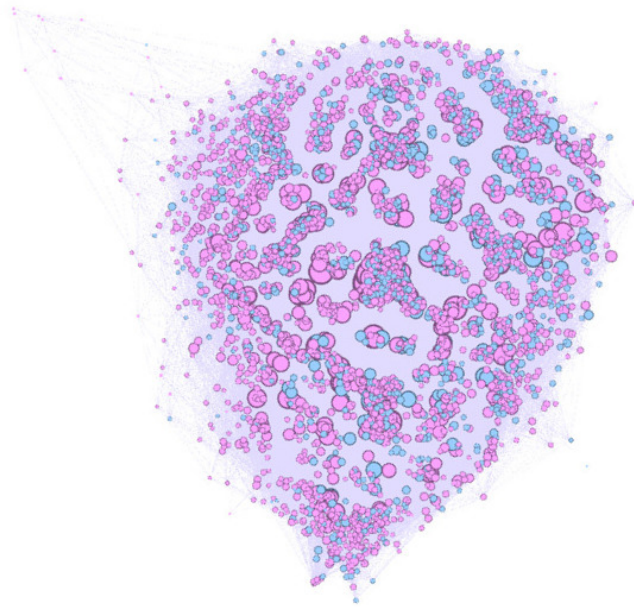# Feminism in Hollywood

A Network Tour of Data Science

Group 31

Fall 2018

Sinan GÖKÇE, Dilara GÜNAY, Sinan BURSA, Othman BENCHEKROUN

January 18, 2019

1   Introduction

The Merriam-Webster dictionary defines feminism as "the theory of the political, economic and social equality of the sexes" [1]. There has been an ongoing revelation all over the world about providing equal rights and opportunities for women that are equal to those for men. Correlatively, among many others, the motion picture industry is also taking a stance against sexism.

One of the very few substantial indicators of feminism in movies might be the Bechdel Test, which is described as *"the measure of the representation of women in fiction"*. For a movie to pass the Bechdel test, it has to feature at least two named women who engage in a conversation on a topic that is not about a man [2]. It is striking that only a measly half of all movies passes the Bechdel test [3] referring the ancient tide of male dominance in the film industry. This statistical information has created an inclination for us to deeply investigate the influence of feminism on actor communities in Hollywood.

With this reason, we have started our data story exploring the impact of feminism in Hollywood using the Kaggle dataset provided in class, which consists of two files with information on movies and the actors who played in them. Later on, we have supplemented this information with the Bechdel scores of each movie using an API [4].

While creating and analysing our actor network, we wanted to observe any possible relation between adopting feminist principles and achieving success in the industry. Overall, the goal of this project is to come to a conclusion with the question: **Does 'Feminism' sell in Hollywood?**

2   Model Selection Process

In order to find a convincing answer to our research question, we decided to approach our exploration with three different adjacency matrix constructions. In each of the approaches, actors have been represented as nodes.

In the first network, actors are connected if they appeared in the same movie. The edges are weighted regarding the maximum number of co-appearances in our network, meaning that the edge weights are relative to the strongest connection of our network.

The second network is based not only on the co-appearance, but also on the gender. This is to say, there is an edge between two actors and between two actresses, separately, if they took part in the same film. As for the first network, edges are weighted regarding the maximum number of same gender co-appearances in our network. This network was built with the intention to analyze the differences in male and female networks.

In the third network, we created an edge between actors if the film in which they acted in together passed the Bechdel test. In this case, the weights are assigned according to the ratio of co-starred Bechdel approved movies over the relative number of co-appearances. As such if two actors played in 4 movies in total, of which 1 passed the Bechdel test, the weight assigned to their connection is 0.25.

To draw results from these networks, which may potentially be answers to our question, we chose to display gender, film genre, rating and profit as signals (features) on each of the previously defined network models. Rating and profit are powerful indicators of the success of a film, while the first 2 signals—gender and film genre—are the other focus points since they may create interesting gender normative clusters on different networks.

The purpose of conducting 12 different analysis was to discover interesting insights about our actor network and therefore formulate another possible fascinating problematic. However, there were not any interesting patterns observed. As such, we continued our data story with the initial question: *"Does 'Feminism' sell in Hollywood?"*. We selected the Bechdel based network since out of the different constructed models, this network appear to be the closest model through which we can analyze feminist connections of actors.
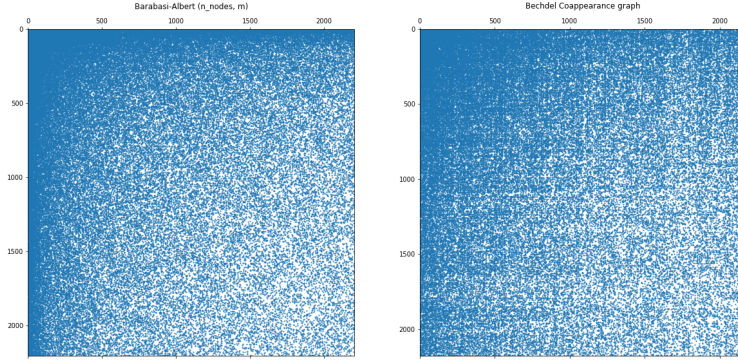
Figure 1: Adjacency Matrices of our Network and the Barabasi-Albert Network

3   Analysis of the Selected Model

The model we have selected—using the co-apparences paired with the Bechdel rating of the movie—is different than the one based on co-appearances that we have analyzed in the milestones. To better understand the structural differences between the networks, we compare below the most important network properties, and derive the random network that represents our own network best.

We cleaned our network to discard disconnected nodes so we have a single connected component in both the current and the previous networks. The first network we studied contains around 1.7 times more nodes (3766 vs 2178) but almost 4 times more links (371044 vs 95686). This is reflected in the average degree, which is 43.93 in the network we use in this project again 98.52 in the network we studied during the semester. However, our network has a larger diameter (4 vs 3), which, when combined with the information above, means our network is a lot more sparse than the one studied before. In parallel, our network also has a larger average clustering coefficient (0.37 against 0.22).

Looking at these results, we can see that our network somewhat behaves like a social network, which makes sense given that our nodes represent the different actors. This intuition is further proved when looking at the Barabasi-Albert network we generate shown in Figure 1. The K-L divergence between the degree distribution of the two networks is 0.37, which is another proof of the similarity between them.

Even though these elements from the milestones allow us to better understand our network, signal processing on the graph doesn't yield much results. Indeed, we see that none of the filters allow us to recover our signal (on the contrary, it introduces even more bias). The main use we can make of signal processing is displaying our ground truth labels (profit) on top of our graph as it can be concluded from Figure 2.
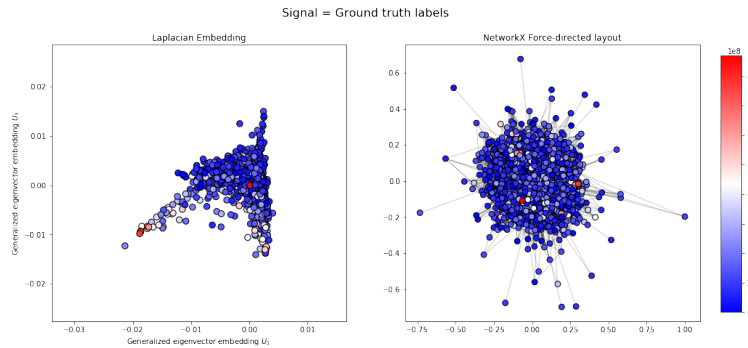

Figure 2: Representation of the Signal on the Network

3.1 Fiedler Vector Correlation

In order to determine the correlation between our signal and the topology of the network, we tried to use the principle of algebraic connectivity, which is based on Fiedler's vector [5]. The Fiedler eigenvalue allows us to determine how connected our network is; it equals to 0.36 in our case, meaning our network is poorly connected. The eigenvector corresponding to this value, the Fiedler vector, allows us to partition our graph into 2 (or more) subgraphs by using its each sign and magnitude. Thus, we use this vector to calculate the correlation of the network with the signal that we try to draw on top of it.

However, when we compute this correlation, we find a value of 0.025, which is very low. When looking more closely at the Fiedler vector, we see that the values are very small (between -0.098 and 0.11). This doesn't allow us to compute a meaningful partition, which has in turn a negative impact on the correlation that we tried to determine. For this reason, we turn to modularity testing in order to find results that will help us answer our problem.

3.2 Bechdel Network

In order to compare the actors that frequently pass the Bechdel test with actors that frequently fail, we constructed two sub-graphs and feature tables as follows:

- Actor network with strictly Bechdel approved movies. This network is referred as the Bechdel-positive network (Figure 3(a)).

- Actor network with strictly non-Bechdel compliant movies. This network is referred as the Bechdel-negative network (Figure 3(b)).

For the purpose of identifying the genders of the actors, two colors were used. The Rose Quartz color is selected to identify male actors and the Serenity (blue) color is selected to identify female actors. Also the size of the node is correlated to the degree of the actor. As such the more the actor plays in movies that pass the Bechdel test, the bigger its node size become. Lastly, the names of the actors with the highest degree are displayed in the constructed graphics.

(a) Bechdel-Positive Network B^+ (b) Bechdel-Negative Network B^-
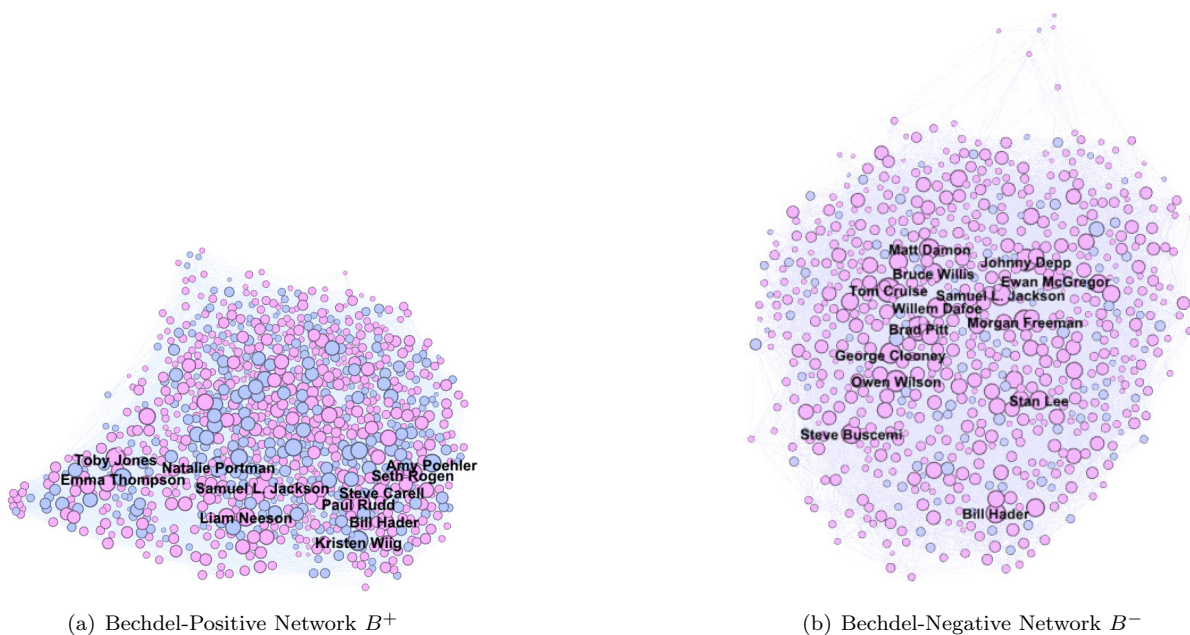
Figure 3: Networks Representation

For the purpose of answering if 'Feminism' sells in Hollywood, we assess the success of the actor by comparing the average rating and the average profit. As we were also curious about the gender patterns, we also included the gender proportions in our results.

We conducted two levels of analysis. The first analysis concentrates on the top ten actors with the highest degree while the second analysis compares the overall network as below:

3.2.1 Comparison of the top ten actors with the highest degree in both networks:

Gender: Interestingly, we have found that the Bechdel-positive sub-network B_{10}^+ is more gender balanced with the ratio of men being 40% where as the Bechdel-negative sub-network B_{10}^- consists entirely men.
Average Rating: The average rating in B_{10}^+ is found to be lower than the B_{10}^-. The results are respectively 6.5 and 6.75.
Profit: For the profit assessment, no significant differences were found between the two networks. The results in B_{10}^+ and B_{10}^- are respectively \$133 million and \$136 million.
Awards and Nominations: For the specific sub-network of 10 actors, the awards and nominations are compared. Our research shows that the actors in B_{10}^+ have $12,4\%$ more awards and $8,3\%$ more nominations compared to the actors in B_{10}^-.

3.2.2 Comparison of all the actors in both networks:

Gender: Similar to the previous gender indications, our results show that the ratio of men in the Bechdel-positive network B^+ is 63.86% whereas the Bechdel-negative network B^- is more male dominant with 83.83% of the actors. For this comparison, it is important to keep in mind that the ratio of men in the total actor network, regardless of the Bechdel test is 70.2%.
**Average Rating:** The average rating in B^+ is again lower than the B^- with 6.42 and 6.54 respectively.
Profit: In this wider profit analysis, we obtained different results compare to the previous sub-network. In fact, contrary to the average rating, the B^+ network is found to be more profitable than the B^- network, with respective profits of \$146 million and \$132 million.

In addition to our problematic, we also wanted to find the communities inside the Bechdel network. Two different community clustering methods are evaluated in Figure 4 below. The first modularity method (Figure 4(a)) is considered unstable as the number of communities differ for different iterations. This is why a more replicable, Clauset-Newman-Moore greedy modularity maximization algorithm is selected. For this method, we obtain four different communities seen in Figure 4(b).
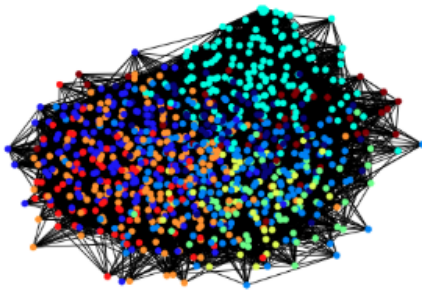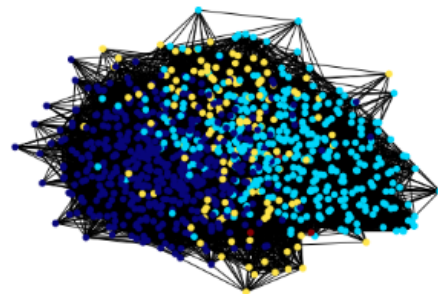
(a) First Modularity Method   (b) Clauset-Newman-Moore Modularity Method

Figure 4: Different modularity methods

More specifically, in this analysis we wanted to find gender patterns in the communities that were formed. As observed in the #MeToo Movement, women may have formed tighter communities between each other. However, our analysis demonstrated no significant gender patterns in each community.

4   Conclusion and Evaluation

According to our results and analysis we can observe the following points:

- The movies that do not pass the Bechdel test cast more men as expected.

- Looking at the top ten actors from the Bechdel-negative network, we can identify actors like Brad Pitt, Matt Damon and Tom Cruise which tend to make movies revolve around the story of a single male character (one man show). Also these actors tend to be good looking male actors, considered as "sex symbols" in Holywood.

- The Bechdel-positive network tends to be more gender balanced.

- The Bechdel-positive network tends to make more profit on movies. With respect to the analysis of top ten Bechdel-positive actors, they are not only nominated for more awards but also have won significantly more than Bechdel-negative ones.

- There are no significant differences in the average rating.

If we define the success of an actor by their talent measured with the number of awards won and the average generated, we can conclude that according to our dataset, **feminism sells just that tiny bit better** is an answer to our initial question "Does feminism sell in Hollywood?". This is probably due to the fact that—what we call—feminist actors are more talented and well-rounded, which in turn increases the movie's quality.

Moreover, since we chose to use the actors as nodes, our network can be assimilated to a social network (especially given our results with the Barabasi-Albert graph). Choosing movies as nodes could yield more interesting results. However, it would require us to determine a new metric to weight the edges, which is why we decided to go on with actors as nodes. This allowed us to explore the feminism issue in the film industry more deeply, and with a new approach.

5   References

[1] https://www.merriam-webster.com/dictionary/feminism

[2] https://en.wikipedia.org/wiki/Bechdel_test

[3] https://bechdeltest.com/

[4] https://bechdeltest.com/api/v1/doc

[5] http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.592.1730rep=rep1type=pdf

6   Appendix

Top 10 Bechdel actors:
vote average = 6.50
profit = $1.33 * 10^8$

$$gender\ ratio(\frac{male}{female}) = 40\%$$

Top 10 No Bechdel actors:
$vote\ average = 6.75$
$profit = 1.36 * 10^8$
$genderratio = 100\%$

Bechdel Top 10 actors have won awards %12.4 times more than No Bechdel Top 10 actors.
Bechdel Top 10 actors have been nominated for awards %8.3 times more than No Bechdel Top 10 actors.

All Bechdel actors:
$vote\ average = 6.42$
$profit = 1.46 * 10^8$
$gender\ ratio = 63.86\%$

All No Bechdel actors:
$vote\ average = 6.54$
$profit = 1.32 * 10^8$
$genderratio = 83.83\%$

Top 10 Bechdel actors:
Samuel L. Jackson: 37 wins and 83 nominations
Paul Rudd: 8 wins and 19 nominations
Meryl Streep: 175 wins and 356 nominations
Jim Broadbent: 23 wins and 54 nominations
Natalie Portman: 85 wins and 131 nominations
Liam Neeson: 22 wins and 40 nominations
Naomi Watts: 55 wins and 88 nominations
Nicole Kidman: 95 wins and 196 nominations
Julianne Moore: 113 wins and 155 nominations
Kristen Wiig: 12 wins and 68 nominations

Top 10 No Bechdel actors:
Tom Cruise: 51 wins and 81 nominations
Brad Pitt: 67 wins and 152 nominations
Matt Damon: 36 wins and 141 nominations
Morgan Freeman: 61 wins and 79 nominations
Robert De Niro: 55 wins and 97 nominations
George Clooney: 90 wins and 198 nominations
Johnny Depp: 76 wins and 143 nominations
Owen Wilson: 14 wins and 49 nominations
Willem Dafoe: 69 wins and 76 nominations
Samuel L. Jackson: 37 wins and 83 nominations