative explanations that could plausibly lead to the differences we observe.

Network Mediation Analysis

Having demonstrated the effect of the LC on student friendships, we turn to testing the proposed mechanism by which the LC operated. This mediation analysis uses the full network of students each year and an exponential random-graph model, or ERGM (Robins et al. 2007). Parameter estimates reflect the likelihood that a friendship will be present, versus absent, in a given dyad based on a given effect. Marginal effects are used to test for mediation (Duxbury 2021).

ERGM findings mirror the causal analysis. LC participants had significantly more friends overall, and significantly more friends in the LC than non-participants. As shown for the Base model in Figure 2.3 (M1), the odds of a friendship were 1.1-1.3 times greater for LC participants vs. non-participants (panel *a*) and LC participants were 8-11 times more likely than non-participants to be friends with LC students (panel *b*).

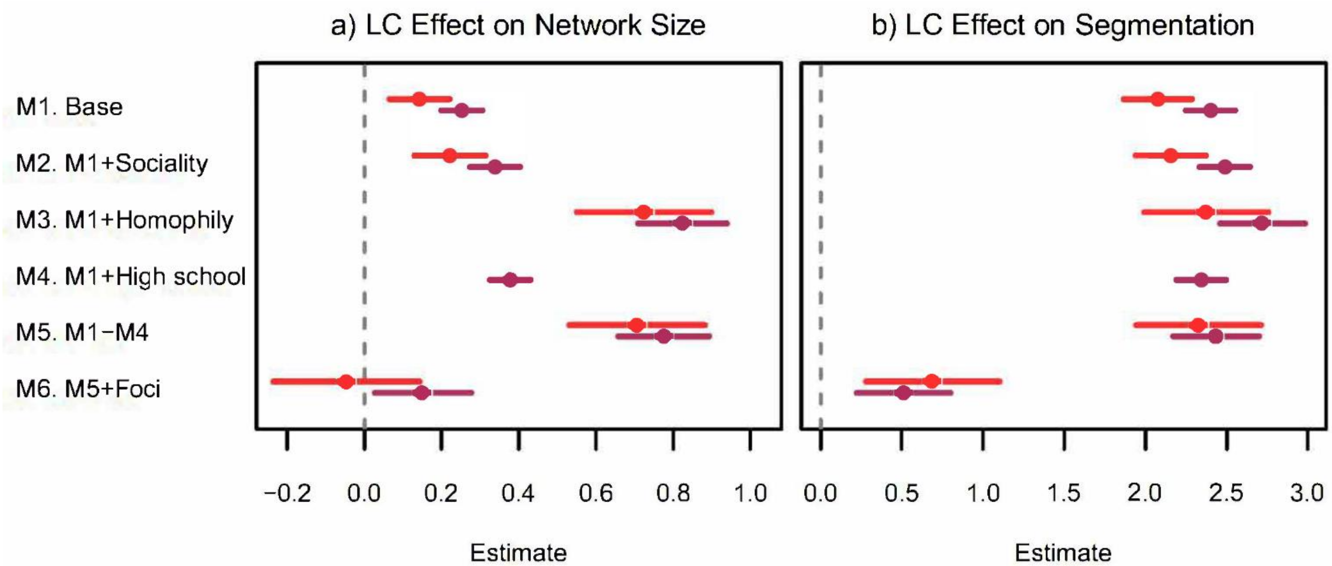


Figure 2.3. ERGM estimates testing alternative explanations and mediation of LC on network size and segmentation. Model 1 is the base model and demonstrates the main LC effect. Models 2-4 introduce measures of sociality, homophily, and same high school separately. Model 5 includes all alternative explanations simultaneously. Model 6 introduces the foci variables while controlling for all alternative mechanisms. (*a*) and (*b*) provide the coefficients from the specified ERGMs for network size and segmentation, respectively (see Supplementary Materials Section 4 for details). Plotted distance from zero corresponds to the direct magnitude of the LC effect on each network outcome across models. Point estimates and 95% confidence intervals correspond to Year 1 (light red) and Year 2 (dark red) in both panels. Full model results in Supplementary Materials S2.16 and S2.17.

While we hypothesize that these effects are driven by the LC serving as a foci for friendship activity, other possibilities exist. **(1)** The LC may have concentrated more sociable groups of students who would otherwise make more friends even in the absence of the LC program (Stadtfeld et al. 2019). **(2)** The LC may have drawn students who shared greater similarities than students outside the LC and hence would be more likely to become friends even in the absence of the LC program. Coupled with the power of homophily in driving friendship (McPherson, Smith-Lovin, Cook 2001), this could have created a more fertile friendship environment within the LC. **(3)** LC participants may have been more likely than non-participants to know one another before entering the LC program.

Models 2-5 test for these possibilities and show that they largely affected friendship in the expected manner (see Supplementary Materials Section 4). In at least one of the years, first-generation, low-income, and commuter students named fewer friends, while higher GPA and female students named more friends, net of LC participation. Students were also likely to befriend peers who were similar in race/ethnicity, first-generation status, high-school GPA, and gender, as well as retain friends from high school. However, none of these alternative explanations accounted for the observed effect of LC participation on friendship (AME results described in Supplementary Materials Section 4 support this inference).

Model 6 introduces effects to account for the assignment of LC students to specific classes and sections. These are powerful forces driving friend selection: students were approximately 1.5-4 times more likely to befriend a classmate, and LC students were 8.5-12 times more likely to befriend someone in their same section (see Supplementary Materials Section 4). As shown in Figure 3, with the introduction of foci effects in M6, the positive estimates of LC participation on network size (panel *a*) and segmentation (panel *b*) disappeared. The marginal effect estimates support this inference and indicate that foci mediated all of the effect of the LC on both network size and segmentation, revealing a suppression effect.

Combined, these findings support our theoretical argument that the LC served as a foci that provided opportunities and support for friendship development.

DISCUSSION

In this study, we combined a quasi-experimental approach with social network analysis to understand how learning communities shape friendships within a STEM major. Based on the regression discontinuity design, our analysis offers evidence that the LC led to an additional friend in the major, although the effect was only marginally significant. In addition, we found that LC participation significantly and substantially increased the segmentation of student friendship networks. By contrast, students developed equally close and group-based friendships regardless of LC participation. The latter may be a product of the first-year environment, where it is paramount for students to rebuild their networks and develop sources of companionship and support. Such a drive may be so fundamental (Kadushin 2012) that it can be met regardless of the added interaction opportunities provided by the LC.

These findings have important policy implications for efforts to facilitate connections among college students (Felten and Lambert 2020; McCabe 2016), specifically within STEM majors (Stadtfeld et al. 2019; Turetsky et al. 2020). Network science demonstrates how the interplay of friend selection (Flashman 2012) and influence processes (Carrell, Sacerdote, and West 2013; Dokuka, Valeeva, and Yudkevich 2020) contribute to academic performance differentials in schools (Stadtfeld et al. 2019), potentially exacerbating gaps among students who enter college at varying achievement levels (Wimmer and Lewis 2010). College administrators should be attentive to these dynamics when designing curricular interventions (Cox 2017; Valente 2012). Namely, the size and segmentation effects found here could have both intended and unintended academic consequences (Jaffee et al. 2008).

On one hand, students are likely to benefit from the additional friend in the major (Hasan and Bagde 2013) that the simple LC helped to promote. Friends provide important help and support with the adjustment during the transition to college (McCabe 2016; Tinto 1987; Tinto 2003). Particularly in competitive STEM majors, friends can be valuable sources of social capital by improving access to academic resources (Dokuka, Valeeva, and Yudkevich 2020; Hasan and Bagde 2013) and fostering a heightened sense of belonging (Nunn 2021). As a result, LCs offer to improve persistence and success in STEM through their impact on the social integration of entering students (Dagley et al. 2016; Stadtfeld et al. 2019; Turetsky et al. 2020).

On the other hand, by using SAT math score as the eligibility criterion, the LC promoted some friendships at the expense of others. Namely, through block-scheduling of coursework, the LC effectively sorted friendships by prior performance, making relationships between higher- and lower-performing students less likely (Hallinan and Sørensen 1985; Kubitschek and Hallinan 1998). Thus, findings from this study have significance for discussions around curricular interventions that actively sort individuals into peer groups, such as remedial education, “ability” grouping, or tracking, by revealing how such interventions may affect patterns of relational ties. Because friendships represent a unique influence on the lives of students (Hallinan 1981; Hasan and Bagde 2013; McCabe 2016)—distinct from roommates (Garlick 2018) and larger classroom or peer groups (Carrell, Fullerton, and West 2009; Lomi et al. 2011; Poldin, Valeeva, and Yudkevich 2016)—the structuring of friendships could magnify STEM academic disparities (Raabe, Boda, and Stadtfeld 2019) by inhibiting diverse networks inside and outside the classroom (Burt 2004; Oakes 2005; Park et al. 2021).

Our analysis is not without limitations. Although our quasi-experimental approach represents a more rigorous investigation relative to past work, the estimated effect is local and only applicable to individuals around the threshold. Future work may wish to extend our analysis by conducting randomized controlled trials that would enable us to estimate the average treatment effects of the LC. In a similar vein, our in-depth analysis relies upon data from one STEM major and at one location. Future research would benefit from focusing on a broader set of majors and across institutional types. Finally, whereas we utilized mutuality as one measure of tie strength, we acknowledge that there may be other ways to capture this construct that future work could explore (Marsden and Campbell 1984; Mattie et al. 2018).

Notwithstanding these limitations, the present study makes important theoretical and empirical contributions regarding the role of foci in structuring networks. First, our results reveal that by constructing foci, network interventions may have both intended and unintended consequences for group outcomes (Sørensen 1970; Valente 2012). Thus, for higher education administrators, manipulating coursework is a powerful form of network engineering that requires attentiveness to potential social as well as academic consequences. Second, we show that even foci with relatively low levels of constraint on interpersonal interaction can shape social relationships in significant ways (Feld 1981). While the current LC design did not impact mutuality or friendship density, future interventions may be able to impact these outcomes by focusing interactions more intensely. Such efforts could include placing students into study partnerships or groups that are even smaller than the classrooms and ~30-person study sections in the observed LC. Such actions could be especially fruitful for fostering network connections and supporting social integration for students from diverse backgrounds and other groups historically at greater risk of STEM attrition.

CHAPTER 2

SUPPLEMENTARY MATERIALS

Section 1: Description of LC Program, Data Collection, Sample Characteristics, and Missing Data

Learning Community Design

Figure S2.5 provides an overview of the learning community program, including eligibility requirements and the following types of program support:

(1) Academic remediation: LC students are required to take an additional developmental chemistry course online the summer prior to college matriculation. This course is designed to prepare potential biological sciences majors for college-level courses in chemistry and biology.

(2) Academic and social support: Students are placed into cohorts where they are matched with a senior biological sciences mentor. Mentors are upperclassman biological sciences majors selected by the department; they have a tutoring background and have excelled in introductory biological sciences courses. The mentors provide increased academic support and serve as students' main guide to campus resources and opportunities. Additionally, LC students participate in a weekly 50-minute seminar led by a LC mentor. Seminar topics are generally academic in nature and focus particularly on study skills, metacognition, and research experience. Also discussed are general first-year issues, such as how best to communicate with professors and TAs and how to manage fast-paced coursework in a quarter system.

We note that patterns of co-enrollment differed across years. In Year 1, non-LC students were allowed to co-enroll in the same introductory biology section with LC students, while in Year 2 they were not. Likewise, in Year 1, non-LC students could co-enroll in the same freshman seminar sections along with the LC students, whereas in Year 2, they could not. These differences across years likely help to explain the stronger size and segmentation effects of the LC in Year 2 as seen in Figures 2.1 and 2.3.

Survey Instruments and Data Collection Design

Electronic surveys were sent to the entire freshman cohort during the final week of Fall term to collect data on friendship ties within the major, for both intervention participants and non-participants, and various aspects of student background to serve as controls. Information on LC participation, as well as additional student demographic data was provided by the university. These surveys took roughly 20 minutes to complete and were administered to all first-year students enrolled as biological sciences majors. Student completion of the fall surveys was tied to course credit, helping generate a response rate of over 93% (see S2.6).

There were some differences in survey design across the two years. In Year 1, the campus learning management system's survey feature was used to administer the survey. This relatively unsophisticated survey tool required all questions to appear on the same page, and was unable to make use of display logic, text piping, or skip logic. As such, the question used to generate each friend's name was immediately followed by questions about that friend (i.e., name interpreter question). That is, students were asked to list their first friend, and then immediately answer one to three follow-up questions about that friend, before moving on to list the next friend, answer follow-up questions about the second friend, and so on. In Year 2, however, Qualtrics was used to administer the survey. With the benefit of display logic and text piping, this survey asked students to first list up to 10 friends, before then moving to subsequent pages in which follow-up questions were asked about those friends.

In addition, the wording and nomination limit varied slightly across cohorts. In Year 1, we stated, “These questions ask you about friends who are also Bio Sci majors. We define a friend as someone who you enjoy spending time with.” In Year 2, we simply stated “Please list people in the Bio Sci major that you consider your friends.” And, we limited the number of friendship nominations to eight in the first year but set the limit at ten for the second year. Supplementary analyses show less than 1% of students named the max number in Year 1, and around 1% named the max in Year 2.

These methodological differences are likely responsible for the observed difference in average number of friends across the two years as evident in Figures 2.1 and S2.6. In particular, the order of the Year 1 friendship questions could have had a “training effect” whereby respondents realized that naming a friend led to several follow-up questions. This can increase the burden of data collection and inhibit the elicitation of additional names [1-2] The design of the Year 2 survey captured friendship nominations (and hence network size) prior to students being asked follow-up, name interpreter questions. Hence, there is no possibility for training within the survey. As a methodological artifact, the difference in number of friends across years means that we cannot compare the two years but does not affect our tests of differences between LC and non-LC students *within* a given year because the methods were consistent across students within the year. To adjust for these differences, we use year fixed effects in the RD analysis and estimate ERGMs separately for each year.

Summary of Cohorts

Figure S2.6 provides a description of the biological sciences freshman cohorts, over the two years of this study in terms of its composition, academics, and networks. These cohorts represent all first-time (i.e., non-transfer), biological sciences freshman students, drawn from administrative records of those who have declared the relevant major and are thus on the departmental roster. Descriptive statistics highlight that across cohorts LC students are disproportionately URM, first-generation college, low income, and female, compared to their non-LC counterparts. Academically, we see that in Year 2, LC students entered with slightly lower SAT math scores compared to Year 1 LC students and ended the fall term with a lower average cumulative GPA. Finally, with respect to networks, we see that mean outdegree increased across cohorts which, as discussed, we attribute to methodological changes.

Sample and Missing Data

Our sample comes from the roster of students identified as a Biological Sciences major each year. Given our high response rate and ability to draw on administrative records, there was a minimal amount of missing data; most covariates were missing less than three percent. For the RD analyses, we utilized listwise deletion to create our analytic sample, resulting in a final pooled sample size of N=1,854. For the ERGMs, we retained all students (Year 1 N=879; Year 2 N=1,083) and used mean imputation for missing covariate data since the ability to keep all cases is critical in network analysis to avoid losing important structural features. Outgoing friendship nominations of non-respondents were specified as missing, which allows them to be imputed during model fitting, but treats them as non-informative to parameter estimates.

Section 2: Description of RD Measures

Dependent Variables

There are many potential ways to conceptualize friendship ties, but we draw upon measures of the size, strength, structure, and composition of friendship networks since these dimensions index important aspects of resources and support [3]. To construct these outcomes, each respondent's friend nominations were matched to data from other respondents and administrative data provided by the university. All measures are constructed at the individual level from complete network (or sociocentric) data gathered at the major cohort level [4-5]. Additionally, our measures are limited to friends who are directly connected to a focal student (i.e., local neighborhood with distance of one).

Network size refers to the number of friends with whom a focal student is connected. In this analysis, we measure the number of ties in terms of total degree, where we take the total number of peers a focal student is connected to, without differentiating who named who as a friend (the focal student or the peer). Figure S2.7 presents a visualization of variation in network size, where three different networks are displayed; the upper left student is an isolate having no friends, the upper right focal student has eight friends, and the lower left focal student has three friends.

Tie strength reflects the idea that relationships vary along dimensions such as closeness, intensity, and meaningfulness. One indicator of strength in a dyad is whether it is significant enough that both members acknowledge the relationship by naming each other as friends. Such relationships are defined as mutual (or reciprocal). To capture how student networks vary in terms of overall strength, we calculate the count of all a students' ties that are reciprocated. Alternative analyses utilizing **1)** the *count of nonreciprocal* ties, **2)** the *proportion of reciprocal* ties, or **3)** the *count of persistent (or stable)* ties from week 2 to week 10 as the outcome, produce substantively similar results. Figure S8 provides a sociogram of two different friendship networks, where the focal student on the left has four mutual ties (i.e., four edges with arrows going in both directions), whereas the focal student on the right has four non-mutual ties (i.e., two incoming arrows and two outgoing arrows).

Network structure encapsulates a more holistic view of friendships by recognizing that students not only have friends, but that those friends may be connected to one another. When one's friends are themselves friends, then a group exists. Such social closure allows for the development of group norms, enhanced social influence, and greater exchange of resources. We use density as our measure of network structure, calculated as the number of observed ties among a respondent's named friends divided by the number of potential ties. In Figure S2.9, we present three different friendship network structures, where each focal student has the same number of friends, but they differ in the number of ties present among their friends.

Finally, network composition refers to the makeup of one's network. Although many characteristics could be considered, given the design of the LC, we expect friendships with LC participants to vary greatly depending upon whether a student belongs to the LC. Thus, we examine the proportion of friends in the learning community. Figure S2.10 provides a visualization of variation in network composition. If we imagine that the colors corresponded with LC participation, then the upper left network would indicate that all friends of the focal student are either LC participants or non-participants, whereas the other two networks have varying degrees of heterogeneity in terms of LC participation.

Control Variables

We account for differences in observed background characteristics by including covariates gathered from administrative and survey data to improve the precision of our estimates. Specifically, we include measures of gender (reference=female), first-generation student status (yes=1), low-income status (yes=1), Fall term cumulative GPA, high school GPA, SAT reading score, and SAT writing score provided through administrative sources. We also include a measure of race/ethnicity (reference=White) that was provided by individual students through administrative sources in Year 1 and primarily through survey data collection in Year 2, with imputation from administrative sources when survey information was missing. For this reason, four categories were used for the Year 1 cohort (i.e., White, Black/African American, Hispanic/Latinx, Asian/Asian American), whereas an additional “other” category was included for the Year 2 cohort that corresponded to students who identified as multiple racial/ethnic categories, or who checked the “other” category in the survey. Finally, we include a variable indicating survey completion status (completed=1), since network measures can be sensitive to missing data.

Section 3: Checking Conditions of RD

As a crucial first step to ensure an accurate estimate of the impact of the program through an RD framework, we must evaluate the main conditions of the model. As directed by Jacob et al. [6], we use a variety of graphical plots to explore the relationship between the rating variable and other measures of interest. Specifically, through these graphical plots and significance tests where applicable, we aim to **1)** examine whether the RD design should be considered “sharp” or “fuzzy”, **2)** assess the internal validity of the RD design by examining potential manipulation of the running variable or differences in baseline characteristics at the cutoff, and **3)** visualize the discontinuity of the outcome variables at the cutoff to help explore potential functional form issues.

First, we used the *rdbwselect* function in STATA to estimate the optimal bandwidth on the pooled sample [7]. Although the optimal bandwidth varied across outcomes, we chose a bandwidth of ±70 points, since this closely matched similar previous studies [8]. Supplementary analyses using half and double the bandwidth size produced substantively similar results. As can be seen in Figures S2.11 and S2.12, we find evidence of a discontinuity in the probability of receiving the treatment at the cutoff. However, this discontinuity is not perfect since there is evidence of noncompliance, indicating that we should pursue a fuzzy RD design [9]. Next, we plot the density of the running variable using STATA's user-written *rddensity* command [10]. Significance tests of a discontinuity at the cutoff indicate that there is a significant jump for Year 2 ($p<.001$) but not Year 1 ($p>.10$)(see Figures S2.13 and S2.14). Finally, we plot the baseline characteristics of the students as a function of the assignment variable (see Figures S2.15 and S2.16). In general, we do not see much evidence of a discontinuity in baseline characteristics for Year 2, but there do seem to be some jumps for Year 1. Supplementary analyses reveal substantively similar effects across years. All RD models control for observed background characteristics to improve precision.

Section 4: Description of ERGMs and Estimated Effects

ERGM Overview

Whereas the RD analysis demonstrates the causal impact of the LC on friendship networks, it does not offer insight to *how* the LC had its demonstrated effect. Toward that end, the social network analysis was performed with the goal of discerning whether the impact of LC participation on friendship was directly due to LC organizational factors (i.e., course scheduling) versus alternative mechanisms that may have resulted from the assignment of students to courses. The alternative mechanisms we investigate reflect different ways that the students assigned to the LC could have reasonably been different from the students who did not participate in the LC. The general logic of our analysis was to **1)** estimate a base ERGM that only included the hypothesized main effect of LC participation, **2)** estimate subsequent models that introduced effects to account for alternative friend selection mechanisms, and **3)** a full model that included foci effects as well as alternative factors simultaneously. ERGMs were estimated using the *statnet* package in R.

Because of scaling issues that arise with nonlinear outcomes, we converted estimated ERGM coefficients to marginal effects to compare models. To determine if an alternative mechanism is responsible for the observed effect of LC participation, we calculated the marginal effect of the respective LC effect in the base model and compared it to subsequent models that introduced potential confounds or mediators. If a confound or mediator were responsible for the observed effect of the LC on friendship, then the marginal effect of the LC would decrease compared to the base model. We also calculated how much of the main LC effect was explained (or mediated) by effects introduced to the model, following the method proposed by Duxbury [11] and implemented in the *ergMargins* package in R.

The ERGM itself is a multivariate network model that estimates the presence, versus absence, of friendships (or "ties") conditional upon effects included in the model. The set of effects included in an ERGM capture particular configurations of ties in the network. All models included an *edges* effect, to control for the overall probability of observing a tie. Additional effects can incorporate individual, dyadic, or network properties as discussed below.

Unlike the RD analysis, the ERGM considered the entire major cohort for each year, with students designated as either LC participants or non-participants. Given our quasi-experimental design, with LC status exogenously assigned to students, we focus on distinguishing the total effect of LC assignment through the assignment of students to cohorts and sections from potential effects due to the manipulation of other population features. The alternative explanations we test are all based on individual and dyadic properties; hence, our specification is also referred to as a dyad-independence model [12].

ERGM Specification of Hypothesized Effects

We estimate a sequence of ERGMs to test each of the two hypotheses supported by the RD design. Each sequence of models begins with a base model that represents the main effect of LC participation on friendship. First, we consider the finding that LC participants had more friends than non-participants. The base model testing this effect uses a *nodecov* effect associated with student LC participation (see ref. 13 for a fuller description of ERGM effects). This effect calculates the sum of the LC participation covariate across the two students in each dyad. This sum evaluates to “0 if neither student is an LC participant, “1” if one student is an LC participant, and “2” if both are LC participants. The associated parameter estimate can be interpreted as the difference in the log-odds of observing a tie for a one-unit increase in this sum (e.g., for a dyad with no LC members to a dyad with one LC member). The results for the base model are reported as M1 in Figures S2.17a and S2.17b. As a robustness test, we estimated an alternate set of models that only considered friendships from the perspective of the student naming a friend (i.e., using the *nodeocov* effect). These models produced substantively similar findings (see Figures S2.19a and S2.19b).

The second series of models evaluates the finding that LC participants were more likely to name friends in the LC than were non-LC participants. This finding is tested with a dyadic effect capturing the combination of student LC participation in each dyad. The *nodemix* effect creates a separate dyadic indicator for all but one combination of LC participation status. Our specification created effects to represent friendships from one LC-participant to another LC-participant (LC → LC), a non-participant to a non-participant (non-LC → non-LC), and an LC-participant to a non-participant (LC → non-LC). This left friendships where a non-participant named an LC-participant as a friend as the reference category (non-LC → LC). Our interest is in whether LC → LC dyads were more likely to exhibit friendships than non-LC → LC dyads (the reference category). Hence, we treat the LC → LC dyad parameter estimate as our indicator of segmentation. Results for this base model are reported as M1 in Figures S2.17a and S2.17b.

ERGM Specification of Alternatives

LC assignment was based on math SAT scores, which are correlated with student background characteristics. Hence, the composition of the LC was different from the overall composition of the major. (see Figure S2.6). One possibility is that the students assigned to the LC were more sociable or more likely to make friends due to background characteristics, such as race and gender [14]. Hence, we consider several factors that were associated with either LC placement or friendship volume, including race/ethnicity, gender, first-generation student status, low-income status, high school GPA, and commuter status. These attributes were entered into the model using *nodecov* effects or, for categorical attributes (i.e., race), a *nodefactor* effect (which specified a separate effect for each level of the factor, excluding one). Full model results are shown as M2 in Figures S2.17a, S2.17b, S2.18a, and S2.18b.

Another explanation for our observed findings is that the composition of the LC friendship pool affected the capacity for students to find friends who were similar to themselves. Homophily is one of the most common patterns found in human social relationships [15], including among university students [14]. Foci such as the LC have the potential to act as a filter, bringing together a set of individuals that is more homogenous than the broader population and promoting relationships among them [16-17]. If this were to occur, then LC students from backgrounds that are over-represented in the LC would have an easier time finding similar peers to befriend than students outside the LC. We test for this possibility using **1)** a set of effects that represent similarity on background factors within each dyad – *nodematch* effects for dichotomous measures and categorical measures (0=different scores, 1=matching scores), *absdiff* effects for continuous measures (reflecting the absolute difference between two students' scores) – and **2)** a set of interactions between similarity and whether both members of a dyad were LC participants ("1"=both LC participants, "0"=at least one student not an LC participant; created as dyadic measures and entered as *edgecov* effects). With this specification, the main effects of similarity capture the strength of homophily for dyads that did not include two LC participants, and the interactions represent how homophily among LC participants deviated from this main effect. Results are presented as M3 in Figures S2.17a, S2.17b, S2.18a, and S2.18b.

A third possibility is that the LC brought together students who were more likely to have a pre-existing friendship. For instance, LC participants may have been more likely to draw from the same high schools. For Year 2 only, we have information on which high school each student in our sample attended. We use this data to construct a dyadic covariate representing, for each pair of students in the sample, whether they attended the same high school (1=yes, 0=no). We included this covariate in the model with an *edgecov* term. Results are presented as M4 in Figures S2.17b and S2.18b.

It is possible that the aforementioned alternative explanations each accounted for a small part of the LC effect observed, but not enough to fully explain it away. Hence, we estimated a composite model that included all of the alternative explanations tested in M2-M4. We present this as M5 in Figures S2.17a, S2.17b, S2.18a, and S2.18b.

Our final models test the hypothesis that the effect of LC participation on friendships operated through assignment to the same classes and LC section (M6 in Figures S2.17a, S2.17b, S2.18a, and S2.18b). We focus on the three courses that all majors were required to take and were subject to the block-scheduling design: “Introduction to Biology”, “Introduction to Chemistry”, and “Freshman Seminar.” For each of these courses, we coded each student dyad “1” if they took the class together, and “0” if they did not. Similarly, dyads were coded “1” if both members were in the same LC section and “0” otherwise (including dyads that included LC non-participants). We used *edgecov* effects to test how these four foci affected friendships. M6 introduces these four effects along with the main LC effect of interest.

Average Marginal Effect (AME) Results

To evaluate the potential confounds represented by the alternative selection mechanisms, as well as the mediating effect of shared foci, we converted ERGM parameter estimates into partial average marginal effects following Duxbury [11]. Partial AMEs represent the direct effect of the LC predictor net of the other effects introduced to each model. Partial AMEs and 95% confidence intervals for each model are presented in Figure S2.20. Figure S2.21 presents these partial marginal effects and standard errors, along with the total AME, indirect AME of the LC effects on friendship (via the effects introduced to the model) and standard errors. We also calculate the percent of the total AME that is mediated by the partial AME.

Supplementary Tables and Figures

S2.1. T-test of Means on Network Outcomes by LC Status and Year

Network Outcome	Year	LC Participant (yes=1)	Mean	SD	P-value
Total Degree					
	1	1	2.94	2.08	
	1	0	2.50	1.98	0.005
	2	1	5.21	3.12	
	2	0	4.06	2.83	<0.001
Count of Mutual Ties					
	1	1	1.03	1.08	
	1	0	0.97	1.06	0.47
	2	1	2.06	1.61	
	2	0	1.73	1.42	0.002
Density					
	1	1	0.27	0.31	
	1	0	0.22	0.27	0.04
	2	1	0.25	0.23	
	2	0	0.25	0.24	0.86
Proportion of Friends in LC					
	1	1	0.76	0.29	
	1	0	0.12	0.25	<0.001
	2	1	0.82	0.23	
	2	0	0.08	0.18	<0.001

S2.2. Logistic Regression of LC Participation on Odds of Isolation within Major Cohort

	M1 (Full sample)	M2 (Within bandwidth)
Learning Community Participant (yes=1)	0.45**	0.49**
	(.104)	(.133)
Individual-level Controls	X	X
Year Fixed Effects	X	X
N	1,854	1,031

†p<.10; *p<.05; **p<.01;***p<.001. Standard errors in parentheses.

NOTES: Results reported as odds ratios.

Models examine isolation as measured at the end of fall term among freshmen.

Covariates include: race/ethnicity (reference=White), gender, first-generation student status, low-income status, Fall term cumulative GPA, high school GPA, SAT reading, SAT writing, and survey completion status.

S2.3. LATE Estimates from Pooled Local Polynomial Regression on Various Network Outcomes, LC Effect: Bandwidth ±70

Network Outcome Measure	M1	M2	M3	M4
Network Size				
Total degree	1.01†			
	(.559)			
Tie Strength				
Count of Reciprocal Ties		0.32		
		(.318)		
Network Structure				
Density			-0.09	
			(.066)	
Network Composition				
Proportion of Friends in LC				0.54***
				(.062)
Individual-level Controls	X	X	X	X
Year Fixed Effects	X	X	X	X
N	1,031	919	772	919

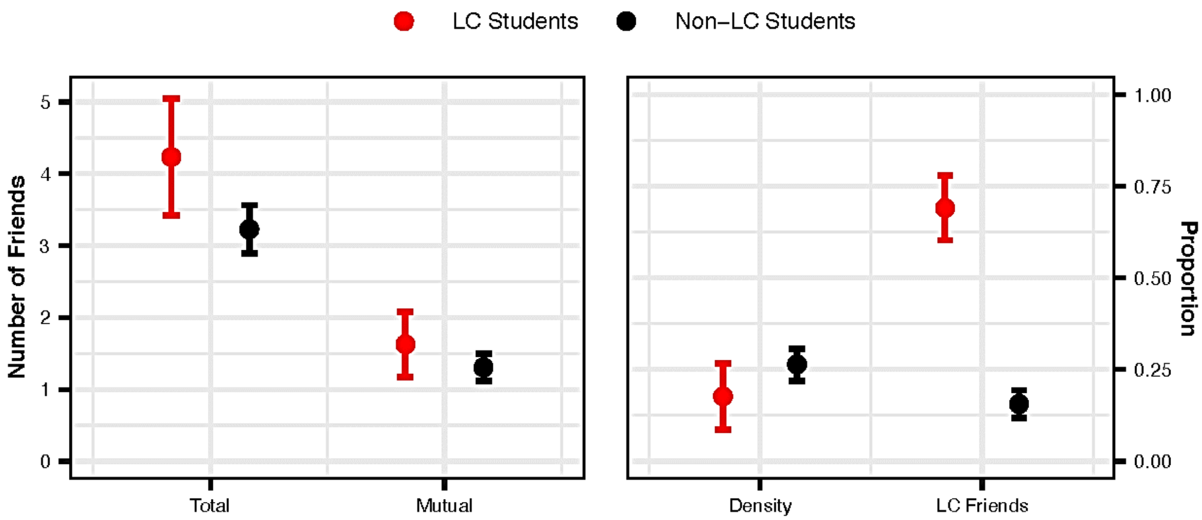
†p<.10; *p<.05; **p<.01;***p<.001. Robust standard errors in parentheses.

NOTES: Estimates adjust for fuzzy RD design.

Covariates adjusted include: race/ethnicity (reference=White), gender, first-generation student status, low-income status, Fall term cumulative GPA, high school GPA, SAT math score distance from cutoff, SAT math score distance squared, an interaction term between SAT math score distance from cutoff and whether the student was below the eligibility threshold, SAT reading, SAT writing, and survey completion status.

Model 1 includes all students within bandwidth; Model 2 is limited to students with at least one tie; Model 3 is limited to students with at least two ties; Model 4 is limited to students with at least one tie.

S2.4 Predicted Network Outcomes Based on LATE Estimates



S2.5. Learning Community Program Description

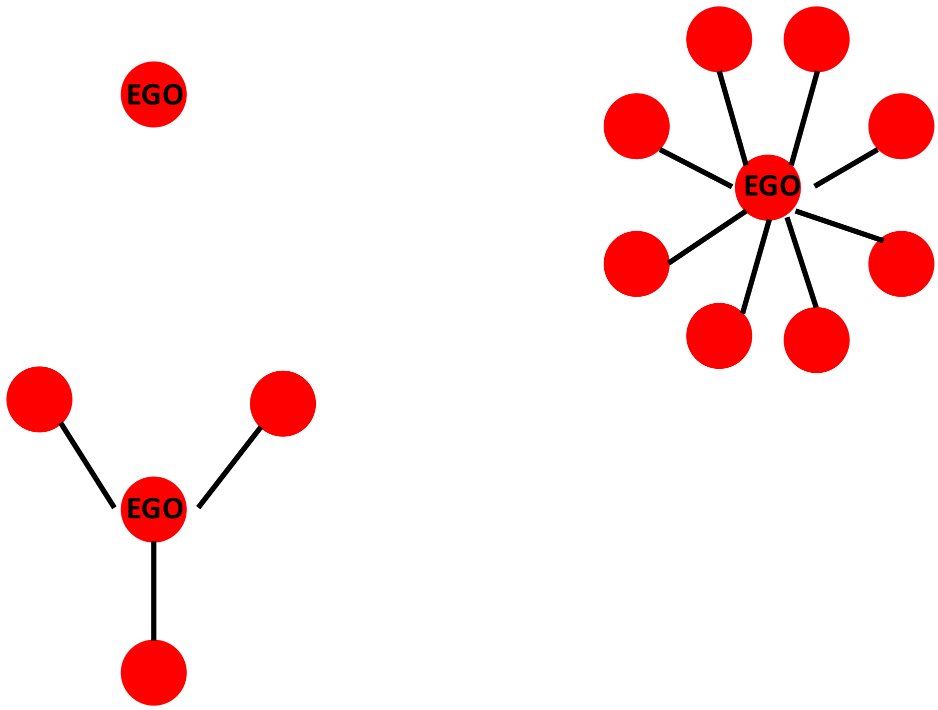
Year	Eligibility	Program Component				Attrition
		Additional Chem Prep	Peer Mentor	LC Seminar (1-hour weekly session)	LC/Non-LC Co-enrollment Practices	
1	<600 SAT Math	No	Yes; holds 1-hour weekly session	Yes; each quarter	Non-LC students allowed to co-enroll in LC intro bio section; Non-LC students allowed to co-enroll in freshman seminar sections	Bio: Students must earn a D- or above to move forward in the sequence. Chem: Students must earn a C- or above to move forward in the sequence.
2	Same as above	Chem 1X (fall)- Required[1]	Same as above	Same as above	Non-LC students not allowed to co-enroll in LC sections	Same as above

[1]Co-requisite with *Introduction to Chemistry* course so LC students remain on-sequence with the rest of the cohort in terms of Chemistry.

S2.6. Descriptive Overview of Biological Sciences Freshman Cohorts

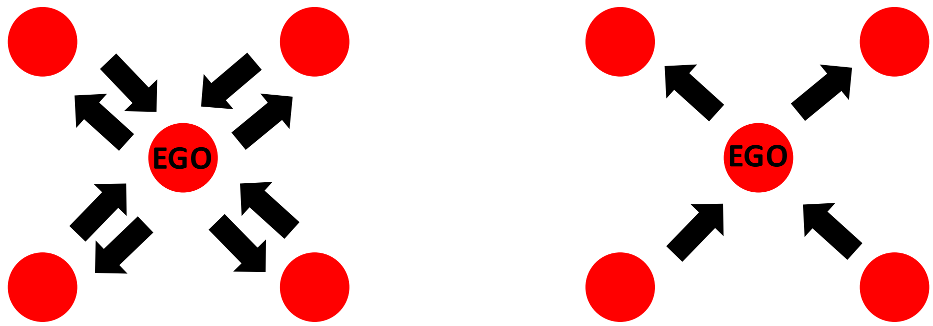
		Year 1			Year 2		
Composition		LC	Non-LC	Overall	LC	Non-LC	Overall
	Size of Cohort (N)	240	637	879	290	784	1083
	% URM	67.1	18.5	31.8	75.2	27.7	40.4
	% First-Generation College	77.4	40.5	50.7	75.0	43.1	51.7
	% Low-Income	51.3	22.8	30.6	60.1	31.0	38.8
	% Female	80.4	63.6	68.2	84.6	63.3	68.9
Academics							
	Mean SAT Math	516	656	617	498	658	616
	Mean SAT Reading	539	608	589	522	609	586
	Mean SAT Writing	514	606	581	502	607	579
	Mean Fall Term GPA	2.43	3.16	2.95	1.95	3.16	2.83
Networks							
	Mean Outdegree	1.88	1.66	1.72	3.42	2.75	2.93
	Mean Indegree	1.93	1.65	1.72	3.54	2.71	2.93
	Mean Total Degree	2.91	2.50	2.61	5.11	3.96	4.27
	Mean Count of Mutual Ties	1.02	0.96	0.98	1.96	1.67	1.75
	Mean Density	0.28	0.22	0.24	0.26	0.26	0.26
	Mean Proportion of Friends in LC	0.76	0.12	0.30	0.82	0.08	0.29
Survey							
	% Survey Completion	92.1	94.2	93.6	94.5	95.3	95.1

S2.7. Examples of Variation in Network Size



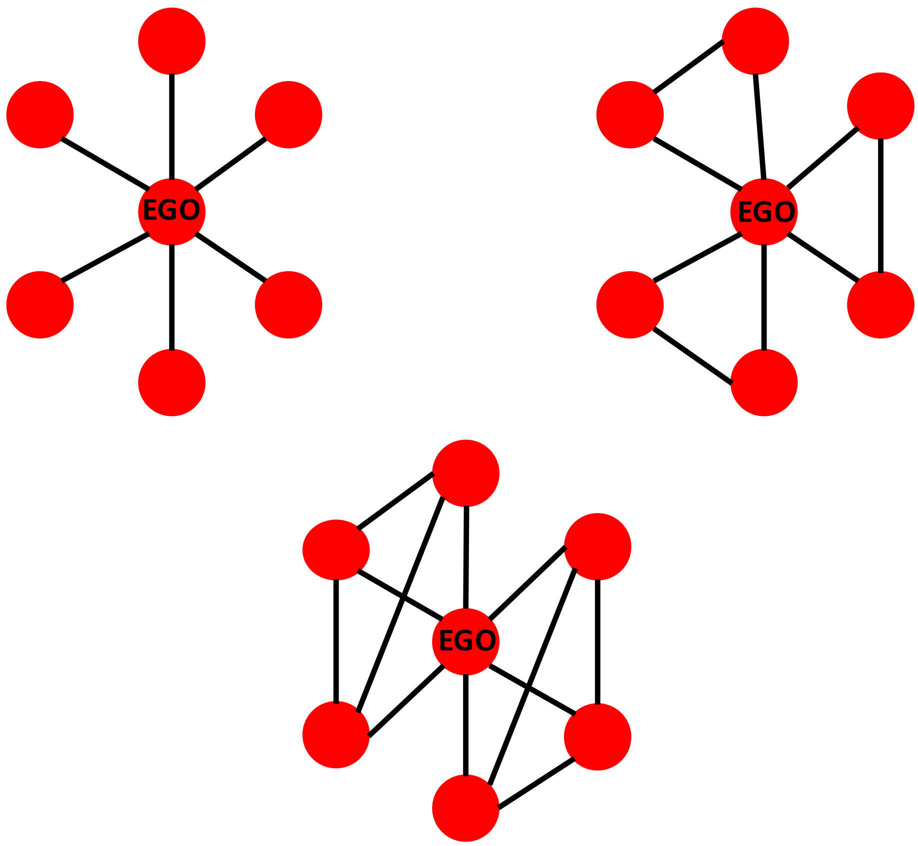
NOTE: Focal individual (ego) represented by middle node (or circle); alters represented by outer nodes.

S2.8. Examples of Variation in Tie Strength



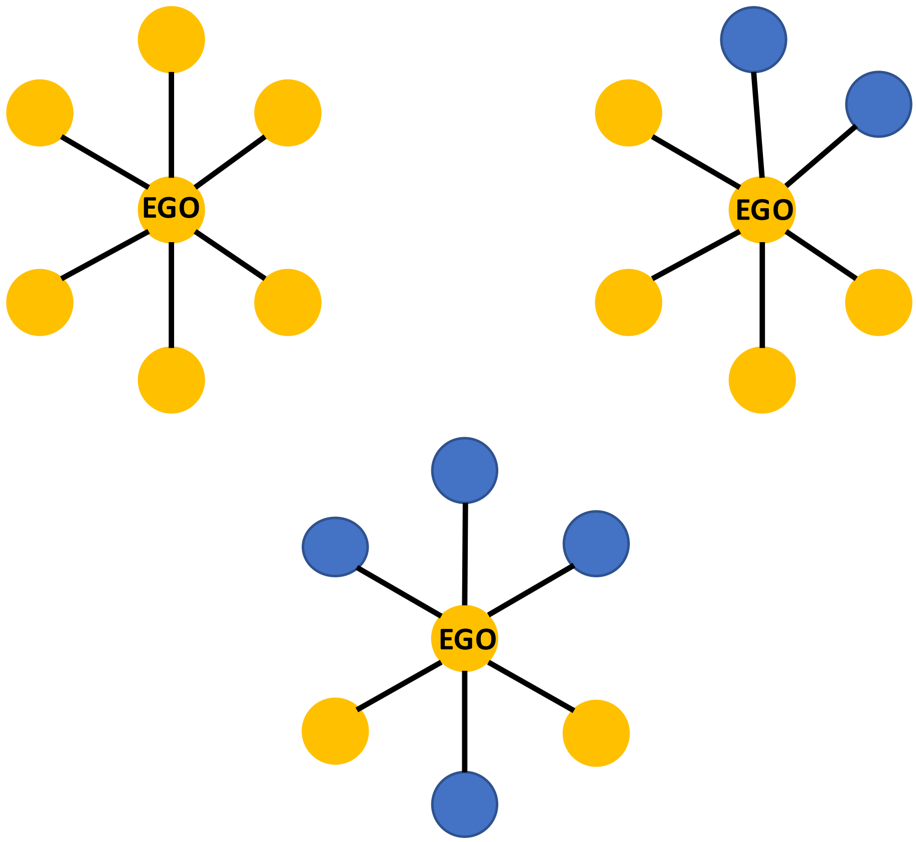
NOTE: Focal individual (ego) represented by middle node (or circle); alters represented by outer nodes; arrows indicate directionality of tie nomination.

S2.9. Examples of Variation in Network Structure



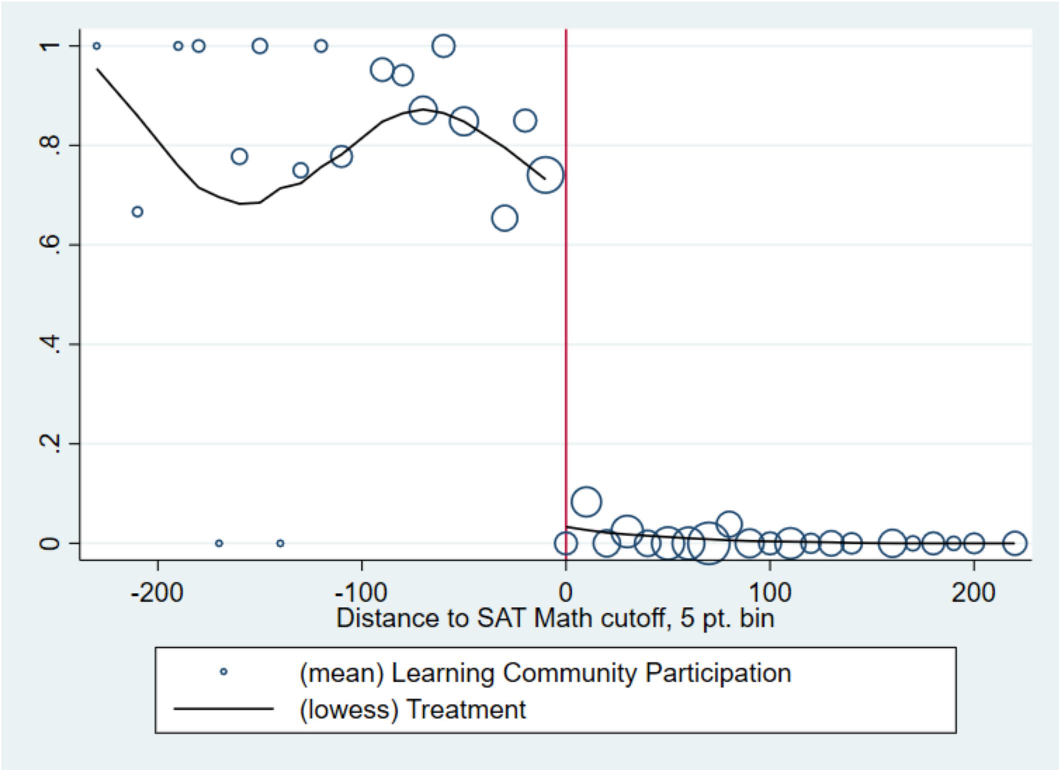
NOTE: Focal individual (ego) represented by middle node (or circle); alters represented by outer nodes.

S2.10. Examples of Variation in Network Composition

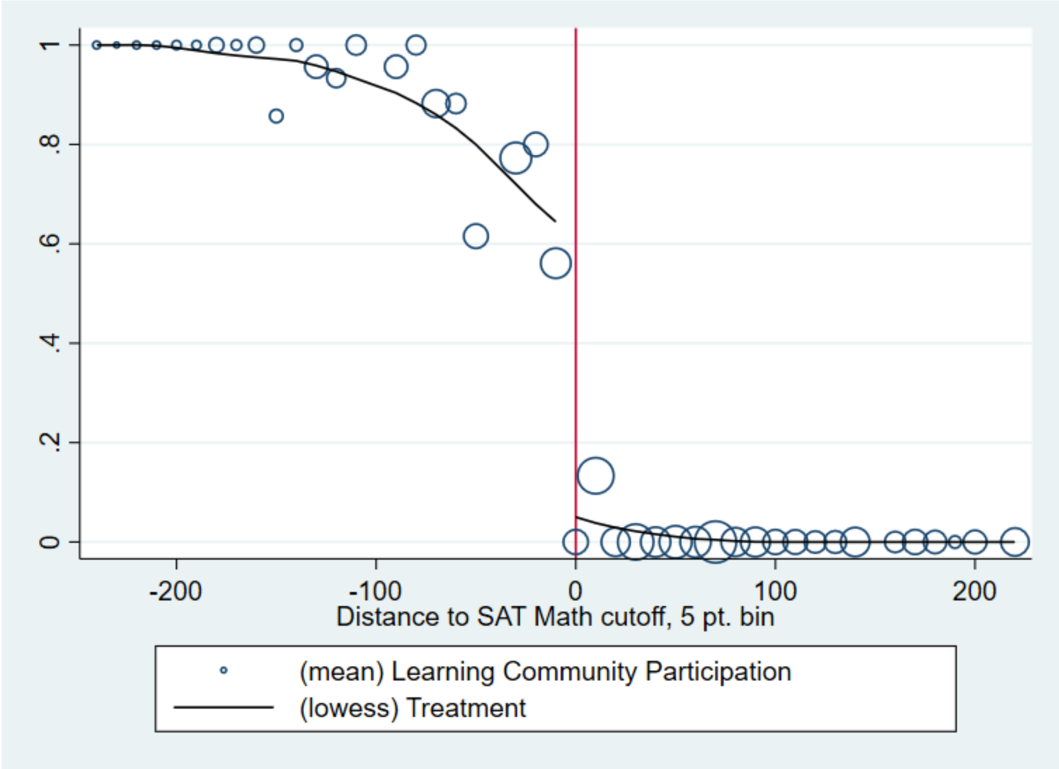


NOTE: Focal individual (ego) represented by middle node (or circle); alters represented by outer nodes; node color specifies group membership based on some characteristic, category, or other differentiating criteria.

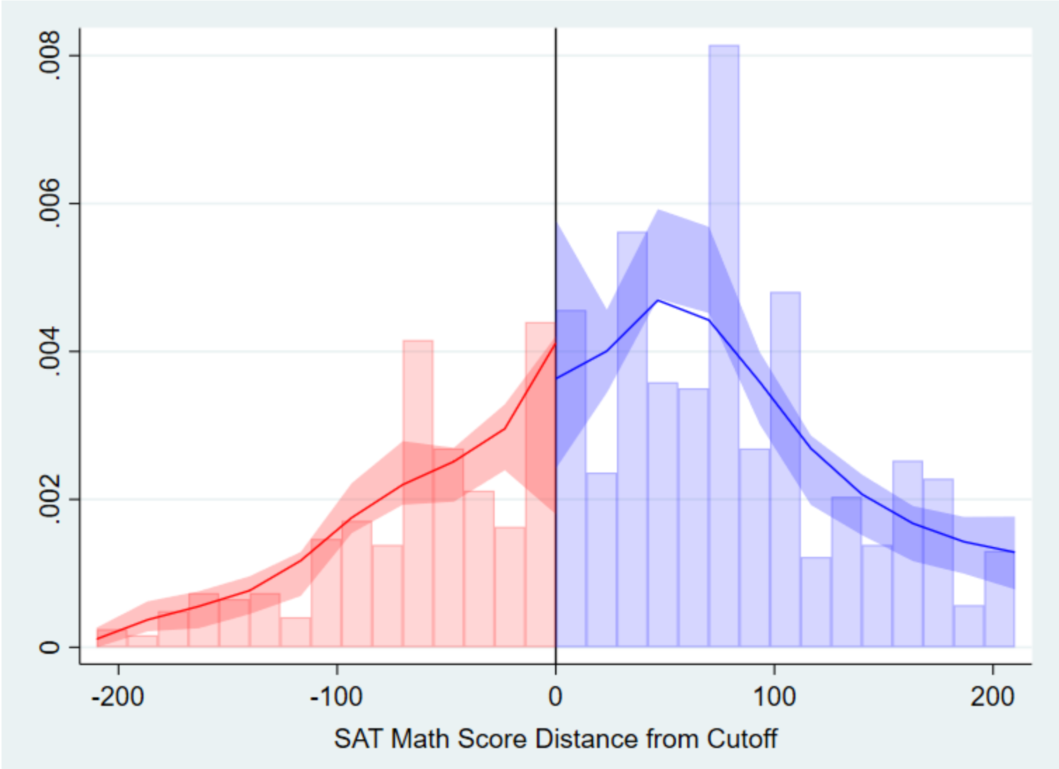
S2.11. Probability of Treatment by Distance from SAT Math Cutoff: Year 1



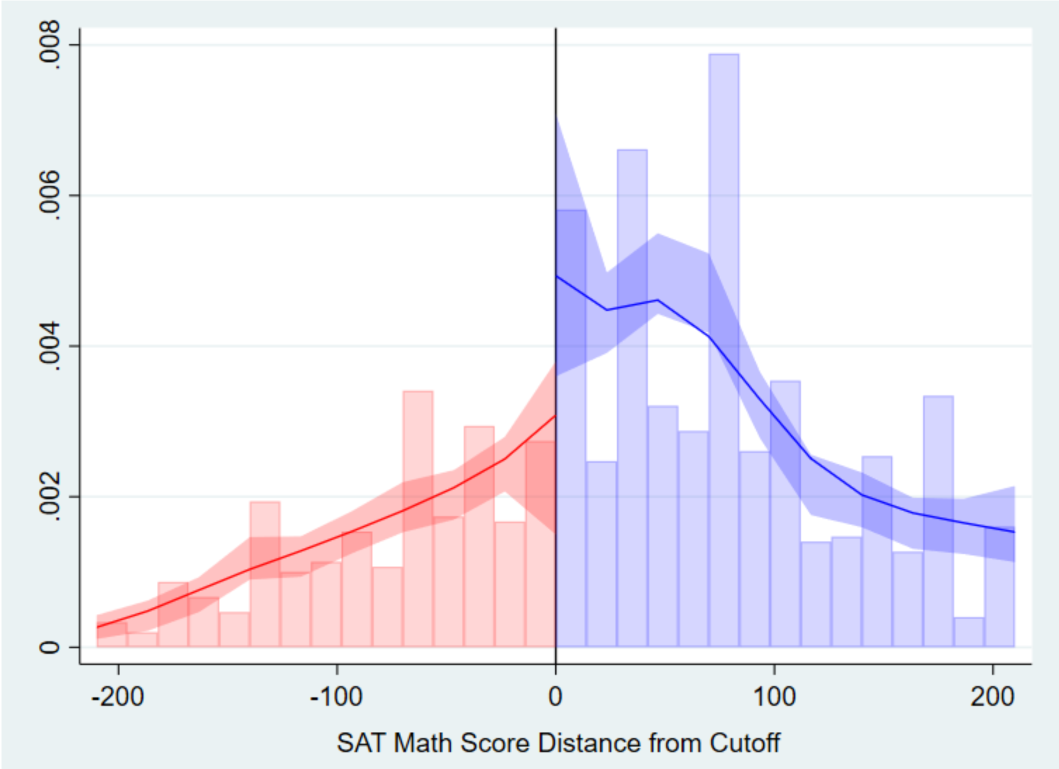
S2.12. Probability of Treatment by Distance from SAT Math Cutoff: Year 2



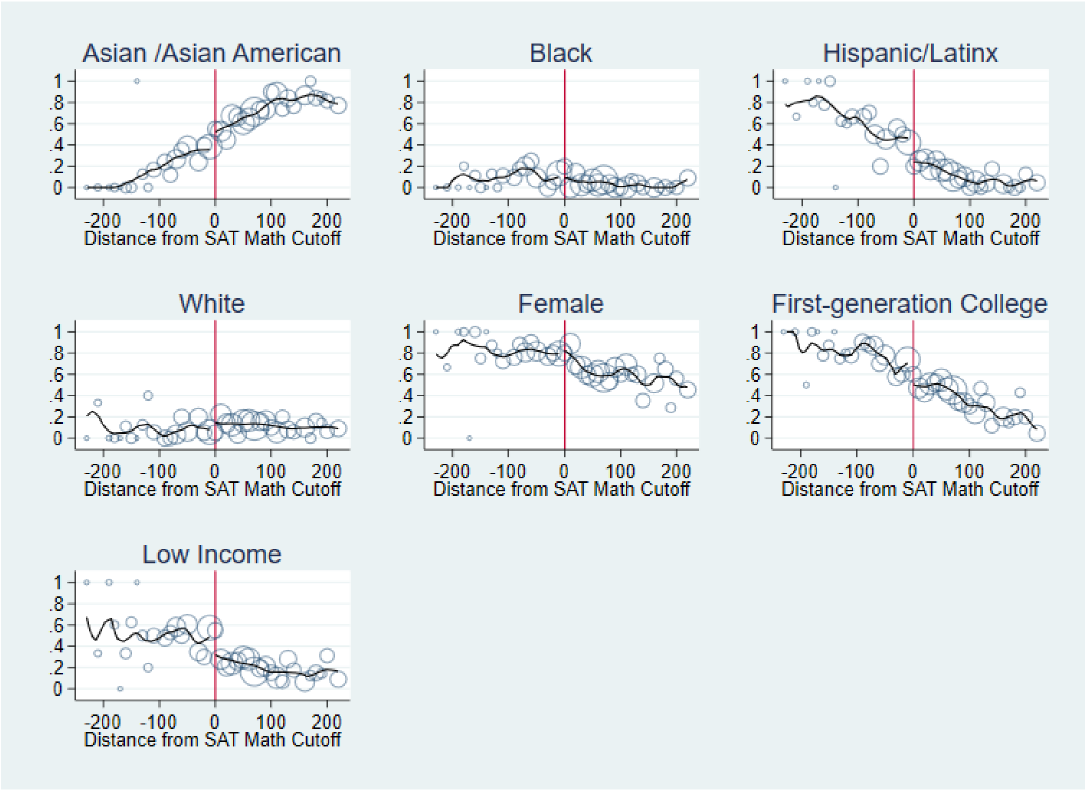
S2.13. RD Manipulation Test using Local Polynomial Density Estimation: Year 1



S2.14. RD Manipulation Test using Local Polynomial Density Estimation: Year 2



S2.15. Distribution of Baseline Characteristics by Distance from SAT Math Cutoff: YR 1



S2.16. Distribution of Baseline Characteristics by Distance from SAT Math Cutoff: YR 2

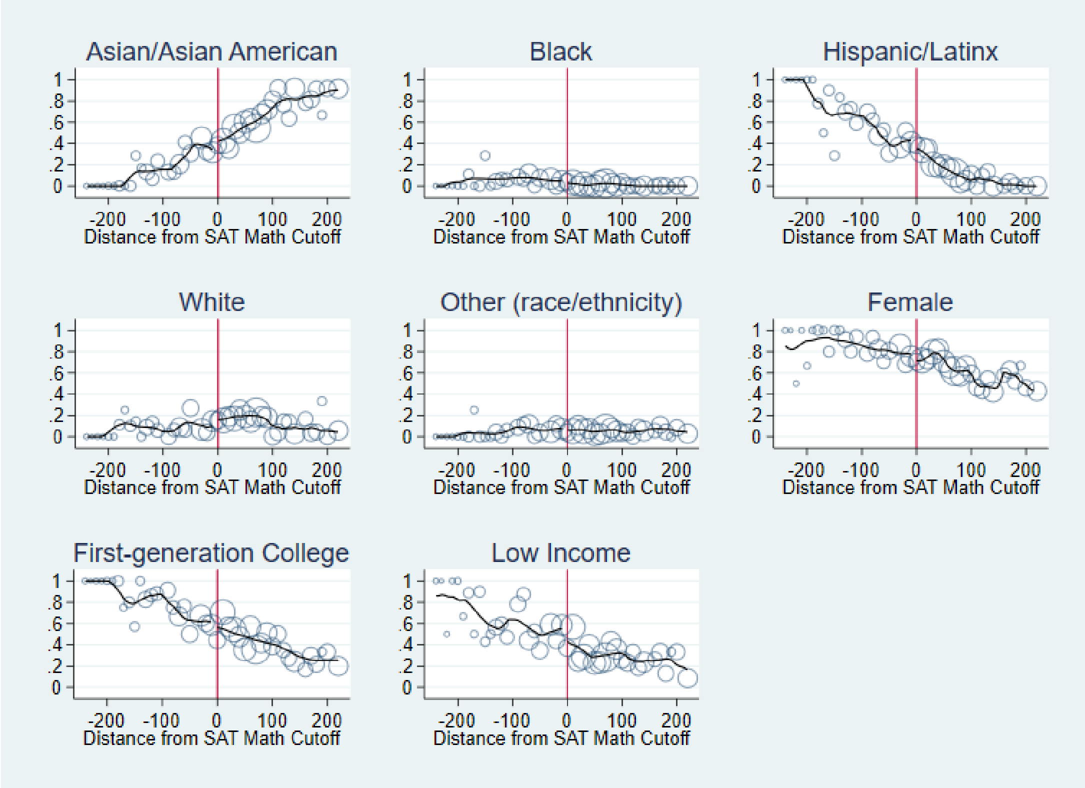


Figure S2.17a. Full ERGM results for models testing the effect of LC-participation on friendship volume (Year 1)

Year 1	M1			M2			M3			M4		M5			M6		
	coeff.	SE		coeff.	SE		coeff.	SE		coeff.	SE	coeff.	SE		coeff.	SE	
edges	-6.313	.035	***	-9.056	.838	***	-8.817	.157	***			-9.732	.967	***	-10.229	.052	***
nodecov																	
LC participant (ref: non-participant)	**.143**	**.040**	*******	**.222**	**.047**	*******	**.725**	**.089**	*******			**.707**	**.090**	*******	**-.482**	**.046**	*******
nodefactor																	
race (ref: Asian)																	
White				-.096	.061							.391	.059	***	.370		***
Latinx				-.204	.051	***						.030	.052		-.003		
Black				-.226	.082	***						.308	.081	***	.283		***
nodecov																	
first-generation (ref: non-first-generation)				-.059	.043							-.025	.044		-.051	.044	
female (ref: male)				.226	.042	***						-.034	.039		-.033	.039	
low-income (ref: not low-income)				.009	.045							.010	.045		-.006	.045	
high school GPA				.325	.102	***						.104	.115		.229	.116	*
commuter (ref: on-campus)				-.185	.047	***						-.198	.047	***	-.201	.047	***
nodematch																	
same race							.894	.064	***			1.096	.073	***	1.057	.073	***
same first generation							.207	.059	***			.183	.060	**	.167	.061	***
same gender							.889	.067	***			.908	.071	***	.916	.071	***
same LC participation status							1.703	.110	***			1.674	.112	***	.295	.126	**
absdiff																	
high school GPA							-.973	.188	***			-.956	.206	***	-.757	.209	***
edgecov																	
LC participant X same race							-.051	.129				-.110	.138		-.060	.138	
LC participant X same first generation							-.442	.130	***			-.339	.138	*	-.277	.138	*
LC participant X abs.diff high school GPA							1.093	.386	***			1.018	.387	**	.978	.391	**
LC participant X same gender							-.493	.143	***			-.489	.144	***	-.464	.145	***
same bio course															1.400	.059	***
same chem course															.947	.066	***
same freshman seminar															.392	.056	***
same LC section															2.245	.115	***
BIC	21941			21964			20884					20922			19418		

Figure S2.17b. Full ERGM results for models testing the effect of LC-participation on friendship volume (Year 2)

Year 2	M1			M2			M3			M4			M5			M6		
	coeff.	SE		coeff.	SE		coeff.	SE		coeff.	SE		coeff.	SE		coeff.	SE	
edges	-5.968	.025	***	-6.102	.553	***	-8.491	.106	***	-6.240	.027	***	-9.425	.640	***	-8.972	.654	***
nodecov																		
LC participant (ref: non-participant)	**.253**	**.027**	*******	**.339**	**.033**	*******	**.824**	**.059**	*******	**.380**	**.027**	*******	**.776**	**.060**	*******	**-.284**	**.066**	*******
nodefactor																		
race (ref: Asian)																		
White				-.077	.041	†							.360	.040	***	.333	.040	***
Latinx				-.076	.035	*							.141	.038	***	.097	.037	***
Black				-.042	.077								.494	.079	***	.483	.080	***
Other				.008	.055								.511	.056	***	.521	.056	***
nodecov																		
first-generation (ref: non-first-generation)				-.003	.030								.020	.030		-.001	.031	
female (ref: male)				-.142	.028	***							-.282	.027	***	-.293	.027	***
low-income (ref: not low-income)				-.118	.029	***							-.136	.030	***	-.131	.030	***
high school GPA				.065	.067								.149	.076	.	.225	.078	***
commuter (ref: on-campus)				-.230	.033	***							-.348	.032	***	-.337	.032	***
nodematch																		
same race							.765	.043	***				.872	.050	***	.873	.051	***
same first generation							.264	.042	***				.158	.043	***	.149	.043	***
same gender							.513	.044	***				.628	.045	***	.625	.046	***
same LC participation status							1.972	.078	***				1.733	.079	***	-.350	.100	***
absdiff																		
high school GPA							-.160	.128					-.066	.139		.086	.141	
edgecov																		
LC participant X same race							-.219	.085	***				-.240	.096	*	-.154	.096	
LC participant X same first generation							-.301	.088	***				-.233	.093	*	-.175	.094	†
LC participant X abs.diff high school GPA							-.083	.219					.105	.222		-.039	.224	
LC participant X same gender							-.283	.096	***				-.083	.099		-.054	.101	
same high school										4.274	.051	***	3.980	.054	***	4.055	.056	***
same bio course																1.471	.048	***
same chem course																.698	.048	***
same freshman seminar																.387	.047	***
same LC section																2.464	.073	***
BIC	43261			43282			41095			39863			37905			35305		

Figure S2.18a. Full ERGM results for models testing the effect of LC-participation on friendship segmentation (Year 1)

Year 1	M1		M2		M3		M4		M5		M6	
	coeff.	SE	coeff.	SE	coeff.	SE	coeff.	SE	coeff.	SE	coeff.	SE
edges	-7.180	.092 ***	-9.849	.843 ***	-8.040	.120 ***			-8.972	.961 ***	-10.658	.970 ***
nodemix												
LC status (reference: non-LC → LC)												
non-LC → non-LC	1.119	.098 ***	1.040	.101 ***	.926	.099 ***			.914	.102 ***	.724	.106 ***
LC → non-LC	-.108	.134	-.108	.134	-.108	.134			-.108	.134	-.108	.135
LC → LC	**2.076**	**.107 *** **	**2.155**	**.110 *** **	**2.375**	**.195 *** **			**2.328**	**.197 *** **	**-.240**	**.220**
nodefactor												
race (ref: Asian)												
White			-.096	.061					.391	.059 ***	.370	.060 ***
Latinx			-.205	.051 ***					.030	.052	-.003	.051
Black			-.226	.082 ***					.308	.081 ***	.283	.082 ***
nodecov												
first-generation (ref: non-first-generation)			-.059	.043					-.025	.044	-.051	.044
female (ref: male)			.226	.042 ***					-.034	.039	-.033	.039
low-income (ref: not low-income)			.009	.045					.010	.045	-.006	.045
high school GPA			.326	.102 ***					.104	.115	.229	.116 *
commuter (ref: on-campus)			-.185	.047 ***					-.198	.047 ***	-.201	.047 ***
nodematch												
same race					.894	.064 ***			1.096	.073 ***	1.057	.073 ***
same first generation					.207	.059 ***			.183	.060 **	.167	.061 ***
same gender					.889	.067 ***			.908	.071 ***	.916	.071 ***
absdiff												
high school GPA					-.973	.188 ***			-.956	.206 ***	-.757	.209 ***
edgecov												
LC participant X same race					-.051	.129			-.110	.138	-.060	.138
LC participant X same first generation					-.442	.130 ***			-.339	.138 *	-.277	.138 *
LC participant X abs.diff high school GPA					1.093	.386 ***			1.018	.387 **	.978	.391 **
LC participant X same gender					-.493	.143 ***			-.489	.144 ***	-.464	.145 ***
same bio course											1.400	.059 ***
same chem course											.947	.066 ***
same freshman seminar											.392	.056 ***
same LC section											2.245	.115 ***
BIC	21321		21344		20896				20935		19419	

Figure S2.18b. Full ERGM results for models testing the effect of LC-participation on friendship segmentation (Year 2)

Year 2	M1		M2		M3		M4		M5		M6	
	coeff.	SE	coeff.	SE	coeff.	SE	coeff.	SE	coeff.	SE	coeff.	SE
edges	-6.920	.069 ***	-6.971	.555 ***	-7.591	.086 ***	-6.978	.069 ***	-8.576	.636 ***	-9.183	.651 ***
nodemix												
LC status (reference: non-LC → LC)												
non-LC → non-LC	1.243	.073 ***	1.158	.075 ***	1.072	.074 ***	1.056	.073 ***	.883	.077 ***	-.140	.087
LC → non-LC	-.153	.102	-.150	.102	-.159	.102	-.152	.102	-.155	.103	-.156	.103
LC → LC	**2.401**	**.078 *****	**2.489**	**.080 *****	**2.721**	**.134 *****	**2.343**	**.078 *****	**2.435**	**.135 *****	**-.709**	**.160 *****
nodefactor												
race (ref: Asian)												
White			-.077	.041 †					.360	.040 ***	.334	.040 ***
Latinx			-.076	.035 *					.142	.038 ***	.097	.037 ***
Black			-.042	.077					.494	.079 ***	.483	.080 ***
Other			.008	.055					.511	.056 ***	.521	.056 ***
nodecov												
first-generation (ref: non-first-generation)			-.003	.030					.020	.030	-.001	.031
female (ref: male)			-.142	.028 ***					-.282	.027 ***	-.293	.027 ***
low-income (ref: not low-income)			-.118	.029 ***					-.136	.030 ***	-.131	.030 ***
high school GPA			.065	.068					.149	.076 .	.225	.078 ***
commuter (ref: on-campus)			-.231	.033 ***					-.348	.032 ***	-.337	.032 ***
nodematch												
same race					.765	.043 ***			.873	.050 ***	.874	.051 ***
same first generation					.264	.042 ***			.158	.043 ***	.149	.043 ***
same gender					.513	.044 ***			.628	.045 ***	.625	.046 ***
absdiff												
high school GPA					-.161	.128			-.066	.139	.085	.141
edgecov												
LC participant X same race					-.219	.085 ***			-.241	.096 *	-.154	.096
LC participant X same first generation					-.301	.088 ***			-.233	.093 *	-.175	.094 †
LC participant X abs.diff high school GPA					-.083	.219			.105	.222	-.039	.224
LC participant X same gender					-.283	.096 ***			-.083	.099	-.054	.101
same high school							4.029	.051 ***	3.980	.054 ***	4.055	.056 ***
same bio course											1.471	.048 ***
same chem course											.698	.048 ***
same freshman seminar											.387	.047 ***
same LC section											2.464	.073 ***
BIC	41587		41609		41106		38468		37916		35614	

Figure S2.19a. Full ERGM results for models testing the effect of LC-participation on friendship volume (alternative specification of attribute effects - Year 1)

Year 1	M1		M2		M3		M4		M5		M6	
	coeff.	SE	coeff.	SE	coeff.	SE	coeff.	SE	coeff.	SE	coeff.	SE
edges	-6.266	.031 ***	-7.578	.591 ***	-8.149	.127 ***			-9.240	.679 ***	-9.639	.673 ***
nodeocov												
LC participant (ref: non-participant)	**.124**	**.056 ***	**.190**	**.066 *****	**.414**	**.114 *****			**.713**	**.090 *****	**-.394**	**.112 *****
nodeofactor												
race (ref: Asian)												
White			-.073	.086					.489	.094 ***	.472	.094 ***
Latinx			-.129	.072 †					.112	.081	.082	.080
Black			-.125	.112					.450	.120 ***	.429	.120 ***
nodeocov												
first-generation (ref: non-first-generation)			-.127	.061 *					-.101	.062	-.109	.063 †
female (ref: male)			.309	.060 ***					.048	.064	.059	.064
low-income (ref: not low-income)			-.025	.064					-.028	.064	-.039	.064
high school GPA			.300	.144 *					.080	.158	.199	.159
commuter (ref: on-campus)			-.156	.066 **					-.169	.066 *	-.178	.066 ***
nodematch												
same race					.812	.062 ***			1.049	.072 ***	1.067	.071 ***
same first generation					.154	.058 ***			.184	.060 **	.204	.060 ***
same gender					.795	.064 ***			.878	.070 ***	.937	.070 ***
same LC participation status					1.256	.099 ***			1.683	.111 ***	.615	.092 ***
absdiff												
high school GPA					-1.168	.188 ***			-.966	.197 ***	-.717	.194 ***
edgecov												
LC participant X same race					.201	.129			-.119	.136	-.209	.129
LC participant X same first generation					-.165	.132			-.345	.134 *	-.444	.126 ***
LC participant X abs.diff high school GPA					1.974	.358 ***			1.042	.386 **	.548	.371
LC participant X same gender					.025	.137			-.504	.144 ***	-.711	.124 ***
same bio course											1.373	.059 ***
same chem course											.903	.066 ***
same freshman seminar											.399	.056 ***
same LC section											2.135	.109 ***
BIC	21949		22005		20929				20945		19455	

Figure S2.19b. Full ERGM results for models testing the effect of LC-participation on friendship volume (alternative specification of attribute effects - Year 2)

Year 2	M1			M2			M3			M4			M5			M6		
	coeff.	SE		coeff.	SE		coeff.	SE		coeff.	SE		coeff.	SE		coeff.	SE	
edges	-5.889	.021	***	-5.661	.385	***	-7.817	.093	***	-6.123	.024	***	-8.504	.436	***	-7.829	.437	***
nodeocov																		
LC participant (ref: non-participant)	**.238**	**.038**	*******	**.267**	**.047**	*******	**.564**	**.084**	*******	**.382**	**.039**	*******	**.809**	**.060**	*******	**-.300**	**.083**	*******
nodeofactor																		
race (ref: Asian)																		
White				-.049	.058								.424	.063	***	.395	.063	***
Latinx				.013	.050								.257	.057	***	.228	.057	***
Black				-.033	.111								.486	.116	***	.479	.118	***
Other				.034	.079								.547	.085	***	.558	.086	***
nodeocov																		
first-generation (ref: non-first-generation)				-.030	.042								-.016	.044		-.038	.044	
female (ref: male)				-.030	.040								-.208	.043	***	-.207	.044	***
low-income (ref: not low-income)				-.106	.041	***							-.137	.042	**	-.125	.042	***
high school GPA				-.033	.094								.016	.101		.066	.103	
commuter (ref: on-campus)				-.124	.045	***							-.265	.046	***	-.250	.047	***
nodematch																		
same race							.688	.042	***				.768	.049	***	.788	.049	***
same first generation							.201	.041	***				.171	.043	***	.177	.043	***
same gender							.428	.042	***				.567	.046	***	.578	.046	***
same LC participation status							1.511	.076	***				1.811	.079	***	-.131	.080	
absdiff																		
high school GPA							-.374	.128	***				-.150	.133		-.005	.133	
edgecov																		
LC participant X same race							-.011	.086					-.271	.093	**	-.253	.091	***
LC participant X same first generation							.004	.089					-.262	.091	**	-.272	.087	***
LC participant X abs.diff high school GPA							.648	.209	***				.050	.221		-.228	.214	
LC participant X same gender							.311	.094	***				-.194	.099	*	-.265	.087	***
same high school										4.216	.050	***	3.920	.053	***	3.993	.055	***
same bio course																1.463	.048	***
same chem course																.675	.047	***
same freshman seminar																.378	.047	***
same LC section																2.423	.071	***
BIC	43308			43415			41220			39962			38151			36326		

Figure S2.20 Average Marginal Effects (AME) Corresponding to LC Size and Segmentation Effects in Figures S2.17 and S2.18

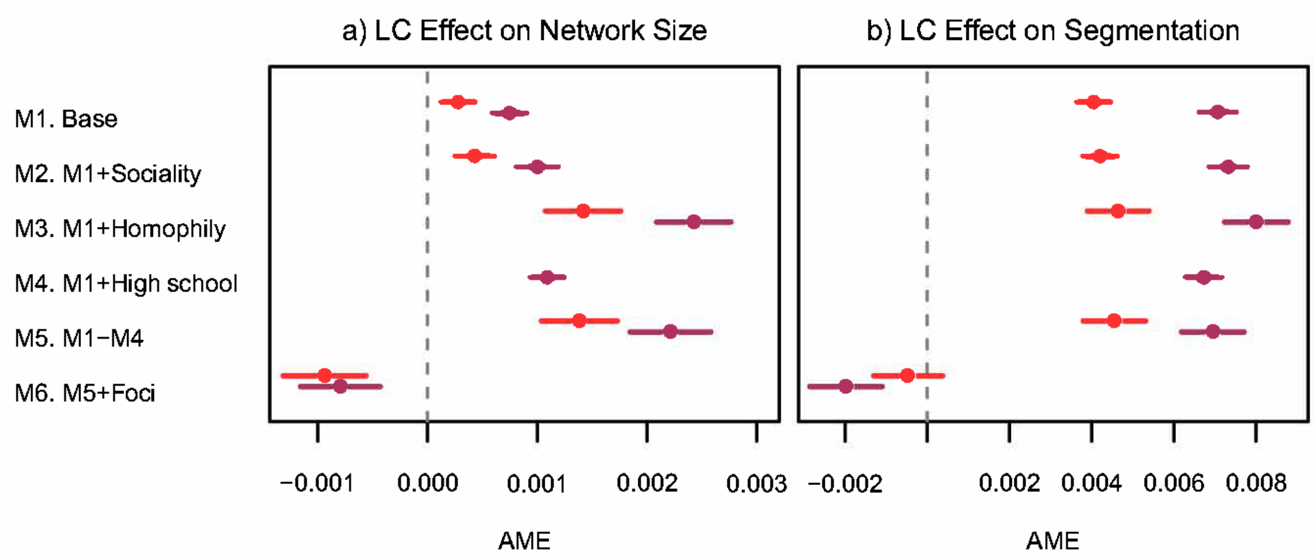


Figure S2.21. Mediation Test of LC Using Series of ERGMs

		Year 1						Year 2					
		Total AME	Partial AME	SE	Indirect AME	SE	Percent Mediated	Total AME	Partial AME	SE	Indirect AME	SE	Percent Mediated
Size													
M1a	Base	.00028	.00028	.00008				.00075	.00075	.00008			
M2a	Sociality	.00028	.00043	.00009	-.00015	.00010	-55.1	.00075	.00100	.00010	-.00025	.00010	-33.8
M3a	Homophily	.00028	.00142	.00017	-.00114	.00018	-405.8	.00075	.00243	.00017	-.00168	.00018	-224.9
M4a	Same HS	--	--	--	--	--	--	.00075	.00110	.00008	-.00035	.00011	-46.9
M5a	M1-M4	.00028	.00138	.00018	-.00110	.00018	-393.4	.00075	.00221	.00019	-.00147	.00019	-196.5
M6a	M5+Foci	.00028	-.00094	.00019	.00122	.00021	434.6	.00075	-.00079	.00019	.00154	.00020	206.3
Segmentation													
M1b	Base	.00407	.00407	.00021				.00707	.00707	.00023			
M2b	Sociality	.00407	.00422	.00021	-.00016	.00008	-3.8	.00707	.00733	.00024	-.00026	.00032	-3.7
M3b	Homophily	.00407	.00465	.00038	-.00058	.00028	-14.3	.00707	.00800	.00039	-.00094	.00045	-13.2
M4b	Same HS	--	--	--	--	--	--	.00707	.00672	.00022	.00034	.00032	4.9
M5b	M1-M4	.00407	.00455	.00038	-.00049	.00029	-12.0	.00707	.00695	.00039	.00012	.00045	1.7
M6b	M5+Foci	.00407	-.00047	.00043	.00453	.00054	111.5	.00707	-.00198	.00045	.00905	.00050	128.0

NOTES: Average Marginal Effect (AME).

Total AME is the average change in tie probability for a one-unit change in the predictor (LC status or LC dyad).

Partial AME is the direct effect of the LC predictor net of effects introduced to the model.

Indirect AME is the average change in tie probability that attributable to the LC indirectly through the confounding or mediating effect introduced.

The LC predictor is student LC status (participant vs. non-participant) in size models and LC -> LC dyad vs. a non-LC participant -> LC participant dyad in segmentation models.

Supplementary References

1. Fischer CS (2009) The 2004 GSS finding of shrunken social networks: An artifact?. *Am Sociol Rev* 74(4):657-669.
2. Paik A, Sanchagrin K (2013) Social isolation in America: An artifact. *Am Sociol Rev* 78(3):339-360.
3. McCabe JM (2016) *Connecting in College: How Friendship Networks Matter for Academic and Social Success* (University of Chicago Press, Chicago, IL).
4. Crossley N, Bellotti E, Edwards G, Everett MG, Koskinen J, Tranmer M (2015) *Social Network Analysis for Ego-Nets: Social Network Analysis for Actor-Centred Networks* (Sage, Thousand Oaks, CA).
5. Perry BL, Pescosolido BA, Borgatti SP (2018) *Egocentric Network Analysis: Foundations, Methods, and Models* (Cambridge University Press, Cambridge, UK).
6. Jacob R, Zhu P, Somers MA, Bloom H (2012) *A Practical Guide to Regression Discontinuity* (MDRC, New York, NY).
7. Calonico S, Cattaneo MD, Farrell MH (2020) Optimal bandwidth choice for robust bias-corrected inference in regression discontinuity designs. *Econom J* 23(2):192-210.
8. Xu D, Solanki S, McPartlan P, Sato B (2018) EASEing students into college: The impact of multidimensional support for underprepared students. *Educ Res* 47(7):435-450.
9. Imbens GW, Lemieux T (2008) Regression discontinuity designs: A guide to practice. *J Econom* 142(2):615-635.
10. Cattaneo MD, Jansson M, Ma X (2018) Manipulation testing based on density discontinuity. *Stata J* 18(1):234-261.
11. Duxbury SW (2021) The Problem of Scaling in Exponential Random Graph Models. *Sociol Methods Res* DOI:10.1177/0049124120986178:1-39.
12. Robins G, Pattison P, Kalish Y, Lusher D (2007) An introduction to exponential random graph (p*) models for social networks. *Soc Networks* 29(2):173-191.
13. Morris M, Handcock MS, Hunter DR (2008) Specification of exponential-family random graph models: terms and computational aspects. *J Stat Softw* 24(4):1548-7660.
14. Lewis K, Kaufman J, Gonzalez M, Wimmer A, Christakis N (2008) Tastes, ties, and time: A new social network dataset using Facebook.com. *Soc Networks* 30(4):330-342.
15. McPherson M, Smith-Lovin L, Cook JM (2001) Birds of a feather: Homophily in social networks. *Annu Rev Sociol* 27(1):415-444.
16. Feld SL (1981) The focused organization of social ties. *AJS* 86(5):1015-1035.
17. Feld SL (1982) Social structural determinants of similarity among associates. *Am Sociol Rev* 47(6):797-801.

CHAPTER 3

Curricular Differentiation and Informal Networks:

How Formal Grouping and Ranking Practices shape Friendships among College Students

ABSTRACT

This study draws upon complete friendship network data on two freshman biological sciences cohorts at a selective university in the United States to investigate *how* and *to what extent* allocating students to curricular groups and grading their performance in class shapes 1) processes of friend selection at the dyadic level and 2) friendship clustering at the network level. Through a set of stochastic actor-oriented models, results show that students tend to befriend peers from the same curricular group versus a different one (i.e., curricular group homophily) as well as befriend higher-performing peers (i.e., performance-based status). Follow-up analyses reveal that friendship clustering by curricular group placement is largely due to course co-enrollment (i.e., proximity), while academic performance-based clustering is primarily the result of students aligning their own performance to match the average performance of their friends (i.e., influence). We discuss implications of these findings for helping to promote learning in higher education.

Keywords: curricular differentiation, friendships, clustering, higher education, learning

INTRODUCTION

Since the pioneering work of Coleman's (1961) *The Adolescent Society*, sociologists of education have recognized the importance of studying the friendships that develop among students within schooling settings. Friends represent a source of peer social capital available to students—they provide access to information and help to cultivate norms and practices related to success in school (Coleman 1988; Hallinan 1982; Hasan and Bagde 2013; McCabe 2016). Friends also impact and reinforce one's identity and can support a sense of belonging in school (Nunn 2021).

Although sociologists have a longstanding interest in the relationships that form among students (Epstein and Karweit 1983), we know relatively less about how curricular practices shape the friendships that emerge. Research in this area has tended to focus on primary and secondary school (Hallinan and Sørensen 1985; Kubitschek and Hallinan 1998), with fewer studies at the postsecondary level. For example, recent research has once again drawn attention to the link between secondary school tracking systems and academically-based sorting within schools (Engzell and Raabe 2023). However, it remains unclear *how* and *to what extent* routine curricular practices at the organizational level may shape the friendships that develop among college students. Understanding friendship networks is important because if friends influence one another, then peer-to-peer connections may have implications for helping to promote learning in college (Felten and Lambert 2020).

In this study, I build upon the concept of *curricular differentiation* to analyze the link between formal curricular practices and the informal networks that arise among college students. Specifically, I refer to curricular differentiation as routine curricular practices that formally stratify students. In this way, curricular differentiation is organizational differentiation specific to the curricular domain (Sørensen 1970; Tyson and Roksa 2016). Two forms of curricular differentiation that are especially relevant within U.S. colleges and majors include: 1) allocating students to curricular groups (e.g., developmental, honors, etc.) and 2) grading student performance in class (Sørensen 1970). Through forms of differentiation such as these, students are placed into closer proximity with certain peers (Feld 1981; Frank, Muller, and Mueller 2013), as well as formally ranked relative to one another (Domina, Penner, and Penner 2017; Jeffrey 2020).

Forms of differentiation, such as those based on curricular practices, can shape friend selection through *homophily* or *status* mechanisms. Academic homophily would occur if students disproportionally developed friendships with peers from the same curricular group or with similar levels of academic performance (McPherson, Smith-Lovin, and Cook 2001). This could arise because students in the same curricular group share coursework (Feld 1981; Frank, Muller, and Mueller 2013) or because students choose to befriend similar peers. For example, having friends who perform similarly or engage in similar academic practices, such as study habits or in-class participation, can offer identity reinforcement regarding one's own performance. With respect to academic status, if curricular group membership or academic performance represent status markers, then there are reasons to suspect that students will have a heightened propensity to befriend higher-ranked peers (Jasso 2001; Kubitschek and Hallinan 1998). In turn, the strength of homophily effects relative to status effects will shape how integrated or segregated students are across curricular groups and performance levels in the network (i.e., level of clustering along these dimensions).

This analysis has the following three aims: 1) to examine how formal curricular practices shape processes of friend selection among college students, 2) to estimate the relative importance of academically-based effects on friend selection at the dyadic level compared to other well-known effects, and 3) to quantify how much academically-based friendship clustering is due to forms of curricular differentiation versus alternative mechanisms. I draw upon complete friendship network data on two freshman biological sciences cohorts at a large, selective, public university in the Western United States. Through a set of stochastic actor-oriented models (SAOMs), I first estimate the effects of curricular group placement and academic performance on friend selection, and then perform two follow-up analyses to gauge the magnitude of these effects.

As such, this study extends the existing literature in three important ways. First, compared to previous research in this area, the current study represents a clearer investigation of how curricular differentiation shapes friend selection with fewer concerns regarding unobserved heterogeneity among students on dimensions such as differential expectations or motivations. Specifically, not only does sorting into a STEM major at a selective college lead to a relatively more academically homogenous group of students compared to the K-12 context, but here I am also able to condition on pre-college factors that have been largely missing in prior work (i.e., high school GPA and SAT score). Second, past work on friend selection among college students has generally only focused on one aspect of differentiation (i.e., either grouping or performance) (Boda et al. 2020; Dokuka, Valeeva, and Yudkevich 2020; Smirnov and Thurner 2017; Van Duijn et al. 2003). Consequently, this is the first study of which I am aware, to test for homophily and status in terms of both curricular grouping and academic performance simultaneously. This is critical in the U.S. context, since grouping practices and performance are likely correlated with one another. Finally, I build upon prior studies (Engzell and Raabe 2023; Flashman 2012; Kubitschek and Hallinan 1998), by not only investigating the presence of specific effects but also estimating their size. Namely, while uncovering the processes and mechanisms behind friend selection among college students is important, from both a theoretical and practical standpoint it is also critical to be able to gauge the magnitude of these effects.

BACKGROUND

Curricular Differentiation within Colleges

Colleges are organizations structured to categorize, rank, sort, and select students (Domina, Penner, and Penner 2017; Jeffrey 2020; Mullen 2011; Stevens, Armstrong, and Arum 2008). As such, colleges enact many forms of organizational differentiation that are likely to impact the friendships that develop among students, such as through housing restrictions and the provision of co-curricular activities (Armstrong and Hamilton 2013; Espenshade and Radford 2009; Lee 2016; Stearns, Buchmann, and Bonneau 2009). In this study I focus on formal practices within the curricular domain—what I refer to as curricular differentiation. Two forms of curricular differentiation common at U.S. colleges include: 1) curricular grouping of students into separate and tiered levels (e.g., honors, remedial, etc.) and 2) the implicit ranking of students that occurs through grading (Sørensen 1970).

First, colleges determine how to assign students to classes for instructional purposes. Because students enter postsecondary schooling with varying levels of familiarity with college-level work (Jack 2019), many institutions in the U.S. adopt forms of curricular grouping akin to tracking in the K-12 context (Sørensen 1970; Tyson and Roksa 2016). For example, postsecondary institutions may utilize some type of remedial or developmental education to help address gaps in understanding for those entering college “academically underprepared” (Long and Boatman 2013). Similarly, colleges may attempt to provide broader and deeper learning opportunities for "academically talented" undergraduates through honors programs (Rinn and Plucker 2019). Despite mixed findings on the educational effectiveness of these programs (Sanabria, Penner, and Domina 2020), the intended goal of these classes and curricular groups is nonetheless to match instructional resources to meet the differential needs of entering students (Bettinger and Long 2008; Bowman and Culver 2018). However, by placing students in separate classes and creating distinct tiers, curricular grouping may unintentionally lead to hierarchical labels being assigned to students (Domina, Penner, and Penner 2016; Scott-Clayton and Rodriguez 2015).

Second, colleges are expected to allocate grades to students based upon their academic performance in the classroom. While this organizational task largely represents a routine, taken-for-granted practice (Meyer and Rowan 1977; Schneider and Hutt 2014), with the expansion of higher education in the U.S. over time, disparate college performance is becoming an important sorting mechanism linking educational and occupational stratification more broadly (Gerber and Cheung 2008). Indeed, within the structure of contest mobility apparent in the American educational system (Turner 1960), unequal academic performance among undergraduates can be viewed as a form of horizontal stratification that selects some students for competition at the next higher level (i.e., graduate school) (Posselt and Grodsky 2017; Xie, Fang, and Shauman 2015).

Curricular Differentiation and Friend Selection

Curricular differentiation within American higher education may thus shape the friendships that emerge through 1) *academic homophily* and 2) *academic status*.

Academic Homophily

Homophily refers to the commonly observed tendency for people with greater similarity on some characteristic to be connected (McPherson, Smith-Lovin, and Cook 2001). Patterns of homophily can arise due to 1) spatial proximity (i.e., foci effects (Feld 1981)) or from 2) choices to befriend individuals who share membership in a socially relevant category (Wimmer and Lewis 2010). First, students in the same curricular group are more likely to share classes than are students in different curricular groups. As a result, curricular group homophily may emerge from course overlap, since this increases opportunities to interact and the chance of forming a friendship (Feld 1981; Jeffrey et al. 2022; Kossinets and Watts 2009; Weber, Schwenzer, and Hillmert 2020). Second, schooling-based forms of distinction created through curricular groups (Domina, Penner, and Penner 2016) or differential academic performance could lead to academic homophily by serving as a basis of trust or commonality among students (Kossinets and Watts 2009; Wimmer and Lewis 2010). For instance, students may perceive co-enrollment in honors as a signal of shared commitment to following institutional norms of academic excellence or pursuing high-status occupations.

Evidence at the K-12 level shows that students from the same "ability group" and academic track are more likely to form friendships (Frank, Muller, and Mueller 2013; Hallinan and Sørensen 1985; Kubitschek and Hallinan 1998). Yet at the postsecondary level, findings are mixed. While some research shows that placement into the same academic program or learning community increases the likelihood of a friendship tie (Boda et al. 2020; Jeffrey et al. 2022; Van Duijn et al. 2003), other studies do not find a significant effect (Brouwer et al. 2018; Brouwer et al. 2022). Missing from prior work in higher education, however, has been an examination of the impact of curricular practices when the groups are *ranked* in some way. While any form of curricular grouping could lead to patterns of homophily, it is less clear what to expect in the presence of some form of tiered tracks. As I explain below, there are reasons to suspect that students may seek out similar peers but also higher-ranked peers.

In terms of academic performance, findings at the K-12 level largely confirm that greater performance similarity predicts friendship development (Flashman 2012; McFarland et al. 2014; Rambaran et al. 2017). But past work outside the United States at the college level has shown mixed results. While some evidence provides support for performance-based homophily (Brouwer et al. 2018; Smirnov and Thurner 2017), other studies show no significant effect (Brouwer et al. 2022; Dokuka et al. 2020). Importantly, much of the work that has failed to find homophily on performance has instead found evidence for status-based effects, implying that in certain settings the drive to befriend higher-performing peers may override the motivation to seek out similar peers.

Academic Status

Sociologists have conceptualized status as “inequality based on differences in honor, esteem, and respect” (Ridgeway 2014:2) and have argued that processes of competition and ranking among individuals can lead to patterns of hierarchy within networks (Jasso 2001; McFarland et al. 2014). Coleman (1961) argued that within schooling settings, competition for grades and other forms of scholastic achievement is a way in which students pursue and gain respect and recognition from those around them. As such, higher curricular group placement and higher academic performance could represent status markers. If so, then students may befriend higher-ranked peers due to the perceived rewards from affiliation (Epstein and Karweit 1983; Hallinan 1982) or the belief that association will produce a “halo” effect increasing one’s own status (Kubitschek and Hallinan 1998). However, it is also possible that students may not approach or attempt to befriend higher-ranked peers for fear of rejection or because they anticipate that their efforts will not be reciprocated (Gould 2002; Kubitschek and Hallinan 1998).

Despite reasons to suspect status-based effects due to curricular grouping, relatively few studies have explicitly examined how this might occur. Kubitschek and Hallinan (1998) tested for status effects based on high school track and found some supportive evidence, although this varied by school size. Specifically, they found that the likelihood of naming a higher-track student as a friend was greater in smaller schools compared to larger ones, presumably due to decreased barriers to interaction in the smallest schools. In other words, the likelihood of cross-track interaction increased as school size decreased and thus greater knowledge of relative track placement and the opportunity to form friendships. At the postsecondary level, a study done in the Netherlands largely failed to find status effects due to enrollment in a short (i.e., 2-year) versus regular (i.e., 4-year) program, though this division of students may not have entailed a formal ranking in the same way as in the U.S. case (Van Duijn et al. 2003).

Past research looking at possible status effects due to academic performance has also shown mixed results. Namely, work outside the United States, at both the K-12 and higher education levels, has provided evidence that academic performance acts as a status indicator, with higher-performing students selected more often as friends (An 2022; Brouwer et al. 2022; Dokuka et al. 2020). In contrast, studies on U.S. K-12 schooling have largely failed to find significant status effects based on GPA (Coleman 1961; Rambaran et al. 2017), although this varies by methodological approach (see Flashman 2012). Despite these largely null findings in U.S. K-12 settings, college differs from high school in fundamental ways likely to make GPA a more salient status marker. First, sorting into higher education selects students who performed relatively better in K-12 schooling compared to those who did not enter college. Second, as more students expect to attend graduate school, and undergraduate grades remain a critical component of graduate admissions (Stevens 2009), high performance is likely salient and recognized among students. For example, recent research shows that higher grades and levels of outspokenness predict greater recognition among peers in undergraduate biology classrooms (Grunspan et al. 2016).

In sum, U.S. colleges employ various forms of curricular differentiation with the intended goal to facilitate teaching and learning. As discussed, these routine practices may shape friend selection through two primary mechanisms: homophily and status. Specifically, curricular group placement and academic performance may not only physically separate and differentiate students (i.e., relevant for homophily effects) but also assign students an ordinal rank (i.e., relevant for status effects). Mixed findings in prior studies could be due to the fact that past work has approached the topic too narrowly, such as failing to examine both grouping and performance or not considering homophily and status jointly. The current analysis overcomes these issues, while also analyzing a much larger sample than often used (e.g., around 30–100 students in Brouwer et al. 2022; Dokuka et al. 2020; Van Duijn et al. 2003) which allows for more statistical power to estimate effects.

The Current Study - A Longitudinal Network Approach

Over the past few decades, there has been substantial growth in the use of network analysis to answer social scientific questions (Rivera, Soderstrom, and Uzzi 2010). In turn, advances to our understanding of network properties and tools with which to analyze them, have led to more sophisticated strategies to unpack the underlying mechanisms that give rise to the networks we observe (Robins et al. 2007). With scholars increasingly acknowledging the importance of connections for success in college (Felten and Lambert 2020), social network analysis provides a rigorous and systematic way to examine friendships in higher education from a relational sociological standpoint (Tierney and Kolluri 2020).

To examine how curricular differentiation shapes friendships among college students in the U.S. context, I draw upon a longitudinal network approach and estimate a set of stochastic actor-oriented models (SAOMs). One of the main purposes of SAOMs is to estimate multiple selection and influence mechanisms in one joint model (see Steglich, Snijders, and Pearson 2010). This is important here since, to reduce bias in estimates of selection on academic performance, we need to account for the possibility of friend influence effects on GPA. As I discuss next, past work has shown that patterns of homophily on academic performance can come about through a mix of selection and influence processes.

Friend Influence

Past sociological work has highlighted the importance of accounting for influence effects when we are estimating selection on a behavioral attribute (Steglich, Snijders, and Pearson 2010). For example, in the current setting, observed patterns of academic performance homophily could result from students selecting peers with similar levels of academic performance or from students influencing one another to perform similarly over time. Additionally, whereas the theory of status effects would lead us to expect higher performance to induce more friendship nominations, the reverse could also be true. Namely, receiving more friendship nominations could lead to higher performance. Previous research at the K-12 level, as well as postsecondary work outside the U.S., has produced mixed results. Specifically, when modeling the joint effects of selection and influence on academic performance homophily, past studies have found evidence of everything from selection but not influence (Flashman 2012; Smirnov and Thurner 2017), influence but not selection (Dokuka et al. 2020), both selection and influence (Rambaran et al. 2017), to neither selection nor influence (Brouwer et al. 2022). While I cannot claim to reconcile these mixed findings in this paper, I do build upon and extend this body of work in a meaningful way by controlling for pre-college factors as well as accounting for relevant sociodemographic covariates when modeling behavioral dynamics, which have been missing in most of the previous research.

The Case of Freshmen in Biological Sciences

In this study I analyze two freshman biological sciences cohorts at a selective college. Studying first-time entering freshmen allows for a clearer estimate of the impact of curricular practices on friend selection with fewer concerns regarding causal order (e.g., students may choose certain classes because of their friends). Where sociological theories of homophily predict that students will befriend peers from the same curricular group or similar performance level, theories of status and friend selection predict that students will have a greater propensity to befriend higher-ranked students. Especially in the case of freshman STEM students, the transition to higher education coupled with the competitive culture of entry-level, "weed-out" courses and gatekeeping practices (Seymour, Hunter, and Harper 2019) could compel students to seek out similar peers who can provide emotional support, but it may also drive them to find higher-performing peers who can aid with academic challenges (Smith 2015).

At the current research site and major, freshman students were allocated to one of three curricular groups: 1) *developmental*, 2) *regular*, or 3) *honors*. Developmental students are those who scored below 600 on their SAT math section. It is critical to note, however, the potentially unique aspects of the developmental group at the current research site. Specifically, developmental students were placed into a freshman learning community where they were block-registered into the same sections for required introductory courses during their first year. They were also provided with additional resources, primarily through means of a weekly seminar that entailed one-hour sessions designed to promote study skills, career advice, and help with navigating the new academic environment. As such, the developmental track was designed with the intended goal to give supplementary aid to students. Honors students were selected by faculty in the major and invited to participate based on their high school achievements (e.g., GPA, SAT/ACT, class rank, etc.). Students in honors were placed into the same introductory honors biology course, received priority enrollment in their other courses, and were provided opportunities to participate in undergraduate research. Regular students represent the largest of the three groups and include all students not in the developmental or honors programs.

As with most college majors, students within biological sciences are required to complete a set of major-specific courses in order to graduate. In turn, due to the sequential nature of many of the required courses, students normally take the same set of classes during their first year. Like tracking in the high school context (Lucas 1999), students are more likely to share course sections with others from the same curricular group in their major-required classes. The coupling of curricular group assignment and placement into different course sections was tighter in the second year compared to the first. For example, in cohort 1 non-developmental students were allowed to enroll in the developmental section if space remained, whereas this option was eliminated in cohort 2. Finally, throughout the first year, students also tend to complete several non-major-specific courses to fulfill general education requirements needed to graduate.

It is also important to note the temporal context of this study. While data on the first cohort was collected prior to the start of the COVID-19 pandemic (i.e., 2018-2019 academic year), the second cohort experienced the disruption of the virus during the study period (i.e., 2019-2020). The most notable change from the Fall term to the Spring term was the transition to remote instruction. Most students during the Spring term of data collection were living at home and taking classes online. However, even though the physical learning environment of the students was altered drastically by the pandemic, preliminary analyses revealed that the social networks of the students remained surprisingly resilient in the face of these challenges. As a result, I include data collected during the pandemic, but discuss potential differences when applicable.

DATA AND METHODS

Data

Table 3.1. Description of Biological Science Cohorts across Terms

Variable Term	Fall 2018	Fall 2019	Spring 2019	Spring 2020
Size (N)	859	1054	812	1010
Composition	Mean / Prop.	Mean / Prop.	Mean / Prop.	Mean / Prop.
Honors (yes=1)	.07	.02	.07	.02
Regular (yes=1)	.66	.72	.67	.73
Developmental (yes=1)	.27	.26	.25	.25
White (yes=1)	.11	.16	.11	.16
Black/African American (yes=1)	.07	.03	.06	.03
Hispanic/Latinx (yes=1)	.25	.27	.24	.26
Asian/Asian American (yes=1)	.57	.54	.59	.55
International student (yes=1)	.06	.04	.06	.05
First-generation college (yes=1)	.50	.51	.49	.50
Low-income student (yes=1)	.31	.39	.29	.38
Female (yes=1)	.68	.69	.68	.68
High school GPA	4.09	4.04	4.09	4.05
SAT score	1204	1200	1210	1206
Fall term GPA	2.96	2.84	--	--
Spring term GPA	--	--	3.05	3.31

Note: Proportions may not sum to 1 due to rounding.

Data come from two first-time entering biological sciences cohorts at a large, selective, R1 public university in the Western United States. As shown in Table 1, the current research site represents a diverse context in terms of race/ethnicity, gender, and socioeconomic background. Namely, most students are nonwhite, are female, and a large share are from low-income or first-generation-college backgrounds. In addition, students come from over four hundred high schools and around three-quarters of them live on campus. During the final weeks of Fall term and Spring term (2018-2019; 2019-2020), electronic surveys were sent to the entire freshman cohort to collect data on friendship ties along with additional background information. Response rates for each wave of data collection were generally high and ranged from 88% – 96%.[9] An important feature of the present study was the ability to link survey information with administrative records (i.e., transcripts) provided by the university.

Variables

Dependent

Friendships

As part of the survey, students were asked to list the first and last names of their friends within the biological sciences major.[10] These nominations were later matched with roster data provided by the university to create the complete (or sociocentric) friendship network for each major cohort (Perry, Pescosolido, and Borgatti 2018). It is important to note that friendship for purposes of this study is meaningfully directional and does not require that both students name each other (Wasserman and Faust 1994). For example, student *i* may nominate student *j* as a friend without student *j* also indicating that student *i* is a friend. Allowing for this directionality to the friendship relationship is essential for estimating status effects.

[9] This data was collected as part of an institutional initiative that surveyed students about their experiences in the major. Importantly, students were given assignment credit for completing the survey.
[10] Due to data collection modifications across cohorts, students could name up to eight friends in Cohort 1 and up to ten friends in Cohort 2.

Independent

Curricular Group Placement

Curricular group placement was coded to align with the ordinal nature of the categories, with higher numbers corresponding to a higher level. Namely, developmental was coded as (1), regular (2), and honors (3). There was some mobility from initial placements over the course of the first year. Specifically, around 11-12% of students changed from their initial placement from the fall to spring terms. In this analysis, I fix the curricular group measure based upon their fall term placement since this indicates their first assignment by the university and likely has enduring impacts on the friendship dynamics among students. Supplementary analyses using the spring term membership instead produced substantively similar results (see S3.1).

Academic Performance

I use the student's term-specific GPA as recorded by the university in administrative records to measure their academic performance. As such, this measure captures the academic performance of each student across all their courses in a given term (i.e., either fall or spring). Because the analyses utilize a stochastic actor-oriented modeling strategy, I recode the GPA measures into ordinal levels as required by the SAOM approach (Ripley et al. 2024). Academic performance is thus recoded into ten ordinal categories that correspond to roughly equal intervals[11] as follows: (1) 0-1.50; (2) 1.50-2.00; (3) 2.00-2.25; (4) 2.25-2.50; (5) 2.50-2.75; (6) 2.75-3.00; (7) 3.00-3.25; (8) 3.25-3.50; (9) 3.50-3.75; (10) 3.75-4.00.

Individual-level Covariates

[11] Due to sparseness at the bottom of the distribution, I draw upon larger intervals below 2.0.

Administrative data provided measures of gender (female=1; male=0), first-generation student status (yes=1; no=0), low-income status (yes=1; no=0), international student status (yes=1; no=0), high school GPA, and SAT score. Race/ethnicity is measured using a set of dummy variables (reference=White). Students' racial/ethnic background was provided through administrative sources which allowed for students to be classified into one of four racial/ethnic categories (i.e., White, Black/African American, Hispanic/Latinx, Asian/Asian American). Finally, I include a measure of same-major ties, which indicates the students' self-reported share of friendships that are within the major.[12]

Dyadic Covariates

I incorporate information on where the students attended high school as well as their housing situation during their first year in college. As a dyadic measure, these variables capture the likelihood of a tie for those who attended the same high school or who share the same living situation (i.e., same on-campus dorm or live off-campus).

Analytic Approach

Stochastic Actor-Oriented Model (SAOM)

I draw upon a set of stochastic actor-oriented models (SAOMs) to account for the co-evolution of friendship ties and academic performance simultaneously. SAOMs accomplish this by modeling network and behavior changes as Markov processes where the network state at time t depends only on the network state at time t-1. Importantly, while the data is collected at discrete time points, the SAOM models the unobserved dynamic feedback interdependencies between

[12] For Cohort 1, this variable was measured categorically as follows: 1=0-20%; 2=21-40%; 3=41-60%; 4=61-80%; 5=81-100%. For Cohort 2, this variable was measured using a 7-point slider scale with the following indicators: 1="none"; 4="half"; and 7="all".

selection and influence processes taking place continuously between observations. Namely, SAOM estimation is done through an agent-based simulation algorithm where the model conditions on the first observation and randomly select actors to do one of the following: 1) change a tie (i.e., add or drop one outgoing tie), 2) change behavior (i.e., increase or decrease by one unit), or 3) make no change; these potential changes are known as "micro steps" (Steglich, Snijders, and Pearson 2010: 348). Both the frequency of opportunities to make changes, as well as the probability of taking these micro steps, are estimated using subcomponents of the model known as the *rate functions* and *objective functions*, respectively. The rate functions thus determine the wait time until an actor gets the opportunity to make a network (i.e., network rate function) or behavior (i.e., behavior rate function) change, while the objective functions determine the changes an actor makes to either the network (i.e., network objective function) or behavior (i.e., behavior objective function) (Steglich, Snijders, and Pearson 2010).

Measuring Homophily Effects

I analyze homophily effects due to curricular group membership through a *sameX* curricular group term, where the *sameX* effect indicates if two students have the same attribute value (i.e., 1=yes; 0=no). [13] This similarity or difference between students is then used to predict whether a friendship exists in the dyad. A significant and positive coefficient would be evidence of homophily based on common curricular group membership.

Because academic performance is an ordinal measure with ten levels, I examine performance-based homophily effects through a *simX* academic performance term. In RSiena, similarity is measured by taking the difference between ego and alters' academic performance

[13] Words in italics refer to the effect names as used in the *RSiena* package in R.

and standardizing it by the range of possible values. As such, the similarity measure equals 1 if two students have the same value and 0 if they are maximally dissimilar (i.e., one has the highest value and the other the lowest possible value) (Ripley et al. 2024). A significant and positive coefficient on the *simX* academic performance term would be evidence of performance-based homophily (i.e., similar students are more likely to be friends).

As discussed, patterns of curricular group homophily may arise in part due to shared coursework. To assess this possibility, I include a dyadic measure of course co-enrollment (i.e., specified through a *coDyadCovar* effect in RSiena) that estimates how each additional class shared between student *i* and student *j* in the spring term contributes to the likelihood of a friendship, conditional on the rest of the network.

Measuring Status Effects

From a network perspective, sociometric popularity is a common way to examine the effect of status among individuals (Brouwer et al. 2022; Martin and Murphy 2020; Snijders and Lomi 2019; Wasserman and Faust 1994). Specifically, popularity centers on the receivers' characteristics as a key driver of tie formation (Sauder, Lynn, and Podolny 2012). Thus, I include an *altX* curricular group term to examine if higher-ranked students are more likely to be named as friends. Likewise, to test for status effects based on GPA, I utilize an *altX* academic performance term, which captures the extent to which students with higher GPAs receive more nominations.

Alternative Mechanisms shaping Friendships

First, to estimate homophily and status effects based on curricular group placement and academic performance, we must adjust for the possibility that these characteristics are related to how many friends students themselves nominate. Thus, to account for differences in outgoing ties I include an *egoX* term for both curricular group membership and academic performance, which capture if higher placed or higher performing students, respectively, name more friends.

Second, since academic and sociodemographic characteristics may be correlated with one another, selection on one could give the appearance of selection on the other (Stearns, Buchmann, and Bonneau 2009; Weber, Schwenzer, and Hillmert 2020). As a result, I account for friendship based on salient background characteristics. For instance, past research has highlighted the tendency to observe same-race as well as same-gender friendships in college (Brouwer et. al 2022; Dokuka et al. 2020; Mayer and Puller 2008). I include *sameX* terms to account for homophily on categorical characteristics. I also account for main covariate effects (i.e., incoming/outgoing ties) using *egoX/altX* terms for dichotomous (and dummy) variables.

Third, to account for the fact that some students attended the same high school, and thus might have known each other prior to college, I include a *sameX* high school term. To adjust for unobserved characteristics that could be correlated with differential academic preparation across students (e.g., expectations, motivations, etc.), I include *simX* high school GPA and *simX* SAT score terms. To control for the known importance of residential life in shaping friendships in college (Armstrong and Hamilton 2013; Stearns, Buchmann, and Bonneau 2009), I include a *sameX* housing term. Finally, I include an *egoX* same-major ties term to control for the share of friends within the biological sciences.

Finally, to account for endogenous network mechanisms that may shape friendship ties, I include several prominent structural factors that have been identified in the social network literature (Rivera, Soderstrom, and Uzzi 2010). These effects include the following: *density* (out-degree), *in-degree popularity (sqrt)*, *out-degree activity (sqrt)*, *in-degree activity (sqrt)*, *reciprocity*, *transitive triplets*, and *transitive reciprocated triplets*. In order, these effects capture the general tendency to form friendships, tendency for actors with high in-degrees to attract extra incoming ties, the tendency for actors with high out-degrees to send out extra outgoing ties, the tendency for actors with high in-degrees to send out extra outgoing ties, the tendency to reciprocate incoming ties, the tendency to become friends with the friends of one's friends, and to account for differential reciprocity effects for open and closed groups, respectively.

Measuring Friend Influence

In terms of the behavioral part of the model, the *linear* and *quadratic shape* parameters control for the distributional features of the outcome. To account for friend influence, I include an *average similarity* academic performance term to examine whether college students adapt their academic performance to become (or stay) similar to the average performance of their friends. This term will assess whether there is an influence effect of friend academic performance on one's own performance. I also include *in-degree* and *out-degree* effects to test whether receiving more incoming ties or sending more outgoing ties, respectively, impacts performance. To estimate the effect of friends on academic performance we need to control for important individual covariates that may also predict GPA (Flashman 2012). In this analysis I include several covariates that may impact grades in college—namely, race/ethnicity, gender, first-generation student status, low-income status, international student status, high school GPA, and SAT score. I also account for the student's curricular group placement.

Missing Data and Attrition

To properly estimate the SAOMs, two additional steps were required to prepare the dataset for analysis. First, around 4-5% of students who began in the biological sciences major in the fall term were no longer in the major by spring term. As such, students who left the major were coded as *structural zeros* for the spring term network. Conceptually, this specifies in RSiena that no incoming or outgoing ties may be present, and thus are considered structurally determined values. Second, while the response rates to each questionnaire survey were very high, there were still some non-respondents. Those who did not complete the survey were treated as missing by coding their outgoing ties as NA.[14] Missing information regarding covariates (less than 3%) was treated by SIENA's default imputation procedures for missing values (see Ripley et al. 2024 for details). Finally, the SAOM was estimated using centered covariates, except for the high school and housing variables.

Modeling Strategy

I begin by estimating a separate SAOM for each first-year student cohort. Each SAOM includes the main academic homophily and academic status effects of interest as well as all covariates and endogenous network effects. The estimated effects can be interpreted as the expected contribution to the log-odds of a student creating (or maintaining) a friendship in a given dyad conditional on all other effects in the model. SAOM parameter estimates for this study are calculated using the RSiena package, version 1.4.1 (Ripley et al. 2024).

To gauge the magnitude of these effects, I perform two follow-up analyses that implement different kinds of "knockout experiments" to examine how friend selection and friendship clustering, respectively, would change if we removed certain effects from the model while holding all other factors constant (see Huang and Butts 2023). First, while SAOM parameter estimates may show a significant effect on friendship, comparing different effects in a

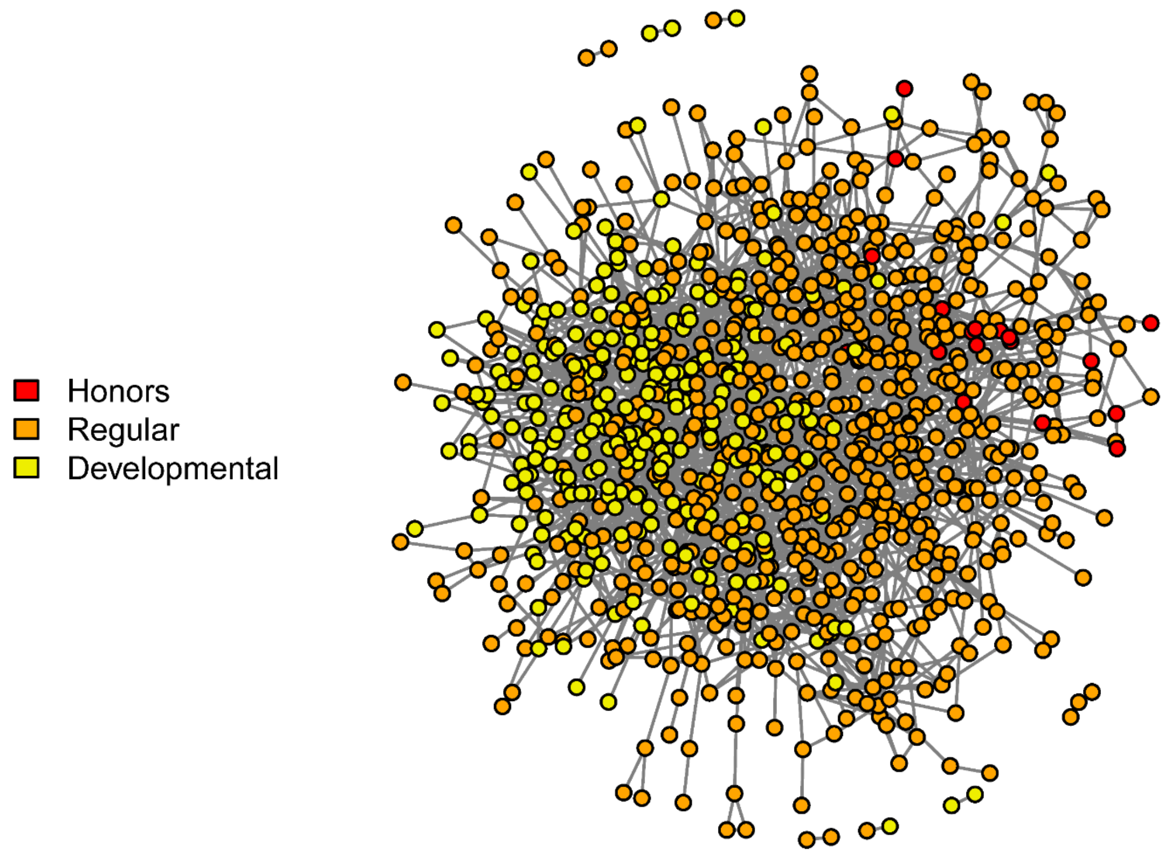
[14] Note that even students who did not take the survey could still be nominated as a friend, and as a result could receive incoming ties.

given model, as well as the same effects across cohorts, can be difficult. For instance, the main analysis may show that being from the same high school significantly increases the likelihood of a friendship. However, if few students attended high school together, then this mechanism would not be expected to explain much of the network that emerged. Thus, to understand the impact each effect has on the network, we need an estimate of the main effect that is adjusted for opportunities for that effect to shape friendship. To accomplish this, I rely upon the *sienaRI* function within RSiena to produce an estimate of the relative importance of effects on the network (Indlekofer and Brandes 2013). This step will allow me to compare, for example, the relative strength of academic-related effects to sociodemographic effects on friendship.

Second, while the SAOM and *sienaRI* analyses focus on the mechanisms behind friendship development, we are also interested in how these processes may contribute to friendship clustering in the network. Namely, from both a theoretical and policy standpoint, it is important to know how much these underlying mechanisms may ultimately matter for patterns of academic clustering that arise in the network. As such, I draw upon the *MEMS* function within the netmediate package in R to produce an estimate of the micro effect on the macro structure (MEMS) as discussed in Duxbury (2024). For the purposes of this paper, this final analysis will look at the overall level of clustering in the network by curricular group placement and academic performance as measured by the Moran's I statistic. Because my main estimates utilize longitudinal network models that account for selection and influence, this step will allow me to investigate how much of the observed academically-based clustering in the network (i.e., network autocorrelation on these dimensions) is due to selection versus influence mechanisms. For example, the MEMS will allow me to quantify how much of the observed clustering in the network by academic performance among students is due to students selecting similar-performing peers versus being influenced to perform similarly over time (Steglich, Snijders, and Pearson 2010).

RESULTS

Figure 3.1. Freshman Biological Sciences Friendship Network Shaded by Curricular Group Placement: Fall 2019



Visualizing the friendship network through sociograms is one tool for understanding basic patterns in a network. For example, from the Fall 2019 friendship network shown in Figure 3.1, we observe stark clustering of friendships along curricular group lines where nodes of the same curricular group (color) are found closer together. Importantly, however, we see that even though the network is segmented by curricular group it is still essentially one component. In other words, it is not split into three distinct parts, which is what we would observe if friendships were solely among students in the same curricular group. This means that most students can reach each other either directly or indirectly (e.g., through friends of their friends). Similar patterns can be observed across terms and cohorts (see Appendices D-F).[15] These patterns thus provide visual evidence of curricular group homophily in the friendship network.

Table 3.2. Descriptive Statistics of Friendship Network

Measure	Fall 2018	Fall 2019	Spring 2019	Spring 2020
Academic homophily effects				
Curricular group (*Moran's I*)	.587***	.683***	.494***	.624***
Academic performance (*Moran's I*)	.231***	.357***	.137***	.116***
Academic status effects				
Curricular group (*Pearson's r* with indegree)	.012	-.134***	-.030	-.102**
Academic performance (*Pearson's r* with indegree)	.036	.060†	.173***	.135***

†p<.10; *p<.05; **p<.01;***p<.001

To quantify these and other network patterns, I draw upon summary measures of homophily and status by forms of curricular differentiation. First, to quantify clustering by curricular group membership and GPA, I calculated the *Moran's I* network autocorrelation coefficient, which ranges from -1 to 1 and measures the degree to which friends display similarity in curricular group level and academic performance. As shown in Table 3.2, observed values for Moran's I range from .49 to .68 for curricular group and from .12 to .36 for GPA, indicating very strong to moderately strong levels of similarity in terms of curricular group membership and academic performance, respectively. Second, to get an overall sense of potential status effects due to curricular differentiation, I estimate *Pearson's r* (i.e., correlation coefficient) between measures of formal rank and network indegree. A positive association indicates that higher curricular group placement or performance level receives more incoming ties. Pearson's r

[15] The 2019-2020 networks exhibit heightened levels of clustering likely due to the programmatic changes across years that led to a stronger separation of students in their Fall term classes.

results show a negative correlation between curricular group membership and incoming ties, and a positive correlation between academic performance and friendship nominations, on average. While these statistics describe overall patterns on each dimension independently, I turn next to the stochastic actor-oriented models to test for homophily and status effects on these dimensions jointly.

Table 3.3. Parameter Estimates from SAOMs Measuring the Co-evolution of Friendship Ties and Academic Performance among Freshmen in Biological Sciences: Fall to Spring

Cohort	2018 – 2019			2019 – 2020		
Effects	Coefficient		SE	Coefficient		SE
Selection: Friendship Ties						
Rate function	4.015	***	(.218)	5.454	***	(.297)
Academic homophily effects						
Shared coursework	.747	***	(.051)	.630	***	(.033)
Curricular group (*same*)	.287	**	(.096)	.341	***	(.082)
Academic performance (*similarity*)	.444		(.465)	.590	†	(.311)
Academic status effects						
Curricular group (*alter*)	-.304	**	(.098)	-.123		(.094)
Academic performance (*alter*)	.088	**	(.032)	.051	*	(.025)
Covariates: Network Outcome						
Outdegree (density)	-7.201	***	(.363)	-6.059	***	(.232)
In-degree popularity (sqrt)	-.139		(.126)	.048		(.061)
Out-degree activity (sqrt)	1.081	***	(.113)	.725	***	(.075)
In-degree activity (sqrt)	-1.077	***	(.198)	-1.015	***	(.130)
Reciprocity	3.983	***	(.252)	4.270	***	(.183)
Transitive triplets	1.912	***	(.129)	1.157	***	(.068)
Transitive reciprocated triplets	-2.129	***	(.226)	-.983	***	(.099)
Curricular group (*ego*)	-.044		(.109)	-.073		(.096)
Academic performance (*ego*)	-.051		(.036)	.007		(.024)
Race/ethnicity (*same*)	.397	***	(.089)	.504	***	(.060)
Black/African American (*ego*) 1=yes	.235		(.247)	.070		(.215)
Black/African American (*alter*) 1=yes	.109		(.221)	-.017		(.192)
Hispanic/Latinx (*ego*) 1=yes	.089		(.193)	.140		(.110)
Hispanic/Latinx (*alter*) 1=yes	.230		(.174)	-.219	*	(.111)
Asian/Asian American (*ego*) 1=yes	-.023		(.172)	.040		(.089)
Asian/Asian American (*alter*) 1=yes	.034		(.156)	-.216	*	(.087)
Gender (*same*)	.311	***	(.094)	.309	***	(.063)

Gender (*ego*) 1=female	.461	***	(.111)	.102		(.071)
Gender (*alter*) 1=female	-.357	***	(.093)	-.186	**	(.067)
First-generation student status (*same*)	.088		(.083)	-.007		(.055)
First-generation student status (*ego*) 1=yes	-.275	**	(.105)	-.084		(.072)
First-generation student status (*alter*) 1=yes	-.097		(.094)	-.060		(.065)
Low-income status (*same*)	.053		(.098)	.019		(.060)
Low-income status (*ego*) 1=yes	-.068		(.123)	-.073		(.073)
Low-income status (*alter*) 1=yes	-.095		(.107)	-.116	†	(.065)
International student status (*same*)	.411	*	(.164)	.079		(.103)
High school (*same*)	.698	***	(.177)	.657	***	(.114)
Housing (*same*)	.619	***	(.080)	.343	***	(.056)
High school GPA (*similarity*)	-.613	†	(.324)	.356		(.517)
SAT score (*similarity*)	.520		(.358)	.438	†	(.247)
Same-major ties (*ego*)	.170	***	(.041)	.030		(.023)
Influence: Academic Performance						
Rate function	11.552	***	(1.038)	11.275	***	(1.042)
Linear shape	-.012		(.037)	.308	***	(.042)
Quadratic shape	-.016	**	(.006)	.005		(.007)
Average similarity	2.406	**	(.752)	1.867	**	(.697)
In-degree	.042	*	(.019)	.012		(.015)
Covariates: Performance Outcome						
Out-degree	.011		(.015)	.022		(.017)
Curricular group placement	-.018		(.040)	-.116	†	(.061)
Race/ethnicity (reference=White)						
Black/African American (1=yes)	-.042		(.078)	-.003		(.095)
Hispanic/Latinx (1=yes)	-.030		(.060)	.032		(.056)
Asian/Asian American (1=yes)	.027		(.056)	.108	*	(.049)
Gender (1=female)	.063	†	(.036)	.094	*	(.038)
First-generation student status (1=yes)	-.015		(.039)	-.130	***	(.039)
Low-income status (1=yes)	.014		(.037)	-.015		(.037)
International student status (1=yes)	.083		(.073)	-.130		(.080)
High school GPA	.150	†	(.085)	-.003		(.069)
SAT score	.000		(.000)	-.000		(.000)
Maximum Convergence Ratio		.159			.167	

† p<.10; *p<.05; **p<.01; ***p<.001

Note: Standard errors in parentheses.

Table 3.3 shows the parameter estimates from the set of SAOMs measuring the co-evolution of friendship ties and academic performance. First, in terms of homophily, we find a positive and significant effect of shared coursework on friendships. Namely, across cohorts, each additional class shared among students significantly increases the likelihood of a friendship developing ($p<.001$). In addition, the positive and significant curricular group (*same*) effect indicates that students are more likely to befriend a peer from the same curricular group versus a different one, net of course overlap. In contrast, when we look at selection by academic performance, we find only a marginally significant effect of academic performance (*similarity*) in the second cohort ($p<.10$). Finally, when testing for friend influence, we find strong and consistent effects across both cohorts. In particular, the positive and significant *average similarity* effect indicates that students tend to become (or stay) the same as their friends with respect to performance ($p<.01$).

In terms of status effects, the SAOM results show that higher-performing students are more likely to receive a friendship nomination compared to lower-performing students ($p<.05$). In contrast, the negative curricular group (*alter*) effect indicates that students in a lower-tiered group are more likely to receive friendship nominations. However, it is important to note that this negative association is net of homophily and only reaches significance in the first cohort ($p<.01$). When looking at potential network effects on performance, we find a positive and significant *in-degree* effect in the first year ($p<.05$) where students who received more friendship nominations tended to perform better over time.

Next, looking at SAOM estimates from background characteristics and other controls, we see several significant effects. As has been highlighted in prior research, we find that students tend to befriend others similar to themselves in terms of race/ethnicity and gender ($p<.001$). There are also ego and alter effects for gender, where female students tend to send out more friendship nominations while male students tend to receive more nominations ($p<.01$).

Additionally, as expected, Table 3.3 shows that students who attended the same high school, or who share the same living situation, are more likely to develop a friendship ($p<.001$).

Finally, in terms of endogenous network mechanisms, we also find several significant effects. For example, the negative outdegree (density) term ($p<.001$) indicates that ties are unlikely in the network unless other factors are present to increase the tendency to become friends. The positive reciprocity effect ($p<.001$) shows that students are more likely to reciprocate incoming friendship ties. Likewise, the positive transitive triplets term ($p<.001$) indicates that ties that create more transitive triads have greater odds of forming than not.

Figure 3.2. Relative Importance of Effects on Friend Selection:
Freshmen in Biological Sciences (Fall – Spring)

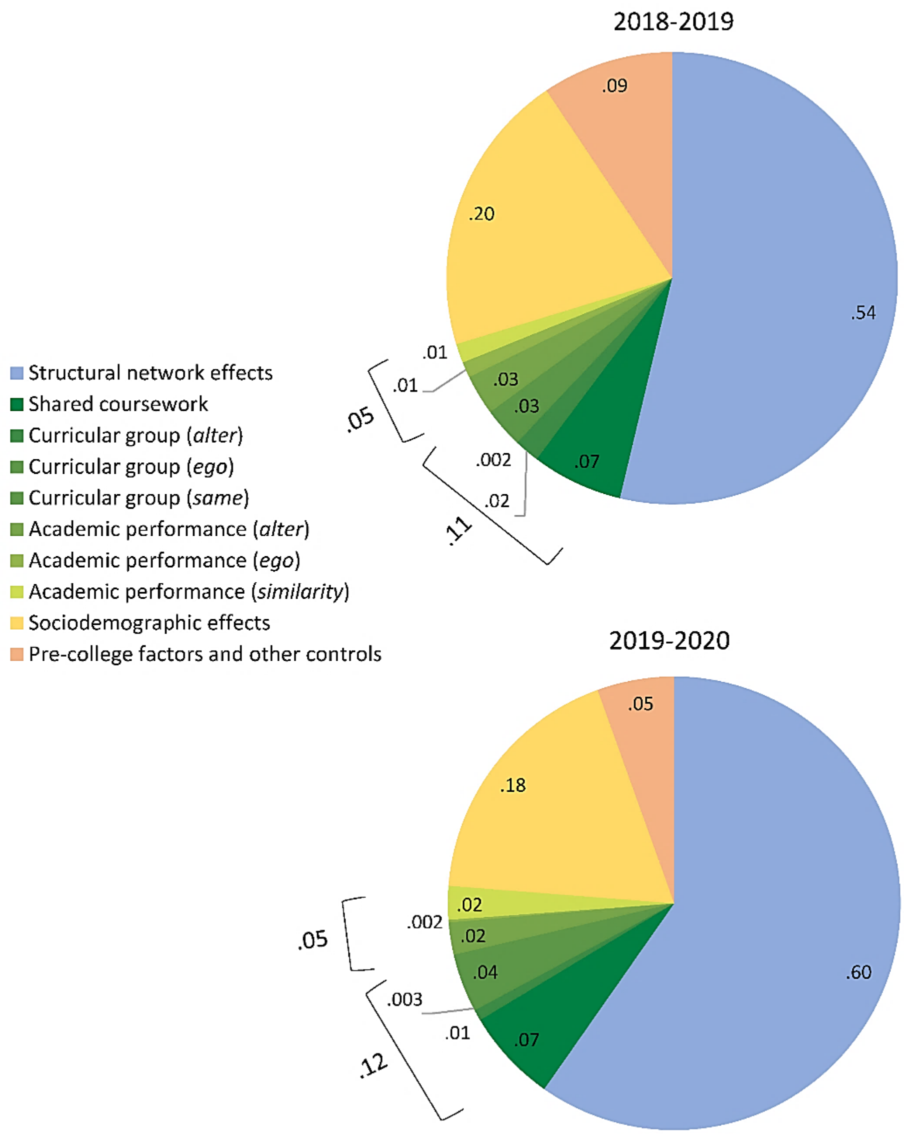


Figure 3.2 shows the relative importance of effects, and groups of effects, on the friendship network as estimated from the SAOMs for each cohort (see Appendix G for full results). As expected, we see that structural network effects have a strong impact on friendships (est. 54–60% of the relative share of importance) and thus act as important controls for our main effects of interest. In turn, there are several key takeaways from this part of the analysis. First, we can see that the relative importance of effects related to curricular group placement are about twice as strong combined as the effects due to academic performance (11-12% vs. 5%). Second,

comparing each academic-related effect, we find that shared coursework has the largest relative contribution to the friendships that emerge (est. 7% of the relative share of importance). Third, Figure 3.2 highlights that the combined total importance of all academically-based selection effects, are roughly the same as all the sociodemographic effects (e.g., race/ethnicity, gender, first-generation student status, etc.) (16-17% vs. 18-20%). Finally, except for a few effects such as housing, it is notable how consistent the measures of relative importance are across the two cohorts.

Table 3.4. Results from the Micro Effect on Macro Structure (MEMS) Estimation using Morans I by Curricular Group Placement and Academic Performance as the Outcomes

	Curricular Group Placement				Academic Performance			
Joint Parameters	MEMS 2018	% Change	MEMS 2019	% Change	MEMS 2018	% Change	MEMS 2019	% Change
Academic mechanisms								
Shared coursework	.135**	18.58	.070*	9.76	.008	1.16	-.001	-2.28
	(.048)		(.034)		(.049)		(.037)	
All curricular group-based selection effects[a]	.174***	24.05	.134**	18.90	.002	-1.23	.004	.24
	(.050)		(.041)		(.050)		(.037)	
All performance-based selection effects[b]	-.003	-.69	-.022	-3.28	.015	2.96	.023	9.30
	(.048)		(.035)		(.056)		(.042)	
All performance-based influence effects[c]	-.002	-.44	-.004	-.60	.100*	39.88	.045	20.86
	(.048)		(.033)		(.056)		(.042)	
All performance-based selection and influence effects[d]	-.001	-.31	-.024	-3.48	.126*	50.06	.081*	39.74
	(.045)		(.037)		(.058)		(.045)	
All academic-based selection and influence effects[e]	.175***	24.20	.125**	17.45	.135*	53.67	.097*	47.86
	(.054)		(.044)		(.064)		(.048)	
Sociodemographic mechanisms								
All sociodemographic selection effects[f]	-.022	-3.30	-.016	-2.41	-.009	-5.74	-.006	-4.71
	(.058)		(.039)		(.055)		(.039)	
Structural network mechanisms								
All endogenous network selection effects[g]	.083†	11.37	.065†	9.07	.022	7.13	.003	-.02
	(.064)		(.050)		(.056)		(.041)	
Pre-college factors and other mechanisms								
All pre-college and other selection effects[h]	-.015	-2.33	-.009	-1.45	-.003	-4.06	-.005	-4.24
	(.053)		(.035)		(.052)		(.037)	

†p<.10; *p<.05; **p<.01;***p<.001

Notes: Standard deviations in parentheses.

Each MEMS was estimated using 1,000 simulations.

a includes: shared coursework + *same/ego/alter* curricular group effects.

b includes: *similarity/ego/alter* academic performance effects.

c includes: *average similarity/in-degree/out-degree* effects.

d includes: all effects from b & c.

e includes: all effects from a – c.

f includes: *same/ego/alter* race/ethnicity, *same/ego/alter* first-generation student status, *same/ego/alter* low-income status, *same/ego/alter* gender, *same* international student status effects.

g includes: *reciprocity, transitive triplets, transitive recipr. triplets, in-degree popularity (sqrt), in-degree activity (sqrt), out-degree activity (sqrt)*.

h includes: *same* high school, *same* housing, *similarity* SAT score, *similarity* high school GPA, *ego* same-major ties.

Whereas Figure 3.2 focuses on mechanisms behind friend selection at the dyadic level, Table 3.4 shifts to how much the modeled effects explain network-level friendship clustering by curricular group placement and academic performance. From the micro effect on macro structure (MEMS) analysis, we gain several important insights. First, looking at how much academic mechanisms contribute to friendship clustering by curricular group placement, the effect of shared coursework stands out. Specifically, I estimate that roughly 10-20% of the clustering by curricular groups can be explained by shared coursework ($p<.05$). In turn, while the SAOMs showed some evidence of a tendency to befriend lower-tiered students, net of homophily, the MEMS indicate that the overall joint impact of all curricular group-based selection effects leads to clustering in the network (.174 and .134; $p<.001$). In other words, the tendencies toward curricular group homophily far outweigh the tendencies to cross curricular group boundaries and thus lead to more segregation than integration (see S3.2). Second, investigating potential drivers of clustering by academic performance, the MEMS estimates highlight that this is dominated by performance-based influence effects. For example, in Cohort 1, our MEMS indicate that about 40% of the Moran's I by academic performance is explained by all performance-based influence effects ($p<.05$). Third, Table 4 shows that both dimensions of academic clustering are driven by their own set of micro processes, where performance-based selection and influence effects are not related to curricular-group clustering, and grouping-based selection effects are not related to performance clustering. Fourth, when testing the potential impact of alternative mechanisms on friendship clustering, we see that only the endogenous network effects significantly contribute. This is somewhat surprising given that we expect several of the other covariates to correlate with our academic measures and thus highlights the importance of running these counterfactual simulations.

DISCUSSION

While sociologists of education have a longstanding interest in the friendships that form among students within schooling settings, we know relatively little about the link between routine curricular practices and the informal networks that emerge. In this study, I utilize data on two freshman biological sciences cohorts at a large, selective college and a set of stochastic actor-oriented models (SAOMs), as well as two follow-up analyses, to: 1) examine how formal curricular practices shape processes of friend selection among college students, 2) estimate the relative importance of academically-based effects on friend selection at the dyadic level compared to other well-known effects, and 3) quantify how much academically-based clustering of friendships in the network is due to forms of curricular differentiation versus alternative mechanisms.

With respect to the first aim, results of the SAOMs provide strong evidence of curricular group homophily. Namely, we see that students tend to befriend peers from the same curricular group versus a different one, and that each additional class shared among students increases the likelihood of a friendship developing. We also find evidence of academic performance-based status effects where students exhibit a heightened tendency to befriend higher-performing peers. While not consistent across cohorts, our results provide some support for academic performance homophily and reveal a tendency among students to befriend peers in a lower-tiered curricular group, net of curricular group homophily. Importantly, the SAOMs also show evidence that students tend to align their own academic performance to the average performance of their friends.

In terms of the second aim, the estimates of relative importance highlight the magnitude of different effects related to curricular differentiation for processes of friend selection among college students. For example, among academic effects, we see that the combined relative importance of effects related to curricular grouping are about twice as large as those for academic performance. Thus, these findings help to quantify the strength of potentially more direct and salient aspects of curricular differentiation on friendships versus arguably more indirect and less visible forms of differentiation (Van Duijn et al. 2003). Notably, based on the combined standardized effects as shown in Figure 3.2, we find that all the academic-related effects have roughly the same relative importance on the friendships that develop as all the common sociodemographic effects discussed in much of the prior sociological research (Armstrong and Hamilton 2013; Espenshade and Radford 2009; Lee 2016; Stearns, Buchmann, and Bonneau 2009). However, it is important to point out that these estimates of relative importance are taken from among students within the same major. This is likely because these forms of grouping and ranking practices are relatively more salient and known among this subset of students. These results should thus be viewed as complementary to other work that focuses on friendships of individual students within colleges (Johnson 2022; McCabe 2016) as well as friendships across majors within a given institution (Mayer and Puller 2008; Wimmer and Lewis 2010).

Finally, regarding the third aim of the study, results of the MEMS shift our focus to drivers of academically-based clustering of friendships at the network level. The MEMS show that patterns of clustering (as measured by Moran's I) by curricular group placement are largely driven by shared coursework. Thus, like research on "local positions" in high schools (Frank, Muller, and Mueller 2013), we see from the MEMS that curricular group homophily among college students is primarily the result of proximity and opportunities to interact as structured by course co-enrollment (i.e., foci)(Feld 1981; Jeffrey et al. 2022). By contrast, the MEMS reveal that friendship clustering by academic performance among college students is dominated by friend influence effects. Compared to recent research (Engzell and Raabe 2023), findings here indicate that patterns of friendship sorting by performance measures can emerge due to influence effects rather than selection and that performance clustering does not result from similarity on other metrics such as standardized test scores (see S3.3). As a result, these findings highlight that proximity mechanisms behind clustering along curricular group lines are separate from the influence mechanisms driving observed clustering by academic performance.

Conceptually, results shown here add to our understanding of the link between forms of curricular differentiation and the informal networks that develop. Specifically, I show that over time, student friendship networks tend to become highly segregated by curricular group placement and moderately clustered by academic performance (see Table 3.2). Importantly, I shed light on some of the processes and mechanisms behind these patterns. First, I demonstrate not only the significance of shared coursework in predicting which friendships develop, but I also reveal that allocating students to distinct classes is *the main driver* of clustering along curricular group lines at the macro (or network) level. In other words, in the absence of this routine curricular practice, we would not expect to see the levels of segregation along these lines that we do. Second, it seems that academic performance-based status effects related to ranking among students, while significant at the dyadic level, has a largely null effect on performance-based clustering in the network. For example, while the academic performance (*alter*) effect is significant across cohorts (see Table 3.3), the joint impact of homophily and status seem to largely offset one another (see S3.4).

Overall, like research on tracking at the K-12 level, we can think about the potential policy implications of these findings in terms of both learning and inequality (Gamoran 1992; Hallinan

1994). With respect to learning, akin to tracking at the K-12 level (Frank, Muller, and Mueller 2013; Kubitschek and Hallinan 1998), we find evidence that higher educational institutions can substantially alter which friendships develop through grouping practices that restrict or constrain who attends classes together. In turn, one of the key takeaways of this paper is the substantial impact that the average performance of one's friends within the major has on one's own performance (Hasan and Bagde 2013). If we view academic performance in college—net of pre-college factors—as a sign of learning, then decisions surrounding course enrollment can have indirect consequences for aggregate levels of learning among students through the friendships that develop. Thus, whereas discussions surrounding grouping practices at the K-12 level have largely focused on what takes place in the classroom (Gamoran 1992; Hallinan 1994), I add here that at the higher education level, we must also acknowledge the social dynamics that occur outside the classroom.

In terms of inequality, the findings presented here are less clear for three reasons. First, although student friendships do tend to sort based on background characteristics, this sorting is not related to the performance-based clustering we see in the network. Indeed, though not significant, the MEMS indicate that selection based on sociodemographic characteristics as well as pre-college and other factors contribute to a decrease in performance-based clustering in the network (see Table 3.4). Second, while here I document friend influence effects, to understand the total effect of grouping practices, we would also need to account for possible teacher effects and the potential impact of differential pedagogical practices. Third, in general, we still lack a good understanding of what predicts success in college (Bowman 2023), and thus a related aspect is exactly how or why friends would matter for performance. For instance, it may be related to study habits or other academic practices (Johnson 2022; McCabe 2020). Better understanding these dynamics is essential for thinking about potential network interventions. For example, Figure 3.1 highlights that most students in the major are connected to one another through some path. However, due to the fast pace of introductory STEM courses, friendship clustering in the network could lead some students to gain access to beneficial information earlier than others and thus gain a competitive advantage in the major.

This study is not without limitations. First, while inferences are based on longitudinal data and capture the temporal ordering of effects, I cannot claim to document causal effects. Second, the status effects shown here call for further research. My data is unable to unpack why we find status effects for performance but not curricular group. I speculate that any tendency among students to befriend lower-ranked peers could be the result of the additional resources the developmental students were provided with in this context. In other words, this finding could stem from instrumental drives among students to access the additional resources offered to this group. It could also be the result of failed attempts to befriend higher-ranked peers if those students are more insular or otherwise less available. As such, better understanding the nature of status effects within educational settings is a critical area for sociologists of education to pursue (Martin and Murphy 2020). Future research should explore how students themselves make sense of the forms of curricular differentiation examined in this study. For instance, it remains unclear exactly how aware students may be of each other's curricular group placement compared to their academic performance (Santoro and Bunte 2022). It may be that academic performance is more salient and known among students and thus able to induce status effects where weaker forms of tracking do not (Dokuka et al. 2020; Grunspan et al. 2016; Smith 2015). Additionally, future work should collect and study alternative measures of status among peers compared to sociometric popularity examined here (see Vörös, Block, and Boda 2019). Ultimately, however, findings here indicate that potential status effects based upon either curricular group membership or academic performance are unlikely to lead to meaningful integration in the network across levels. Institutions wanting to facilitate these connections would likely have to be more intentional about structuring opportunities to interact.

CHAPTER 3

SUPPLEMENTARY MATERIALS

S3.1. Main Effects from SAOMs using Spring Term Curricular Group Membership

Selection: Friendship Ties	2018 – 2019	2019 – 2020
Academic homophily effects		
Shared coursework	.744***	.607***
	(.049)	(.037)
Curricular group (*same*)	.178†	.319***
	(.101)	(.086)
Academic performance (*similarity*)	.444	.783*
	(.477)	(.328)
Academic status effects		
Curricular group (*alter*)	-.390***	-.410***
	(.103)	(.093)
Academic performance (*similarity*)	.073*	.045†
	(.031)	(.025)
Influence: Academic Performance		
Average similarity	2.085**	1.298*
	(.705)	(.656)
In-degree	.046*	.026
	(.019)	(.017)
Maximum Convergence Ratio	.168	.210

†p<.10; *p<.05; **p<.01;***p<.001

Notes: Standard errors in parentheses.

SAOMs also account for: (1) Network dynamics on friendship: *density*, *reciprocity*, *transitive triplets*, *transitive recipr. triplets*, *in-degree popularity (sqrt)*, *in-degree activity (sqrt)*, *out-degree activity (sqrt)*, *ego* curricular group, *ego* academic performance, *same/ego/alter* race/ethnicity, *same/ego/alter* first-generation student status, *same/ego/alter* low-income status, *same/ego/alter* gender, *same* international student status, *same* high school, *same* housing, *similarity* SAT score, *similarity* high school GPA, *ego* same-major ties and (2) Behavioral dynamics on GPA: *linear shape*, *quadratic shape*, *out-degree*, race/ethnicity, first-generation student status, low-income status, gender, international student status, curricular group placement, SAT score, and high school GPA.

S3.2. Results from the Micro Effect on Macro Structure (MEMS) Estimation using Morans I by Curricular Group Placement as the Outcome

Joint Parameters	MEMS 2018	% Change	MEMS 2019	% Change
Shared coursework + curricular group (*same*) + curricular group (*alter*) selection effects	.177***	24.53	.127**	17.79
	(.052)		(.041)	

†p<.10; *p<.05; **p<.01;***p<.001
Notes: Standard deviations in parentheses.
MEMS were estimated using 1,000 simulations.

S3.3. Results from the Micro Effect on Macro Structure (MEMS) Estimation using Morans I by Academic Performance as the Outcome

Parameter	MEMS 2018	% Change	MEMS 2019	% Change
SAT score (*similarity*) selection effects	.001	-2.61	.002	-1.21
	(.048)		(.039)	

†p<.10; *p<.05; **p<.01;***p<.001
Notes: Standard deviations in parentheses.
MEMS were estimated using 1,000 simulations.

S3.4. Results from the Micro Effect on Macro Structure (MEMS) Estimation using Morans I by Academic Performance as the Outcome

Joint Parameters	MEMS 2018	% Change	MEMS 2019	% Change
Academic performance (*similarity*) + academic performance (*alter*) selection effects	.021	5.81	.028	12.10
	(.056)		(.039)	

†p<.10; *p<.05; **p<.01;***p<.001
Notes: Standard deviations in parentheses.
MEMS were estimated using 1,000 simulations.

CONCLUSION

My dissertation has examined various ways in which individuals are ranked and sorted in higher education. Theoretically, I have shown that students are ranked not only across institutions in terms of selectivity, but also within a given institution (e.g., curricular grouping, grades, etc.). I have also drawn attention to not only the ways in which routine curricular practices may shape the networks that emerge, but also the significant influence friends have on one another's performance in college.

Together the studies presented here highlight the importance of understanding the hierarchical nature of the educational system in the US. Specifically, we need to continue to interrogate the stratification of educational credentials in the US context. This stratification could be conceptualized in terms of both its vertical and horizontal dimensions. The vertical dimension would focus on how far students go in their studies, and the role of not only undergraduate but also graduate degrees. The horizontal dimension would study differentiation across a given level of schooling, such as institutional selectivity, major or field of study, and performance level.

A more systematic investigation of educational credentials in the US would allow for two fruitful areas of future research. First, within the subfield of stratification and mobility, we need a better grasp of how much of the intergenerational transmission of advantage is due to the signaling power of a given educational credential and if this has changed over time. This understanding would allow us to see how much skills, credentials, and occupational attainment are aligned. In turn, if the goal is to expand access to skill development in college, then it seems more studies are needed within the "sociology of learning"(Boocock 1966). Namely, we still know relatively little about how to leverage interpersonal ties to maximize learning in higher education. With increasing doubts about the value of a college degree, as well as the rapid growth of artificial intelligence, understanding the unique contribution of peer-to-peer connections and their influence on learning seems like a promising potential avenue.

REFERENCES

Alon, Sigal. 2009. The Evolution of Class Inequality in Higher Education: Competition, Exclusion, and Adaptation. *American Sociological Review* 74(5):731-755.

An, Brian P. 2010. The relations between race, family characteristics, and where students apply to college. *Social Science Research* 39(2):310-323.

An, Weihua. 2022. "Friendship Network Formation in Chinese Middle Schools: Patterns of Inequality and Homophily." *Social Networks* 68(1):218-228.

Andrade, Maureen S. 2007. “Learning Communities: Examining Positive Outcomes.” *Journal of College Student Retention: Research, Theory & Practice* 9(1):1-20.

Armstrong, Elizabeth A., and Laura T. Hamilton. 2013. *Paying for the Party: How College Maintains Inequality.* Cambridge, MA: Harvard University Press.

Bastedo, Michael N., and Ozan Jaquette. 2011. “Running in Place: Low-income Students and the Dynamics of Higher Education Stratification. *Educational Evaluation and Policy Analysis* 33(3):318-339.

Bettinger, Eric P., and Bridget T. Long. 2008. "Addressing the Needs of Underprepared Students in Higher Education does College Remediation Work?." *Journal of Human Resources* 44(3):736-771.

Blau, Peter M., and Otis D. Duncan. 1967. *The American Occupational Structure*. John Wiley & Sons.

Blinder, Alan S. 1973. “Wage Discrimination: Reduced Form and Structural Estimates.” *Journal of Human Resources* 8(4):436-455.

Boda, Zsófia, Timon Elmer, András Vörös, and Christoph Stadtfeld. 2020. "Short-term and Long-term Effects of a Social Network Intervention on Friendships among University Students." *Scientific Reports* 10(1):1-12.

Boocock, Sarane S. 1966. "Toward a Sociology of Learning: A Selective Review of Existing Research." *Sociology of Education* 39(1):1-45.

Boudon, Raymond. 1974. *Education, Opportunity, and Social Inequality: Changing Prospects in Western Society*. John Wiley & Sons, Inc.

Bowen, William G., Michael McPherson, and Matthew M. Chingos. 2009. *Crossing the Finish Line: Completing College at America's Public Universities*. Princeton, NJ: Princeton University Press.

Bowman, Nicholas A., ed. 2023. *How College Students Succeed: Making Meaning across*

Disciplinary Perspectives. Sterling, Virginia: Stylus Publishing.

Bowman, Nicholas A., and K. C. Culver. 2018. "When do Honors Programs Make the Grade? Conditional Effects on College Satisfaction, Achievement, Retention, and Graduation." *Research in Higher Education* 59(3):249-272.

Brewer, Dominic J., Eric R. Eide, and Ronald G. Ehrenberg. 1999. “Does it Pay to Attend an Elite Private College? Cross-cohort Evidence on the Effects of College Type on Earnings.” *Journal of Human Resources* 34(1):104-123.

Brouwer, Jasperina, Andreas Flache, Ellen Jansen, Adriaan Hofman, and Christian Steglich. 2018. "Emergent Achievement Segregation in Freshmen Learning Community Networks." *Higher Education* 76(3):483-500.

Brouwer, Jasperina, Carlos A. de Matos Fernandes, Christian EG Steglich, Ellen PWA Jansen, WH Adriaan Hofman, and Andreas Flache. 2022. "The Development of Peer Networks and Academic Performance in Learning Communities in Higher Education." *Learning and Instruction* 80:101603.

Burt, Ronald S. 2004. "Structural Holes and Good Ideas." *American Journal of Sociology* 110(2):349-399.

Cabrera, Alberto F., and Steven M. La Nasa. 2000a. “Overcoming the Tasks on the Path to College for America's Disadvantaged.” *New Directions for Institutional Research* 2000(107):31-43.

Cabrera, Alberto F., and Steven M. La Nasa. 2000b. “Understanding the College-choice Process.” *New Directions for Institutional Research* 2000(107):5-22.

Calonico, Sebastian, Matias D. Cattaneo, and Rocio Titiunik. 2015. "Optimal Data-driven Regression Discontinuity Plots." *Journal of the American Statistical Association* 110(512):1753-1769.

Carrell, Scott E., Richard L. Fullerton, and James E. West. 2009. "Does Your Cohort Matter? Measuring Peer Effects in College Achievement." *Journal of Labor Economics* 27(3): 439-464.

Carrell, Scott E., Bruce I. Sacerdote, and James E. West. 2013. "From Natural Variation to Optimal Policy? The Importance of Endogenous Peer Group Formation." *Econometrica* 81(3):855-882.

Chen, Xianglei. 2013. “STEM Attrition: College Students' Paths into and out of STEM Fields. *National Center for Education Statistics*. Retrieved from http://ies.ed.gov/pubsearch/pubsinfo.asp?pubid=2014001rev. Accessed 22 September 2021.

Chetty, Raj, David J. Deming, and John N. Friedman. 2023. “Diversifying Society’s Leaders? The Causal Effects of Admission to Highly Selective Private Colleges.” No. w31492. *National Bureau of Economic Research.*

Chetty, Raj, John N. Friedman, Emmanuel Saez, Nicholas Turner, and Danny Yagan. 2020. "Income Segregation and Intergenerational Mobility across Colleges in the United States." *The Quarterly Journal of Economics* 135(3):1567-1633.

Coleman, James S. 1961. *The Adolescent Society: The Social Life of the Teenager and Its Impact on Education*. Glencoe, IL: Free Press.

Coleman, James S. 1988. "Social Capital in the Creation of Human Capital." *American Journal of Sociology* 94:S95-S120.

Cox, Amanda Barrett. 2017. “Cohorts "Siblings", and Mentors: Organizational Structures and the Creation of Social Capital.” *Sociology of Education* 90(1):47-63.

Dagley, Melissa, Michael Georgiopoulos, Amber Reece, and Cynthia Young. 2016. "Increasing Retention and Graduation Rates through a STEM Learning Community." *Journal of College Student Retention: Research, Theory & Practice* 18(2):167-182.

DiMaggio, Paul, and Filiz Garip. 2012. "Network Effects and Social Inequality." *Annual Review of Sociology* 38:93-118.

Dokuka, Sofia, Diliara Valeeva, and Maria Yudkevich. 2020. "How Academic Achievement Spreads: The Role of Distinct Social Networks in Academic Performance Diffusion." *Plos one* 15(7): e0236737.

Domina, Thurston, Andrew M. Penner, and Emily K. Penner. 2016. "'Membership Has Its Privileges': Status Incentives and Categorical Inequality in Education." *Sociological Science* 3:264-295.

Domina, Thurston, Andrew Penner, and Emily Penner. 2017. "Categorical Inequality: Schools as Sorting Machines." *Annual Review of Sociology* 43:311-330.

Domizi, Denise. 2008. “Student Perceptions about their Informal Learning Experiences in a First-year Residential Learning Community.” *Journal of the First-Year Experience & Students in Transition* 20(1):97-110.

Duxbury, Scott W. 2023. "The Problem of Scaling in Exponential Random Graph Models." *Sociological Methods & Research* 52(2):764-802.

Duxbury, Scott W. 2024. "Micro Effects on Macro Structure in Social Networks." *Sociological Methodology* 54(1):1-26.

Engzell, Per, and Isabel J. Raabe. 2023. "Within-school Achievement Sorting in Comprehensive and Tracked Systems." *Sociology of Education* 96(4):324-343.

Epstein, Joyce L., and Nancy Karweit, eds. 1983. *Friends in School: Patterns of Selection and Influence in Secondary Schools*. New York: Academic Press, Inc.

Espenshade, Thomas J., and Alexandria Walton Radford. 2009. *No Longer Separate, Not Yet Equal: Race and Class in Elite College Admission and Campus Life*. Princeton, NJ: Princeton University Press.

Feld, Scott L. 1981. "The Focused Organization of Social Ties." *American Journal of Sociology* 86(5):1015-1035.

Felten, Peter, and Leo M. Lambert. 2020. *Relationship-rich Education: How Human Connections Drive Success in College*. Baltimore, MD: Johns Hopkins University Press.

Flashman, Jennifer. 2012. "Academic Achievement and its Impact on Friend Dynamics." *Sociology of Education* 85(1):61-80.

Ford, Karly Sarita, and Jason Thompson. 2016. “Inherited Prestige: Intergenerational Access to Selective Universities in the United States.” *Research in Social Stratification and Mobility* 46(2):86-98.

Frank, Kenneth A., Chandra Muller, and Anna S. Mueller. 2013. "The Embeddedness of Adolescent Friendship Nominations: The Formation of Social Capital in Emergent Network Structures." *American Journal of Sociology* 119(1):216-253.

Freeman, Tierra M., Lynley H. Anderman, and Jane M. Jensen. 2007. "Sense of Belonging in College Freshmen at the Classroom and Campus Levels." *The Journal of Experimental Education* 75(3):203-220.

Gamoran, Adam. 1992. "Synthesis of Research: Is Ability Grouping Equitable?." *Educational Leadership* 50(2):11-17.

Garlick, Robert. 2018. "Academic Peer Effects with Different Group Assignment Policies: Residential Tracking Versus Random Assignment." *American Economic Journal: Applied Economics* 10(3):345-369.

Gerber, Theodore P., and Sin Yi Cheung. 2008. "Horizontal Stratification in Postsecondary Education: Forms, Explanations, and Implications." *Annual Review of Sociology* 34(1): 299-318.

Gould, Roger V. 2002. "The Origins of Status Hierarchies: A Formal Theory and Empirical Test." *American Journal of Sociology* 107(5):1143-1178.

Goyette, Kimberly A. 2008. "College for Some to College for All: Social Background, Occupational Expectations, and Educational Expectations Over Time." *Social Science Research* 37(2):461-484.

Grodsky, Eric, and Erika Jackson. 2009. “Social Stratification in Higher Education.” *Teachers College Record* 111(10): 2347-2384.

Grunspan, Daniel Z., Sarah L. Eddy, Sara E. Brownell, Benjamin L. Wiggins, Alison J. Crowe, and Steven M. Goodreau. 2016. "Males Under-estimate Academic Performance of their Female Peers in Undergraduate Biology Classrooms." *PloS one* 11(2): e0148405.

Haller, Archibald O., and Alejandro Portes. 1973. “Status Attainment Processes.” *Sociology of Education* 46(1):51-91.

Hallinan, Maureen T. 1982. "The Peer Influence Process." *Studies in Educational Evaluation* 7(3):285-306.

Hallinan, Maureen T. 1994. "Tracking: From Theory to Practice." *Sociology of Education* 67(2): 79-84.

Hallinan, Maureen T., and Aage B. Sørensen. 1985. "Ability Grouping and Student Friendships." *American Educational Research Journal* 22(4):485-499.

Handcock, Mark S., David R. Hunter, Carter T. Butts, Steven M. Goodreau, and Martina Morris. 2008. “Statnet: Software Tools for the Representation, Visualization, Analysis and Simulation of Network Data. *Journal of statistical software* 24(1):1548-7660.

Hartup, Willard W., and Nan Stevens. 1997. “Friendships and Adaptation in the Life Course.” *Psychological Bulletin* 121(3):355-370.

Hasan, Sharique, and Surendrakumar Bagde. 2013. "The Mechanics of Social Capital and Academic Performance in an Indian College." *American Sociological Review* 78(6): 1009-1032.

Hays, Robert B., and Diana Oxley. 1986. "Social Network Development and Functioning during a Life Transition." *Journal of Personality and Social Psychology* 50(2):305-313.

Heckman, James J. 1979. "Sample Selection Bias as a Specification Error." *Econometrica: Journal of the Econometric Society* 47(1):153-161.

Hirschauer, Stefan. 2023. "Telling People Apart: Outline of a Theory of Human Differentiation." *Sociological Theory* 41(4):352-376.

Holland, Megan M. 2014. “Navigating the Road to College: Race and Class Variation in the College Application Process.” *Sociology Compass* 8(10):1191-1205.

Holzman, Brian, Daniel Klasik, and Rachel Baker. 2020. “Gaps in the College Application

Gauntlet.” *Research in Higher Education* 61:795-822.

Hossler, Don, and Nick Vesper. 1999. *Going to College: How Social, Economic, and Educational Factors Influence the Decisions Students Make*. JHU Press.

Hotchkiss, Julie L., Robert E. Moore, and M. Melinda Pitts. 2006. "Freshman Learning Communities, College Performance, and Retention." *Education Economics* 14(2):197-210.

Hoxby, Caroline M., and Christopher Avery. 2012. “*The Missing "One-offs": The Hidden Supply of High-achieving, Low-income Students.”* (No. w18586). National Bureau of Economic Research.

Hoxby, Caroline, and Sarah Turner. 2013. “Expanding College Opportunities for High-achieving, Low-income Students.” *Stanford Institute for Economic Policy Research Discussion Paper* 12(014):7.

Huang, Gary, Nebiyu Taddese, and Elizabeth Walter. 2000. “Entry and Persistence of Women and Minorities in College Science and Engineering Education.” *National Center for Education Statistics*. Retrieved from https://eric.ed.gov/?id=ED566411. Accessed 22 September 2021.

Huang, Peng, and Carter T. Butts. 2023. "Rooted America: Immobility and Segregation of the Intercounty Migration Network." *American Sociological Review* 88(6):1031-1065.

Hurtado, Sylvia, Kevin Eagan, and Mitchell Chang. 2010. *Degrees of Success: Bachelor’s Degree Completion Rates among Initial STEM Majors*. UCLA, CA: Higher Education Research Institute.

Imbens, Guido W., and Joshua D. Angrist. 1994. “Identification and Estimation of Local Average Treatment Effects.” *Econometrica* 62(2):467-475.

Imbens, Guido W., and Thomas Lemieux. 2008. “Regression Discontinuity Designs: A Guide to Practice. *Journal of Econometrics* 142(2):615-635.

Indlekofer, Natalie, and Ulrik Brandes. 2013. "Relative Importance of Effects in Stochastic Actor-oriented Models." *Network Science* 1(3):278-304.

Ingels, Steven J., Daniel J. Pratt, Deborah R. Herget, Jill A. Dever, Laura Burns Fritch, Randolph Ottem, James E. Rogers, Sami Kitmitto, and Steve Leinwand. 2013. “High School Longitudinal Study of 2009 (HSLS: 09) Base Year to First Follow-Up Data File Documentation. Appendixes. NCES 2014-361.” *National Center for Education Statistics*.

Jack, Anthony A. 2019. *The Privileged Poor: How Elite Colleges Are Failing Disadvantaged Students*. Cambridge, MA: Harvard University Press.

Jackson, Michelle, ed. 2013. *Determined to Succeed?: Performance Versus Choice in Educational Attainment*. Stanford, CA: Stanford University Press.

Jacob, Robin, Pei Zhu, Marie-Andrée Somers, and Howard Bloom. 2012. *A Practical Guide to Regression Discontinuity.* New York, NY: MDRC.

Jaffee, David, Adam Carle, Richard Phillips, and Lucy Paltoo. 2008. "Intended and Unintended Consequences of First-year Learning Communities: An Initial Investigation. *Journal of the First-Year Experience & Students in Transition* 20(1):53-70.

James, Sylvia M., and Susan R. Singer. 2016. "From the NSF: The National Science Foundation's Investments in Broadening Participation in Science, Technology, Engineering, and Mathematics Education through Research and Capacity Building." *CBE—Life Sciences Education* 15(3):1-8.

Jann, Ben. 2008. "The Blinder–Oaxaca Decomposition for Linear Regression Models." *The Stata Journal* 8(4):453-479.

Jasso, Guillermina. 2001. "Studying Status: An Integrated Framework." *American Sociological Review* 66(1):96-124.

Jeffrey, Wesley. 2020. "Crossing the Finish Line? A Review of College Completion Inequality in the United States by Race and Class." *Sociology Compass* 14(5):1-19.

Jeffrey, Wesley, David R. Schaefer, Di Xu, Peter McPartlan, and Sabrina Solanki. 2022. "STEM Learning Communities Promote Friendships but Risk Academic Segmentation." *Scientific Reports* 12(1):1-9.

Jencks, Christopher. 1972. *Inequality: A Reassessment of the Effect of Family and Schooling in America.* New York, NY: Basic Books Inc.

Johnson, Anthony M. 2019. "''I Can Turn It On When I Need To'': Pre-college Integration, Culture, and Peer Academic Engagement among Black and Latino/a Engineering Students." *Sociology of Education* 92(1):1-20.

Johnson, Anthony M. 2022. "Collaborating in Class: Social Class Context and Peer Help-Seeking and Help-Giving in an Elite Engineering School." *American Sociological Review* 87(6):981-1006.

Kadushin, Charles. 2012. *Understanding Social Networks: Theories, Concepts, and Findings*. New York, NY: Oxford University Press.

Kolluri, Suneal, and William G. Tierney. 2020. "Introduction: Education, Equity, and the

Promise of Seeing Relationally." Pp. 1-12 in *Relational Sociology and Research on Schools, Colleges, and Universities*, edited by William G. Tierney and Suneal Kolluri. SUNY Press.

Kossinets, Gueorgi, and Duncan J. Watts. 2009. "Origins of Homophily in an Evolving Social Network." *American Journal of Sociology* 115(2):405-450.

Kraemer, Barbara A. 1997. "The Academic and Social Integration of Hispanic Students into College." *The Review of Higher Education* 20(2):163-179.

Kubitschek, Warren N., and Maureen T. Hallinan. 1998. "Tracking and Students' Friendships." *Social Psychology Quarterly* 61(1):1-15.

Lee, David S., and Thomas Lemieux. 2010. "Regression Discontinuity Designs in Economics." *Journal of Economic Literature* 48(2):281-355.

Leszczensky, Lars, and Sebastian Pink. 2015. "Ethnic Segregation of Friendship Networks in School: Testing a Rational-choice Argument of Differences in Ethnic Homophily between Classroom-and Grade-level Networks." *Social Networks* 42:18-26.

Lomi, Alessandro, Tom AB Snijders, Christian EG Steglich, and Vanina Jasmine Torló. 2011. "Why are Some More Peer Than Others? Evidence from a Longitudinal Study of Social Networks and Individual Academic Performance." *Social Science Research* 40(6):1506-1520.

Long, Bridget Terry, and Angela Boatman. 2013. "The Role of Remedial and Developmental Courses in Access and Persistence." Pp. 89-107 in *The State of College Access and Completion*, edited by Laura W. Perna and Anthony P. Jones. Routledge.

Lor, Yang Va. 2023. *Unequal Choices: How Social Class Shapes Where High-achieving Students Apply to College*. New Brunswick, NJ: Rutgers University Press.

Lucas, Samuel R. 1999. *Tracking Inequality: Stratification and Mobility in American High Schools. Sociology of Education Series*. New York, NY: Teachers College Press.

Lucas, Samuel R. 2001. "Effectively Maintained Inequality: Education Transitions, Track Mobility, and Social Background Effects." *American Journal of Sociology* 106(6):1642-1690.

Marsden, Peter V., and Karen E. Campbell. 1984. "Measuring Tie Strength." *Social Forces* 63(2):482-501.

Martin, John Levi, and James P. Murphy. 2020. "Networks, Status, and Inequality." Pp. 98-118 in *The Oxford Handbook of Social Network Analysis*, edited by R. Light and J. Moody. New York, NY: Oxford University Press.

Maton, Kenneth I., Shauna A. Pollard, Tatiana V. McDougall Weise, and Freeman A. Hrabowski. 2012. "Meyerhoff Scholars Program: A Strengths-based, Institution-wide Approach to Increasing Diversity in Science, Technology, Engineering, and Mathematics." *Mount Sinai Journal of Medicine: A Journal of Translational and Personalized Medicine* 79(5):610-623.

Mattie, Heather, Kenth Engø-Monsen, Rich Ling, and Jukka-Pekka Onnela. 2018. "Understanding Tie Strength in Social Networks using a Local “Bow Tie” Framework." *Scientific Reports* 8(1):1-9.

Mayer, Adalbert, and Steven L. Puller. 2008. "The Old Boy (and Girl) Network: Social Network Formation on University Campuses." *Journal of Public Economics* 92(1-2):329-347.

McCabe, Janice M. 2016. *Connecting in College: How Friendship Networks Matter for Academic and Social Success*. Chicago, IL: University of Chicago Press.

McCabe, Janice M. 2020. "Why Study with Friends?: A Relational Analysis of Students' Strategies to Integrate Social and Academic Life." Pp. 135-156 in *Relational Sociology and Research on Schools, Colleges, and Universities,* edited by William G. Tierney and Suneal Kolluri. Albany: SUNY Press.

McDonough, Patricia M. 1997. *Choosing Colleges: How Social Class and Schools Structure Opportunity*. Albany, NY: Suny Press.

McFarland, Daniel A., James Moody, David Diehl, Jeffrey A. Smith, and Reuben J. Thomas. 2014. "Network Ecology and Adolescent Social Structure." *American Sociological Review* 79(6):1088-1121.

McPherson, Miller, Lynn Smith-Lovin, and James M. Cook. 2001. "Birds of a Feather: Homophily in Social Networks." *Annual Review of Sociology* 27:415-444.

Meyer, John W. 1970. “High School Effects on College Intentions.” *American Journal of Sociology* 76(1):59-70.

Meyer, John W., and Brian Rowan. 1977. "Institutionalized Organizations: Formal Structure as Myth and Ceremony." *American Journal of Sociology* 83(2):340-363.

Mollenhorst, Gerald, Beate Volker, and Henk Flap. 2014. "Changes in Personal Relationships: How Social Contexts Affect the Emergence and Discontinuation of Relationships." *Social Networks* 37:65-80.

Morgan, Stephen L. 2018. "A New Social Psychological Model of Educational Attainment." Pp. 202-207 in *Inequality in the 21st Century: A Reader,* edited by David Grusky and Jasmine Hill. Routledge.

Mullen, Ann L. 2011. *Degrees of Inequality: Culture, Class, and Gender in American Higher*

Education. Baltimore, MD: Johns Hopkins University Press.

Mullen, Ann L., and Kimberly A. Goyette. 2019. "Aiming High: Social and Academic Correlates of Applying to and Attending 'Reach' Universities." *British Journal of Sociology of Education* 40(8):1072-1089.

National Science Foundation, Broadening Participation Working Group. 2014. "Pathways to Broadening Participation in Response to the CEOSE 2011-2012 Recommendation." *National Science Foundation*. Retrieved from https://www.nsf.gov/pubs/2015/nsf15037/nsf15037.pdf. Accessed 22 September 2021.

National Survey of Student Engagement. 2015. *Engagement Insights: Survey Findings on the Quality of Undergraduate Education - Annual Results 2015* (Bloomington, IN).

Niu, Sunny Xinchun, and Marta Tienda. 2008. "Choosing Colleges: Identifying and Modeling Choice Sets." *Social Science Research* 37(2):416-433.

Nora, Amaury. 1993. "Two-year Colleges and Minority Students' Educational Aspirations: Help or Hindrance." *Higher Education: Handbook of Theory and Research* 9(3):212-247.

Nunn, Lisa M. 2021. *College Belonging: How First-year and First-generation Students Navigate Campus Life*. New Brunswick, New Jersey: Rutgers University Press.

Oakes, Jeannie. 2005. *Keeping Track: How Schools Structure Inequality*. New Haven, CT: Yale University Press.

Oaxaca, Ronald. 1973. "Male-Female Wage Differentials in Urban Labor Markets." *International Economic Review* 14(3):693-709.

Odle, Taylor, and Jennifer Delaney. 2023. "Experimental Evidence on "Direct Admissions" from Four States: Impacts on College Application and Enrollment." *Available at SSRN 4548344*.

Odle, Taylor, and Preston Magouirk. 2023. " 'Non-Submitters': Evidence on Students Who Start but Don't Complete a College Application." *Available at SSRN 4527340*.

Otto, Sheila, Mary A. Evins, Michelle Boyer-Pennington, and Thomas M. Brinthaupt. 2015. "Learning Communities in Higher Education: Best Practices." *Journal of Student Success and Retention* 2(1):1-20.

Park, Julie J., Young K. Kim, Kristyn Lue, Jia Zheng, Roshan Parikh, Cinthya Salazar, and Arman Liwanag. 2021. "Who Are You Studying With? The Role of Diverse Friendships in STEM and Corresponding Inequality." *Research in Higher Education* 62(8):1146-1167.

Perna, Laura W. 2006. "Studying College Access and Choice: A Proposed Conceptual Model."

Pp. 99-157 in *Higher Education: Handbook of Theory and Research*, edited by John C. Smart. Dordrecht: Springer Netherlands.

Perna, Laura Walter, and Marvin A. Titus. 2005. “The Relationship between Parental Involvement as Social Capital and College Enrollment: An Examination of Racial/Ethnic Group Differences.” *The Journal of Higher Education* 76(5):485-518.

Perry, Brea L., Bernice A. Pescosolido, and Stephen P. Borgatti. 2018. *Egocentric Network Analysis: Foundations, Methods, and Models*. Cambridge, UK: Cambridge University Press.

Persell, Caroline Hodges, Sophia Catsambis, and Peter W. Cookson Jr. 1992. “Differential Asset Conversion: Class and Gendered Pathways to Selective Colleges.” *Sociology of Education* 65(3):208-225.

Poldin, Oleg, Diliara Valeeva, and Maria Yudkevich. 2016. "Which Peers Matter: How Social Ties Affect Peer-group Effects." *Research in Higher Education* 57:448-468.

Posselt, Julie R., and Eric Grodsky. 2017. "Graduate Education and Social Stratification." *Annual Review of Sociology* 43:353-378.

Raabe, Isabel J., Zsófia Boda, and Christoph Stadtfeld. 2019. "The Social Pipeline: How Friend Influence and Peer Exposure Widen the STEM Gender Gap." *Sociology of Education* 92(2):105-123.

Radford, Alexandria Walton. 2013. *Top Student, Top School? How Social Class Shapes Where Valedictorians Go to College*. Chicago, IL: University of Chicago Press.

Rambaran, J. Ashwin, Andrea Hopmeyer, David Schwartz, Christian Steglich, Daryaneh Badaly, and René Veenstra. 2017. "Academic Functioning and Peer Influences: A Short-term Longitudinal Study of Network–behavior Dynamics in Middle Adolescence." *Child Development* 88(2):523-543.

Reardon, Sean F., Rachel Baker, and Daniel Klasik. 2012. *Race, Income, and Enrollment Patterns in Highly Selective Colleges, 1982-2004.*

Reay, Diane, Miriam E. David, and Stephen J. Ball. 2005. *Degrees of Choice: Class, Race, Gender and Higher Education*. Sterling, US: Trentham Books.

Ridgeway, Cecilia L. 2014. "Why Status Matters for Inequality." *American Sociological Review* 79(1):1-16.

Rinn, Anne N., and Jonathan A. Plucker. 2019. "High-ability College Students and Undergraduate Honors Programs: A Systematic Review." *Journal for the Education of the Gifted* 42(3):187-215.

Ripley, Ruth M., Tom A.B. Snijders, Zsofia Boda, Andras Voros, and Paulina Preciado. 2024. *Manual for RSiena* (version January 17, 2024). Oxford, UK: University of Oxford, Department of Statistics; Nuffield College.

Rivera, Mark T., Sara B. Soderstrom, and Brian Uzzi. 2010. "Dynamics of Dyads in Social Networks: Assortative, Relational, and Proximity Mechanisms." *Annual Review of Sociology* 36:91-115.

Robins, Garry, Pip Pattison, Yuval Kalish, and Dean Lusher. 2007. "An Introduction to Exponential Random Graph (p*) Models for Social Networks." *Social Networks* 29(2):173-191.

Robinson, Karen Jeong, and Josipa Roksa. 2016. “Counselors, Information, and High School College-going Culture: Inequalities in the College Application Process.” *Research in Higher Education* 57:845-868.

Roderick, Melissa, Vanessa Coca, and Jenny Nagaoka. 2011. “Potholes on the Road to College: High School Effects in Shaping Urban Students’ Participation in College Application, Four-year College Enrollment, and College Match.” *Sociology of Education* 84(3):178-211.

Roksa, Josipa, and Denise Deutschlander. 2018. “Applying to College: The Role of Family Resources in Academic Undermatch.” *Teachers College Record* 120(6):1-30.

Rumberger, Russell W., and Scott L. Thomas. 1993. "The Economic Returns to College Major, Quality and Performance: A Multilevel Analysis of Recent Graduates." *Economics of Education Review* 12(1):1-19.

Sanabria, Tanya, Andrew Penner, and Thurston Domina. 2020. "Failing at Remediation? College Remedial Coursetaking, Failure and Long-term Student Outcomes." *Research in Higher Education* 61(4):459-484.

Sauder, Michael, Freda Lynn, and Joel M. Podolny. 2012. "Status: Insights from Organizational Sociology." *Annual Review of Sociology* 38(1):267-283.

Schneider, Jack, and Ethan Hutt. 2014. "Making the Grade: A History of the A–F Marking Scheme." *Journal of Curriculum Studies* 46(2):201-224.

Scott-Clayton, Judith, and Olga Rodriguez. 2015. "Development, Discouragement, or Diversion? New Evidence on the Effects of College Remediation Policy." *Education Finance and Policy* 10(1):4-45.

Sewell, William H., Archibald O. Haller, and Alejandro Portes. 1969. “The Educational and Early Occupational Attainment Process.” *American Sociological Review* 34(1):82-92.

Seymour, Elaine, and Anne-Barrie Hunter, eds. 2019. *Talking About Leaving Revisited:*

Persistence, Relocation, and Loss in Undergraduate STEM Education. Springer.

Shavit, Yossi, and Hyunjoon Park. 2016. "Introduction to the Special Issue: Education as a Positional Good." *Research in Social Stratification and Mobility* 43:1-3.

Smirnov, Ivan, and Stefan Thurner. 2017. "Formation of Homophily in Academic Performance: Students Change their Friends Rather than Performance." *PloS one* 12(8):e0183473.

Smith, Rachel A. 2015. "Magnets and Seekers: A Network Perspective on Academic Integration Inside Two Residential Communities." *The Journal of Higher Education* 86(6): 893-922.

Smith, Barbara Leigh, Jean MacGregor, Roberta Matthews, and Faith Gabelnick. 2004. *Learning Communities: Reforming Undergraduate Education.* San Francisco, CA: Jossey-Bass.

Smith, Jonathan, Matea Pender, and Jessica Howell. 2013. "The Full Extent of Student-college Academic Undermatch." *Economics of Education Review* 32:247-261.

Snijders, Tom AB, and Alessandro Lomi. 2019. "Beyond Homophily: Incorporating Actor Variables in Statistical Network Models." *Network Science* 7(1):1-19.

Sørensen, Aage Bøttger. 1970. "Organizational Differentiation of Students and Educational Opportunity." *Sociology of Education* 43:355-376.

Stadtfeld, Christoph, András Vörös, Timon Elmer, Zsófia Boda, and Isabel J. Raabe. 2019. "Integration in Emerging Social Networks Explains Academic Failure and Success." *Proceedings of the National Academy of Sciences* 116(3):792-797.

Stearns, Elizabeth, Claudia Buchmann, and Kara Bonneau. 2009. "Interracial Friendships in the Transition to College: Do Birds of a Feather Flock Together Once They Leave the Nest?." *Sociology of Education* 82(2):173-195.

Steglich, Christian, Tom AB Snijders, and Michael Pearson. 2010. "Dynamic Networks and Behavior: Separating Selection from Influence." *Sociological Methodology* 40(1):329-393.

Stevens, Mitchell L. 2009. *Creating a Class: College Admissions and the Education of Elites*. Cambridge, MA: Harvard University Press.

Stolzenberg, Ross M., and Daniel A. Relles. 1997. "Tools for Intuition about Sample Selection Bias and its Correction." *American Sociological Review* 62(3):494-507.

Taylor, Barrett J., and Brendan Cantwell. 2019. *Unequal Higher Education: Wealth, Status, and Student Opportunity*. New Brunswick, NJ: Rutgers University Press

Thomas, Reuben J. 2019. "Sources of Friendship and Structurally Induced Homophily across the Life Course." *Sociological Perspectives* 62(6):822-843.

Thomas, Scott L. 2000. "Ties that Bind: A Social Network Approach to Understanding Student Integration and Persistence." *The Journal of Higher Education* 71(5):591-615.

Tierney, William G., and Suneal Kolluri, eds. 2020. *Relational Sociology and Research on Schools, Colleges, and Universities*. Albany, NY: SUNY Press.

Tinto, Vincent. 1987. *Leaving College: Rethinking the Causes and Cures of Student Attrition*. Chicago, IL: University of Chicago Press.

Tinto, Vincent. 2003. "Learning Better Together: The Impact of Learning Communities on Student Success." *Higher Education Monograph Series* 1(8):1-8.

Tinto, Vincent, and Anne Goodsell. 1994. "Freshman Interest Groups and the First-year Experience: Constructing Student Communities in a Large University." *Journal of The First-Year Experience & Students in Transition* 6(1):7-28.

Turetsky, Kate M., Valerie Purdie-Greenaway, Jonathan E. Cook, James P. Curley, and Geoffrey L. Cohen. 2020. "A Psychological Intervention Strengthens Students' Peer Social Networks and Promotes Persistence in STEM." *Science Advances* 6(45):eaba9221.

Turley, Ruth N. López. 2009. "College Proximity: Mapping Access to Opportunity." *Sociology of Education* 82(2):126-146.

Turner, Ralph H. 1960. "Sponsored and Contest Mobility and the School System." *American Sociological Review* 25(6):855-867.

Tyson, Will, and Josipa Roksa. 2016. "How Schools Structure Opportunity: The Role of Curriculum and Placement in Math Attainment." *Research in Social Stratification and Mobility* 44:124-135.

Valente, Thomas W. 2012. "Network Interventions." *Science* 337(6090):49-53.

Van Duijn, Marijtje AJ, Evelien PH Zeggelink, Mark Huisman, Frans N. Stokman, and Frans W. Wasseur. 2003. "Evolution of Sociology Freshmen into a Friendship Network." *Journal of Mathematical Sociology* 27(2-3):153-191.

Vaquera, Elizabeth, and Grace Kao. 2008. "Do You Like Me As Much As I Like You? Friendship Reciprocity and its Effects on School Outcomes among Adolescents." *Social Science Research* 37(1):55-72.

Von Hippel, Paul T. 2007. "Regression with Missing Ys: An Improved Strategy for Analyzing Multiply Imputed Data." *Sociological Methodology* 37(1):83-117.

Vörös, András, Per Block, and Zsófia Boda. 2019. "Limits to Inferring Status from Friendship Relations." *Social Networks* 59:77-97.

Wasserman, Stanley, and Katherine Faust. 1994. *Social Network Analysis: Methods and Applications.* Cambridge, UK: Cambridge University Press.

Weber, Hannes, Marc Schwenzer, and Steffen Hillmert. 2020. "Homophily in the Formation and Development of Learning Networks among University Students." *Network Science* 8(4):469-491.

Whalen, Donald F., and Mack C. Shelley. 2010. "Academic Success for STEM and Non-STEM Majors." *Journal of STEM Education: Innovations and Research* 11(1):1-16.

Wimmer, Andreas, and Kevin Lewis. 2010. "Beyond and Below Racial Homophily: ERG Models of a Friendship Network Documented on Facebook." *American Journal of Sociology* 116(2):583-642.

Wolniak, Gregory C., Tricia A. Seifert, Eric J. Reed, and Ernest T. Pascarella. 2008. "College Majors and Social Mobility." *Research in Social Stratification and Mobility* 26(2):123-139.

Xie, Yu, Michael Fang, and Kimberlee Shauman. 2015. "STEM Education." *Annual Review of Sociology* 41:331-357.

Xu, Di, Sabrina Solanki, Peter McPartlan, and Brian Sato. 2018. "EASEing Students into College: The Impact of Multidimensional Support for Underprepared Students." *Educational Researcher* 47(7):435-450.

Zumbrunn, Sharon, Courtney McKim, Eric Buhs, and Leslie R. Hawley. 2014. "Support, Belonging, Motivation, and Engagement in the College Classroom: A Mixed Method Study." *Instructional Science* 42:661-684.

APPENDICES

Appendix A. Chapter 1 Descriptive Statistics

Measure	Mean/ prop.	% imputed
Dependent variable		
Application selectivity (most selective based on IPEDS data)	32.19	0
Applied (yes=1)	0.89	0
Main independent variable		
Socioeconomic composite score (-2 to 2)	-0.07	0
Rational action model		
Performance differentials		
11th-grade GPA	2.92	9.9
AP coursework (yes=1)	0.42	5.6
Standardized test score (SAT or equivalent)	987	1.0
Choice differentials		
Informational resources		
Attended program at or taken tour of college campus (yes=1)	0.55	1.8
Searched for college options (yes=1)	0.87	1.9
Talked w/ high school counselor (yes=1)	0.65	1.9
Talked w/ college admission's counselor (yes=1)	0.12	2.1
Took preparatory course for college admission exam (yes=1)	0.48	2.0
College considerations		
Being close to home (very important=1)	0.23	3.7
Cost of attendance (very important=1)	0.65	3.9
Academic quality/reputation (very important=1)	0.78	3.9
Family/friend recommendations (very important=1)	0.22	3.8
Family legacy (very important=1)	0.08	4.0
Degree program (very important=1)	0.77	3.9
Graduate school placement (very important=1)	0.57	4.0
Job placement (very important=1)	0.73	3.9
Play school sports (very important=1)	0.21	3.8
Campus social life / school spirit (very important=1)	0.58	3.6
Status attainment model		
Educational expectations (*level* ranging from 1 <HS to 6 MA+)	4.98	10.1
Educational expectations (*type* based on IPEDS data)	33.18	42.1
Number of applications submitted	2.81	4.6
Controls		
Race/ethnicity (*reference=white*)		3.9
White	0.57	
Black / African American	0.10	
Hispanic / Latinx	0.15	

Asian / Asian American	0.10	
Multiracial / other	0.09	
Gender (*reference=female*)	0.51	<1
School control (*reference=public*)		1.8
Public	0.81	
Catholic	0.12	
Other private	0.07	
School type (*reference=regular*)		4.2
Regular	0.93	
Charter school	0.02	
Special program school	0.03	
Other school type	0.02	
School urbanicity (*reference=city*)		1.8
City	0.29	
Suburb	0.30	
Town	0.12	
Rural	0.29	
Geographic region (*reference=New England*)		1.8
New England	0.04	
Middle Atlantic	0.12	
East North Central	0.21	
West North Central	0.07	
South Atlantic	0.22	
East South Central	0.09	
West South Central	0.09	
Mountain	0.05	
Pacific	0.11	
High school size (enrollment number)	1224	4.2
Percent free / reduced price lunch	0.35	5.5

NOTE: Values may not sum to one due to rounding.
SOURCES: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSLS), 2012, 2013.
U.S. Department of Education, National Center for Education Statistics, Integrated Postsecondary Education Data System (IPEDS), 2011-2012.

Appendix B. Results of the Blinder-Oaxaca Decomposition Analysis of the Gap in College Application Selectivity between Students in the Top and Bottom SES Quintiles

Component	Coef.	Std. Err.	Sig.
Outcome differential			
Low SES (*prediction*)	19.702	(2.08)	***
High SES (*prediction*)	43.641	(.524)	***
Difference	-23.939	(2.11)	***
Endowments effect			
Controls	-1.040	(.663)	
GPA	-2.889	(.445)	***
AP coursework	-0.550	(.200)	**
Standardized tests	-6.934	(.573)	***
Information	-0.849	(.207)	***
Considerations	-0.917	(.228)	***
Expectations: level	-1.037	(.399)	*
Expectations: type	-3.109	(.256)	***
Number of applications	-3.233	(.231)	***
Total	-20.557	(1.09)	***
Coefficients effect			
Controls	-0.784	(2.46)	
GPA	2.232	(4.54)	
AP coursework	2.361	(.863)	**
Standardized tests	-11.811	(4.72)	*
Information	1.090	(2.03)	
Considerations	3.402	(1.79)	†
Expectations: level	4.689	(5.24)	
Expectations: type	-3.640	(1.30)	**
Number of applications	5.314	(.999)	***
Constant	-6.565	(10.48)	
Total	-3.712	(1.38)	**
Interaction			
Controls	-0.195	(1.07)	
GPA	-0.374	(.763)	
AP coursework	-1.057	(.391)	**
Standardized tests	2.470	(.989)	*
Information	0.042	(.341)	
Considerations	1.092	(.383)	**

Expectations: level	-0.619	(.693)	
Expectations: type	0.934	(.338)	**
Number of applications	-1.963	(.384)	***
Total	0.330	(1.87)	

†p<.10; *p<.05; **p<.01;***p<.001

NOTE: Controls include: *race/ethnicity, gender, school control, school type, school urbanicity, geographic region, high school size, and percent free/reduced price lunch*. Information includes: *attended program at or taken tour of college campus, searched for college options, talked w/ high school counselor, talked w/ college admission's counselor, took preparatory course for college admission exam*. Considerations include: *being close to home, cost of attendance, academic quality/reputation, family/friend recommendations, family legacy, degree program, graduate school placement, job placement, play school sports, campus social life/school spirit*.

SOURCES: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSLS), 2012, 2013.

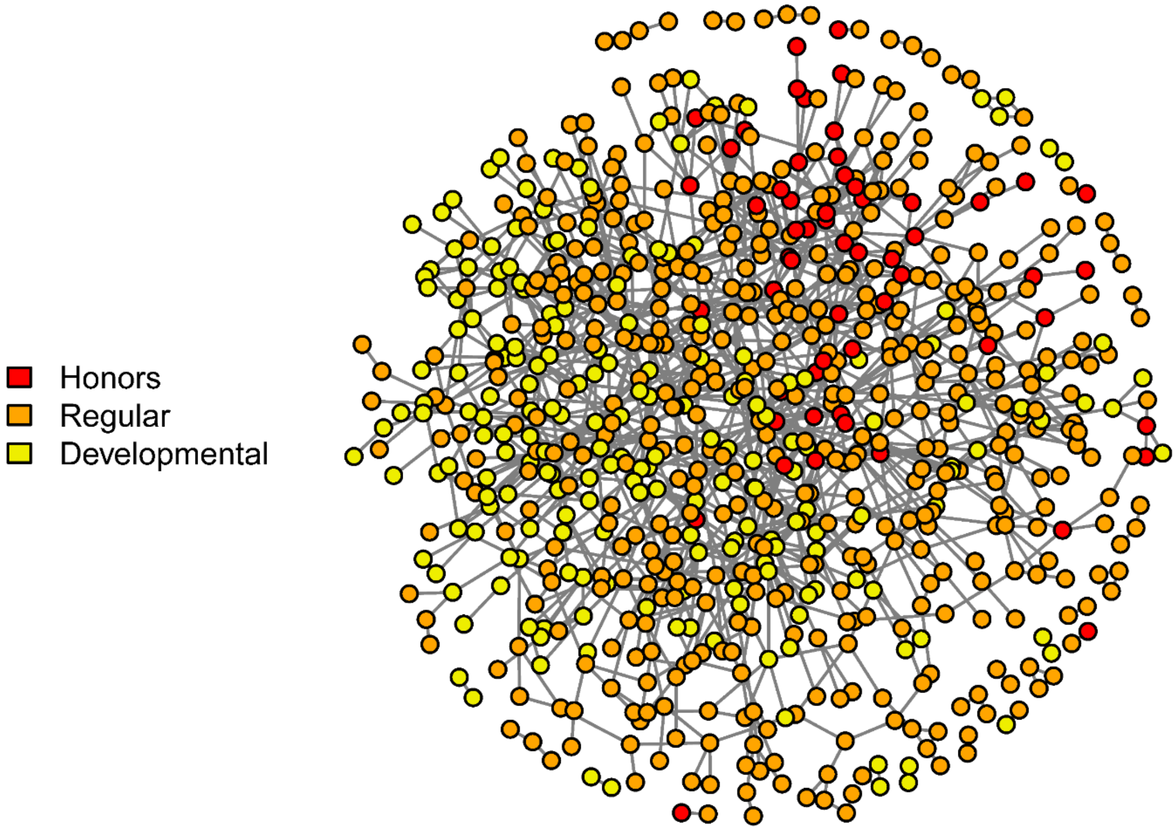
U.S. Department of Education, National Center for Education Statistics, Integrated Postsecondary Education Data System (IPEDS), 2011-2012.

Appendix C. Rating Criteria for Barron's Competitiveness Index 2011-2012

Most Competitive	These colleges require high school rank in the top 10-20% and grade average A to B+. Median freshman test scores at these colleges are generally between 655 and 800 on the SAT and 29 and above on the ACT. In addition, many of these colleges admit only a small percentage of those who apply-usually fewer than one third.
Highly Competitive	Colleges in this group generally look for students with grade averages of B+ to B and accept most of their students from the top 20-35% of the high school class. Median freshman test scores at these colleges generally range from 620 to 654 on the SAT and 27 or 28 on the ACT. These schools generally accept between one third and one half of applicants.
Very Competitive	Colleges in this category generally admit students whose averages are no less than B- and who rank in the top 35-50% of their graduating class. They generally report median freshman test scores in the 573 to 619 range on the SAT and from 24-26 on the ACT. These schools generally accept between one half and three quarters of their applicants.
Competitive	This category is a very broad one, covering colleges that generally have median freshman test scores between 500 and 572 on the SAT and between 21 and 23 on the ACT. Some of these colleges require that students have high school averages of B- or better, although others state a minimum of C+ or C. Generally, these colleges prefer students in the top 50-65% of the graduating class and accept between 75% and 85% of their applicants.
Less Competitive	Included in this category are colleges with median freshman test scores generally below 500 on the SAT and below 21 on the ACT; some colleges that require entrance examinations but do not report median scores; and colleges that admit students with averages generally below C who rank in the top 65% of the graduating class. These colleges usually admit 85% or more of their applicants.
Noncompetitive	The colleges in this category generally only require evidence of graduation from an accredited high school (although they may also require completion of a certain number of high school units.) Some require that entrance examinations be taken for placement purposes only, or only by graduates of unaccredited high schools or only by out-of-state students. Generally, if a college accepts 98% or more of its applicants, it automatically falls in this category. Colleges are also rated as Noncompetitive if they admit all state residents, but have some requirements for nonresidents.

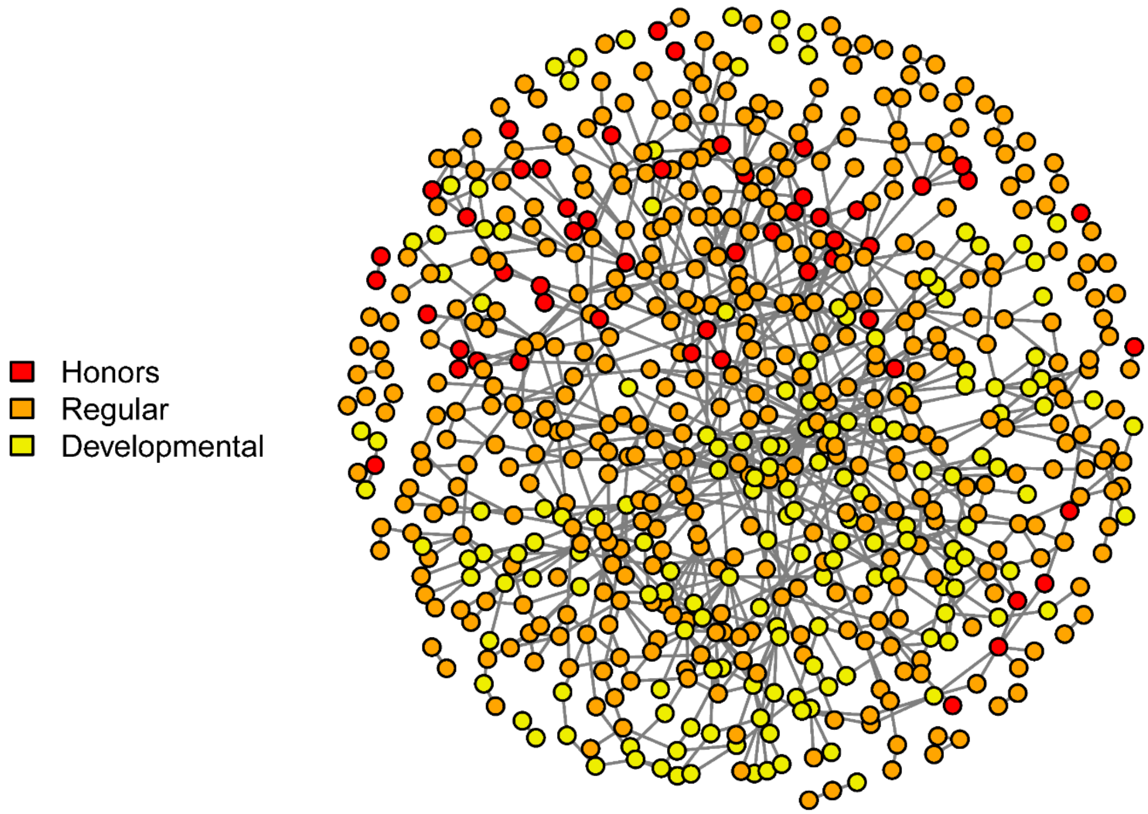
SOURCE*:* Barron's Profiles of American Colleges 2011

Appendix D. Freshman Biological Sciences Friendship Network Shaded by Curricular Group Placement: Fall 2018



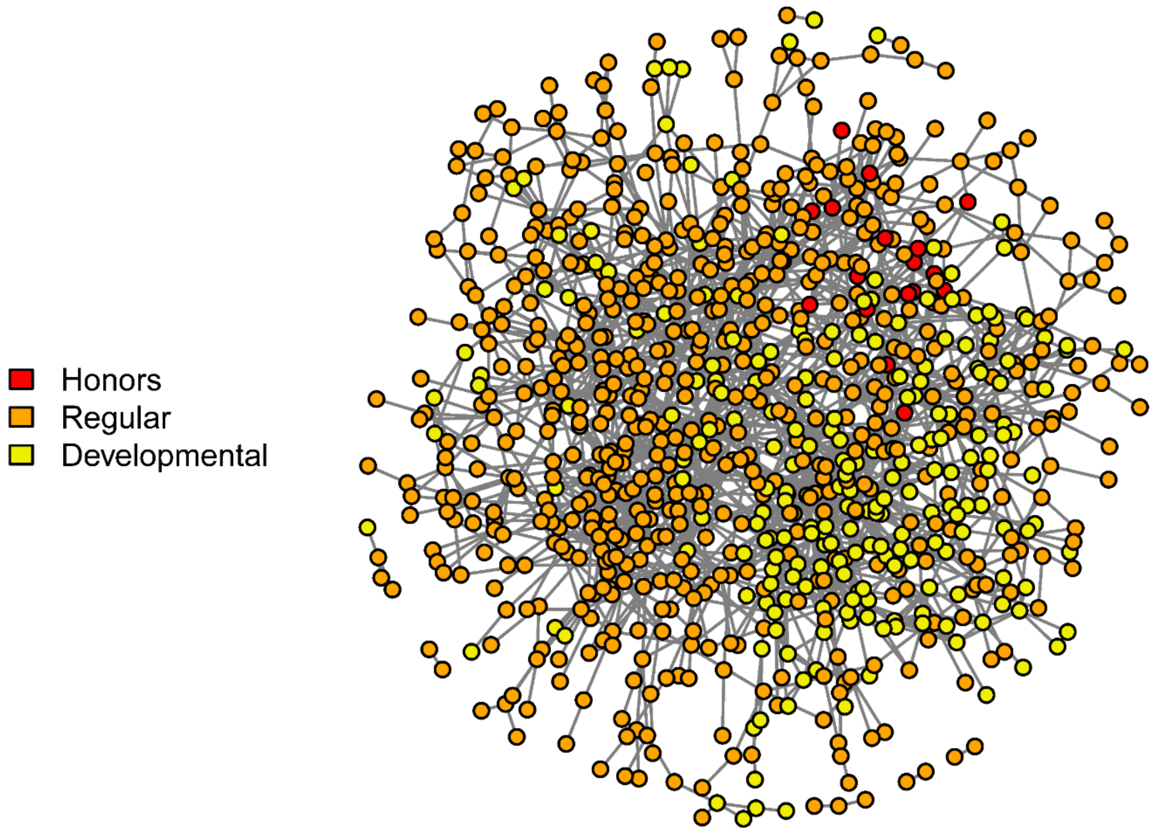
Note: Isolates are not displayed.

Appendix E. Freshman Biological Sciences Friendship Network Shaded by Curricular Group Placement: Spring 2019



Note: Isolates are not displayed.

Appendix F. Freshman Biological Sciences Friendship Network Shaded by Curricular Group Placement: Spring 2020



Note: Isolates are not displayed.

Appendix G. Expected Relative Importance of Effects on Friend Selection: Full Estimates

Effect	2018	2019
Outdegree (density)	.209	.186
Reciprocity	.064	.111
Transitive triplets	.030	.054
Transitive reciprocated triplets	.014	.019
In-degree popularity (sqrt)	.022	.011
In-degree activity (sqrt)	.085	.114
Out-degree activity (sqrt)	.114	.104
Shared coursework	.065	.067
Curricular group (*alter*)	.018	.008
Curricular group (*ego*)	.002	.003
Curricular group (*same*)	.027	.039
Academic performance (*alter*)	.029	.023
Academic performance (*ego*)	.011	.002
Academic performance (*similarity*)	.014	.024
Race/ethnicity (*same*)	.031	.050
Black/African American (*alter*) 1=yes	.002	.0002
Black/African American (*ego*) 1=yes	.003	.001
Hispanic/Latinx (*alter*) 1=yes	.012	.015
Hispanic/Latinx (*ego*) 1=yes	.003	.006
Asian/Asian American (*alter*) 1=yes	.002	.019
Asian/Asian American (*ego*) 1=yes	.001	.002
Gender (*alter*) 1=female	.023	.015
Gender (*ego*) 1=female	.019	.005
Gender (*same*)	.030	.036
First-generation student status (*alter*) 1=yes	.007	.006
First-generation student status (*ego*) 1=yes	.013	.004
First-generation student status (*same*)	.008	.001
Low-income status (*alter*) 1=yes	.006	.010
Low-income status (*ego*) 1=yes	.003	.004
Low-income status (*same*)	.005	.002
International student status (*same*)	.037	.008
High school (*same*)	.007	.010
Housing (*same*)	.050	.030
High school GPA (*similarity*)	.009	.002
SAT score (*similarity*)	.009	.010
Same-major ties (*ego*)	.019	.004