

2020

STUDENT COLLABORATIONS IN INTRODUCTORY STATISTICS COURSES; A NETWORK STUDY

Isabel S. Nowinowski
University of Rhode Island, inowinowski@gmail.com

Follow this and additional works at: <https://digitalcommons.uri.edu/theses>

Terms of Use

All rights reserved under copyright.

Recommended Citation

Nowinowski, Isabel S., "STUDENT COLLABORATIONS IN INTRODUCTORY STATISTICS COURSES; A NETWORK STUDY" (2020). *Open Access Master's Theses*. Paper 1843.
<https://digitalcommons.uri.edu/theses/1843>

This Thesis is brought to you by the University of Rhode Island. It has been accepted for inclusion in Open Access Master's Theses by an authorized administrator of DigitalCommons@URI. For more information, please contact digitalcommons-group@uri.edu. For permission to reuse copyrighted content, contact the author directly.

STUDENT COLLABORATIONS IN INTRODUCTORY STATISTICS
COURSES; A NETWORK STUDY

BY

ISABEL S. NOWINOWSKI

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE
IN
STATISTICS

UNIVERSITY OF RHODE ISLAND

2020

MASTER OF SCIENCE THESIS
OF
ISABEL S. NOWINOWSKI

APPROVED:

Thesis Committee:

Major Professor Natallia Katenka
 Jing Wu
 Ashley Buchanan
 Nasser H. Zawia

DEAN OF THE GRADUATE SCHOOL

UNIVERSITY OF RHODE ISLAND

2020

ABSTRACT

Generation Z, also known as iGen, (individuals born between the mid-1990s and early 2010s), characterized as tech-savvy, independent, and visual, is beginning to graduate college and enter the workforce. While significant research effort has focused on understanding the learning preferences of the preceding Millennial generation (individuals born between the early 1980s and mid-1990s), less is known about the way technology has influenced the educational expectations and learning preferences of Generation Z. A deeper and broader understanding of the way this generation learns would allow universities to modify and enhance course structures and teaching methodologies to suit this incoming generation of students better. In this thesis, we used secondary survey and performance data collected in all undergraduate statistics courses at the University of Rhode Island in Spring 2017 to distinguish the learning preferences of this new generation. Data collected contained student demographics, study habits, learning preferences, pre- and post-course attitudes, stress levels, and the names of student collaborators.

The goals of this study were to understand the main drivers of collaboration among Generation Z students taking introductory statistics courses and to identify differences in demographics, study habits, learning preferences, performance, and attitudes towards statistics between collaborators and independent learners. We used Network and Classical methods to characterize the network of students who collaborate and to distinguish collaborators from independent learners. Of the two courses explored as part of this study, the focus was on data collected in course, Introductory Biostatistics (STA 307), given the high response rate and collaborative structure of the network.

Descriptive statistics suggest that students enrolled in the same major are more likely to connect than students in disparate majors, perhaps because they

have had opportunities to connect in other courses. Exponential Random Graph Models (ERGMs) were used to gain insight into and make inferences about the effects of endogenous and exogenous factors on the determinants of ties within a network. ERGMs fitted to the network of student collaborators indicate that students are more likely to collaborate with classmates in their recitation section and with students who share similar characteristics, namely other athletes, students living in the same type of housing, in-state students, and out-of-state students. Male students are also more likely to collaborate with other male students than females are to collaborate with one another. The significance of the geometrically weighted edge-wise shared partnerships (GWESP) statistic in the model suggests the presence of transitivity, meaning that there is a significant proportion of students studying in groups of three.

The results of the comparison between independent learners and student collaborators show that collaborators are more likely to complete practice exams. This is expected as students may be working through practice exams with their peers. Independent learners value the instructor's knowledge of the material to their learning, likely because they are more reliant on the instructor for understanding. At the same time, collaborators lean on their peers for knowledge sharing and support. Evidence does not suggest that student collaborators outperform independent learners in STA 307. While independent learners do not appear to be at risk of underperforming relative to collaborators, partnership with other students provides a natural support system, giving students an additional learning tool by which to learn.

ACKNOWLEDGMENTS

I would like to first and foremost thank Natallia Katenka for her unending patience and guidance throughout this whole process. I would also like to thank my committee members, Jing Wu and Ashley Buchanan, for their feedback and support. I want to acknowledge Mustafa Kulenovic for his willingness to serve as my defense chair, and for his guidance throughout my tenure at the University of Rhode Island. A special thanks to Kaitlin Dio for the collection and cleaning of the raw data sets. I want to thank Marjana Catanzaro for her collaboration and encouragement over the last five years. I'd also like to recognize Scott Capista, Eric Nowinowski, Kristina Lopes, and Vanessa Martins. I couldn't have done this without you. Lastly, thank you to the rest of the Department of Computer Science and Statistics for a memorable experience.

TABLE OF CONTENTS

ABSTRACT	ii
ACKNOWLEDGMENTS	iv
TABLE OF CONTENTS	v
LIST OF FIGURES	vii
LIST OF TABLES	ix
CHAPTER	
1 Introduction	1
1.1 Background	4
1.2 Data Description	8
1.3 Network Application	11
2 Methods	14
2.1 Network Methods	14
2.2 ERGM	17
2.3 Tests of Significance	23
3 Results	25
3.1 Network Data Description	25
3.2 Assortativity	28
3.3 ERGMs	31
3.4 Comparison of Collaborators and Independent Learners	40
4 Main Findings	48

	Page
4.1 Limitations	50
4.2 Future Research	51
LIST OF REFERENCES	53
APPENDIX	
BIBLIOGRAPHY	57

LIST OF FIGURES

Figure		Page
1.1	The pie chart represents the proportion of students in each population by major, STA 307 is on the left and STA 308 on the right. Pharmacy students constitute 77.0% of the STA 307 population. On the right, Biology students make up 50.0% of the STA 308 population and Medicine/Pre-Medicine students represent the second largest group (20.6%).	10
3.1	The STA 307 student collaboration network is represented in the top row; the full network is on the left, the connected network in the center, and the LCC on the right (LCC). There is a sizeable connected component in the network of STA 307 students. Visuals reveal that a greater proportion of student responders enrolled in lecture with Instructor 2. The network of STA 308 student collaborators is represented in row 2, on the left is the full network and the connected network on the right. The full network shows that many students are working independently, while numerous isolated groups of students characterize the connected graph.	26
3.2	Goodness-of-fit plots measuring the quality of ERGMs fitted to the full network, connected, and LCC networks of STA 307 students. We expect the observed data in bold to fall between the 10th and 90th percentiles obtained from the simulated networks. The full network shows the best fit while the LCC shows the poorest. The full network model underestimates the shortest distance between students, and does not perfectly capture the number of neighbors shared between two connected students or the number of collaborators per student.	38
3.3	STA 308 Full Network ERGM degree goodness-of-fit plot shows that the model fits the number of collaborators per student relatively well compared to the actual network.	40
.1	A greater proportion of students in STA 307 responded that they sometimes prefer to study in groups than in STA 308. . . .	55
.2	The majority of students in both STA 307 and STA 308 responded that they sometimes prefer to study in the library. . . .	55

.3	A greater proportion of students in STA 307 responded that they always complete practice tests before exams than in STA 308.	56
----	--	----

LIST OF TABLES

Table		Page
1.1	Description of Student Demographics, Study Habits, Learning Preferences, and Attitudes Collected in Introductory Statistics Courses at the University of Rhode Island in Spring 2017	7
1.2	Description of SATS-36 Attitudes: Affect, Cognitive Competence, Value, Difficulty, Interest, and Effort	8
1.3	STA 307 and STA 308: Course Enrollment and Survey Participants	9
1.4	STA 307 and STA 308 Student Athlete Status, Residence, Gender, Age, Stress Level, and Course Performance	11
1.5	Pre-Course Attitudes Towards Statistics	11
2.1	Network Descriptive Measures: Density, Degree, Assortativity, Transitivity, Local Clustering Coefficient, Local Betweenness Centrality	17
3.1	Network Description	27
3.2	STA 307 Recitation Day Assortativity	30
3.3	STA 308 Recitation Day Assortativity	31
3.4	STA 307 ERGM Results: Full Network, Connected, and LCC	34
3.5	STA 308 Full Network ERGM Results	40
3.6	Differences in Attitudes Towards Statistics between Collaborators and Independent Learners in STA 307	43
3.7	Differences in Attitudes Towards Statistics between Collaborators and Independent Learners in STA 308	44
3.8	STA 307 Student Learning Preferences and Demographics between Collaborators and Independent Learners	46
3.9	STA 308 Student Learning Preferences and Demographics between Collaborators and Independent Learners	47

CHAPTER 1

Introduction

As the youngest Millennials (individuals born between the early 1980s and mid-1990s) are transitioning out of Universities, students of Generation Z, also known as iGen (individuals born between the mid-1990s and early 2010s) are beginning to graduate and enter the workforce (Seemiller and Grace, 2016). But who are the students of Generation Z, and what characteristics best define these students, allowing us to understand their learning preferences and study habits better? To gain insight into this new generation, we must first discuss the advancements in technology made in the last two decades and highlight the influence of technology on both Millennials and Generation Z. Older Millennials were witness to the development of new technologies, observing the transition from dial-up to high-speed Internet, the introduction of the first smartphone, and the advent of social media sites like MySpace, Facebook, Instagram, and Snapchat. These advances empowered users to ingest and share information rapidly (Strauss and Howe, 2003). Unlike the Millennial generation, Generation Z was born into a world where the rapid development of technology and the availability of smartphones by which to share information was both commonplace and expected (Shatto and Erwin, 2016; Seemiller and Grace, 2016).

How Generation Z and Millennials were introduced to technology differs and has had a considerable impact on the learning preferences and study habits of each generation. Several characteristics differentiate Generation Z from the Millennial Generation. Howe and Strauss (2003) describe the Millennial Generation as special, sheltered, confident, team-oriented, conventional, pressured, and achieving (Strauss and Howe, 2003). Millennials are special in that they feel their existence

is vital to the nation. They take a twist to traditional social norms, have developed strong peer connections, and are pressured to excel.

Millennials desire to be recognized for their hard work, prefer to work in groups rather than independently, require encouragement but enjoy working out a problem before asking for help, and seek course structures employing the most current technology to augment lecture (Strauss and Howe, 2003). Their successors, Generation Z, have been described as pragmatic, individualistic, cautious, open-minded, heavily dependent on technology, and requiring immediacy (Shatto and Erwin, 2016; Seemiller and Grace, 2016; Chicca and Shellenbarger, 2018; Igel and Urquhart, 2012; Loveland, 2017; Rickes, 2016; Spears et al., 2015; Turner, 2015; Twenge, 2017). These qualities contribute to a learning style where students prefer to work out problems independently, at their own pace, only after seeking collaboration with peers (Seemiller and Grace, 2016). Seemiller and Grace (2017) believe that this self-directed but not isolated learning style is a by-product of participation in independent learning assignments in primary and secondary school. Generation Z students also prefer logic-based approaches that allow students to learn from trial and error, and experiential learning forms offering real-world experience (Seemiller and Grace, 2016). Despite numerous character differences, the Millennial generation and Generation Z share a common desire for the continued integration of technology in their educational experiences and collaboration with peers as a means of learning.

Significant research effort focused on understanding the learning preferences of the preceding Millennial generation, including a study performed by Toothaker and Taliaferro (2017), which explored the learning styles and preferences of Millennial nursing students pursuing a Bachelor's in Nursing. Interview and analysis of

13 Millennial students showed that learners sought a hands-on classroom experience and additional guidance from their instructors to gain confidence in practicing their nursing skills. Findings also revealed that many students were disengaged during lectures and did not feel comfortable participating in classroom activities. Toothaker and Taliaferro (2017) prescribe a blend of traditional as well as active methods to enhance student engagement among Millennial nursing students. In a second study, Goldman and Martin (2016) used qualitative methods to examine how technology and social contexts influence the teaching techniques to which Millennials are most likely to respond. To improve student performance and augment learning, researchers suggest instructors devise assignments that allow for student creativity as opposed to scripted tasks, promoting students to become co-creators of ideas.

Less is known about how technology has shaped the educational expectations and learning preferences of Generation Z students. Our research was motivated by the desire to gain additional insight into the study habits and preferences of this generation, focusing on the way these students collaborate. Findings would allow universities to enhance course structures, teaching methodologies, and classroom activities to suit this incoming generation of students better. Aforementioned, Millennials and Generation Z students share an affinity for group work as part of their learning experience but collaborate in different manners. Millennials prefer to work in groups while Generation Z students are partial to working out problems independently and only after, talking through results with others (Seemiller and Grace, 2016). To investigate collaboration among Generation Z students, we explored secondary survey and performance data collected in two similar undergraduate statistics courses at the University of Rhode Island, in Spring 2017. Introductory Biostatistics 307 (STA 307) and Introductory Statistics 308

(STA 308) cover related topics, the key difference being that STA 307 focuses on health science applications while STA 308 provides a general overview of statistical topics. It should be noted that while the focus of this study was on Generation Z students, there are several students in each course formally classified as part of the Millennial generation.

We used both Network and Classical methodologies to understand how Generation Z students in two similar undergraduate statistics courses collaborate. We focused on the following two research questions: (1) What are the drivers of student collaborations in undergraduate statistics courses? (2) What characteristics differentiate collaborators from independent learners among students enrolled in introductory statistics courses? In the following sections, we describe in detail the data used, methods, results, and future research.

1.1 Background

The secondary survey and performance data used in this project were collected from two undergraduate statistics courses, STA 307: Introductory Biostatistics and STA 308: Introductory Statistics. Both courses covered the basics of probability, the central limit theorem, one and two sample inference, correlation, and regression, among other topics. Data were collected in two additional undergraduate statistics courses at the University of Rhode Island, omitted from this study due to a low response rate. Each undergraduate course, STA 307 and STA 308, consisted of multiple lecture and recitation sections. Recitations led by teaching assistants (TAs) provided students with the opportunity to ask questions and work through problem sets in a setting with a smaller instructor to student ratio.

Data collection consisted of two waves, the first at the beginning of the semester and the second at the end. Introductory and pre-course attitude surveys were administered during wave 1, and exit and post-course attitude surveys

in wave 2. Data collected included student attitudes, study habits, learning preferences, and demographics.

The names of collaborators reported by students were captured in the exit survey. Partnerships between students were used to both generate the network of student collaborators and to distinguish student collaborators from independent learners. Students listed the first and last names of peers in the course with whom they partnered with over the semester. Two students were linked as collaborators if, at minimum, one student reported the other on the exit survey.

Also, demographic information collected through the introductory survey included whether the student was living on-campus or off-campus, athlete status, self-identified gender, and in-state or out-of-state residence. Students were asked to indicate whether or not they felt stressed, indicating yes or no. Lastly, the introductory survey asked students to provide the lecture and recitation sections they attended (Table 1.1).

On both the introductory and exit surveys, students rated on a five-point scale (1 = never and 5 = always) their study habits and learning preferences to study in the library, at home, in groups, and alone. Students also rated the frequency at which they completed all homework assignments and practice tests. Students rated learning aids like weekly quizzes, note-taking, and homework, on an eight-point scale with 1 = least beneficial and 8 = most useful. Finally, course grades were not captured in the surveys administered to students but were provided by instructors at the end of the semester.

The Survey of Attitudes Towards Statistics (SATS-36) was used to capture student attitudes towards statistics and mathematical background. The SATS-36 is a standard tool developed by Candace Schau, a statistics professor of over 25 years, to understand better the effect of student attitudes on teaching and learning.

The SATS-36 instrument is comprised of 36 Likert-type questions used to evaluate students' cognitive competence, affect, value, difficulty, effort, and interest in statistics (Table 1.2). Nolan et al. (2012) compared the validity and reliability measures of several surveys that focus on capturing attitudes. Through the evaluation of published evidence, they found that the SATS-36 showed the strongest construct validity and internal consistency. The internal consistency of student attitudes captured in introductory statistics courses at the University of Rhode Island was assessed by Kaitlin Dio (2018) using the Omega value. Computation of the Omega values across pre-course and post-course attitudes revealed that all attitudes except pre-test difficulty measured the appropriate constructs, falling above the acceptable threshold (above 0.7). Omega values ranged between 0.77 and 0.93 for all measures except pre-test difficulty (0.64).

Many studies have reported results of the SATS-36 survey across a variety of student populations. Most notably, Schau and Emmioglu (2012) performed a nationwide study of 2200 students, finding that, on average, students entered courses with positive effort, cognitive competence, value, and interest, and neutral difficulty. At the end of the semester, cognitive competence, affect, and difficulty attitudes remained about the same while a decrease in value, interest and effort was observed.

Table 1.1: Description of Student Demographics, Study Habits, Learning Preferences, and Attitudes Collected in Introductory Statistics Courses at the University of Rhode Island in Spring 2017

Variable Name	Description	Source
Lecture/ Recitation Section	Lecture instructor ID and recitation day section	Introductory
Housing	On-campus/Off-campus	Introductory
Athlete Status	Yes/No	Introductory
Self-Identified Gender	Male/Female	Introductory
In-State/ Out-of-State	In-State: Rhode Island Resident Out-of-State: Non-Rhode Island Resident	Introductory
Stress	Yes/No	Introductory
Timely Email	How beneficial is timely email response from instructors to your learning? (1 = least beneficial, 8 = most-beneficial)	Introductory
Instructor Knowledge	How beneficial is the instructors in-depth knowledge of material to you as a learner? (1 = least beneficial, 7 = most beneficial)	Introductory
Names of Collaborators	List any students you have collaborated with from this class throughout the semester	Exit
Group Study	Did you prefer to study in groups? (1 = Never, 5 = Always)	Exit
Library Study	Did you prefer to study in a library? (1 = Never, 5 = Always)	Exit
Practice Test Completion	Did you complete practice tests before exams? (1 = Never, 5 = Always)	Exit
Age	Student age	Attitudes
Past Math	How well did you do in mathematics courses you have taken in the past? (1 = very poorly, 7 = very well)	Attitudes
Math Proficiency	How good are you at math? (1 = very poor, 7 = very good)	Attitudes
Final Grade Number	Numerical final course grade	Instructor

Table 1.2: Description of SATS-36 Attitudes: Affect, Cognitive Competence, Value, Difficulty, Interest, and Effort

Attitude	Description	Example Question
Affect	Feelings concerning statistics	I will like statistics
Cognitive Competence	Attitude about their intellectual knowledge/skills applied to statistics	I will understand statistics equations
Value	Attitudes about the usefulness, relevance, and worth of statistics in personal and professional life	I use statistics in my everyday life
Difficulty (Easiness)	Attitude about the easiness of statistics	Statistics formulas are easy to understand
Interest	Level of individual interest in statistics	I am interested in learning statistics
Effort	Amount of work the student expends to learn statistics	I plan to work hard in my course

In the next section, we describe the inclusion criteria for students to be included in this study, as well as the attitudes, study habits, learning preferences, and demographics of these students.

1.2 Data Description

In STA 307, most students that completed wave 1 surveys also completed wave 2. In STA 308, 41 of the 102 students that consented to the use of their data completed wave 1 surveys but did not complete wave 2. In order to preserve the number of student collaborations, these 41 students were included in the study in addition to the 61 that completed both wave one and wave two surveys.

The focus of this research is on STA 307, yielding both a high survey response rate and a more tight-knit student collaborative network than observed in STA 308. Of the 153 students enrolled in STA 307, 128 students provided consent, and 100 completed all four surveys. Of the 248 students enrolled in STA 308, 170 provided consent, and 102 of the 170 students completed the pre-course surveys

(Table 1.3). Of all students enrolled in STA 307 and STA 308, 65.4% and 41.1%, respectively, are represented in this study.

Table 1.3: STA 307 and STA 308: Course Enrollment and Survey Participants

	STA 307	STA 308
Total Enrollment	153	248
Consenting Totals	128 (83.7%)	170 (68.5%)
Survey Participants	100 (65.4%)	102 (41.1%)

Some students meeting the criteria to be included in this study were excluded for one of two reasons. They either collaborated with students that did not consent to the use of their data or had a high proportion of missing data. The networks are therefore, not reflective of all reported collaborations in each course, and it should be noted that inclusion of these students could alter outcomes. There were six students in STA 307 that were excluded for these reasons and 21 in STA 308.

Students in both courses were predominantly between 18 and 20 years old, making up 92.0% of students in STA 307 and 83.3% of students in STA 308 (Table 1.4). On average, students in STA 307 academically outperformed students in STA 308. Students in STA 307 achieved an average grade of 92.6 and median of 95.0. Students in STA 308 achieved an average grade of 90.1 and median of 90.5.

The majors in which students in STA 307 and STA 308 enrolled differentiate the two populations (Figure 1.1). Pharmacy students in STA 307 constitute 77.0% of all students; the remainder of students were studying Medicine/Pre-Medicine, Biology, Arts/Humanities, Business, Sociology/Social Work, and Mathematics. Of STA 308 students, 50.0% were studying Biology, and 20.6% studied Medicine/Pre-Medicine (Figure 1.1). The other students spanned across the fields of Business, Chemistry, Engineering, Arts/Humanities, Mathematics, Psychology, and Economics.

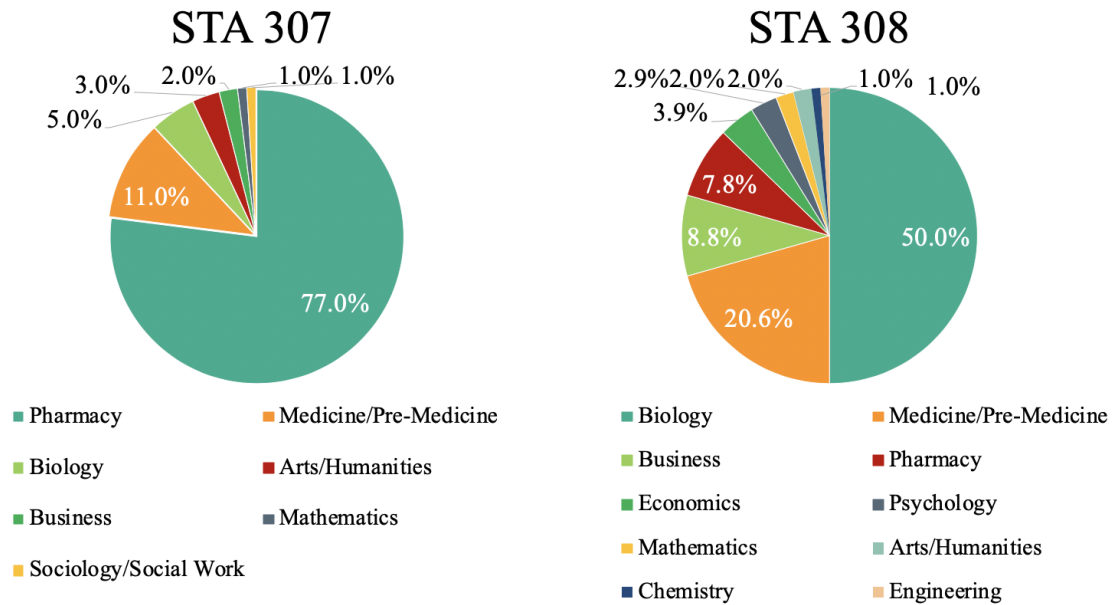


Figure 1.1: The pie chart represents the proportion of students in each population by major, STA 307 is on the left and STA 308 on the right. Pharmacy students constitute 77.0% of the STA 307 population. On the right, Biology students make up 50.0% of the STA 308 population and Medicine/Pre-Medicine students represent the second largest group (20.6%).

In STA 307, 67.0% of students and 77.5% of students in STA 308 reported feeling stressed. STA 307 consisted of 23.0% athletes, 30.0% in-state students, and 76.0% on-campus residents, whereas 16.7% of students in STA 308 were athletes, 39.2% were in-state, and 63.7% lived on campus. In addition, 28.0% of students in STA 307 and 32.4% of students in STA 308 self-identified as male. Mean and standard deviation attitudes towards statistics shown in Table 1.5 are similar between students enrolled in STA 307 and STA 308.

Table 1.4: STA 307 and STA 308 Student Athlete Status, Residence, Gender, Age, Stress Level, and Course Performance

Attribute	STA 307 %	STA 308 %
Athlete	23.0%	16.7%
In-State	30.0%	39.2%
On Campus	76.0%	63.7%
Male	28.0%	32.4%
Age 18-20	92.0%	83.3%
Pre-Course Stress	67.0%	77.5%
Final Grade	Mean: 92.6 Median: 95.0 StDev: 8.4	Mean: 90.1 Median: 90.5 StDev: 8.3

Table 1.5: Pre-Course Attitudes Towards Statistics

Attitude	STA 307		STA 308	
	Mean	StDev	Mean	StDev
Cognitive Competence	5.54	0.95	5.54	1.02
Affect	5.00	1.01	4.83	1.14
Effort	6.58	0.64	6.51	0.58
Difficulty	4.04	0.58	4.04	0.64
Interest	4.78	1.33	4.49	1.22
Value	5.27	1.03	5.15	1.03

1.3 Network Application

Multiple studies have been published that are similar in spirit to ours, using Classical and Network methods to investigate factors associated with student connections and interactions. In the following section, we walk through the similarities and differences between three such studies.

In the first study, Brewe et al. (2012) utilized social network analysis, specifi-

cally hierarchical multiple regression models to investigate the factors contributing to participation in a Physics Learning Center at Florida International University. Similar to the data collected in our study, Brewe et al. (2012) surveyed undergraduate students enrolled in one or more Physics courses, asking students to share the names of students with whom they worked with in the center (Brewe et al., 2012). The similarities end there; Brewe et al. employed hierarchical multiple regression modeling to predict the centrality of students in the Physics Learning Center, showing that centrality was independent of gender and ethnicity. Findings also revealed that students who frequented the center were more likely to be central participants of the community. The focus of our study differs in that we did not seek to identify central participants in the network, but rather, the factors associated with student collaboration.

In a second study performed at the University of Melbourne, Australia, Gallagher and Robins (2015) sought to address the cultural and ethnic influences on interlocution among a cohort of English for Academic Purposes (EAP) students using ERGMs. The study most resembles ours in methodology and data collection. Researchers asked 75 students enrolled in a mandatory EAP course to cite the names of students with whom they had engaged in the two weeks prior. Researchers fit an ERGM to understand the drivers of communication between participants better. Gallagher and Robins (2015) found that students from different cultures who engaged in conversation with one another were often introduced to each other by a mutual third party. Additionally, students that shared a classroom were more likely to interact than students that did not.

We apply similar methods in our research to those observed in the large scale study performed by An and Vanderweele (2019) in which they used descriptive statistics and ERGMs to measure the treatment diffusion of smoking prevention

and cessation in approximately four-thousand students. Each of the 90 participating classes was assigned to one of four treatment types: (1) Students did not receive intervention (2) Students were selected at random to receive the intervention (3) Central students most likely to spread information received intervention (4) Groups of friends received treatment. Treated students received topical brochures and attended a workshop on smoking prevention. Afterward, students were asked to report up to six students with whom they had shared the brochure or discussed contents of the workshop. Results from ERGMs revealed that popular and high performing, male students, are the most likely to spread smoking prevention brochures. While we did not explore the diffusion of information in our study, in the future, we should consider extending our research to measure the dissemination of knowledge among students.

CHAPTER 2

Methods

2.1 Network Methods

In this study, we were particularly interested in exploring the factors associated with student collaboration, in two undergraduate statistics courses. We focused on the influence of similar study habits, learning preferences, and attitudes towards statistics on student connections. Using the data in which students reported the names of collaborators, we created a collaboration network of students that connected and those that did not. In this section, we formally introduce network graphs, and review both descriptive network statistics and Exponential Random Graph Models (ERGMs).

Networks are complex systems representing the relational structure of data. Network analysis, in its simplest form, is founded on the concept of a network graph, $G = (V, E)$. Graph G consists of vertices (V), also referred to as nodes, actors, or members and edges (E) regarded as links or ties (Kolaczyk, 2009; Kolaczyk and Csárdi, 2014). Unordered pairs u, v make up the elements of the edges, E , such that distinct vertices $u, v \in V$. A dyad is a pair of nodes, while a triad is a set of three nodes, both of which can be linked or unlinked (Kolaczyk and Csárdi, 2014). Networks are either undirected or directed; undirected graphs imply that there is no direction between the nodes that link an edge. In directed graphs, edges between linked nodes in a graph G are directed from vertex v to u , u to v , or are mutual. In this study, we assume the graph to be undirected. A simple example of a network is represented by two students collaborating on a homework assignment together; each student represented as a node connected by an edge.

Visualization of the network allows us to initially evaluate the network of stu-

dent collaborators across STA 307 and STA 308. In addition to visualizations, descriptive measures help us to understand the structure of the network and relationships between nodes. In this study, we used six network descriptive statistics to describe the system of student collaborators. The six statistics include density, degree, assortativity, transitivity, local clustering coefficient, and local betweenness centrality (Kolaczyk and Csárdi, 2014).

The density describes the number of observed connections between students relative to all possible connections (Kolaczyk and Csárdi, 2014). Within an undirected graph G , the metric is expressed as

$$den(G) = \frac{|E_G|}{|V_G|(|V_G| - 1)/2} \quad (2.1)$$

The degree d_v of a vertex v in a graph $G = (V, E)$ is the number of peers a student collaborated with. The degree distribution is defined as $(f_d)_{d \geq 0}$ where f_d is a fraction of vertices $v \in V$.

Assortativity is a measure of network mixing patterns, specifically the selective linking among vertices based on a particular attribute (Table 2.1). In the instance of a categorical vertex attribute in graph G , each vertex is designated one of M categories. The measure is given as

$$r_a = \frac{\sum_i f_{ii} - \sum_i f_{i+} f_{+i}}{1 - \sum_i f_{i+} f_{+i}} \quad (2.2)$$

where i and j are categories of M . The variable f_{ij} is the fraction of edges in G joining a vertex with attribute i to a vertex with attribute j . In matrix f , f_{i+} and f_{+i} represent the i th marginal row and column sums. Assortativity takes a value between -1 and 1, and is interpreted much like the pearson correlation coefficient. Assortativity is applicable to continuous variables as well, as described by

$$r = \frac{\sum_{x,y} xy(f_{xy} - f_{x+}f_{+y})}{\sigma_x \sigma_y} \quad (2.3)$$

where (x, y) are all observed unique pairs. In the continuous case, σ_x and σ_y are the standard deviations associated to the distributions of f_{x+} and f_{+y} . The global degree assortativity summarizes the degree-degree correlation (Kolaczyk and Csárdi, 2014).

Transitivity quantifies the occurrence of connected triples closing to form triangles within a network, given as

$$clT(G) = \frac{3\tau_{\Delta}(G)}{\tau_3(G)} \quad (2.4)$$

Here, $3\tau_{\Delta}$ is the number of triangles in graph G and $\tau_3(G)$ is the number of connected triples. In the network of student collaborators, transitivity translates to the frequency of students working in groups of three. The local measure of transitivity is also referred to as the local clustering coefficient. This measure summarizes the extent to which each node is a complete graph, and is defined as

$$clT(v) = \frac{\tau_{\Delta}(v)}{\tau_3(v)} \quad (2.5)$$

where v is the vertex of interest. The local betweenness centrality measures the extent to which a vertex is positioned between other pairs of vertices and is denoted as follows

$$c_B(v) = \sum_{s \neq t \neq v \in V} \frac{\sigma(s, t|v)}{\sigma(s, t)} \quad (2.6)$$

In this instance, $\sigma(s, t|v)$ is the total number of shortest paths between s and t passing through v . Then, $\sigma(s, t)$ is the total number of shortest paths between s and t .

Table 2.1: Network Descriptive Measures: Density, Degree, Assortativity, Transitivity, Local Clustering Coefficient, Local Betweenness Centrality

Measure	Description
Density	The number of observed collaborations between students relative to all possible collaborations
Degree	The number of collaborators a student partners with
Assortativity	Selective linking among students with similar attributes
Transitivity	The proportion of triangles relative to all triples in the network
Local Clustering Coefficient	The extent to which each student connects to their neighbors
Local Betweenness Centrality	The extent to which each student is positioned in between two other students

2.2 ERGM

In this section, we describe in detail the classical ERGM, focusing on the analytical form, dependence among the edges, estimation, and limitations (Cranmer and Desmarais, 2011).

The fully specified class of ERGMs was proposed by Wasserman and Pattison (1996) and has since evolved through the work of Park and Newman (2004), Snijders et al. (2006), among others. (Park and Newman, 2004; Wasserman and Pattison, 1996; Snijders et al., 2006). Exponential Random Graph Models (ERGMs) offer a flexible way to gain insight and make inferences about the effects of endogenous and exogenous factors on the determinants of ties within a network (Cranmer et al., 2012). ERGMs are used to model the prominence and significance of structural dependencies and measure the influ-

ence of node and edge attributes on the structure of the network. ERGMs address the problem encountered in applying classical methods by accounting for the dependency between ties upon covariates and other ties present in the network (Cranmer and Desmarais, 2011).

In a standard regression model, a series of observations y , are drawn from a univariate distribution where independence is assumed between all possible y values. In the context of an ERGM, independence is not assumed between all possible values of y . In an ERGM, a single observation y consists of the entire network of ties drawn from a multivariate distribution. The intent is to select features of the observed network that differentiate it from a random draw of all other possible networks with the same number of nodes (Cranmer et al., 2012). The researcher chooses features like network structures and node covariates to test specific hypotheses about how certain structures or attributes influence tie formation. In this study, we considered the influence of shared student attributes on the formation of a tie between students. For example, we expected student-athletes to partner with other student-athletes, given a shared common interest in athletics. These distinguishing features are included as a set of statistics computed on the network. Model parameters are then estimated to maximize the likelihood of observing the network of interest, y . Estimates describe the covariates and network structures that drive the formation of the network (Cranmer and Desmarais, 2011). Covariate parameter estimates can be interpreted as the odds of observing an edge, thus quantifying the effect of some covariate on the presence of a tie (Cranmer and Desmarais, 2011). Large positive covariate values will increase the probability of observing an edge while large negative values decrease the probability.

Consider Y , a random network of n nodes. The probability of observing the

network y instead of all other possible networks conditional on a realization x of a random vector X is

$$P_{\theta,\beta}(Y = y|X = x) = \frac{1}{k(\theta, \beta)} e^{\theta_1 S_1(y) + \theta_2 AKT_1(y) + \beta^T g(y,x)} \quad (2.7)$$

where $k(\theta, \beta)$ is a normalization constant and θ is the coefficient of the network statistic taking a non-zero value when Y_{ij} are dependent for vertex pairs i and j (Kolaczyk and Csárdi, 2014). The number of edges are denoted by S_1 . The network statistic $AKT_1(y)$, also known as the geometrically weighted edgewise shared partnerships (GWESP) statistic accounts for the presence of transitivity. The statistic is given as

$$AKT_1(y) = 3T_3 + \sum_{k=2}^{N_v} (-1)^{k+1} T_k(y) \quad (2.8)$$

where N_v is the number of vertices and T_k is the number of k -triangles where k is the set of k individual triangles sharing a common base (Kolaczyk and Csárdi, 2014). Next, $g(y, x)$ represents the vector of statistics based on y and network attribute information x , and β is the corresponding vector of parameters. Often, $g(y, x)$ follows

$$g(y, x) = \sum_{1 \leq i < j \leq N_v} y_{ij} h(x_i, x_j) \quad (2.9)$$

where h is a measure of similarity in attributes for vertices i and j representing both main and second-order effects. Main effects, which represent numerical and categorical attributes, take the form

$$h(x_i, x_j) = x_i + x_j \quad (2.10)$$

while second-order effects, which represent homophily or the tendency for students to collaborate with others within the same attribute class are given as

$$h(x_i, x_j) = I(x_i = x_j) \quad (2.11)$$

The estimated parameters obtained from an ERGM can be interpreted similarly to regression coefficients. If the parameter estimate is significantly different from zero, we conclude that the corresponding statistic influences the probability of observing a particular instance of the network. Parameter estimates are often interpreted as the odds-ratio of an increase or decrease in the formation of a tie conditional on all other ties in the network.

The likelihood function obtained by taking the log of equation (1) is maximized in order to estimate the parameters of the network for each statistic. Maximization requires summation over all possible configurations of the network (Cranmer and Desmarais, 2011). In an undirected network with n nodes, this translates to $2^{\binom{n}{2}}$ combinations. Computation is extremely demanding requiring that the likelihood function be approximated. There are two types of approximation methods, the first is Markov Chain Monte Carlo (MCMC) maximum likelihood (Geyer and Thompson 1992) and the second is maximum pseudolikelihood (Frank and Strauss 1986) (Frank and Strauss, 1986; Geyer and Thompson, 1992). At present, MCMC is the default in most statistical packages (Cranmer and Desmarais, 2011). The `ergm` R package used in this study employs the Metropolis-Hastings MCMC algorithm (Hunter et al., 2008). In the iterative algorithm, the sum of the denominator of the likelihood function is approximated using a series of networks. These networks are sampled from a distribution informed by the parameters which maximized the likelihood in the prior sample of networks. This process continues until the approximated likelihood values show little variability. In a simple example,

$$\min\left(1, \frac{P_{\theta_0, y}(Y = y_{proposed})q(y_{current}, y_{proposed})}{P_{\theta_0, y}(Y = y_{current})q(y_{proposed}, y_{current})}\right) \quad (2.12)$$

where at each step, a random choice to either stay at $y_{current}$ for an additional step or change to $y_{proposed}$ where $y_{proposed}$ is chosen from an auxiliary distribution

dependent on $y_{current}$.

In the second method, pseudolikelihood, the product over the conditional probability of each tie given the other ties in the network replaces the joint likelihood of the ties in the network (Cranmer and Desmarais, 2011). The maximum pseudolikelihood is computed using a hill-climbing algorithm to estimate the vector of parameters to maximize. If two analysts are running separate simulations on the same study, MCMC can produce different results. Computation can be a challenge when it comes to large networks as the number of draws must be finite. On the other hand, pseudolikelihood estimation is quite fast. The disadvantage is that the loss of efficiency as a result of replacing the joint likelihood with a product over conditionals is not quantified.

In this study, ERGMs allow us to determine if the formation of student collaborations is influenced by student demographics, study preferences, and attitudes, or if these collaborations are not associated with these measured factors. They also allow us to quantify the influence of a particular attribute on the formation of a tie between two students. Though ERGMs help us to understand whether the observed network can be derived from member attributes and network structures, a limitation of ERGMs should be noted. ERGMs are at risk of model degeneracy, a scenario in which network structures are not adequately captured. In the event of degeneracy, the model is informed by a subset of nodes and is therefore not representative of the entire observed network (Harris, 2014). Higher-order dependence models address this limitation by proposing partial conditional dependence which assumes that the absence of a tie between two nodes is dependent on other ties in the network. Hunter and Handcock (2008) proposed three higher-order dependence model specifications to the general ERGM: GWESP referenced above, geometrically weighted degree distribution (GWD), and geometrically weighted

dyad wise shared partnerships (GWDSF).

The fit of ERGMs are evaluated using the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and goodness-of-fit tests. The AIC and BIC estimate model quality and are each used to compare the fit of models. The AIC is defined as

$$AIC = 2k - 2\ln\hat{L} \quad (2.13)$$

for which k is the number of parameters in the model and \hat{L} is the maximum value of the likelihood function. The secondary criterion, BIC, is defined as

$$BIC = \ln(n)k - 2\ln\hat{L} \quad (2.14)$$

where n is the number of data points in the observed network, k is the number of parameters in the model, and \hat{L} is the maximum value of the likelihood function (Harris, 2014).

Goodness-of-fit methods are applied to evaluate how well the model captures the original network. Networks are simulated through MCMC using model statistics, and the features of the simulated networks are compared against those of the observed (Hunter et al., 2008), through of a series of plots. The distribution of a statistic of interest from the observed data is plotted against the distribution from the simulated networks. Statistics of interest include minimum geodesic distance, edge-wise shared partners, degree, and triad census. The minimum geodesic distance shows the proportion of nodes relative to the shortest distance between two students. Edge-wise shared partners can be interpreted as the number of neighbors shared by two connected students. Degree shows the distribution of collaborators per student. Lastly, triad census describes the proportion of groups with zero, one, two, and three connections (Hunter et al., 2008).

2.3 Tests of Significance

To understand better the differences in demographics, attitudes, study habits, and learning preferences between independent learners and student collaborators, an adjusted form of the Welch t-test was adopted. The Welch t-test is used to determine if there is a difference in means between two independent and identically distributed populations with unequal variance (Welch, 1947). The network data used in this study is intrinsically dependent and interconnected. To draw conclusive results, we must introduce a naive adjustment addressing the dependency between student collaborators. The difference in means will not be affected but the possibility of correlation requires we adjust the standard deviation.

X_i is an independent observation i from the population of collaborators and Y_i is an independent observation i from the population of independent learners. Assuming that the two populations are normally distributed where $\bar{X} \sim N(\mu_x, \frac{\sigma_x^2}{n_x})$ and $\bar{Y} \sim N(\mu_y, \frac{\sigma_y^2}{n_y})$, the t-statistic follows

$$t = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}}} \quad (2.15)$$

where \bar{X} is the estimated mean of collaborators and \bar{Y} is the estimated mean of independent learners. The estimated variance is represented by s^2 and n is the number of students in the population. The degrees of freedom are given as

$$df = \frac{(\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y})^2}{\frac{(\frac{s_x^2}{n_x})^2}{n_x - 1} + \frac{(\frac{s_y^2}{n_y})^2}{n_y - 1}} \quad (2.16)$$

Confidence intervals are expressed as

$$(\bar{X} - \bar{Y}) \pm t_{\frac{\alpha}{2}, df} \sqrt{\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}} \quad (2.17)$$

We assume student collaborators are defined by $\bar{X} \sim N(\mu_x, \frac{\sigma_x^2 + \tau\mu_d}{n_x})$ and independent learners in the network follow $\bar{Y} \sim N(\mu_y, \frac{\sigma_y^2}{n_y})$. As with independent

populations, we use connections between collaborative links to estimate the covariance of certain numerical characteristics by $\hat{\tau} = cov(X_i, X_j)$ between all pairs i and j and zero otherwise. The estimated average degree of all connected nodes is represented by $\hat{\mu}_d$. The adjusted t-test statistic then follows

$$t = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{s_x^2 + \hat{\tau}\hat{\mu}_d}{n_x} + \frac{s_y^2}{n_y}}} \quad (2.18)$$

and the degrees of freedom are updated to

$$df = \frac{\left(\frac{s_x^2 + \hat{\tau}\hat{\mu}_d}{n_x} + \frac{s_y^2}{n_y}\right)^2}{\frac{\left(\frac{s_x^2 + \hat{\tau}\hat{\mu}_d}{n_x}\right)^2}{n_x - 1} + \frac{\left(\frac{s_y^2}{n_y}\right)^2}{n_y - 1}} \quad (2.19)$$

The updated confidence interval reflecting the adjustment is expressed as

$$(\bar{X} - \bar{Y}) \pm t_{\frac{\alpha}{2}, df} \sqrt{\frac{s_x^2 + \hat{\tau}\hat{\mu}_d}{n_x} + \frac{s_y^2}{n_y}} \quad (2.20)$$

In the following section, we examine the network descriptive statistics applied to the STA 307 and STA 308 student collaborative networks, we present results of the ERGMs fitted, as well as the outcomes of the adjusted t-tests. Results of the Welch t-test will be discussed for comparative purposes only.

CHAPTER 3

Results

3.1 Network Data Description

Visualizing the STA 307 and STA 308 networks in Figure 3.1, the top row shows the full network, connected, and largest connected component (LCC) from left to right while the full and connected STA 308 networks are displayed in row 2. Node color represents the recitation section each student attended. Each recitation section corresponds to a lecture section meaning that students in a particular recitation section were all enrolled in lecture with the same instructor.

Blue nodes indicate the student took lecture with instructor 1, while nodes marked in green represent students that took lecture with instructor 2. This logic also applies to the STA 308 networks in which red nodes represent students taking lecture with instructor 3 and purple represent students taking lecture with instructor 4. STA 307 and STA 308 consisted of two lecture sections, each taught by different instructors, delineated by instructor 1, 2, 3, and 4.

There is a sizeable connected component in the network of STA 307 students. Visualizations reveal that a higher proportion of student responders enrolled in lecture with instructor 2 than instructor 1. There do not appear to be many student collaborators not connected to the largest connected component.

The full STA 308 network shows that many students are working independently while the connected graph is characterized by numerous isolated groups of students. It is visually apparent that the STA 308 collaborative network is extremely sparse compared to STA 307. Unlike in STA 307, there is not a significant largest connected component.

The full network of student responders in course STA 307 consists of 100 nodes and 136 edges, 82 nodes and 136 edges in the connected network, and 67 nodes

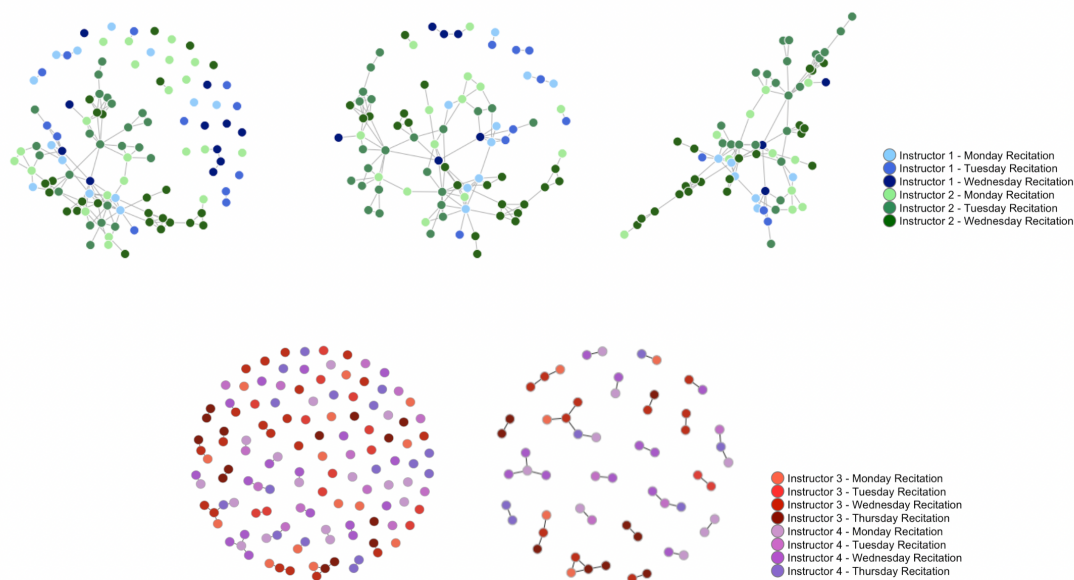


Figure 3.1: The STA 307 student collaboration network is represented in the top row; the full network is on the left, the connected network in the center, and the LCC on the right (LCC). There is a sizeable connected component in the network of STA 307 students. Visuals reveal that a greater proportion of student responders enrolled in lecture with Instructor 2. The network of STA 308 student collaborators is represented in row 2, on the left is the full network and the connected network on the right. The full network shows that many students are working independently, while numerous isolated groups of students characterize the connected graph.

and 127 edges in the largest connected component (LCC) (Table 3.1). Global degree assortativity is 0.33 in the full and connected network, and 0.23 in the LCC, suggesting a propensity for students of a high degree to collaborate with other students of high degree. The average degree is 2.72 in the full network, 3.32 in the connected network, and 3.79 in the LCC. On average, students collaborate with 2, 3, or 4 partners across the full, connected, and LCC networks. The density in the LCC (0.06) is suggestive of a sparse network. Transitivity, which summarizes the relative frequency with which connected triples close to form triangles, demonstrates that about 42% of connected triples close this way in the network of STA 307 students. The mean local clustering coefficient, which measures the degree to

which neighbors form a closed graph or a clique, is 0.52 in the full and connected network, and 0.54 in the LCC. The local clustering coefficient implies that, on average, 52% - 54% of connections are realized among neighboring students, indicative of a moderately dense graph. The betweenness centrality for each vertex is the number of shortest paths that pass through the vertex. The mean betweenness centrality is 69.62 in the full network, 84.90 in the connected network, and 103.84 in the LCC, suggesting that students who collaborate are often collaborating with more than one student.

Table 3.1: Network Description

	STA 307			STA 308	
	Full Network	Connected Network	LCC	Full Network	Connected Network
Nodes	100	82	67	102	57
Edges	136	136	127	35	35
Global Degree Assortativity	0.33	0.33	0.23	-0.02	-0.02
Degree Mean	2.72	3.32	3.79	0.69	1.23
Degree StDev	2.45	2.31	2.29	0.73	0.54
Density	0.03	0.04	0.06	0.01	0.02
Transitivity	0.42	0.42	0.42	0.19	0.19
Clustering Coefficient Mean	0.52	0.52	0.54	0.23	0.23
Betweenness Centrality Mean	69.62	84.90	103.84	0.17	0.30

The STA 308 network consists of 102 nodes and 35 edges in the full network, and 57 nodes and 35 edges in the connected network. The LCC consists of five students, and was omitted from analysis due to small sample size. The global

assortativity in the full and connected network is -0.02 , which suggests there is an extremely low propensity for students of a high degree to collaborate with other students of high degree, or for students with a low degree to collaborate with other students of low degree. It should be noted that the global assortativity may be influenced by low survey response rates.

The mean degree is 0.69 and 1.23 in the full network and connected network, respectively, implying that on average, students in the full network either do not collaborate or partner with one other student. On average, students in the connected network connect with one other student. STA 308 is comprised mainly of isolates and linked dyads- pairs of students working together (Figure 3.1). The density in the full network is 0.01 and 0.02 in the connected, further supporting sparse connectivity between students.

The transitivity shows that about 19% of connected triples close to form triangles. To illustrate, if student A collaborates with student B and student B collaborates with student C, more often than not, student C also collaborates with student A. The local clustering coefficient mean is 0.23 in the full and connected STA 308 networks, suggesting that, on average, 23% of connections are realized among neighboring students. The betweenness centrality mean, 0.17 in the full network and 0.30 in the connected, shows that a moderate number of students link two other students.

3.2 Assortativity

Assortativity is used to measure the propensity for students in the same recitation section to collaborate. In the following section, we continue to describe the STA 307 and STA 308 networks, focusing on assortativity.

TA B shows the highest assortativity compared to other TAs with 0.273 in the Tuesday section and 0.420 in the Wednesday section. This TA could have promoted

a collaborative environment, motivating students to work together (Table 3.2). Students are most likely to collaborate in the Wednesday recitation, taught by TA B, attending lecture taught by Instructor 2. This observation is consistent across the full network/connected network and LCC with assortativity of 0.420, 0.420, and 0.416, respectively. Students also have a high propensity to collaborate within the Tuesday recitation section led by TA A for students enrolled in lecture with Instructor 2. The assortativity is 0.267 in the full/connected networks and 0.244 in the LCC.

In the full and connected networks, students are least likely to collaborate within their recitation sections on Monday's relative to all other lecture sections (Table 3.2). The low assortativity among Monday recitation sections compared to Tuesday and Wednesday may be attributed to the cadence at which new material was introduced to students. Often, students had not yet been exposed to the material covered in Monday's recitation in the lecture. Students may have been less motivated to collaborate in recitation on assignments covering brand new material. During Tuesday and Wednesday recitations, the material was fresh, perhaps incenting students to partner on assignments.

The assortativity in the Wednesday recitation section for students taking lecture with instructor 1 shows the lowest assortativity with 0.079. This may, in part, be skewed by a small sample size ($n = 3$).

Table 3.2: STA 307 Recitation Day Assortativity

Instructor	Recitation Day	Teaching Assistant	Full Network	Connected Network	LCC
1	Monday	A	0.168 (n = 12)	0.168 (n = 9)	0.199 (n = 6)
1	Tuesday	B	0.273 (n = 11)	0.273 (n = 8)	0.265 (n = 3)
1	Wednesday	C	0.244 (n = 9)	0.244 (n = 6)	0.079 (n = 3)
2	Monday	C	0.121 (n = 21)	0.121 (n = 14)	0.146 (n = 11)
2	Tuesday	A	0.267 (n = 23)	0.267 (n = 23)	0.244 (n = 23)
2	Wednesday	B	0.420 (n = 24)	0.420 (n = 22)	0.416 (n = 21)

Assortativity measures were calculated across the recitation sections on the network of STA 308 students to understand if students within these sections were more likely to connect with one another. Findings in Table 3.3 reveal that there is a low propensity for students to collaborate with other students in the Monday recitation sections. The Tuesday recitation for students taking lecture with instructor 3 is skewed in that only two students connect with each other, resulting in an assortativity figure of 1. The Thursday recitation section for students taking lecture with instructor 3 shows the highest assortativity across all sections at 0.461. Perhaps students in this section were simply more willing to partner with one another.

Table 3.3: STA 308 Recitation Day Assortativity

Instructor	Recitation Day	Teaching Assistant	Full Network	Connected Network
3	Monday	D	-0.094 (n = 14)	-0.094 (n = 5)
3	Tuesday	F	1 (n = 10)	1 (n = 2)
3	Wednesday	G	0.190 (n = 16)	0.190 (n = 11)
3	Thursday	F	0.461 (n = 12)	0.461 (n = 9)
4	Monday	E	0.029 (n = 13)	0.029 (n = 9)
4	Tuesday	D	-0.077 (n = 9)	-0.077 (n = 4)
4	Wednesday	E	0.029 (n = 13)	0.029 (n = 11)
4	Thursday	D	0.153 (n = 15)	0.153 (n = 6)

3.3 ERGMs

Now that we have described the networks, we dive into the ERGMs fitted to the STA 307 and STA 308 networks of student collaborators. We focus specifically on the attributes chosen in each model and interpretation.

In fitting an ERGM to the full network of student collaborators in STA 307 using the package `ergm` in R Studio, we include two network statistics, one main effect, and 14 second-order effects (Hunter et al., 2008). Each parameter estimate can be interpreted as the increase in odds of a tie for a unit change in the predictor, given the realized attributes and composition of the network. A positive parameter estimate implies the attribute influences tie formation while a negative coefficient hinders tie formation. A network statistic describes network structures present in

the network, like degree and triangles. The first network statistic included in the model represents the number of edges and can be thought of as the density. The second variable geometrically weighted edgewise shared partners (GWESP) captures the presence of transitivity in the network (Harris, 2014). GWESP translates to the odds of collaboration between two students i and j when they have partners in common, and each student i and j is in at least one other triangle with each of those partners (Hunter et al., 2008).

Main effects can be numerical or categorical in nature. The benefit of timely email response to learning was included as a main ordinal effect in the ERGM. A positive parameter estimate suggests that students responding more positively are more likely to collaborate.

Second-order effects represent homophily or the tendency for students to collaborate with others within the same attribute class. The five attributes which translate to 14 classes include lecture/recitation section, gender, athlete status, on-campus/off-campus resident, and in-state/out-of-state.

To begin model fitting, the network statistic edges, was included to indicate the number of ties. Then, an iterative process testing the significance of student attributes on tie formation was performed. Results from the assortativity analysis prompted us to include recitation section as a second-order effect. Next, student attributes including athlete status, major, age, gender, in-state/out-of-state, on-campus/off-campus residency, employment status, interest in hobbies, presence of stress, and citizenship status were tested as both main effects and second-order effects. Network statistics describing the network of student collaborators were also tested. The inclusion of statistics like transitivity, degree, and k-stars resulted in degenerate models. The thought was that these models did not converge due to sparsity of the network, motivating us to test statistics GWESP, GWD, and

GWDSF that specifically address model degeneracy.

This process of model fitting was also applied to the connected network and LCC of STA 307 student collaborators. Fitting models to each network allows us to understand the factors associated with collaboration among both collaborators and independent learners, all collaborators, and collaborators connected to the LCC. Output in Table 3.4 shows the parameter estimates, standard error or level of uncertainty, p-value, and odds of a tie for each of the three models. Across all attributes and models, edges was the only attribute with a negative coefficient (Table 3.4). A negative edge coefficient tells us that the network is quite sparse. Table 3.4 also shows that students within the same recitation section, housing type, gender, athlete status, and in-state/out-of-state residence are factors associated with student collaboration.

We first deep-dive into the ERGM fitted to the full network. We see that the parameter estimate associated with Instructor 1 - Tue, 1.20, is positive and significant. Converting the parameter estimate to the odds, we can say that students in the Tuesday recitation section taking lecture with Instructor 1 are 3.33 times more likely to collaborate with others in the same section. We see that variables representing Tuesday and Wednesday recitation sections are significant. Students in these sections are likely to connect with other students in the same recitation section. Students in the Wednesday recitation taking lecture with Instructor 1 are 7.32 times more likely to connect with one another. The strikingly high odds compared to those observed in the other Wednesday recitation section (2.64), can be attributed to small sample size. Only nine students in this section consented to the use of their data, while 24 students in the other Wednesday recitation section were included in this study. In summary, the significance of the recitation section in the model implies that students may be more willing to collaborate in a more

Table 3.4: STA 307 ERGM Results: Full Network, Connected, and LCC

Variables	Full Network					Connected					LCC					
	Est.	SE	P	Odds	Est.	SE	P	Odds	Est.	SE	P	Odds	Est.	SE	P	Odds
Network Structures																
Edges	-7.17	0.42	0.00	0.00	-6.18	0.31	0.00	0.00	-6.57	0.45	0.00	0.00				
GWESP (fixed, decay = 0.5)	1.41	0.12	0.00	4.11	1.23	0.12	0.00	3.43	1.21	0.13	0.00	3.34				
Main Effects																
Pre-Course Timely Response to Email	0.07	0.03	0.02	1.07												
Pre-Course Preference to Study in Library									0.10	0.05	0.03	1.11				
Second-Order Effects																
Instructor 1 - Mon	0.71	0.40	0.07	2.03	1.04	0.48	0.03	2.82	1.41	0.46	0.00	4.08				
Instructor 1 - Tue	1.20	0.55	0.03	3.33	1.58	0.53	0.00	4.86	3.27	0.89	0.00	26.31				
Instructor 1 - Wed	1.99	0.35	0.00	7.32	2.21	0.52	0.00	9.14	2.64	1.13	0.02	13.95				
Instructor 2 - Mon	0.27	0.39	0.49	1.30	0.72	0.46	0.12	2.05	0.84	0.41	0.04	2.31				
Instructor 2 - Tue	1.09	0.17	0.00	2.97	0.86	0.18	0.00	2.35	0.72	0.18	0.00	2.05				
Instructor 2 - Wed	0.97	0.15	0.00	2.64	0.94	0.16	0.00	2.57	0.93	0.17	0.00	2.52				
On-Campus Resident	1.14	0.29	0.00	3.12	1.31	0.31	0.00	3.72	1.59	0.39	0.00	4.93				
Off-Campus Resident	0.77	0.19	0.00	2.17	0.76	0.20	0.00	2.14	0.52	0.20	0.01	1.67				
Female	0.17	0.15	0.25	1.18	0.23	0.16	0.15	1.25	0.33	0.15	0.03	1.39				
Male	1.11	0.20	0.00	3.03	1.17	0.22	0.00	3.23	1.19	0.20	0.00	3.27				
Non-Athlete	0.29	0.14	0.05	1.33	0.39	0.16	0.01	1.48	0.38	0.15	0.01	1.47				
Athlete	0.87	0.25	0.00	2.40	0.85	0.29	0.00	2.33	0.99	0.27	0.00	2.68				
In-State	0.31	0.38	0.41	1.37	0.56	0.40	0.17	1.75	0.66	0.45	0.14	1.93				
Out-of-State	0.46	0.16	0.00	1.58	0.44	0.17	0.01	1.56	0.46	0.17	0.01	1.59				

intimate classroom environment offered by recitation.

On-campus residents are 3.12 times more likely to connect with other on-campus residents ($p < 0.01$), and male students are 3.03 times more likely to connect with other male students ($p < 0.01$). There is no evidence supporting that females are likely to connect with other females ($p = 0.25$) or that in-state students are likely to connect with other in-state students ($p = 0.41$). Students who value timely email responses from Instructors and recitation leaders are more likely to collaborate ($p < 0.05$). Out-of-state students are 1.58 times more likely ($p < 0.01$) to connect with other like students.

The network structure, GWESP, used to capture patterns of transitivity shows a positive coefficient, is significant ($p < 0.01$), and shows low standard error (0.12). If two students share study partners and each of them is in at least one other triangle, then the odds of them becoming friends is 4.11. The GWESP coefficient suggests that the model accounts for the influence of a network attribute on collaboration not captured in the demographics or learning preferences.

The ERGM fitted to the connected network, which includes students that collaborated but did not connect to the LCC, shows similar results to the estimates in the full network ERGM. The edges coefficient is -6.18, indicative of a somewhat sparse network ($p < 0.01$). The Tuesday and Wednesday recitations for students taking lecture with Instructor 1 show the highest odds of a tie at 4.86 ($p < 0.01$) and 9.14 ($p < 0.01$) respectively. For students taking lecture with Instructor 2, students in the Tuesday recitation are 2.35 ($p < 0.01$) times more likely to connect with one another and students in the Wednesday recitation are 2.57 ($p < 0.01$) times more likely. There is no evidence to suggest that students in the Monday recitation section for Instructor 2 are more likely to collaborate with one another ($p = 0.12$). Out-of-state students ($p < 0.05$), self-identified males ($p < 0.01$), non-athletes ($p <$

0.05), and athletes ($p < 0.01$) are more likely to connect with other like students. Most notably, self-identified males are 3.23 times more likely to collaborate with other males. As observed in the STA 307 full network ERGM, there is a non-trivial network structure apparent indicating the presence of connected triples.

Drivers of student collaboration in the LCC are consistent with those chosen in the full and connected network models. The recitation section, housing type, in-state/out-of-state residency, self-identified gender, athlete status, and GWESP term accounting for transitivity, are drivers of collaboration in STA 307 across all three networks.

The odds of students taking lecture with Instructor 1 and enrolled in the Tuesday recitation are extremely high. We see in Table 3.4 that students in this section are 26 times more likely ($p < 0.01$) to connect with one another. Students enrolled in the Wednesday recitation section and taking lecture with Instructor 1 are about 14 times more likely to connect with each other ($p < 0.05$). The odds for these two recitation sections are sizable compared to the odds observed across the other four sections but are likely skewed by small sample size ($n=3$ each).

Not observed in the two ERGMs fitted to the STA 307 full and connected networks, both Instructor 1- Mon and Instructor 2 - Mon variables are significant in this model. This suggests that students are more likely to connect with one another in the Monday recitation sections if they are connected to the LCC.

A covariate not included in the full and connected network ERGMs but present in the LCC model is the preference for students to study in the library. Students that prefer to study in the library are 11% more likely to connect with other students ($p < 0.05$). Students sharing a preference to study in the library may be more apt to partner and collaborate on assignments.

Goodness-of-fit diagnostics referenced in Table 3.2 were used to evaluate the

fit of the three ERGMs. Plots show that the ERGM fitted to the full network is a relatively decent fit. We expect the observed data in bold to fall between the 10th and 90th percentiles obtained from the simulated networks. The full network model underestimates the shortest distance between students, and does not perfectly capture the number of neighbors shared between two connected students or the number of collaborators per student. The ERGM fitted to the LCC shows the poorest fit across the three models. Across all models, the triad census plots show that the models adequately represent the number of collaborations in each set of 3 students.

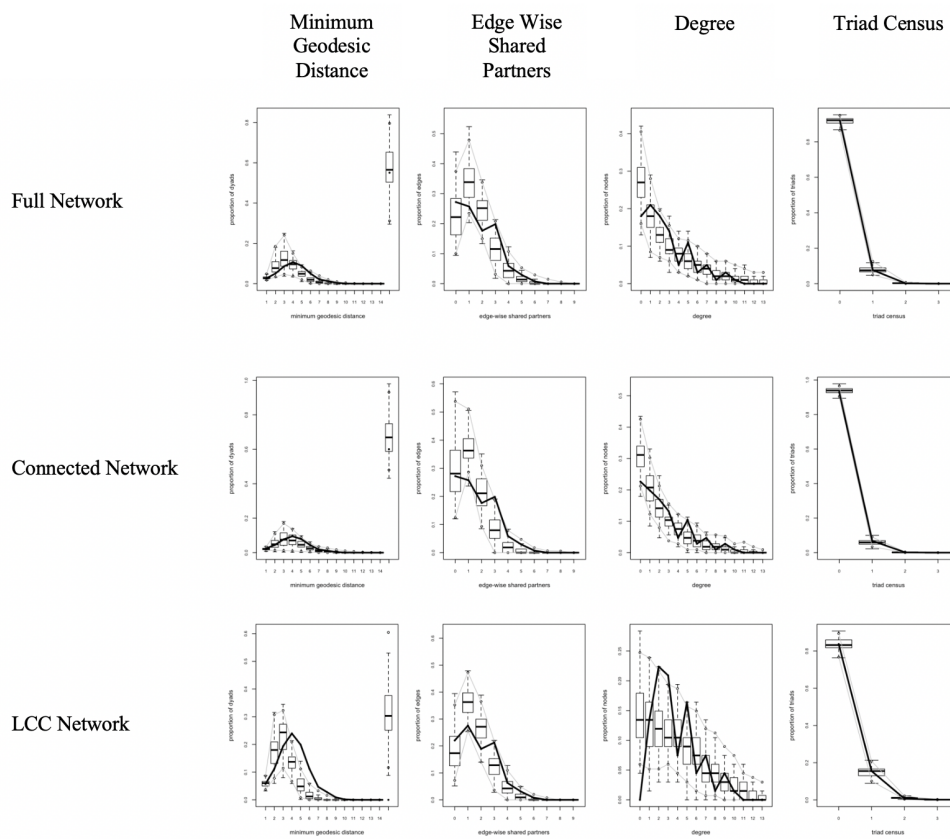


Figure 3.2: Goodness-of-fit plots measuring the quality of ERGMs fitted to the full network, connected, and LCC networks of STA 307 students. We expect the observed data in bold to fall between the 10th and 90th percentiles obtained from the simulated networks. The full network shows the best fit while the LCC shows the poorest. The full network model underestimates the shortest distance between students, and does not perfectly capture the number of neighbors shared between two connected students or the number of collaborators per student.

An ERGM was fitted to the STA 308 full network to compare the drivers of student collaboration in each of the two undergraduate statistics courses evaluated in this study. The network is significantly more sparse than the STA 307 student collaboration network; therefore, the ERGM is not the best fit. There are significantly fewer network statistics and covariates present in this model than in the model fitted to the full STA 307 network.

The estimates of the model showing the best fit are present in Table 3.5. The

coefficient associated with the main effect housing (1.22), for which the reference variable is on-campus residence, suggests that students residing off-campus are more likely than on-campus residents to collaborate ($p < 0.01$). The number of past mathematics courses a student has taken is also a driver of student collaboration in this network. Students taking a greater number of courses have a higher propensity to collaborate ($p < 0.01$). Students taking lecture with Instructor 3 are 11.45 times more likely to collaborate with other students in the same lecture ($p < 0.01$). In contrast, students are 10.15 times more likely to collaborate with other students taking lecture with Instructor 4 ($p < 0.01$). The variable degree, for which the reference number of degrees was set to 1, was included to offset the number of isolates. The coefficient 0.75 suggests that there are students in the network that collaborate with one other student.

Despite a poor fit generally speaking, the goodness-of-fit degree plot in Figure 3.3 shows that the model fits the number of collaborators per student relatively well compared to the actual network. This is likely a result of few collaborations present in the observed network.

Table 3.5: STA 308 Full Network ERGM Results

Variables	Est.	SE	P	Odds
Network Structures				
Edges	-9.73	1.00	0.00	0.00
Degree(1)	0.75	0.24	0.00	2.11
Main Effects				
On/Off Campus	1.22	0.37	0.00	3.39
Number of College Math Courses Taken	0.18	0.07	0.01	1.20
Second-Order Effects				
Instructor 3	2.44	0.66	0.00	11.45
Instructor 4	2.32	0.67	0.00	10.15

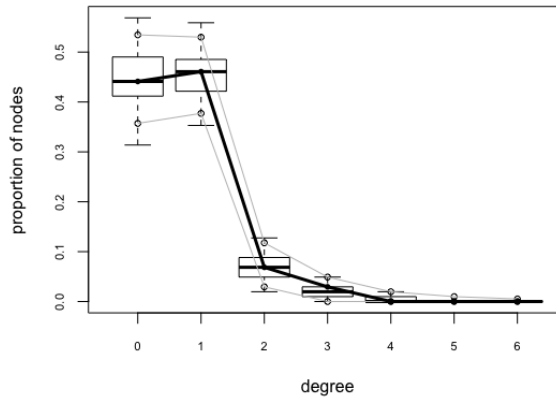


Figure 3.3: STA 308 Full Network ERGM degree goodness-of-fit plot shows that the model fits the number of collaborators per student relatively well compared to the actual network.

3.4 Comparison of Collaborators and Independent Learners

To determine if there was a statistically significant difference in the means of demographics, study habits, learning preferences, performance, and attitudes between collaborators and independent learners in STA 307 and STA 308, we

utilized both the Welch and adjusted t-tests.

In our hypothesis, H_o assumes that there is no difference in means between students who collaborate and students that do not, and H_a indicates that there is a difference in means between students who collaborate and students that do not.

We initially explored the difference in attitudes between collaborators and independent learners in STA 307 (Table 3.6). Interestingly, both the general and adjusted t-tests produce virtually identical results, which was expected given the naive adjustment to the Welch t-test. Going forward, we discuss only the adjusted t-test results.

Examining the results of the adjusted t-tests performed on pre-course cognitive competence, the t-statistic is given as -0.36, the p-value as 0.72, and the 95% confidence interval as (-0.61, 0.43). The negative t-statistic tells us that the mean of collaborators is less than that of independent learners. When the t-statistic is positive, the inverse is true. We can conclude that independent learners share a more positive disposition on their intellectual knowledge and skills applied to statistics than students who collaborate. In looking at the p-value, 0.72, we fail to reject the null hypothesis at the $p = 0.05$ significance level. There is not sufficient evidence to conclude that there is a difference in students' cognitive competence towards statistics between collaborators and independent learners. The confidence interval, based on the t-distribution, tells us the range we can expect the difference in means to fall. Here, we are 95% confident that the difference in means between the two cohorts falls between -0.61 and 0.43. The null hypothesis assumes that the difference in means between student collaborators and independent learners is zero. If zero falls between the confidence interval bounds, we accept the null hypothesis. If zero does not fall within the confidence interval, we accept the alternative hypothesis, concluding that there is a difference in means.

Adjusted t-tests performed on the remaining attitudes collected at the beginning and end of the course do not reveal any significant differences in means between student collaborators and independent learners ($p > 0.05$). The t-statistics associated with student attitudes about the easiness of statistics show that on average, student collaborators felt statistics was easier than independent learners at both the beginning and end of the course. This is supported by the t-statistic 1.69 corresponding to pre-course difficulty (easiness) and 1.57 to post-course difficulty. Findings also show that independent learners were on average more interested in statistics than collaborators at both the beginning and end of the course.

Table 3.6: Differences in Attitudes Towards Statistics between Collaborators and Independent Learners in STA 307

t-test	Welch			Adjusted		
Attitude	t	P	95%	t	P	95%
Pre-Course						
Cognitive Competence	-0.36	0.72	(-0.61, 0.43)	-0.36	0.72	(-0.61, 0.43)
Affect	0.10	0.92	(-0.50, 0.55)	0.10	0.92	(-0.50, 0.55)
Effort	0.29	0.77	(-0.39, 0.52)	0.29	0.78	(-0.40, 0.53)
Difficulty (Easiness)	1.68	0.11	(-0.08, 0.75)	1.69	0.11	(-0.08, 0.75)
Interest	-0.67	0.51	(-1.05, 0.54)	-0.7	0.49	(-1.02, 0.51)
Value	-1.28	0.21	(-1.01, 0.24)	-1.26	0.22	(-1.02, 0.25)
Post-Course						
Cognitive Competence	0.70	0.49	(-0.40, 0.83)	0.70	0.49	(-0.40, 0.83)
Affect	0.62	0.54	(-0.46, 0.85)	0.61	0.55	(-0.47, 0.86)
Effort	-0.02	0.99	(-0.45, 0.44)	-0.02	0.99	(-0.46, 0.45)
Difficulty (Easiness)	1.60	0.12	(-0.10, 0.78)	1.57	0.13	(-0.10, 0.78)
Interest	-1.50	0.15	(-1.49, 0.24)	-1.51	0.15	(-1.49, 0.23)
Value	-1.41	0.17	(-1.02, 0.19)	-1.38	0.18	(-1.04, 0.20)

We performed a similar exercise on the STA 308 data, comparing the pre-course attitudes between collaborators and independent learners. Results in Table 3.7 reveal that there do not appear to be any significant differences in attitudes between collaborators and independent learners at the $p = 0.05$ significance level. Pre-course attitude value, which measures attitudes about the usefulness, relevance, and worth of statistics in personal and professional life, is on the cusp of 0.05, showing a p-value of 0.07.

In summary, there is no difference in the means of attitudes towards statistics

between collaborators and non-collaborators in STA 307 or STA 308. In general, students do not appear to be motivated to collaborate based on affect, value, interest, difficulty, cognitive competence, or effort towards statistics at the beginning or end of the course. Post-course attitudes among STA 308 students were omitted from this analysis due to missing data.

Table 3.7: Differences in Attitudes Towards Statistics between Collaborators and Independent Learners in STA 308

t-test		Welch		Adjusted		
Attitude	t	P	95%	t	P	95%
Pre-Course						
Cognitive Competence	-1.73	0.09	(-0.75, 0.05)	-1.72	0.09	(-0.75, 0.05)
Affect	-0.63	0.53	(0.60, 0.31)	-0.62	0.53	(0.60, 0.31)
Effort	-0.72	0.48	(-0.31, 0.14)	-0.71	0.48	(-0.31, 0.15)
Difficulty	-0.30	0.76	(-0.30, 0.22)	-0.30	0.77	(-0.30, 0.22)
Interest	0.17	0.86	(-0.44, 0.52)	0.18	0.86	(-0.43, 0.51)
Value	-1.80	0.08	(-0.76, 0.04)	-1.85	0.07	(-0.75, 0.03)

In addition to exploring differences in attitudes between collaborators and independent learners, we examined the differences in demographics, study habits, learning preferences, and performance. Reported in Table 3.8 are the attributes for which statistically significant differences were observed between collaborators and independent learners among STA 307 students. While all study habits and learning preferences were tested, only a handful proved to be different between the groups.

There is a difference in means between collaborators and independent learners in their past performance in college math courses ($t = 2.26, p = 0.04$) and self-reported mathematics skills ($t = 2.33, p = 0.03$). In each case, the t-statistic is positive, which implies that collaborators performed better in past math courses

and felt more confident about their skills than independent learners. Evidence suggests that independent learners value the instructor's knowledge more than students who collaborate, evidenced by the negative t-statistic, -2.19 and $p < 0.03$.

Collaborators responded that they complete practice tests before exams more often than independent learners. Not unexpectedly, students that collaborate have a higher preference for group study than independent learners ($t = 2.68, p = 0.01$). Finally, there was no evidence supporting that students who collaborate outperform independent learners in STA 307 ($p = 0.10$).

Table 3.8: STA 307 Student Learning Preferences and Demographics between Collaborators and Independent Learners

t-test	Welch			Adjusted		
Description	t	P	95%	t	P	95%
Background						
How well did you do in past mathematics courses?	2.27	0.03	(0.06, 1.56)	2.26	0.04	(0.06, 1.56)
How good are you at mathematics?	2.38	0.03	(0.12, 1.80)	2.33	0.03	(0.11, 1.81)
Pre-Course						
How beneficial is the instructors in-depth knowledge of material to you?	-2.26	0.03	(-1.22, 0.06)	-2.19	0.03	(-1.24, -0.05)
Do you complete practice tests before exams?	2.35	0.03	(0.08,1.26)	2.32	0.03	(0.07,1.26)
Post-Course						
Did you complete practice tests before exams?	2.20	0.04	(0.03, 1.22)	2.22	0.04	(0.04, 1.22)
Did you prefer to study in groups?	2.64	0.02	(0.17,1.44)	2.68	0.01	(0.18,1.43)
Did you prefer to study in a library?	2.29	0.03	(0.07, 1.21)	2.13	0.04	(0.03, 1.25)
Performance						
Final Course Grade	1.77	0.09	(-0.90, 10.86)	1.73	0.10	(-1.00, 10.97)

The results of t-tests applied to the population of STA 308 students did not show any significant differences between the two cohorts across all measured demographics, study habits, and learning preferences ($p > 0.05$). A sample representation of results in Table 3.9 shows a $p > 0.05$ for performance in past mathematics courses and mathematical ability, study habits, and learning preferences.

Table 3.9: STA 308 Student Learning Preferences and Demographics between Collaborators and Independent Learners

t-test	Welch			Adjusted		
Description	t	P	95%	t	P	95%
Background						
How well did you do in past mathematics courses?	0.11	0.91	(-0.56, 0.62)	0.12	0.91	(-0.53, 0.60)
How good are you at mathematics?	0.20	0.84	(-0.49, 0.60)	0.20	0.85	(-0.49, 0.60)
Pre-Course						
How beneficial is the instructors in-depth knowledge of material to you?	-0.22	0.83	(-0.53, 0.43)	-0.21	0.83	(-0.55, 0.44)
Do you complete practice tests before exams?	1.56	0.12	(-0.08, 0.65)	1.48	0.14	(-0.10, 0.67)
Do you prefer to study in the library?	0.17	0.86	(-0.39, 0.46)	0.17	0.87	(-0.40, 0.48)
Do you prefer to study in groups?	0.63	0.53	(-0.25, 0.49)	0.61	0.54	(-0.26, 0.50)

CHAPTER 4

Main Findings

The goal of this study was to identify the drivers of student collaboration in two undergraduate statistics courses and to distinguish any defining characteristics differentiating collaborators from independent learners.

Descriptive statistics concluded that there were fewer students partnering with their peers in STA 308 than in STA 307, in part attributed to the make-up of the students enrolled. In STA 307, three-quarters of the students were enrolled in the College of Pharmacy while in STA 308, there was a more diverse pool of majors represented. Students enrolled in the same major have the opportunity to connect in other courses which may explain the greater number of connections among this cohort of students. Instructors need to place more significant effort on fostering collaboration in classes like STA 308 perhaps by assigning additional group projects. Instructors should also promote the use of collaborative digital platforms to encourage collaboration among students unable to meet face-to-face.

Assortativity and ERGMs show that students are most likely to collaborate with others in the same recitation section, emphasizing the value of recitation to student collaboration. On average, students in STA 307 collaborate with two, three, or four students while STA 308 students typically collaborate with one other student or work independently. Transitivity in STA 307 revealed a significant portion of students working in groups of three. If group work is used to promote partnership in recitation or lectures, groups between two and four students are recommended. Assortativity also revealed that partnership between students is more likely on recitation days later in the week. It is thought that students in the Monday recitation are less willing to partner on assignments covering topics that

have not yet been introduced in the lecture. Instructors should consider scheduling recitation sections later in the week or wait until the material is covered in the lecture to distribute assignments.

ERGMs fitted to the full STA 307 population revealed the presence of homophily or the likelihood that students are more likely to collaborate with other like students. More specifically, students living on-campus had a greater propensity to connect with other students living on-campus. The reverse is also true of students living off-campus who are more likely to partner with one another. In general, it may be easier for students living on campus to meet-up post lecture or recitation, in a dorm, the dining hall, or even the library. ERGMs revealed that males, athletes, and out-of-state residents were each more likely to connect with other like students. In a diverse class, it may be beneficial for instructors to encourage partnership among students sharing different qualities. The variability in odds of a tie between students in the three ERGMs fitted to the STA 307 full, connected, and LCC networks implies that strong inference cannot be made about the drivers of collaboration across all students enrolled in STA 307 in Spring 2017.

T-tests applied to the STA 307 population reveal that students completing practice tests before exams are more likely to collaborate than students that do not. Collaboration with peers may motivate students to complete practice tests before exams. Not unexpectedly, students that collaborate have a higher preference for group study than independent learners ($t = 2.68, p = 0.01$). Finally, one of the main findings revealed that there was no significant difference in course performance between student collaborators and independent learners. We would expect collaboration with peers to boost achievement among students. Findings drawn from this study are representative only of the students evaluated and cannot be attributed to all students enrolled in STA 307. It is possible that results differ

when looking at all students enrolled in STA 307. In addition, the simple adjusted t-test used to address the dependency in data could be enhanced to account for the degree of nearest neighbors, which may strengthen the precision of the results.

4.1 Limitations

A primary limitation of this study is the nonresponse bias of students that did not consent to the use of their in this study or did not complete the surveys. Findings are limited to participants and cannot be applied to all students enrolled in each course. Of the 153 total students enrolled in STA 307, 128 provided consent, and only 100 completed all four surveys. The response rate is strikingly lower in STA 308, 170 of the 248 total students consented, and of the 170, 102 completed both pre-course surveys. We cannot assume that there were fewer collaborations among all students in STA 308 than in STA 307 given that we can only draw conclusions from the observed responses. Students enrolled but omitted from this study may have substantially different learning preferences and propensity to collaborate than participants. Non-participating students may have either not been present on the days the surveys were administered or opted out.

Nonresponse bias is observed not only at the course level but also at the lecture level in STA 307. While the two lecture sections were relatively balanced in class size, 36 students from the first section and 70 from the second consented and completed the surveys. The instructor of the latter section continued to promote the value of these surveys, which may have motivated students to participate.

In addition to nonresponse bias, the nature of survey data is prone to response bias. Students may not have been truthful in completing the survey questions. Perhaps students were rushing to respond or answered in a way they felt they should or in a way that is socially desirable. Lastly, students may not have listed all collaborators they partnered with throughout the semester. Students could have

listed the names of students they worked with most often, neglecting to list the students with whom they partnered with sparingly. There were several instances in which the student entered a collaborator's first name or a nickname, which did map back to an enrolled student. Un-identified collaborator names were strictly omitted from this study.

Missing data is the primary limitation to this study. The inclusion of students with missing data may result in different outcomes. Consenting students that worked with peers excluded from this study, either because they did not consent or had a high proportion of missing data, were excluded from this study. The network is, therefore, not reflective of all reported collaborations in each course.

4.2 Future Research

In future work, we would like to expand this study across a variety of undergraduate courses spanning multiple disciplines to understand the drivers of collaboration in other courses. If repeated, surveys used in this study should be enhanced to reduce errors in capturing the names of student collaborators. A more robust entry method would allow for greater precision in identifying the network of student partnerships.

We would also like to design an experiment in which group work is heavily promoted and encouraged in one cohort and not the other. We would select an undergraduate statistics course and randomly split the population in two. The first cohort would receive encouragement to collaborate in recitation while the second would not. At the end of the semester, comparison using t-tests would allow us to determine if there is a difference in student performance, stress levels, and post-course attitudes towards statistics between students receiving encouragement to collaborate and those not receiving treatment. We would also like to make predictions to understand which students are most likely to work together. This could

be informative to instructors to know at the beginning of the semester. Instructors could capture student demographics, study habits, and learning preferences to identify students most likely to work independently and perhaps encourage those students to engage in group work.

It is known that students of Generation Z value collaboration as a learning tool, further study of the way in which students seek out collaborators would allow universities and instructors to refine course design and the delivery of information. Improved course design targeted towards these students could bolster performance and deepen students' understanding of the subject matter.

LIST OF REFERENCES

- Brewe, E., Kramer, L., and Sawtelle, V. (2012). Investigating student communities with network analysis of interactions in a physics learning center. *Physical Review Special Topics-Physics Education Research*, 8(1):010101.
- Chicca, J. and Shellenbarger, T. (2018). Connecting with generation z: Approaches in nursing education. *Teaching and Learning in Nursing*, 13:180–184.
- Cranmer, S. J. and Desmarais, B. A. (2011). Inferential network analysis with exponential random graph models. *Political Analysis*, 19(1):66–86.
- Cranmer, S. J., Desmarais, B. A., and Menninga, E. J. (2012). Complex dependencies in the alliance network. *Conflict Management and Peace Science*, 29(3):279–313.
- Frank, O. and Strauss, D. (1986). Markov graphs. *Journal of the American Statistical Association*, 81(395):832–842.
- Geyer, C. J. and Thompson, E. A. (1992). Constrained monte carlo maximum likelihood for dependent data. *Journal of the Royal Statistical Society: Series B (Methodological)*, 54(3):657–683.
- Harris, J. (2014). *An introduction to exponential random graph modeling*. Quantitative Applications in the Social Sciences. SAGE Publications.
- Hunter, D. R., Handcock, M. S., Butts, C. T., Goodreau, S. M., and Morris, M. (2008). Ergm: A package to fit, simulate and diagnose exponential-family models for networks. *Journal of statistical software*, 24(3):nihpa54860.
- Igel, C. and Urquhart, V. (2012). Generation z, meet cooperative learning: Properly implemented cooperative learning strategies can increase student engagement and achievement. *Middle school journal*, 43(4):16–21.
- Kolaczyk, E. (2009). *Statistical analysis of network data: Methods and models*. Springer.
- Kolaczyk, E. D. and Csárdi, G. (2014). *Statistical analysis of network data with R*, volume 65. Springer.
- Loveland, E. (2017). Instant generation. *Journal of College Admission*, 235:34–38.
- Park, J. and Newman, M. E. (2004). Statistical mechanics of networks. *Physical Review E*, 70(6):066117.

- Rickes, P. C. (2016). Generations in flux. *Planning for Higher Education*, 44(4):21–45.
- Seemiller, C. and Grace, M. (2016). *Generation Z goes to college*. John Wiley & Sons.
- Shatto, B. and Erwin, K. (2016). Moving on from millennials: Preparing for generation z. *Journal of continuing education in nursing*, 47:253–254.
- Snijders, T. A., Pattison, P. E., Robins, G. L., and Handcock, M. S. (2006). New specifications for exponential random graph models. *Sociological methodology*, 36(1):99–153.
- Spears, J., Zobac, S. R., Spillane, A., and Thomas, S. (2015). Marketing learning communities to generation z: The importance of face-to-face interaction in a digitally driven world. *Learning Communities Research and Practice*, 3(1):7.
- Strauss, W. and Howe, N. (2003). Millennials go to college: Strategies for a new generation on campus. *American Association of Collegiate Registrars, Washington, DC*.
- Turner, A. (2015). Generation z: Technology and social interest. *The Journal of Individual Psychology*, 71(2):103–113.
- Twenge, J. M. (2017). *iGen: Why today's super-connected kids are growing up less rebellious, more tolerant, less happy—and completely unprepared for adulthood—and what that means for the rest of us*. Simon and Schuster.
- Wasserman, S. and Pattison, P. (1996). Logit models and logistic regressions for social networks. i. an introduction to markov graphs and p . *psychometrika* 61 401–425. *Mathematical Reviews (MathSciNet): MR1424909 Digital Object Identifier: doi*, 10.
- Welch, B. L. (1947). The generalization of student's' problem when several different population variances are involved. *Biometrika*, 34(1/2):28–35.

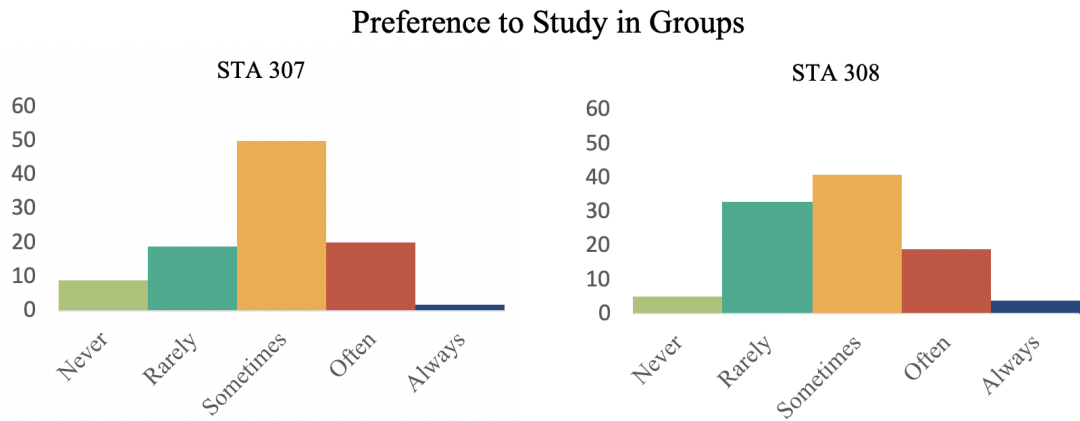


Figure .1: A greater proportion of students in STA 307 responded that they sometimes prefer to study in groups than in STA 308.

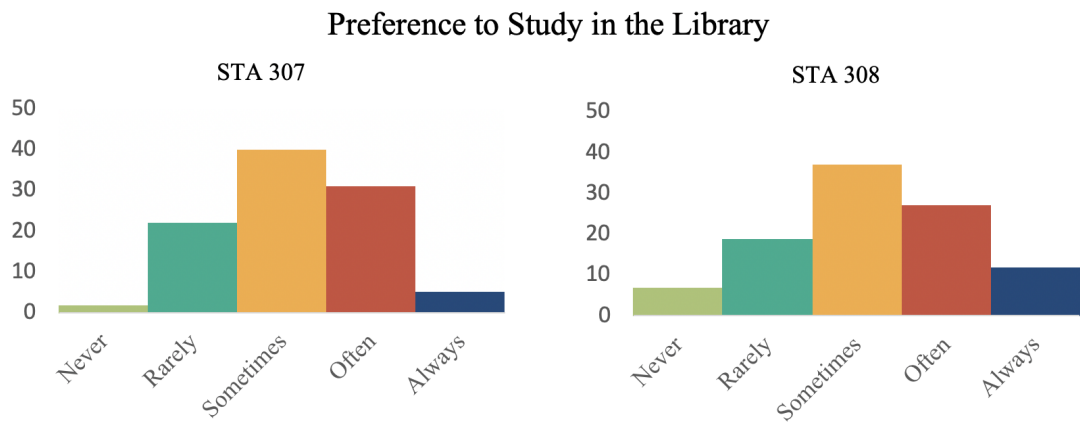


Figure .2: The majority of students in both STA 307 and STA 308 responded that they sometimes prefer to study in the library.

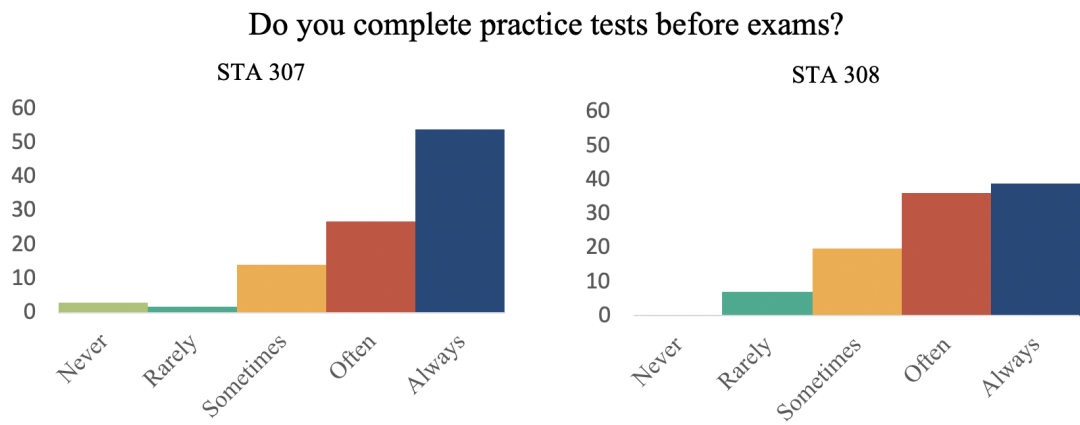


Figure .3: A greater proportion of students in STA 307 responded that they always complete practice tests before exams than in STA 308.

BIBLIOGRAPHY

- Brewe, E., Kramer, L., and Sawtelle, V., “Investigating student communities with network analysis of interactions in a physics learning center,” *Physical Review Special Topics-Physics Education Research*, vol. 8, no. 1, p. 010101, 2012.
- Chicca, J. and Shellenbarger, T., “Connecting with generation z: Approaches in nursing education,” *Teaching and Learning in Nursing*, vol. 13, pp. 180–184, 07 2018.
- Cranmer, S. J. and Desmarais, B. A., “Inferential network analysis with exponential random graph models,” *Political Analysis*, vol. 19, no. 1, pp. 66–86, 2011.
- Cranmer, S. J., Desmarais, B. A., and Menninga, E. J., “Complex dependencies in the alliance network,” *Conflict Management and Peace Science*, vol. 29, no. 3, pp. 279–313, 2012.
- Dalmolin, A. C., Mackeivicz, G. A. O., Pochapski, M. T., PILATTI, G. L., and SANTOS, F. A., “Learning styles preferences and e-learning experience of undergraduate dental students,” *Revista de Odontologia da UNESP*, vol. 47, no. 3, pp. 175–182, 2018.
- Desmarais, B. A. and Cranmer, S. J., “Statistical inference for valued-edge networks: The generalized exponential random graph model,” *PloS one*, vol. 7, no. 1, 2012.
- Frank, O. and Strauss, D., “Markov graphs,” *Journal of the American Statistical Association*, vol. 81, no. 395, pp. 832–842, 1986.
- Gallagher, H. C. and Robins, G., “Network statistical models for language learning contexts: Exponential random graph models and willingness to communicate,” *Language Learning*, vol. 65, no. 4, pp. 929–962, 2015.
- Geyer, C. J. and Thompson, E. A., “Constrained monte carlo maximum likelihood for dependent data,” *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 54, no. 3, pp. 657–683, 1992.
- Goldman, Z. W. and Martin, M. M., “Millennial students in the college classroom: Adjusting to academic entitlement,” *Communication Education*, vol. 65, no. 3, pp. 365–367, 2016.
- Harris, J., *An introduction to exponential random graph modeling*, ser. Quantitative Applications in the Social Sciences. SAGE Publications, 2014. [Online]. Available: https://books.google.com/books?id=scQ_AwAAQBAJ

- Hunter, D. R., Handcock, M. S., Butts, C. T., Goodreau, S. M., and Morris, M., “Ergm: A package to fit, simulate and diagnose exponential-family models for networks,” *Journal of statistical software*, vol. 24, no. 3, p. nihpa54860, 2008.
- Igel, C. and Urquhart, V., “Generation z, meet cooperative learning: Properly implemented cooperative learning strategies can increase student engagement and achievement,” *Middle school journal*, vol. 43, no. 4, pp. 16–21, 2012.
- Kolaczyk, E., *Statistical analysis of network data: Methods and models*. Springer, 01 2009.
- Kolaczyk, E. D. and Csárdi, G., *Statistical analysis of network data with R*. Springer, 2014, vol. 65.
- Loveland, E., “Instant generation.” *Journal of College Admission*, vol. 235, pp. 34–38, 2017.
- Park, J. and Newman, M. E., “Statistical mechanics of networks,” *Physical Review E*, vol. 70, no. 6, p. 066117, 2004.
- R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2013, ISBN 3-900051-07-0. [Online]. Available: <http://www.R-project.org/>
- Rickes, P. C., “Generations in flux,” *Planning for Higher Education*, vol. 44, no. 4, pp. 21–45, 2016.
- Schau, C. and Emmioğlu, E., “Do introductory itatistics courses in the united states improve students’ attitudes?” *Statistics Education Research Journal*, vol. 11, no. 2, 2012.
- Schau, C., Stevens, J., Dauphinee, T. L., and Vecchio, A. D., “The development and validation of the survey of antitudes toward statistics,” *Educational and psychological measurement*, vol. 55, no. 5, pp. 868–875, 1995.
- Seemiller, C. and Grace, M., *Generation Z goes to college*. John Wiley & Sons, 2016.
- Seemiller, C. and Grace, M., “Generation z: Educating and engaging the next generation of students,” *About Campus*, vol. 22, no. 3, pp. 21–26, 2017.
- Shatto, B. and Erwin, K., “Moving on from millennials: Preparing for generation z,” *Journal of continuing education in nursing*, vol. 47, pp. 253–254, 06 2016.
- Snijders, T. A., Pattison, P. E., Robins, G. L., and Handcock, M. S., “New specifications for exponential random graph models,” *Sociological methodology*, vol. 36, no. 1, pp. 99–153, 2006.

- Spears, J., Zobac, S. R., Spillane, A., and Thomas, S., “Marketing learning communities to generation z: The importance of face-to-face interaction in a digitally driven world,” *Learning Communities Research and Practice*, vol. 3, no. 1, p. 7, 2015.
- Strauss, W. and Howe, N., “Millennials go to college: Strategies for a new generation on campus,” *American Association of Collegiate Registrars, Washington, DC*, 2003.
- Turner, A., “Generation z: Technology and social interest,” *The Journal of Individual Psychology*, vol. 71, no. 2, pp. 103–113, 2015.
- Twenge, J. M., *iGen: Why today’s super-connected kids are growing up less rebellious, more tolerant, less happy—and completely unprepared for adulthood—and what that means for the rest of us*. Simon and Schuster, 2017.
- Wasserman, S. and Pattison, P., “Logit models and logistic regressions for social networks. i. an introduction to markov graphs and p ,” *psychometrika* 61 401–425,” *Mathematical Reviews (MathSciNet): MR1424909 Digital Object Identifier: doi*, vol. 10, 1996.
- Welch, B. L., “The generalization of student’s’ problem when several different population variances are involved,” *Biometrika*, vol. 34, no. 1/2, pp. 28–35, 1947.