Table of Contents

Introduction 1

The Data 3

Imputation of Excused Absences on Tests 4

Analysis of the Network 5

Descriptive Statistics of Networks 5

Evolution of the Network 6

Decrease in Assortativity 6

Decrease in Number of Components 8

Charity INCLUDE?? 8

Stability of Collaboration: 9

Conclusion of Network Analysis: 10

Grade Regressions 11

Baseline Models without Student’s Identity 12

Unrestricted Models 14

Expectation Maximization 15

Help Received from Collaboration by Gender 15

Regressing Tests: 16

Simulating Grade Data to Test Null-Hypothesis 17

Conclusion of Predictive Modeling 17

Conclusion and Recommendations 17

Acknowledgements 19

Appendix 20

# 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 reach to obtain full credit. This contrasts with humanities and social science classes where every student’s essay will likely be distinct or plagiarism will be much clearer. Additionally, problems may require a key insight which is easily communicated to other people. Furthermore, as in other disciplines, group work and brainstorming may lead more quickly to the correct answer than if people work in isolation. Finally, the importance of collaboration is enhanced by the relatively large portion of final grades that depend on homework.

Another characteristic of STEM classes is the large gender disparity. There is a wide body of research cataloging the differences in career choices and trajectories 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 while the difference model focuses more on internal goals and behavior (Sonnert and Holton 1996). My research investigates the question of whether collaboration is one of these structural barriers preventing women from succeeding in STEM fields. 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).

Computer Science has become an overwhelmingly male field. Since the middle of the 1980s, the percentage of women working in computer science related professions has steadily decreased. Women have earned only about 15% of the undergraduate degrees in computer science in recent years despite earning 37% in 1985 (Women in Computer Science 2017). While there is evidence that as much as 82% of the gap in STEM bachelor’s degrees is attributable to aspects of education in high school or before (Legewie and DiPrete Pathways 2014), the college experience is still worth studying. Many initiatives including “Girls Who Code” and “NASA Women” have attempted to combat the dearth of women in computer science by increasing women’s opportunity, participation, and interest.

This project explores and analyzes grades and collaboration from a computer science class at Yale 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 in this context and determining its relationship to grades and learning. The class studied is a particularly good example of the importance of collaboration. The class has a reputation for being difficult, the class has many students, and the homework makes up the majority (75%) of 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 will make this class especially suitable for analysis.

The second goal is to explore the gender dynamics of the collaboration and grades to determine if and how different genders might be granted opportunities unequally. This essay aims to help understand and explain these trends by examining granular data from a computer science class. Focusing on the in-class dynamics that most organizations do not have access to, can expose barriers that women face inside the classroom that may be unrelated to external factors. A clearer understanding of the intersection of collaboration and gender can help make classes more data-driven and fair.

# The Data

The data analyzed consists of the grades (on both homework assignments and tests) and self-reported collaboration on homework by students in an upper-level computer science class at Yale University. There were seven problem sets assigned as homework for the course. There were also two in-class tests. While tests posed similar questions as homework problems, the students had to complete them individually, despite being able to collaborate on homework. 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 whom they had received help from, including the option to state that you did not receive help from any other students. The professor strongly encouraged student to complete the forms by stating that failing to do so was 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 up to 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 lists 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), it is possible for only one of these to be in the graph. A network was made for each homework assignment since each homework assignment had different collaboration information.

Some students dropped the class after enrolling. Because the official drop deadline coincided with the due date of the fourth problem set, I will consider that any students who received no points after this date to have dropped the class. After the drop deadline, the class had 86 men, 23 women, and 1 non-binary student exemplifying 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 to be learned from a single data point.[[1]](#footnote-1)

To protect the privacy of the students and to comply with regulations, students were assigned random identifiers. Additionally, since I was a student in the class, to take further precautions, my grades and collaborations were removed from the dataset. This a source of error as removing me changes the network. To further safeguard students’ identities, the grades on homework and tests were perturbed in a hidden manner by the professor before transferring the data. The perturbed grades will be taken as is, but this perturbation is a potential source of error.

There are other potential limitations of the data set. For example, students may not have been completely honest in their reporting. While there are seven instances of students listing more than the allowed number of collaborators, this behavior 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 the students who list no collaborators also have no students list them as collaborators and the remaining 28% may 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 and was removed from the dataset. Additionally, there are instances of students filling out the surveys multiple times. At the advice of the professor, I used students last response to the form. Finally, there are many other attributes that are potentially interesting and important but which are not available. These include overall GPA, class year, residential college affiliation, major, and whether individuals were taking the class pass/fail. These could potentially be confounding variables. Despite these potential problems, the data is rich enough and robust enough to support analysis.

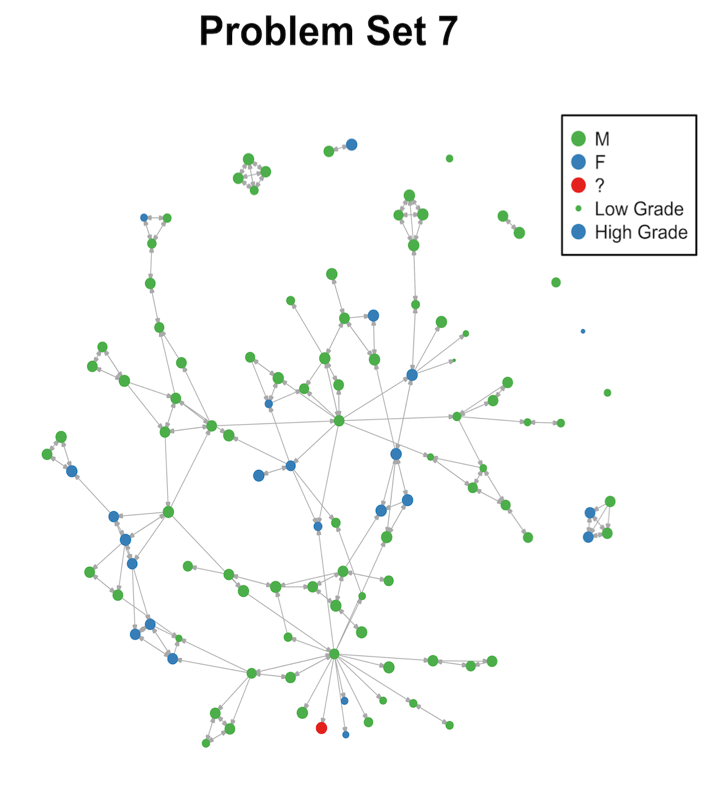
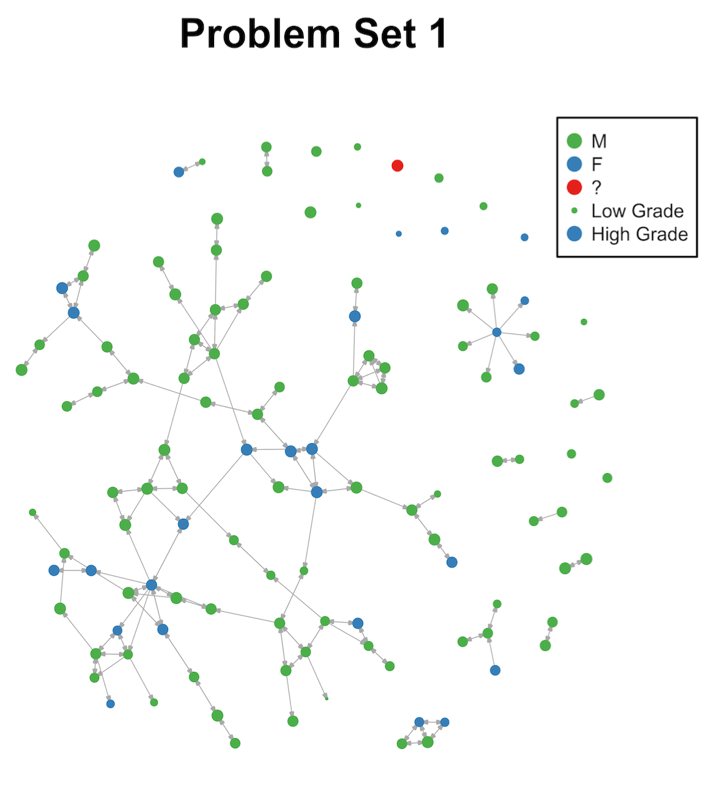
## Imputation of Excused Absences on Tests

There were excused absences that left three out of 220 test grades blank. There are two main strategies for dealing with missing data. Remove the observations that have missing data in any column or impute the value. Because each student is important to the network structure, I did not remove these students from the data. I used the technique of computing conditioned means which leads to unbiased estimates of means but underestimates variance and overestimates covariance (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 by insignificance and collaboration data was not included as predictors. Finally, although I will proceed with the imputed data, it will be clear which grades were imputed and which were not as the imputed grades are not integers.

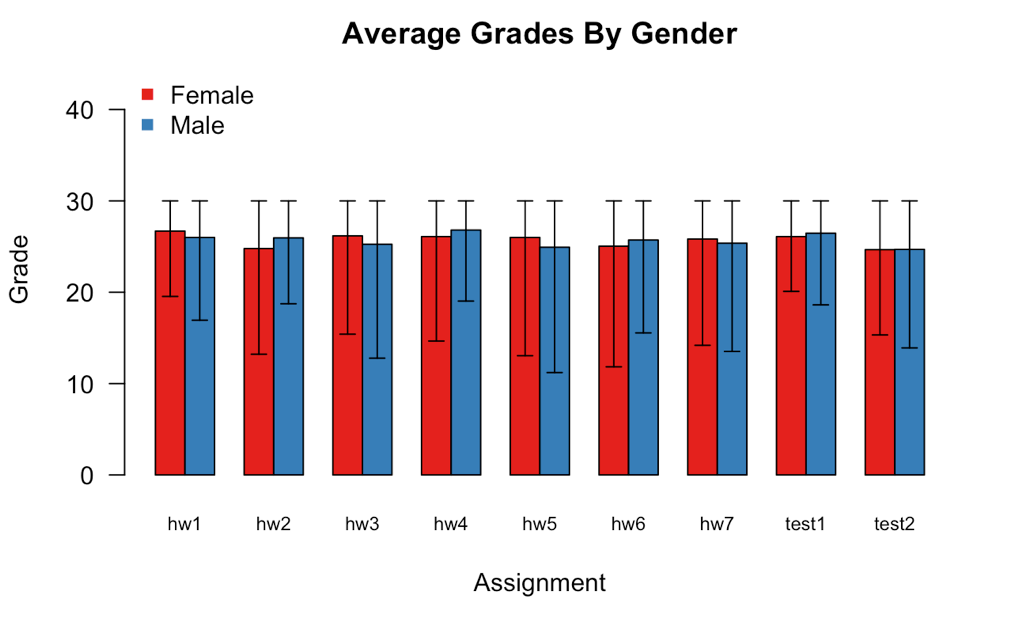
# Analysis of the Network

## Descriptive Statistics of Networks

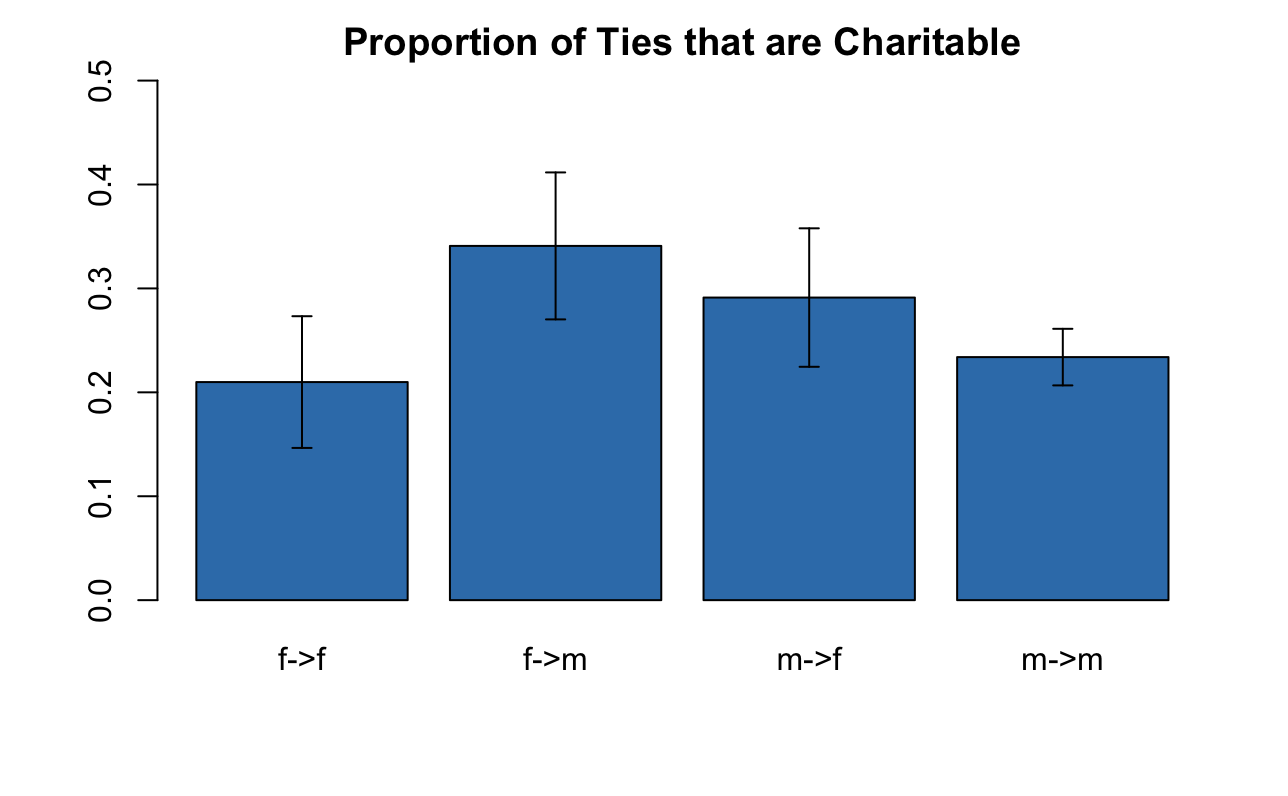
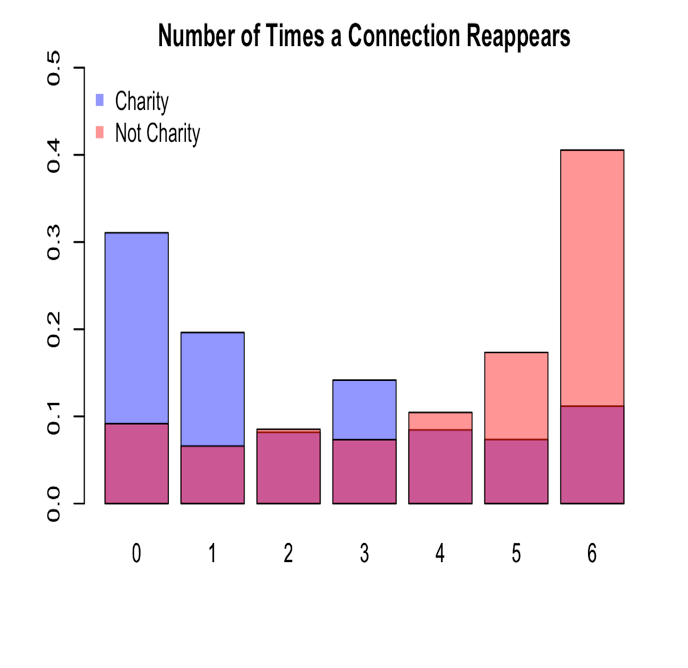
An examination of the network diagrams illustrates a few key trends that motivate further exploration. While network diagrams are not unique, the Fruchterman-Reingold algorithm used to create the position of nodes uses the properties of the graph to place nodes that are connected closer to each other and to minimize edge crossings by assigning forces to edges. Nodes of the same genders are clustered because people tended to work with other students of the same gender. Still, there are quite a few inter-gender edges. The overrepresentation of male students in the class is also apparent from these diagrams. An additional property is the highly active core of the network, where most students reside. The shape of the largest connected component the disperse information sharing network that exists. This sharply contrasts with small, independent, fully-connected working groups. The largest connected component was at least 75 students in every problem and grew as the course went on, reaching 94 of the 110 students in the last problem set. Finally, as nodes are sized by grade on the assignment, the higher grades (larger nodes) tend to be concentrated in the highly-connected core while the worse grades are the disconnected nodes on the periphery.



The most basic 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. Women had higher average in-degree (1.95 v. 1.75) and out-degree (1.92 to 1.77), but both these differences are insignificant. Over 99% of the in-degrees are less than 4 and 94% of the out-degrees are less than 5. The highest out-degree observed was 27. The degree distributions of the graphs do not follow a power law even though most real-world graphs’ degree distributions do follow a power law (Liljeros et al.). This difference is likely caused by the rule that you could only receive help from three other students.



Scores on homework and problem sets did not differ significantly by gender. In every assignment, the grades by gender are indistinguishable and there are no clear patterns. Additionally, grades in this class are typically high, with averages of about student A received at least as good a grade as student B. I grouped edges by the gender of the helper and the helpee and compared the proportions. After applying the Holm correction for multiple comparisons, inter-gender ties were significantly more likely to be charitable than intra-gender ties. While the conclusion that women were in general more likely to give charity than men was found in the prior study was not supported, there are interesting results. One reason that inter-gender ties were more likely to be charitable is that while people mostly worked with the same gender when groups were stuck they may have asked for help solving the problems from someone who they didn’t typically work with. In fact, charitable ties appeared again in only 2.15 other problem sets on average while non-charitable ties appeared in 4.17 other problem sets on average. This difference is highly statistically significant.



## Stability of Collaboration:

Despite the significant evolution of the networks, collaboration is largely consistent over time. 48% of edges are in part of edges that appear either in 6 or 7 of the 7 networks. Additionally, no two networks have less than 51% overlap in edges and the average overlap is 60% between networks. I categorized each edge that appeared any network by the number of times the edge appeared as well as structural and grade information associated with the endpoints. I then used a linear regression to determine which features led to more edge stability.

Cross-gender edges were less likely to remain, despite there being no difference based on the genders involved or direction, only that edges spanned different genders. When transitivity, measured by the Jaccard index of the set of individuals that each of the collaborators collaborated with during the semester, was greater than or equal to el for the simplest model but was removed for the model that allowed grade information to be considered. 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 was included in both models is constraint. Constraint measures access to disparate sources of information which might explain why it is so predictive here. Constraint is defined as

cting 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 exact 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 grades of collaborators, and specifically, the 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 scores 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 simply 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 sufficient natural experiments. Nevertheless, the network structure is crucially linked with grades. Diversity of collaborators appears to be advantageous because lower constraint consistently appears in a variety of models as a good predictor of better grades. Additionally, gender was never a useful or significant predictor of grades in any model. This shows that while there may be differences in collaboration by gender, there were no differences in ability by gender. An important unobserved impact of collaboration may be the time saved in completing the assignment, something that may be equally as valuable to students as better grades. Finally, while there are instances of free-riding on the network and while these students can get higher grades on problem sets, the tests serve as a check against this behavior.

# Conclusion and Recommendations

Collaboration in this class is characterized by a diffuse network in which most activity is concentrated in the largest connected component. Gender is an important feature in the networks but neither side is significantly more active. The average grades received by men and women were indistinguishable. Despite women making up less than a quarter of the class, both the amount of collaboration and the estimates of help received from collaboration do not significantly differ by gender. There is no evidence of either gender being at a disadvantage. Intra-gender edges were more stable and became more prevalent as the course continued. While inter-gender edges were more likely to be charitable, these ties were short-lived. The other important norms enforced by students in the network were reciprocity and transitivity. Relationships that lacked these features were less likely to continue.

Collaborating with classmates is correlated with higher grades. One network feature that was consistently predicted higher grades is lower constraint, a measure of the redundancy of a student’s ties. This highlights the advantage to 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, this could be improved with network information. There is some evidence of free-riding on problem sets. After conditioning on homework grades, test scores are lower for more active collaborators, especially 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, especially because the first collaboration is so important. Encouraging more inter-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 still 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 and reward the students who are learning the most.

NEED ONE FINAL LINE???

# Acknowledgements

# Appendix

The R code used in this project is attached. The attached code is edited to remove redundancy across files and to remove exploratory code. The full code is available on GitHub: <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.

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.

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. [↑](#footnote-ref-1)