There’s no ‘I’ in Class:

The Role of Collaboration and Gender on Success in a Collegiate Computer Science Class

Evan Green

Senior Thesis Submitted to the Department of Applied Mathematics of Yale University in partial fulfillment of the requirements for the degree of Bachelor of Arts

Advisor: Jay Emerson

May 4th, 2017

TABLE OF CONTENTS

Introduction

Collaboration is an important aspect of science, technology, engineering and mathematics (STEM) classes. In STEM classes, assignments typically have a small range of acceptable answers that every student must provide to obtain full credit. This increases the temptation to copy solutions from others and decreases the chances of getting caught plagiarizing. This contrasts with humanities and social science classes where every student's essay is distinct and plagiarism, if it occurs, would be more obvious. Additionally, STEM problems may require a key insight that can be easily communicated to others. Furthermore, group work and collaborative brainstorming may produce the correct answer more quickly than working in isolation. Finally, the importance of collaboration is magnified by the large portion of final course grades that depend on homework.

Students in STEM classes tend be overwhelmingly male, and there is a lot of research cataloging the differences in academic and career choices between men and women in STEM fields. Two overarching theories have been posited to explain these patterns: the deficit model and the difference model (Sonnert and Holton 1996). The deficit model emphasizes structural obstacles to success such as discrimination while the difference model focuses on the impact of people's internal goals and behavior such as differential interest in STEM fields (Sonnert and Holton 1996). My research investigates the question of whether collaboration is a structural barrier preventing women from succeeding in STEM classes. In fact, prior research has found that classroom climate, including group working conditions, may reduce women's confidence, especially in male-dominated fields (Hall 1982).

These gender disparities are heightened in Computer Science, which has become an overwhelmingly male field. Since the 1980s, the percentage of undergraduate degrees in computer science awarded to women has decreased from 37% to 17%, despite women making gains in many other STEM fields (Planet Money 2016). While there is evidence that as much as 82% of the gap in STEM bachelor's degrees can be attributed to differences before college (Legewie and DiPrete 2014), the college experience is still worth studying.

This project explores and analyzes students' grades and collaboration from a computer science class at Yale University to better understand the internal dynamics of the class. There are two major goals of this project. The first goal is to understand collaboration better by describing how it operates and determining its relationship to grades and learning. The class studied is a good example to use to investigate the importance of collaboration. The class has a reputation for being difficult, has over 100 students, and depends primarily on homework for calculating the final grade. In fact, many of the student reviews of the class mention the importance of collaboration. One student warned that “as long as you find a group to collaborate with, it’s a doable class” (Anonymous 2015). This heightened importance of collaboration makes this class especially suitable for analysis.

The second goal is to explore the gender dynamics of collaboration and grades to determine if and how different genders might be unequally granted opportunities. Focusing on the in-class dynamics can determine whether there are barriers to success that women face inside the classroom. A clearer understanding of the intersection of collaboration and gender can help make classes fairer either by cautioning against the acceptance of collaboration if collaboration is found to reinforce existing inequalities, or by encouraging collaboration if collaboration is found to reduce existing inequalities.

The Data

The data analyzed consists of students' grades and self-reported collaboration on homework in an upper-level computer science class at Yale University. The course consisted of seven problem sets assigned as homework and two tests. While students could collaborate on homework, the students had to complete tests individually. I scaled tests and problem sets to both be out of 30 points.

Information on collaboration was collected after each homework assignment via an online form. The professor instructed the students to identify from whom they had received help, including the option to state that you did not receive help from any other students. The professor strongly encouraged students to complete the forms by stating that failing to do so would be considered a violation of academic honesty and by offering points for completion of the forms. Per class rules, students were only allowed to receive help from a maximum of three other students.

The data began as adjacency lists and undertook the following transformations to become a network. Students were represented as nodes, and collaboration was represented as directed edges. When student A cites student B as a collaborator, there exists an edge B→A. While many of the edges are reciprocal (meaning A→B and B→A both exist), others are non-reciprocal. A network was made for each homework assignment since each assignment had different collaboration information.

Some students dropped the class after initially enrolling. Because the official drop deadline coincided with the due date of the fourth problem set, I will assume that any students who received no points after this assignment dropped the class. After the drop deadline, the class had 86 men, 23 women, and 1 non-binary student. This gender makeup exemplifies the large gender divide typical of STEM classes. When comparing performance in different metrics across gender, I exclude the non-binary student because there is not much that can be inferred from a single data point.[1]

To protect the privacy of the students and to comply with regulations, students were assigned random identifiers. Because I was a student in the class, my grades and collaborations were removed from the dataset. This deletion changes the network and is a source of error. To further safeguard students' identities, the grades on homework and

[1] The non-binary student was one of the top performers in the class but was not an active collaborator. The student received help from the same student five different times but never helped another student.

tests were perturbed in a hidden manner by the professor before he gave me the data for this project. The perturbed grades will be taken as is, but this perturbation is a further source of error.

The dataset also has other potential limitations. For example, students may not have been completely honest in their voluntary reporting of collaboration. While there are seven instances of students listing more than three collaborators, working with more students that allowed may have been underreported because these disclosures could have been treated as a violation of academic policies. It is also difficult to determine whether students who did not list any collaborators actually did not collaborate with anyone. However, 72% of students who list no collaborators have no students listing them as collaborators. The remaining 28% could have given help without receiving it. There was one example of a student listing everyone as a collaborator, which indicates students did not always take the online forms seriously; this response was removed from the dataset. Additionally, there are instances of students filling out the forms multiple times. At the advice of the professor, I used the last response to the form in such an instance. Furthermore, there are many other attributes that are potentially interesting and important but that are not available. These include overall GPA, class year, residential college affiliation, major, and whether individuals were taking the class pass/fail. Similarly, time spent on the assignment may be related to collaboration and important to students, but there was no time-related data Finally, I did not observe the selection process by which students chose to take this course, which biases the students left in the course. Despite these potential problems, I will proceed to analyze the data because of its uniqueness and granularity.

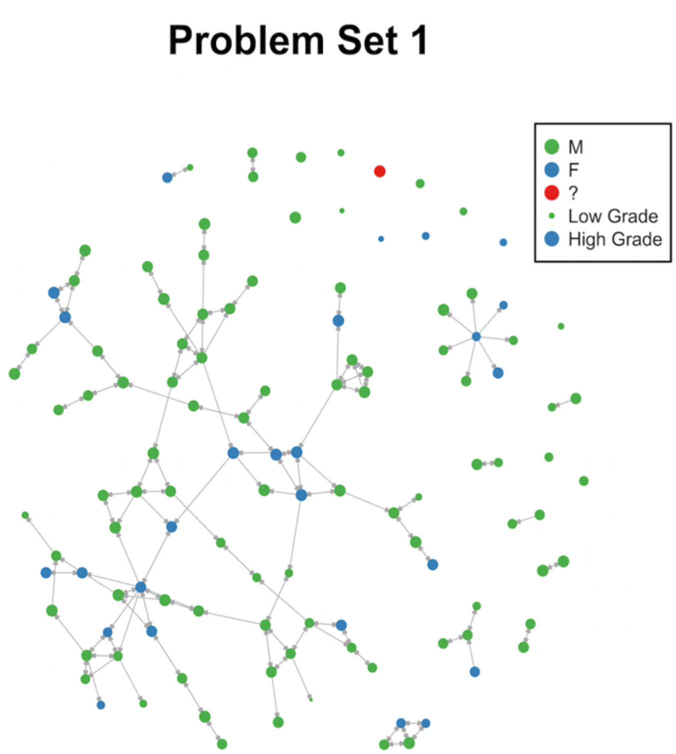
Imputation of Excused Absences on Tests

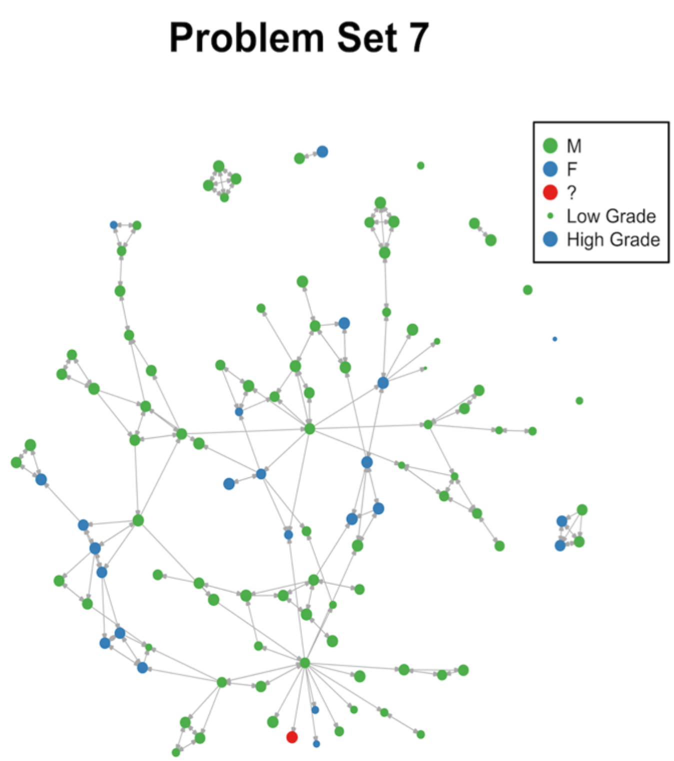
There were excused absences that left three out of 220 test grades with no data. There are two main strategies for dealing with missing data: remove the observations that have any missing data or impute the values. Because each student is important to the network structure, I did not remove these students. I used the technique of computing conditioned means. This method leads to unbiased estimates of means but underestimates variance and overestimates covariance if the missing data comes from the same distribution as the observed data (Huisman 2014). Because of the relatively few instances of missing data, these effects are small. The best model for predicting each test was a linear model that was pruned using a combination of the Akaike information criterion (AIC) and removing predictors that were insignificant. AIC is asymptotically equivalent to leave-one-out cross-validation (Stone 1977). Gender was ruled out as a predictor, and collaboration data was not considered as part of this imputation process. Finally, although I will proceed with the imputed data, it will be clear which test grades were imputed – unlike real grades, the imputed grades are not integers.

Analysis of the Network

Descriptive Statistics of Networks

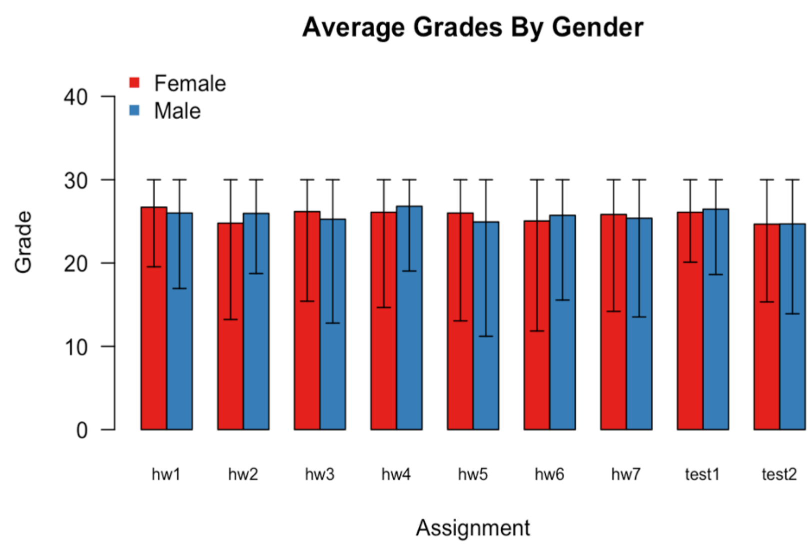
The networks were visualized with the Fruchterman-Reingold algorithm which places nodes by assigning physical forces to edges. This method creates visualizations that mimic the structure of the networks and illustrates a few key trends that are worth exploring. Nodes of the same gender are clustered because people tended to work with other students of the same gender, even with the overrepresentation of male students. Additionally, the highly active, largest connected components of the networks contain most of the students. The shape of each of the largest connected component demonstrates a disperse information-sharing network, implying there are not small, independent, fully-connected working groups. The largest connected component was at least 75 students in every problem set network and grew as the course went on, reaching 94 of the 110 students by the end of the course. Furthermore, the largest connected component tends to constrain students with higher grades, visualized by larger nodes.





The most fundamental properties in these directed graphs are in-degree, the number of students who helped you, and out-degree, the number of students whom you helped. The average in-degree and out-degree across all networks are both 1.8. While women had higher average in-degree (1.95 vs. 1.75) and out-degree (1.92 vs. 1.77) than men, both these differences are insignificant. Over 99% of the in-degrees are less than or equal to 3 and 94% of the out-degrees are less than or equal to 4. The highest out-degree was 27. The degree distributions of the graphs do not follow a power law, even though most real-world graphs' degree distributions do (Liljeros et al. 2001). This difference is likely caused by the class rule that students could only receive help from three others even though there was no restriction on the number they helped.

Grades on homework and problem sets did not differ significantly by gender. The grades by gender are indistinguishable on every assignment. Grades in this class were high with average grades around 26 out of 30 on most assignments, and median grades were one or two points higher than mean grades because of long tails.



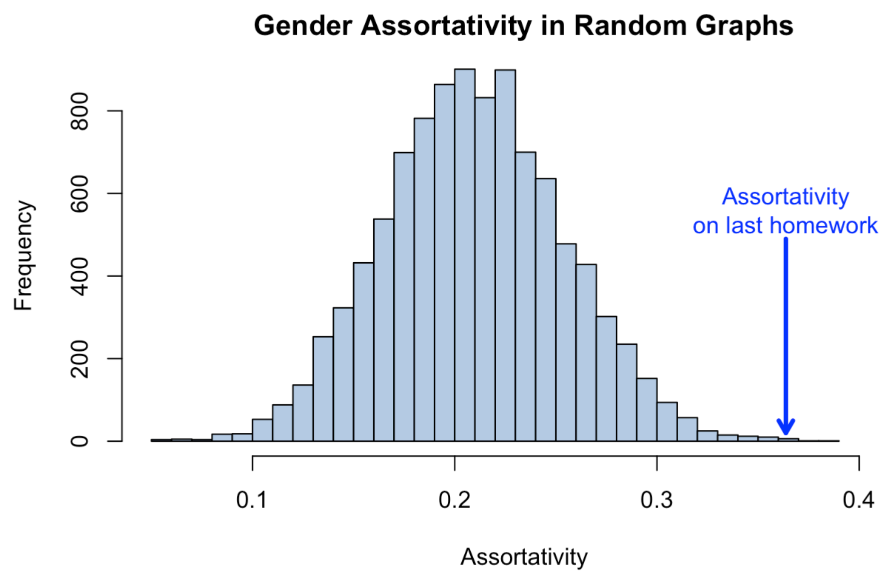
Evolution of the Network

Increase in Assortativity

Gender homophily increased as the course progressed. Homophily is the tendency for nodes that are similar in an external property (i.e. gender) to connect to each other. The assortativity coefficient measures homophily, is positive when homophily is present, and ranges from -1 to 1. Assortativity is calculated as $\frac{Trace(M)-sum(M)}{1-sum(M)}$ where M is the mixing matrix. Assortativity increased from .26 in the first problem set to .36 in the last problem set. This increase was caused both by fewer cross-gender edges and more within-gender edges.

To test whether this increase could have happened by chance, I took the core network, the set of edges that appeared in both the first and last problem, and randomly added edges until there was the same number of edges as in the graph for the last problem set. Randomly creating networks is a common technique in network studies to help determine significance (Bearman et al. 2004).

Out of the 10,000 simulated networks, only three had an assortativity coefficient higher than the observed assortativity coefficient in the last problem set which is strong evidence of inbreeding homophily.[2] This finding gives credence to the idea that gender is important to the students and the way in which the network evolved.[3]



This finding that within-gender edges became more prevalent as the course continued may be related to the different ways that men and women interact in group situations. Men have more negative perceptions of potential working partners, particularly women (Kuhn and Villeval 2013). These negative perceptions may have contributed to a greater lack of cross-gender ties than would have been expected by random chance.

---

[2] It is important to distinguish between baseline homophily and inbreeding homophily. Baseline homophily is the homophily you would expect from random ties due to the prevalence of different groups while inbreeding homophily is the deviation above that random model (McPherson et al. 2001). Inbreeding homophily is more interesting in this analysis because it corresponds to the differential treatment of in-group and out-group members.

[3] Other network features that evolved over time, such as the increase in size of the largest connected component and a decrease in the number of isolated students, were consistent with the evolution of the same features in random networks. These trends were driven by the increase in overall collaboration, not a specific norm in the class.

Additionally, because men tend to talk more during group work and are more likely to interrupt (Hall 1982), mixed-gender working groups may be less effective and enjoyable, reinforcing the tendency to work with one's own gender.

Stability of Collaboration

Although the networks evolved during the course, much of the collaboration was consistent. 48% of edges appear either in six or seven of the seven homework collaboration networks. Additionally, no two networks have less than 51% overlap in edges, and the average overlap between networks was 60%.[4] To determine the features that led to more edge stability, I categorized each edge that ever appeared by the number of networks the edge appeared in as well as structural and grade information associated with the students. I then used linear regression to predict the count of the edges.

Many different aspects were correlated with edge persistence. Cross-gender edges were slightly less likely to continue as the coefficient for Same_Gender was small and only significant to the 0.1 level. Transitivity being great than or equal to $\frac{1}{4}$ led to significantly more collaboration.[5] This suggests that groups are more stable than one-off partnerships. Additionally, both measures of reciprocity, the fact that the reciprocal tie exists at all (Is_Ever_Reciprocal) and the percentage of times the tie is reciprocal (Reciprocal_PCT), were significant predictors of more collaborations perhaps illustrating an aversion to free-riding by the students giving help without receiving any back.

Table 1: Predicting Count of Edges

	Dependent variable:
	Count
Same_Gender	0.288* (0.164)
Is_Ever_Reciprocal	1.111*** (0.279)
Reciprocal_PCT	0.784*** (0.300)
Is_Helper's_Grade_Below_Average_During_Collaboration	−1.280*** (0.193)
Is_Helpee's_Grade_Below_Average_During_Collaboration	−1.064*** (0.198)
Helper's_Course_Average_Grade	0.131*** (0.025)
Helpee's_Course_Average_Grade	0.171*** (0.034)
Helper's_Average_Grade_During_Collaboration	−0.122*** (0.032)
Helpee's_Average_Grade_During_Collaboration	−0.157*** (0.037)
Transitivity_PCT >= $\frac{1}{4}$	1.198*** (0.209)
Constant	2.302*** (0.200)
Observations	520
R^2	0.443
Adjusted R^2	0.432
Residual Std. Error	1.684 (df = 509)
F Statistic	40.515*** (df = 10; 509)
Note:	*p<0.1; **p<0.05; ***p<0.01

[4] Overlap between networks was calculated via the Jaccard index of the edges of the networks.
[5] Transitivity was measured as the Jaccard index of the set of individuals that each of the students collaborated with during the semester

Another important factor was the success of the collaboration. When grades for students while collaborating were worse than those students' average course grade, there was a strong aversion to continuing to collaborate. This is shown by the large, significant, negative coefficients on Is_Helper's_Grade_Below_Average_During_Collaboration and Is_Helpee's_Grade_Below_Average_During_Collaboration. Furthermore, students who received better grades tended to have more stable collaboration (see Helper's_Course_Average_Grade and Helpee_Course_Average_Grade). Whether this was because more stable collaboration led to better grades, because the students who received better grades preferred more stable collaboration, or because of another factor is impossible to determine.

Conclusion of Network Analysis

The collaboration networks of this class are clustered by gender and dominated by a single connected component. Over time, more students become participants in the network. Although prior research has found that women are generally more attracted to collaborative working environments than men (Kuhn and Villeval 2003), there is no convincing evidence here to support that claim because there are no significant differences by gender in the amount of collaboration or in grades received. Collaboration in the network appears to be controlled by several norms. Collaborative relationships that are reciprocal and transitive persist at much higher rates than those relationships that are not. Additionally, partnerships that are single-gender and that receive high grades are more likely to stay together than those across gender or those with lower grades. These findings highlight the focus students may place on fairness and reciprocity as well the preference for working with others of the same gender.

Grade Regressions

Introduction

This project strives to not just describe the norms and dynamics of collaboration but to also understand its connection with grades. When collaboration occurs in a course, it can be difficult to determine what is truly collaborative, in which students are working together to solve the problems, and what is essentially one student copying the answer from another. Students copying is worrisome since these students will presumably not learn the material as well. Additionally, there is the worry that students who collaborate have a big advantage over those who do not. One course reviewer complained, cautioning future students that “if you don't have friends in the class, that just means that your grades will be lower than everyone else who is able to collaborate - beware of that!” (Anonymous 2016). I will analyze and explore these two phenomena, both demonstrating the effects of collaboration and analyzing the potential abuses.

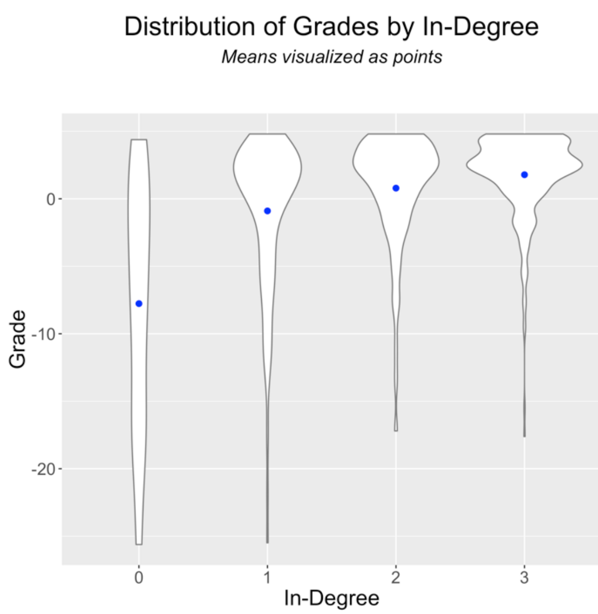
Measuring the influence of collaboration is difficult. It is hard to distinguish the effects that come from individual students’ abilities and efforts when the only individual assignments, tests, may be influenced by the increased or decreased learning that occurs when collaborating with others. Additionally, because collaboration is quite consistent throughout the semester, grades on other problem sets reflect both the students’ abilities and their level of collaboration on those other problem sets. A final difficulty in predicting grades is that there are a few students for each problem set whose grades are very disparate from their grades on other assignments. This may be due to random events that were happening concurrently in students’ lives.

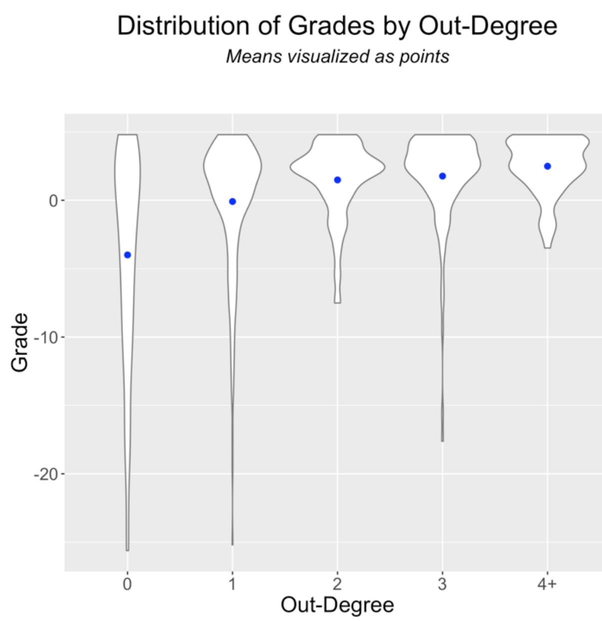
Methods

Because each assignment had a different grade mean, I analyzed grades after subtracting out the means. Additionally, while there were some time dependent trends such as neighboring problem set grades and collaboration being slightly more correlated, I found better results by ignoring which problem set a grade came from. However, to avoid assuming independence when it does not exist, cross-validation on models predicting problem sets occurred by holding out each problem set, training on the rest of the data, and then measuring performance on the held-out data. Grades within each problem set are not independent since grades between collaborators are correlated. All model selection and accuracy over baselines are reported using the held-out data while model coefficients and measures of significance are reporting using the full data set to get the most accurate representation possible.

Initial Exploration

To get a first approximation of the relationship between grades and collaboration, I created a series violin plots which are like boxplots but feature rotated kernel density plots to better illustrate the distribution of data across categories. There are several important observations from these plots. For both in-degree and out-degree, there is a positive correlation between grades and collaboration but also diminishing marginal returns to more collaboration, especially after going from zero collaborators to one. Furthermore, the skew of the data is apparent. While most grades are around zero (the mean), there are long tails in the negative direction. While this analysis does not illustrate any causal information, these plots suggest that more collaboration is related to higher grades.





Models with only Network Structure

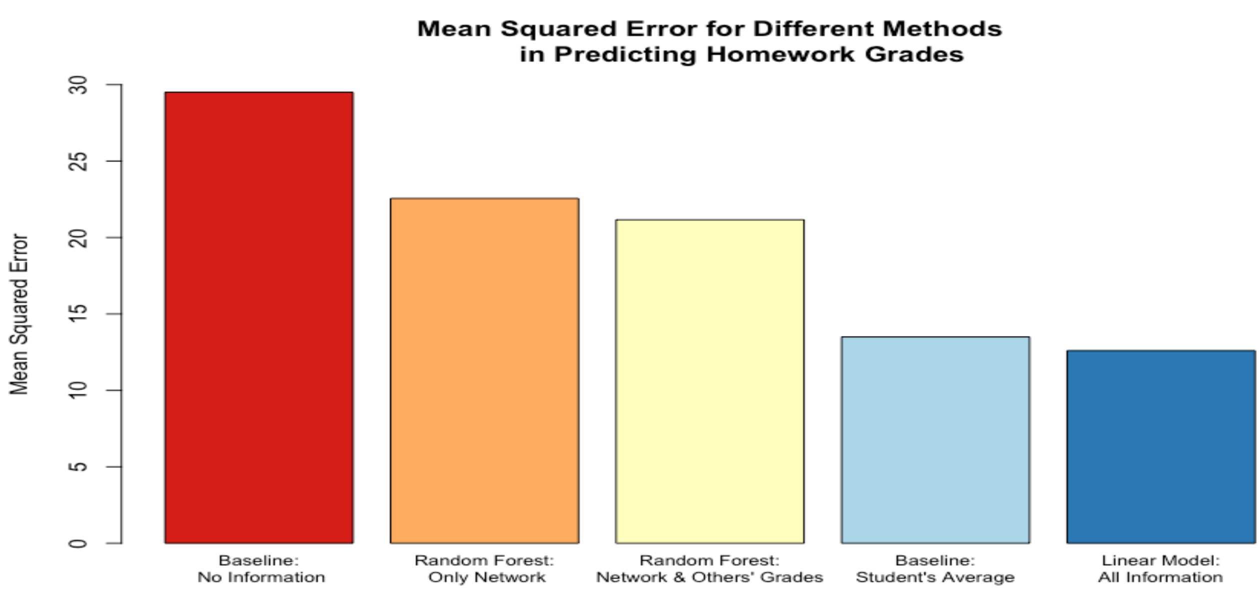
I then fit a variety of models with access to different information. Comparing the fits of these models will give some insight into which of these features are most important for predicting grades, and perhaps success in the class.

I first made models for two subsets of the available information: the first only had access to network information while the second also had extra information about the grades of collaborators. The best models for both scenarios were random forests. For both models, the baseline is a naïve model that predicts the class mean for every person. Both models drastically outperformed the baseline but performed similarly to each other despite the increase in information available in the second model. This similar performance is likely because collaboration and good grades are correlated in a way that allows the simpler model without collaborators' grade information to get an approximation of this information. This explanation is supported by the fact that membership in the largest component was helpful for the network first model was removed from the second. Membership in the largest component is related to higher grades but in a weaker way than collaborators' grades.

Additionally, the measure of centrality that was most helpful for these predictions and that was included in both models is constraint. Constraint is defined as

$$C_i = \sum_{j \in V_i, j \neq i} \left[\sum_{q \in V_i, q \notin \{i,j\}} (p[i,j] + p[i,q]p[q,j]) \right]^2$$

where V_i is set of nodes connected to i, and $p[i,j]$ represents elements in the row-normalized adjacency matrix. Constraint is lower for students with less redundant ties and ranges from 0 to 1. Because constraint is undefined for isolated students, I set constraint to 1.5 for these students to overcome this difficulty. Since random forests were used for these models, the important thing is that there is a cut point that would allow the random forests to disambiguate this arbitrary value from the real values of constraint. Nevertheless, the value of 1.5 also fits well with the linear trend. Constraint measures access to disparate sources of information which might explain its predictive strength. Connections to different groups of students might give you exposure to more ideas and a better chance of getting the correct answer to a homework question. It also may be the



case that better students have lower constraint because many others seek to collaborate with them. Lower constraint has been linked in professional settings to the formulation of better ideas, higher salary, and promotions (Burt 2004). This corroborates the finding that constraint can be an important determiner of success, especially for complex tasks such as computer science assignments.

Unrestricted Models

Linear Model for Problem Sets

I then created a model to predict grades with access to all available information. Importantly, this model had access to individuals' grades on other assignments and used student identities as a factor variable as well as all the network information. This method creates "ability scores" for each student which are an estimate of each student's individual strength after accounting for collaboration.[6] The baseline for this model was an individual's grades on other assignments. Students' grades were a great predictor and made achieving further improvements difficult but possible. The mean squared errors of both the baseline and the trained model are about half as big as the errors of the models that did not have access to the identity of individual students.

Table 2: Predicting Grades without Restriction

	Dependent variable:
	Grade
Avg_Grade_of_Helpees	0.334***
	(0.045)
Out_Degree	0.172**
	(0.073)
Observations	990
R^2	0.644
Adjusted R^2	0.599
Residual Std. Error	3.334 (df = 878)
F Statistic	14.301*** (df = 111; 878)
Note:	*p<0.1; **p<0.05; ***p<0.01
Note:	Ability scores are removed for brevity

Nevertheless, there were significant improvements to be made by including information from the network. These improvements were smaller because the effectiveness of the baseline. Here, the challenge is not finding the

[6] I also attempted to use expectation maximization to simultaneously fit coefficients for each individual's own ability, α_i, as well as coefficients β and γ that measured the effect of the edges coming in and out respectively according to the following formula.

$$Grade = \alpha_i + \beta\left(\frac{1}{d\_in(i)}\sum_{k \in in\_neighbors(i)} \alpha_k\right) + \gamma\left(\frac{1}{d\_out(i)}\sum_{j \in out\_neighbors(i)} \alpha_j\right) + \epsilon$$

However, this method did not produce any improvements from the baseline model of individuals' average grades. Furthermore, the estimates of ability changed very little from this baseline of ability scores being set equal to students' average grade. I suspect the failing of this method is the reliance more on the source of the collaboration than whether collaboration existed. In more successful models, the most helpful information is whether collaboration occurs. Additionally, the reliance on ability scores neglects fine-grained information such as the grade of collaborators on a specific problem set. This method doesn't allow for learning when a group of students may be performing worse than their average on a given problem set.

best students, but predicting deviations from a student's personal class average. Interestingly, the model's network features (Avg_Grade_of_Helpees and Out_Degree) are related to the help given out by a student, not the help received. This ignores the distinction between reciprocal and non-reciprocal ties because out-degree alone is a better predictor. I have two possible theories for why help given is so important. One hypothesis, is that by acting as a teacher to someone seeking help, a student learns the material better and this extra learning is reflected in higher grades. The second possibility, which I find more compelling, is that students confident in their answers will give more help. Therefore, looking at the help given by students gives a proxy for their confidence. This also assumes that students are good judges of the accuracy of their answers, an assumption I believe to be correct from experience.[7]

Benefits of Collaboration for Each Gender

To assess the impact of collaboration on men and women, I compared the percentiles of grades to the percentiles of "ability scores," which were created by the unrestricted linear model. Those with higher grades than ability scores were helped by collaboration while those with higher ability scores than grades may have benefitted from collaboration not being allowed. While women were more likely to receive help (61% vs. 49% for men), this difference is insignificant. This is parallel to the finding that women collaborate more than men in a statistically insignificant manner. Contrary to my expectations that men might receive more help from collaboration because there are so many more men in the class and the historical dominance of STEM by men, both genders received similar amounts of help. One possible explanation for this phenomenon is that men are more confident in their abilities (Kuhn and Villeval 2003; Hall 1982) and believe that they don't need to collaborate, even when collaboration could help them. Another possibility is that women are active in the network because of selection effects. Since people expected collaboration to be an important part of the class, perhaps only those with friends in the class may have chosen to stay. As the minority, this effect may have been stronger on the women, causing the women who stay in the class to more well-connected than one might expect given the underrepresentation of women and tendency to work with others of the same gender.

[7] These conclusions are dampened because this linear model, chosen for the best performance on the held-out data and AIC, is one of many similar models that can achieve similar performance with different features.

Linear Model for Test Grades

I modeled test grades as a function of an individual's average homework grades and collaboration activity to test for evidence of collaboration leading students to copy others' answers without learning the material. Because tests are individual assignments, they serve as a window into what a student can do without the chance to receive any help. I used individuals' average homework grade as a baseline for this model. To prevent overfitting, I used 10-fold cross-validation that randomly chooses which students to include. Test grades for different people are independent since there is no collaboration on tests.

A linear model outperformed the baseline by reducing cross-validated mean squared error by 15% from 12 to 10.2. Average homework grades were a particularly good predictor of test grades probably because of the nature of the class's tests. The tests did not require students to solve new problems but instead had students solve problems that previously appeared in lecture or on a homework assignment. The remaining coefficients suggest an interesting story. After conditioning on homework grades, collaboration led to predictions of lower test grades, particularly of those free-riding. While collaboration led to lower test grades, the extent to which a students' average in-degree was higher than average reciprocal degree led to a larger expected penalty in test grades. It seems that receiving help on homework from students with high grades helps you on a test, possibly because of learning from these peers. However, there is a larger negative effect from the grades of those whom you help because of the larger coefficient on Avg_HW_Grade_of_Helpees than Avg_HW_Grade_of_Helpers. This model provides evidence for the concern that students who collaborate more will not learn as much. The fact that collaboration, and particularly non-reciprocal collaboration, is associated with lower test grades is an important trend to recognize.

Table 3: Predicting Test Grades

	Dependent variable:
	Grade
Avg_HW_Grade_of_Helpers	0.726***
	(0.202)
Avg_In_Degree	−1.885***
	(0.510)
Avg_HW_Grade_of_Helpees	−0.874***
	(0.197)
Avg_Reciprocal_Degree	1.703***
	(0.505)
Avg_HW_Grade	0.794***
	(0.065)
Constant	1.243*
	(0.685)
Observations	220
R^2	0.546
Adjusted R^2	0.536
Residual Std. Error	3.149 (df = 214)
F Statistic	51.562*** (df = 5; 214)
Note:	*p<0.1; **p<0.05; ***p<0.01

Simulating Grade Data to Test the Null Hypothesis

To measure the robustness of these models, I created 1,000 simulated data sets under the null hypothesis that homework and test grades are equal to people's average grades plus a random error term.

This reduced the performance of the random forest predicting grades solely with network information by the least, but still significant, amount. This illustrates that the features included in that model (constraint, reciprocal degree, in-degree, out-degree, and membership in the biggest component) are features that mark higher grades but still are dependent on the dynamics of the network. The random forest that predicted grades based on the network information and the grades of collaborators declined much more in performance but still outperformed the baseline. Since this model relied heavily on the performance of collaborators and now performs worse, this illustrates that grades of collaborators, and specifically, deviations from their average grades, are more correlated than what would happen by chance alone.

In both linear models, the one predicting grades on homework and tests conditioned on all available information and the one predicting test grades based on average level of collaboration, not a single coefficient had a 95% confidence interval that excluded 0. These models were now equivalent to the baseline models. This finding gives more confidence to the results that the two linear models examined are detecting important features instead of following noise in the data and helps to reject the null hypothesis that the network structure is not connected to deviations in grades.

Conclusion of Predictive Modeling

There are several main takeaways from these different models. While more collaboration is predictive of and correlated with higher grades, it is not possible to determine causal impacts. The data does not contain any natural experiments. Nevertheless, the network structure is inextricably linked with grades. Diversity of collaborators appears to be advantageous because lower constraint was a good predictor of higher grades. Furthermore, help given by a student appears to be a more helpful predictor than help received. Additionally, gender was never a useful or significant predictor of grades in any model. Finally, while there is evidence of free-riding, tests can expose this behavior.

Conclusion and Recommendations

Collaboration was an important part of the class and was correlated with many features of interest. While gender is a salient feature in the networks, neither side is significantly more active. Despite women making up less than a quarter of the class, the amount of collaboration, grades received, the estimates of help received from collaboration do not significantly differ by gender. While there is no evidence of either gender being disadvantaged by collaboration, within-gender edges were more stable, more likely to be reciprocal, and became more prevalent as the course continued. The other important norms enforced by students in the network were reciprocity and transitivity. Relationships that lacked these features were less likely to persist.

Collaboration in this class was characterized by a diffuse network, with most activity occurring in the largest connected component. Collaborating with classmates is correlated with higher grades. One network feature that predicted higher grades is lower constraint, a measure of the redundancy of a student's ties. This highlights the advantage of accessing multiple information sources. However, causal effects are elusive. There were no natural experiments and the consistency of collaboration made untangling the impact difficult. Although the best predictor of grades was the grades on other assignments, models could be improved by adding network information. There is some evidence of free-riding on problem sets. After conditioning on homework grades, test grades are lower for more active collaborators, especially for those who received more help than they gave.

Overall, the collaboration system seems to be working well, but there are ways for it to be improved. While most students collaborate, an effort to ensure that all students have access to collaborators could help students who don't know other students in the class. Encouraging more cross-gender collaboration could reduce the gender segregation in the network and lead to a more open and welcoming class. Although all genders succeeded equally in the class, the overwhelming majority of the students are male so attracting more non-male students to the class is another potential improvement. Finally, putting a larger emphasis on the individual tests could reduce the incentive to free-ride on homework and would reward the students who are learning the most.

Acknowledgements

Thank you to Jay Emerson for advising me on this thesis, challenging my ideas, reminding me not to infer causation, and thinking of ways to measure different effects.

Thank you to Dan Spielman for collecting and allowing me to use this unique data set, pushing me in interesting directions on this project, and advising me as an Applied Mathematics major.

Thank you to Andrew Papachristos for teaching me about network theory and brainstorming ideas to explore in my thesis. Thanks also to my group in “Social Networks and Society" made up of George Gemelas, Sophie Deans, Stella Shannon, and Tiana Wang who worked with me on a related class project in the Fall of 2016 that inspired this project.

Thank you to Joscha Legewie for giving me another perspective on gender dynamics in classrooms.

Thank you to my family for supporting me and encouraging me to explore my interests. A special thanks to Ben Green for giving me a reading list of network studies and to Jenny Altshuler for help with edits.

Thank you to Kira Tebbe for being a sounding board for questions about my thesis, helping me organize and edit the final product, and for lending me an artistic eye for visualizations.

Works Cited

Anonymous. (2015). Overview of Evaluations: CPSC 365 01 Design & Analysis of Algorithms.

Anonymous. (2016). Overview of Evaluations: CPSC 365 01 Design & Analysis of Algorithms.

Bearman, P. S., Moody, J., & Stovel, K. (2004). Chains of Affection: The Structure of Adolescent Romantic and Sexual Networks. *American Journal of Sociology*, *110*(1), 44–91. https://doi.org/10.1086/386272

Burt, R. S. (2004). Structural Holes and Good Ideas. *American Journal of Sociology*, *110*(2), 349–399. https://doi.org/10.1086/421787

Hall, R. M. |Sandler. (1982). The Classroom Climate: A Chilly One for Women?. Retrieved from https://eric.ed.gov/?id=ED215628

Hlavac, M. (2015). stargazer: Well-Formatted Regression and Summary Statistics Tables (Version R package version 5.2).

Huisman, M. (n.d.). Imputation of Missing Network Data. *ResearchGate*. Retrieved from https://www.researchgate.net/publication/264896130_Imputation_of_Missing_Network_Data

Kuhn, P. J., & Villeval, M.-C. (2013). *Are Women More Attracted to Cooperation Than Men?* (Working Paper No. 19277). National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w19277

Legewie, J., & DiPrete, T. A. (2014). Pathways to Science and Engineering Bachelor's Degrees for Men and Women. *Sociological Science*, *1*, 41–48. https://doi.org/10.15195/v1.a4

Liljeros, F., Edling, C. R., Amaral, L. A. N., Stanley, H. E., & Åberg, Y. (2001). The web of human sexual contacts. *Nature*, *411*(6840), 907–908. https://doi.org/10.1038/35082140

McPherson, M., Smith-Lovin, L., & Cook, J. (2001). Birds of a Feather: Homophily in Social Networks. *Annual Review of Sociology*, *27*(1), 415–444. https://doi.org/10.1146/annurev.soc.27.1.415

Planet Money. (2016). Episode 576: When Women Stopped Coding. Retrieved April 24, 2017, from http://www.npr.org/sections/money/2016/07/22/487069271/episode-576-when-women-stopped-coding

Shirado, H. (2016). *analysis_advise_networks.R*.

Sonnert, G., & Holton, G. (1996). Career Patterns of Women and Men in the Sciences. *American Scientist*, *84*(1), 63–71.

Stone, M. (1977). An Asymptotic Equivalence of Choice of Model by Cross-Validation and Akaike's Criterion. *Journal of the Royal Statistical Society. Series B (Methodological)*, *39*(1), 44–47.

Appendix

The R code used in this project can be found on GitHub at
https://github.com/evanrgreen/Senior_Thesis.

To run the R files on your local computer, you should first clone the repository. Then, you will need to change the path in the first chunk of each file to the indicate where you have put the folder entitled "Grade Data." Additionally, if you do not have all the packages that are imported in the first chunk installed, you will need to install those before the code will run. You can also run Install_Packages.R in the R Files folder.

There are 4 R files that were integral to the project. They are all in the R Files folder. The PDF versions of the files are also in this folder. The PDF files of the code are edited to remove redundancy across files and to remove exploratory code.

Analyze_Grade_Networks.RMD

The file performs analysis about the network properties. This includes creating random graphs to test for the likelihood of different observed properties, calculating and visualizing the average grades and degree distribution by gender and analyzing the factors that lead to more stable edges.

Grade_Regressions.RMD

This file tests and validates various models that predicted the grades of students based on different assumptions and available information.

Imputing_Missing_Tests.RMD

This file tests different models for imputing the excused absences on tests and outputs a new file that has a complete, imputed set of grades.

Visualize_Network.RMD

This file uses a few different packages to create information visuals of the network itself.