Social Networks and Society
Group Three: George Gemelas, Sophie Deans, Evan Green, Stella Shannon, Tiana Wang
Final Project Report

GENDER AND COLLABORATION PATTERNS IN CS

INTRODUCTION

The twenty-first century: an age characterized by widgets, digits, and new gender norms. An age of great progress for both science and the sexes: unfortunately, progress *within* these categories has not translated to progress *between* them. Despite the fact that technology has disrupted nearly every major industry over the past few decades, bringing computers into everything from home offices to watches and even to eyeglasses, the presence of women on technology's influential frontier is waning. Since the 1990s, the percentage of women working in computer-science related professions has steadily decreased. Of undergraduate degrees awarded for computer science, women earn only 18%. In 1985, women earned 37%. At top employers in Silicon Valley, 70% of the workforce is male; at Twitter that number rises to 90%.[1]

Why is the number of women in computer science declining? While many studies suggest that reversing this trend requires engaging the interest of more women in computer-science through groups such as "Girls Who Code" and programs like "NASA Women," such a strategy is an external approach: it supposes that the barrier for women in computer-science lies in a dearth of opportunities for promotion rather than some structural or tactical obstacle within the field itself. This study considers the notion that there are internal

[1] "Women in Computer Science | ComputerScience.org." *ComputerScience.org*. Web. 2 Dec. 2016.

barriers to a woman's success in computer science as well: that in addition to simply increasing female interest in the field we must undertake the effort to improve gender inefficiencies or barriers to success within computer science programs in order to support the sustained success of women in the field. In order to address this question, this study examines the structure of computer science classes and the relationship between gender and those factors that serve as important aspects of classroom success. We look specifically at access to collaboration networks, which we find is important to success, and whether access to such networks differs between genders.

We chose to examine collaboration because it is a key component of computer science coursework and STEM work in general. We wanted to examine a unique aspect of computer science classrooms that, holding other factors constant, might be creating an exceptionally difficult environment for women and thus driving the increasing gender imbalance within the field. Unlike courses in the humanities or the social sciences in which much work is individual, relying on one's introspection, writing skills, and individual intellectual pursuits, we observe that computer science courses often require group work for success. So, rather than just focus on promoting computer science to women in order to decrease the gender imbalance, should we consider the possibility that even if such interest is gained, it will not be sustained unless barriers to success (structural inequities within computer science classrooms) are addressed as well? If computer science classes do indeed rely on group work for the successful completion of coursework, then observing gendered patterns with respect to group collaboration will offer insight into women's success in computer science classes. This insight will help us re-imagine drivers of the declining presence of women in the field of computer-science and will allow us to craft new strategies for reversing this trend that assume a more critical approach of the methods of computer science education.

The notion that necessitating group work provides a barrier to women is not new, nor is it limited to computer science. Heather Sarsons, an economist at Harvard, has devoted part of her dissertation to the examination of inequities in group work and gender in the male-dominated field of Economics.[2] She finds that women are given less recognition when their co-authors are male as opposed to female. This suggests that women are dis-incentivized to participate in co-gender group work because they are in fact disadvantaged by collaborating with men. In computer science classrooms, does co-gender collaboration similarly hinder success? Are groups formed along rather than across gender lines? How does the gendered character of each group relate to the students' success as measured by grades?

Should re-orienting the approach to engaging women in computer science from external to internal problem analysis prove successful, women will benefit from finding success in a field that treats women more equally than most. Women earn equal pay as men in the same computer science roles while in other professions they earn on average 78 cents to the dollar.[3] Income equality between genders exists, but the proportion of genders within computer science jobs is not equal. But this economic promise may only be achieved if women feel valued and supported in the course of study of computer science. Male-dominated classrooms may prevent women from completing coursework in the first place. While advertising computer science to women through summer development programs and gender-sensitive conferences is important for

[2] Wolfers, Justin. "When Teamwork Doesn't Work for Women." *The New York Times*. The New York Times, 09 Jan. 2016. Web. 3 Dec. 2016.
[3] Doucet, Christine. "Pay Structure, Female Representation and the Gender Pay Gap among University Professors." *Relations Industrielles / Industrial Relations* 67.1 (2012): 51-75. Web.

attracting initial interest, improving curriculum and classroom dynamics may also lead to a more equal proportion of women in the computer science field.

We are interested in how social networks form when computer science students collaborate and what effect that has on grade performance in regards to each gender. Collaboration is a key component to computer science classes, and specifically to the class we have analyzed. We have obtained data from a computer science at Yale University that is one of the most rigorous courses of the Computer Science major. This data set outlines grades and collaboration networks that formed for each of seven unique problem sets (p-sets) over the course of the 2016 spring semester. The data suggests that men and women on a whole perform relatively equal in terms of grade, but women who don't receive collaborative help do worse than men who don't receive collaborative help, and women who do receive collaborative help actually outperform their male counterparts.

PROBLEM/ PAST RESEARCH

To contextualize our study within the broader literature on gendered career patterns, we conducted a literature review in which we examine primary theories on what drives career choices and gendered trajectories. We briefly survey two different theories relating to external versus internal causal models before discussing collaboration, gender discrimination, and team-based preferences among women. This brief overview of how our study is thematically situated within past research and experience shows that while we are analyzing the results from only one computer science class at Yale, the patterns we observe are not isolated.

Research has comprehensively established differences in career choices and trajectories between men and women, particularly in science and technology fields. Specifically, two theories are used to develop causal accounts for these patterns: the deficit model and the difference model (Sonnert and Holton 1996).[4] The deficit model focuses on external factors, whereas the difference model focuses on internal, intrinsic elements that cause gendered career patterns. External factors that inhibit women in science have been continuously identified and measures have been taken to rectify them. Discrimination against women in the sciences has been outlawed for more than two decades in the United States (Sonnert and Holton 1996). However, it is necessary to investigate barriers that may be internalized that are holding back women-- intrinsic behavior patterns or internalized cultures and values. The deficit model identifies structural explanations for gender differences in science careers, through mechanisms that formally or informally exclude females from the sciences. It supports the idea of external obstacles to success as women are shown to receive fewer opportunities, hence affecting their career outcomes. Our research aims to contribute to the investigation of internal factors that cause different trajectories (within the framework of the difference model), by investigating differences in behaviors and outcomes between genders in the college classroom through focus on patterns, preferences and outcomes of collaboration behaviors.

The issue of fewer opportunities for females to collaborate in an academic setting is not new. Researchers have long investigated how classroom climate impacts the education and careers of women. A study published in 1982 — ten years after the passage of Title IX, which prohibits sex discrimination in education — describes the "chilly" learning climates of higher-education institutions for women.[5] In fact, Yale University students were among the sample population that informed the findings. To improve the classroom atmosphere for women,

[4] Sonnert, Gerhard. "Career Patterns of Women and Men in the Sciences." *American Scientist* 84.1 (1996): 63-71. *JSTOR*. Web. 17 Dec. 2016.
[5] Hall, Roberta M., and Bernice R. Sandler. "The Classroom Climate: A Chilly One for Women?." (1982).

the authors of the study recommended that faculty not group students according to sex, especially in a way which implies that women students are not as competent as or do not have status equal to men (Hall 1982). Isolating female students would potentially harm their self-esteem, as well as reinforce negative prejudices among male students. Whether the classroom atmosphere has changed to be less "chilly" towards women in the years since the publication of the study — namely, in terms of forming gender-diverse collaboration groups — is a primary focus of this research project. This atmosphere is important because a lack of encouragement and distorted expectations for female students might drive them to go into traditionally female careers, instead of pursuing their interests (Hall 1982).

Further, recent research has demonstrated that women are more attracted to cooperative work environments (Kuhn 2015).[6] This is shown by the result that women are more likely to choose team-based compensation over individual pay. In addition, the same experiment demonstrates that the way in which teams are formed strongly affects women's participation, as when teams are formed by mutual consent, women's team formation rate increases dramatically (Kuhn 2015). Another significant finding of this research is that voluntarily formed female teams outperform self-selected male teams. Cooperation in the workforce matters.

Tomorrow's workforce is created from today's students. In this study, we explore the internal dynamics of a college computer science class by following homework collaboration, with the aim of contributing to the existing academic discussion regarding gender issues in STEM. We look to highlight how cooperation and collaboration is important for performance in this computer science class, computer science as a field, and STEM in general.

---

[6] Kuhn, Peter, and Marie-Claire Villeval. "Are Women More Attracted to Cooperation Than Men?" (2013): Web.

DATA

The data consists of self-reported collaboration on homework by students in an upper-level computer science class at Yale University called "Design and Analysis of Algorithms." There were seven problem sets (p-sets) assigned as homework for the course, and per the instructions of the professor, students could collaborate with up to three other students to complete each p-set.

Information on collaboration was collected by the professor of the class after each homework assignment via an online quiz. On the quiz was a list of all the names of students in the class. The professor instructed the students to identify which students he/she collaborated with. Though the quizzes were not mandatory, the professor strongly encouraged providing collaboration data by stating that not filling out the data was a violation of academic honesty. The professor instructed the students to limit the number of collaborators for each p-set to no more than three, yet some students actually list more than three collaborators. Over-listing perhaps indicate that some students preferred to be truthful to showing that they followed the rules when completing the collaboration quiz. A limitation of the collaboration data provided by the professor is our inability to distinguish between individuals who didn't report collaboration data and those who did report the information but did not actually collaborate. Both cases were reported as having no collaboration in the data the professor provided. We think that the emphatic requests for collaboration data and the perceived honesty responses are reason to believe that the number of non-reporting students was low.

Nodes in the data represent the individual students in the class, of which there were 134. The edges are the collaboration ties that emerged for each p-set. The edges represent directed collaborative ties. When a student lists another student, we find a directed, out-going tie to the listed collaborator. Since collaboration was self-reported (and not observed), it is possible that one student could list another who did not list them back.

Information on gender was also collected, but done retroactively. Gender wasn't reported by students, but instead was identified by the professor when our research team requested the data. There were 100 men identified, 32 women identified, and two students whose gender the professor did not identify. We acknowledge that the professor may have misidentified the gender of the students since gender was not self-reported, and we also recognize gender was identified on a binary.

To protect the identity of students, the professor anonymized the data by assigning students random identifiers. The professor also wanted to protect the identities by limiting grade data. Exact grades weren't given, but instead the professor made a proxy for the grades on the four of the homework assignments. The students receiving 1s did better than those receiving 0s. Moreover, the professor only provided data on four of the seven p-sets as another attempt to prevent backtracking of students based on the information provided.

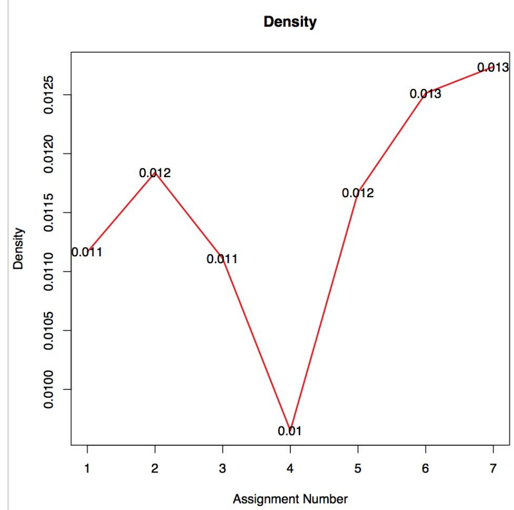
Thanks to the rare initiative of a professor, we were able to collect information on the inner workings of a computer science class. We were able to see gender, grades, and collaboration over over time from the collected data.

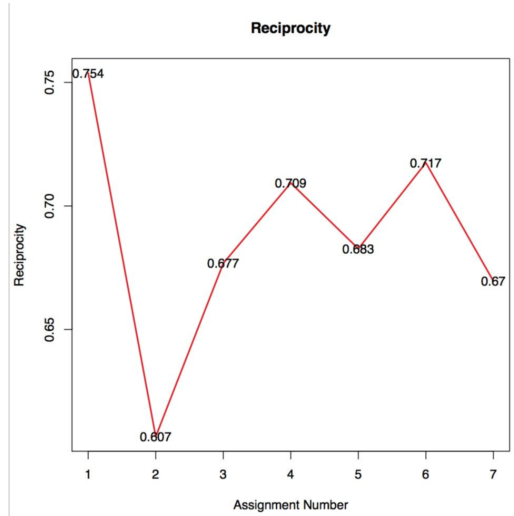
However, this data set is not perfect. Students could have completed the form dishonestly in order to prevent the professor from knowing if they had collaborated with more than the maximum three students. Though there are many instances where students listed collaboration with more than three students, students still may have underreported their actual collaboration. We are unable to understand the full effect of collaboration networks because of our limited access to grades. We had the grade information for only four of the seven problem sets, though we do believe that the other p-sets were not exceptionally different to the extent that they would skew our findings. Moreover, the binary grading information the professor provided us prevents us from more accurately understanding the relationship between grade and collaboration. Are those with the very highest grades the ones who collaborated the most, or they simply the ones whose grade sat just above the proxy threshold? We cannot tell without more complete grade information. Finally, there are many other attributes that are potentially important that we do not have access to. These include overall GPA, class year, major, and whether or not individuals were taking the class 'pass/fail'. These could potentially be confounding variables. Despite these potential problems, we believe the data is extensive and indicative enough to support our findings.

ANALYSIS

Our main findings from analysis of the gender and collaboration data are as follows: 1) collaboration was important to the class 2) that collaboration led to better grades, 3) that given equal levels of collaboration, women slightly outperformed women, 4) that women however were more likely to have no collaboration ties, 5) that there is gender homophily, and 6) that centrality in the network led to better grades.

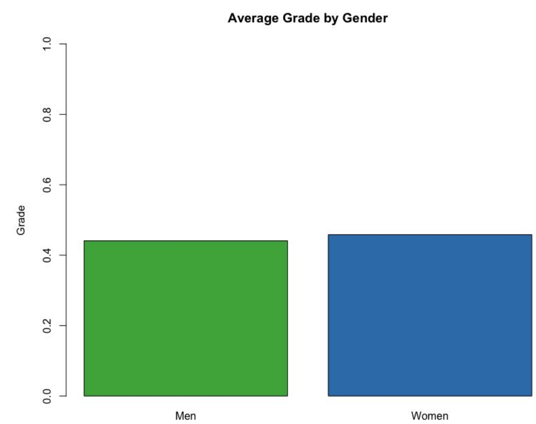
Density of collaborative relationships was higher at the end of the semester as compared to the beginning. The low density of the fourth p-set might be explained by the fact the fourth p-set was markedly easier than the others, according to qualitative understanding about that p-set. Collaboration in the concept of the professor, where people work together and help one another, should be indicated by a reciprocal relationship. A non-reciprocal relationship might indicate that a person simply passed along information and did not actually do constructive work with another. Reciprocity on the sociogram was indicated by a directed connection that went both ways, where an arrow went from node X to node Y and vice-versa. Reciprocity was high throughout all p-sets, but there was some fluctuation. After the initial p-set, there was a relatively sharp drop in collaboration for the second p-set and a jump back up to higher reciprocity levels thereafter. Perhaps the quick drop in the second p-set suggests that some tried to take advantage of the collaboration system the second time around but were quickly excluded from collaboration. In order to receive collaboration help, a person had to also give collaboration help. Reciprocation, an important



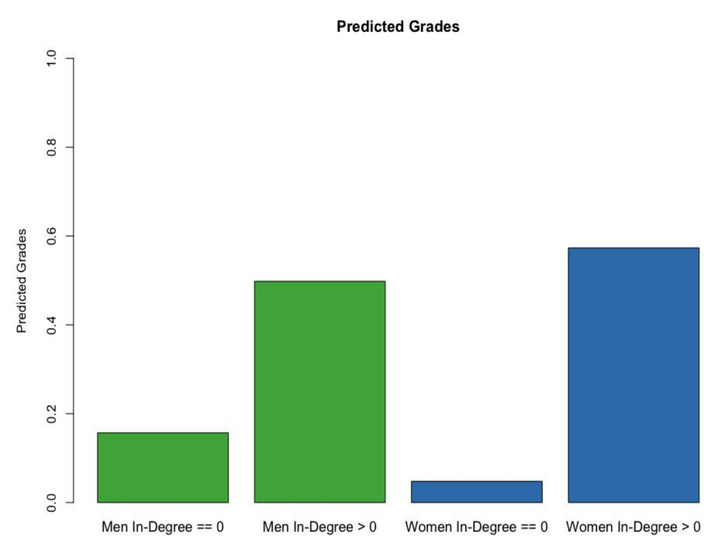


component for collaboration in the conception of the professor, seemed to be a network norm in this study.

The average grade amongst men was .44. The average grade for women was .46. These averages are very similar and, in fact, a t-test shows that there is no statistically significant difference between the averages of men and women. This suggests that men and women perform equally on the whole, thus countering any claims that any one gender is superior in the computer science field.

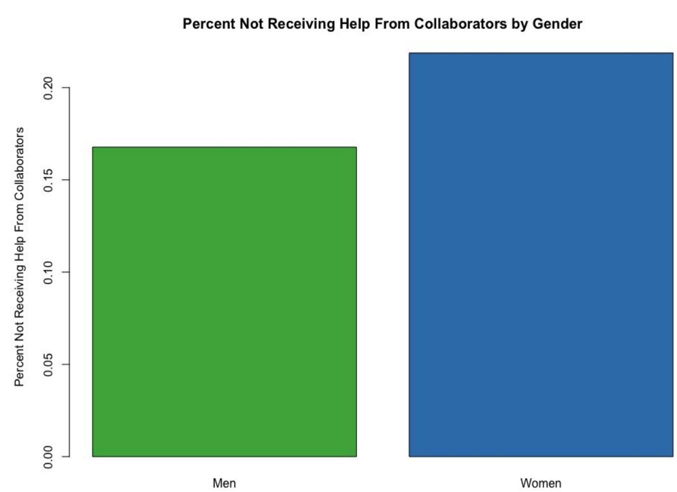


Although the average grades were nearly the same for both men and women, there was variety between the genders when collaboration was taken into account. In this study if a man received no collaboration-- that is, they had an in-degree of 0-- he received an average score of .16. However, if a woman received no collaboration-- had an in-degree of 0-- she received an average score of .05. Without any sources of collaborative help, women tended to get lower test scores. The contrary happened when an individual did receive collaborative help. If a man received collaboration help--if his in-degree was 1 or greater-- he received an average

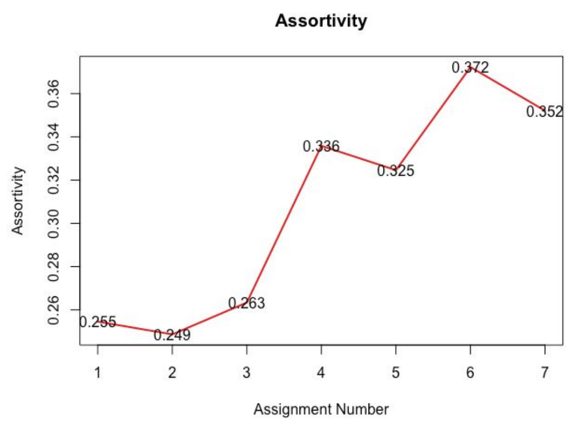


score of .5. Women with in-degrees received on average a score of .57. Women's scores with collaboration was higher than men's. In this study, access to collaboration was more important for a woman to receive a higher grade because without it, she performed worse in comparison to her male counterpart. When she did receive collaboration help, she actually outperformed men. We performed a logistic regression on these results and found that the effect of not receiving collaboration had a large, negative, statistically significant coefficient. The other two coefficients, for being male, and for the interaction of being male and not receiving help, were both insignificant at the .25 and .15 levels respectively. The male coefficient was small and negative while the interaction of being male and not receiving help was larger and positive. These directions are exactly the impacts that we observed. Additionally, the relatively small sample size of the grade information means that with more data these effects might have been significant.

We believe that women doing better than men when both collaborate may be caused by the self-selecting nature of women in this specific computer science class but also of computer science and STEM in general. Women, traditionally seen as underperformers in this discipline, may be less likely to enter a field that is structurally stacked against them, so only the most talented women of the STEM/CS field will enter it.



Women and men on average have similar in-degree, but a larger number of women are likely to have 0 collaboration connections. The percent of men with no collaborators was 17% and the percent of women with no collaborators was 22%. This difference was only significant at the .25 level but with more data it may have been more significant as it is a fairly large difference. For comparison, the p-value for the difference in grades was .76 which is high.



Gender homophily was also observed in the data set. Homophily meant the tendency of men to collaborate with men and women to collaborate with women. In the sociogram, we observed clustering based on the gender of the nodes (indicated by color), and so we tested assortativity. Over the course of the semester, assortativity increased-- .255 for p-set 1 and .352 for p-set 7-- suggesting that over time the men and women end up preferring to collaborate with the same gender.

We next looked at the effects of centrality on grades because we noticed that there was a large core connected component. The sociogram reveals the advantage of being part of this large component. Nodes are sized by the average grade received on problem sets (1s are bigger, 0s are smaller), and it seems like the larger nodes are mostly in the large central component. To test this hypothesis, we added eigen-centrality and betweenness-centrality to the logistic model predicting grades solely based on collaboration and gender. Both eigen-centrality and between-centrality had positive coefficients and were significant at the .06 and .04 level respectively. Higher centrality values indicated better grades. The addition of these new predicting variables did not change the other coefficients and significance scores of the original logistic model. Centrality scores are not simply collinear with in-degree or gender and enrich our understanding of the relationship between collaboration and grade. The finding that both of these types of centrality led to higher grades should not be surprising. Higher values for each of these centrality measures means

```
Call:
glm(formula = Grade ~ as.numeric(inD == 0) * Gender, family = binomial(),
    data = full_degree_summary)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.3052  -1.1741  -0.5842   1.1808   2.4676

Coefficients:
                             Estimate Std. Error z value Pr(>|z|)
(Intercept)                    0.2955     0.2335   1.266  0.20567
as.numeric(inD == 0)          -3.2912     1.0509  -3.132  0.00174 **
Genderm                       -0.3034     0.2652  -1.144  0.25260
as.numeric(inD == 0):Genderm   1.6173     1.1262   1.436  0.15099
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 549.67  on 399  degrees of freedom
Residual deviance: 505.43  on 396  degrees of freedom
AIC: 513.43

Number of Fisher Scoring iterations: 5
```

```
Call:
glm(formula = Grade ~ as.numeric(inD == 0) * Gender + Eigen +
    Between, family = binomial(), data = full_degree_summary)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.9749  -1.0695  -0.5837   1.1569   2.4676

Coefficients:
                              Estimate Std. Error z value Pr(>|z|)
(Intercept)                  0.0483681  0.2457061   0.197  0.84394
as.numeric(inD == 0)        -3.0441003  1.0536659  -2.889  0.00386 **
Genderm                     -0.3232985  0.2690101  -1.202  0.22944
Eigen                        1.1299903  0.6005739   1.882  0.05990 .
Between                      0.0017080  0.0008298   2.058  0.03956 *
as.numeric(inD == 0):Genderm 1.6357113  1.1271501   1.451  0.14673
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 549.67  on 399  degrees of freedom
Residual deviance: 491.32  on 394  degrees of freedom
AIC: 503.32

Number of Fisher Scoring iterations: 5
```

access to more information was helpful in finding the correct answers to the problems.

CONCLUSION

In this study, we looked at the collaboration that took place in a rigorous computer science class at Yale University and analyzed its effect on grades in relation to gender. We believe that the internal dynamics of computer science courses affect the male-dominated composition of the computer science job field.

We analyzed the collaboration data collected after each of the seven problem set homework assignments. Collaboration was important to the success of students in the class. Though we only have grades for four of the assignments, we found that on average, men and women have almost the same average grade. When the dynamics collaboration was taken into account, differences in grade existed between men and women. Women without any source of collaboration help did worse off than their male counterparts who did not receive any sources of collaboration help, but women who had at least one collaborator slightly outperformed men with at least one collaborator. We hypothesize that self-selection of women into computer science might have led to only the brightest women computer science students to be in the class that had a male population with a wider range of academic ability. Women were also more likely to have no collaboration help. In this study, collaboration was more important for a woman's grade.

These findings suggest that the internal dynamics of computer science and perhaps other STEM classes are important for the performance of different genders. Because of the homophilic nature of gender for for collaboration, it may be important to manipulate collaboration dynamics in classroom dynamics to give counteract structural challenges for women. Ensuring women have collaboration groups may have lowered the number of women who received no collaboration help in this class. Having men and women in groups also may be important in the long-run for the computer science and STEM field by creating a precedent where both genders work together. After all, women who received collaborative help actually performed better in this study. To respond to the gender imbalances of computer science, a professor might form collaboration groups herself to ensure all people of all gender have someone to collaborate with and to ensure women and men work together.

Further extensions of this study could include examining collaboration ties among academics in a certain STEM-related field. Although research has been done to analyze co-authorship, co-authorship isn't necessarily the only form of collaboration — creating an advice network similar to the one constructed for our classroom might be a fruitful project. Another direction would be to explore the attributes of the collaboration clusters among male and female students over a longer period of time, or yet another possibility could be to ask students how comfortable or satisfied they were within a collaborative relationship, or ask them to provide a reason for collaborating with a specific individual or group in general.

<u>Works Cited</u>

Doucet, Christine. "Pay Structure, Female Representation and the Gender Pay Gap among University Professors." *Relations Industrielles / Industrial Relations* 67.1 (2012): 51-75. Web.

Hall, Roberta M., and Bernice R. Sandler. "The Classroom Climate: A Chilly One for Women?." (1982).

Kuhn, Peter, and Marie-Claire Villeval. "Are Women More Attracted to Cooperation Than Men?" (2013): Web.

Sonnert, Gerhard. "Career Patterns of Women and Men in the Sciences." *American Scientist* 84.1 (1996): 63-71. *JSTOR*. Web. 17 Dec. 2016.

Wolfers, Justin. "When Teamwork Doesn't Work for Women." *The New York Times*. The New York Times, 09 Jan. 2016. Web. 3 Dec. 2016.

"Women in Computer Science | ComputerScience.org." *ComputerScience.org*. Web. 2 Dec. 2016.