

# Improvements in Gender Representation in Movies Over the Past Century

**Shirley Wang**

University of Toronto

Vector Institute

shirleywang@cs.toronto.edu

## Abstract

Diversity in media has become a central focus in recent years as minorities protest for more equal representation. This paper presents an exploration of the representation of women in movies over the past century. We find that while the gender ratio and number of movies passing the Bechdel test have shown positive improvements, the reasons behind it and actual screentime of women may not actually be improving in a productive manner.

## 1 Introduction

Media plays a big part in how people perceive both the world, and the people around them. There has been a huge push to get more diverse representation and to remove harmful stereotypes as a result. However, it has also become a big marketing strategy now for companies to announce how progressive they are while doing the bare minimum possible (e.g. the gay representation in the live action *Beauty and the Beast* (Lawler, 2017)). It's important for media to have diverse representations that are actually meaningful, and not just background characters you could blink and miss, so we can normalize attitudes of equality in our society.

The feminist movement has been around for a very long time, and its influence can be felt in popular media today, where women are no longer always delegated to damsels in distress, although it can also be argued that media tries to overcompensate in this direction as well. However, there are still issues where there is only one woman in a movie cast, who interacts exclusively with other men. There's even a popular measurement for misrepresentation of women in media, known as the Bechdel test. In this paper, we will be investigating if representation of women in media

has improved since the popularization of movies at the start of the 1900s. We will be looking at how the gender ratio of characters has changed in movies over the last century, if women are getting more scenes in movies and with who, as well as if movies are passing the Bechdel test more often.

## 2 Related Works

Previous analyses on gender in movies tend to focus on a single movie's message or effects on society (Payne, 2022) (Primo, 2018) (Streiff and Dundes, 2017), or compare two movies (Dundes and Streiff, 2016). There is a lot of emphasis on Disney movies and princesses, since Disney movies are a common part of childhood for children who have grown up in the Western world. (Hine et al., 2018b) focuses on comparing how children responded to the princess stereotype after viewing the 'old' *Sleeping Beauty* vs the 'new' *Moana* princess movies. (Hine et al., 2018a) is similar, while it compares the difference in portrayals of prince and princess characters in old and new Disney princess movies. Overall, Disney princess movies have had a push to focus less on traditional romance and marriage endings in recent years, and more towards female independence. However, adults watch movies too, so it's important to also understand if movies aimed to teenagers and adults are also adopting more diverse gender representation. The authors of (Kumar et al., 2022) have a similar topic. They did an analysis using NLP on thousands of Hollywood movie synopses, to see how many perpetuated the usual gender stereotypes. They found that female characters are primarily associated with themes of romance, and while associations with physical beauty and marriage were declining, associations with sexual relationships and weddings were increasing.

### 3 Data

We will be using the MovieGalaxies (Kaminski et al., 2018), Bechdel Test (Bec), and Kaggle Movie (Banik, 2017) datasets. The MovieGalaxies dataset stores information about interactions between characters in 773 movies in graphs, where each node represents a character, and the weight of each edge between two nodes represents how many times the two corresponding characters appeared in a scene together. However, it does not store the gender of those characters. We will be using the Kaggle Movie dataset to find genders of characters for labelling purposes, as well as additional features such as genre of the movie. The Bechdel Test dataset is a dataset managed by its titular website, where anyone can submit if a movie passes the Bechdel test or not. The Bechdel Test has 3 criteria for a movie:

1. It has at least two women in it.
2. Two women have a conversation with each other.
3. The women have a conversation about something other than a man.

This can be used as a scale to represent how much a movie passes the basic progressive bar. However, it does not have any information past this about the conversation. Combining the data together creates three different levels.

1. All movies in the Kaggle Movie dataset (45263 movies)
2. Movies in the Kaggle Movie dataset that also have a Bechdel test rating (7722 movies)
3. Movies in the Kaggle Movie dataset that also have a Bechdel test rating and a MovieGalaxies graph (596 movies)

### 4 Methodology

For the first level of data, we will mainly be looking at if the number of female characters and the gender ratio of movies have improved over the past century. For the second and third levels, we will be looking at what factors tend to affect the Bechdel test rating, and if it seems to be improving over time. I will be using linear and logistic regression to find significant factors affecting gender representation. The Bechdel test dataset and MovieGalaxies dataset are also most likely not a completely random sample, since the data in them is most likely influenced by the personal preferences of the people who put time and effort into

labelling them. As such, comparing general trends with the trends on the full Kaggle Movie dataset will help confirm consistency.

The release date of movies is very left skewed towards recent movies. To account for the skew, we create indicator variables on the Kaggle Movie dataset for if the movie was released after 1978 (the 25th percentile), 2001 (the 50th percentile), and 2010 (the 75th percentile).

To use MovieGalaxies as a part of this analysis, we will be labelling the gender of every node, by getting some ground truth labels from the Kaggle Movie dataset, and labelling the rest through training a Transformer, which will take in node-level features and output its gender. We will be attempting to analyze the connections present in movies in the MovieGalaxies dataset through aggregating these node and edge features at a movie level, and use them as additional predictors for movies in a simple logistic model for predicting Bechdel test rating. Aggregating node-level features will occur by simply counting the number of male, female, and the fraction of female characters amongst all characters in the movie. Aggregating edge-level features will occur in two different ways between the three different possible connections (female-female, female-male, male-male): the number of connections that fall in that group, and the max weight of any edge in that group.

## 5 Results

### 5.1 Improvements in Gender Ratio

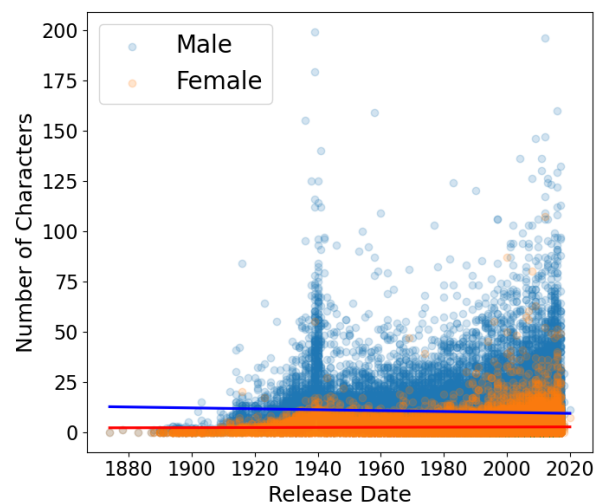


Figure 1: Number of characters appearing in movies over time.

We used robust linear regression to model the relationship between the release date to the number of female characters, male characters, and ratio of female characters in the movie. In Figure 1, while visually, it appears that the number of characters is increasing for both male and female characters, that is actually the minority. The models note a statistically significant decrease in number of female characters by 0.002 per year, and a significant decrease in number of male characters by 0.024 per year. The number of male characters is decreasing faster than number of females, and correspondingly, the model predicts a significant increase in ratio of female characters by 0.0001 per year. These are all extremely small effects, so despite being statistically significant, it would not feel right to claim that gender representation is indeed improving. However, these decreases may be due to how acting is now a much more popular and lucrative occupation than in the past. For most movie studios it may not be feasible to hire many actors, and since it is also common for men to be paid more than women, perhaps movie studios also tend to drop one or two male actors when they need to conserve budget, hence accidentally improving the gender ratio in movies. It seems that the gender ratio in movies may have either improved by factors not related to improving representation, or it has not improved at all.

## 5.2 Factors Affecting Bechdel Rating

For the second level of the data, we fit three logistic regressions: one for each requirement of the Bechdel test. The first model is trained on all 7722 movies to predict if a movie will pass requirement 1. The second model is trained on all movies that pass requirement 1, and predicts if a movie will pass requirement 2. The third model is trained on all movies that pass requirement 2, and predicts if a movie will pass requirement 3. When doing prediction on a movie's final Bechdel test rating, we use the models sequentially. The final prediction accuracy is 59.4%. The model is extremely skewed towards predicting a movie has a rating of 3, which is most likely due to how over 55% of the data has a rating of 3. Figure 2 shows all the effects of potential factors on the Bechdel test.

For the effects of genres, the results tend to correspond with common stereotypes. Action, Adventure, Drama, SciFi, Thriller, War, and Western tend to have male-centric stories, and are all gen-

erally worse at including women. Comedy movies may still be affected by the stereotype of how women are not funny. Crime movies may have women in scattered minor roles or victims, rather than major roles. Historic movies may try to be historically accurate, and have women in positions of little power with little else to talk about.

The genres that tend to do better on the Bechdel test, other than Romance, are Family, Fantasy, Horror, and Mystery movies. Family movies do have a certain subset about sisters, or mother-daughter centric movies. Fantasy movies require large casts, and they have lots of exposition to give. Mystery movies also require a diverse cast of potential suspects. Horror movies usually have characters talking to about how to not die, rather than talking about men. Romance movies tend to meet requirements 1 and 2 of the Bechdel test more often, but requirement 3 less often. Since a majority of romance movies portray heterosexual relationships, and are also aimed towards women, they usually have a woman as the POV character, and give her some female friends or female rivals. But the model predicts a negative effect for meeting requirement 3, which may mean that many romance movies still rely on the main source of conversation between any women in these movies to be about the male love interest.

There have been statistically significant increases in movies that tend to meet the requirements of the Bechdel test over time. It's worth noting that the previous linear regression predicted a slight decrease in number of female characters over time, while this one predicts an increase in number of movies that have at least two female characters. So while movies don't seem to be hiring more female actors, they do tend to hire at least two more often nowadays. This does seem to imply a lot of movie studios trying to hit a basic diversity quota.

Other significant factors that may affect a movie's Bechdel rating are runtime, revenue, and number of votes. Runtime and revenue both have a slightly positive effect for movies meeting requirement 1, which may also tie back to the budget issue. Movies with longer runtimes have higher budgets, and so they can also hire more actors who may be female. Vote count has a negative effect for all three requirements, which seems to suggest that wildly popular movies are less likely to pass any of the Bechdel test requirements.

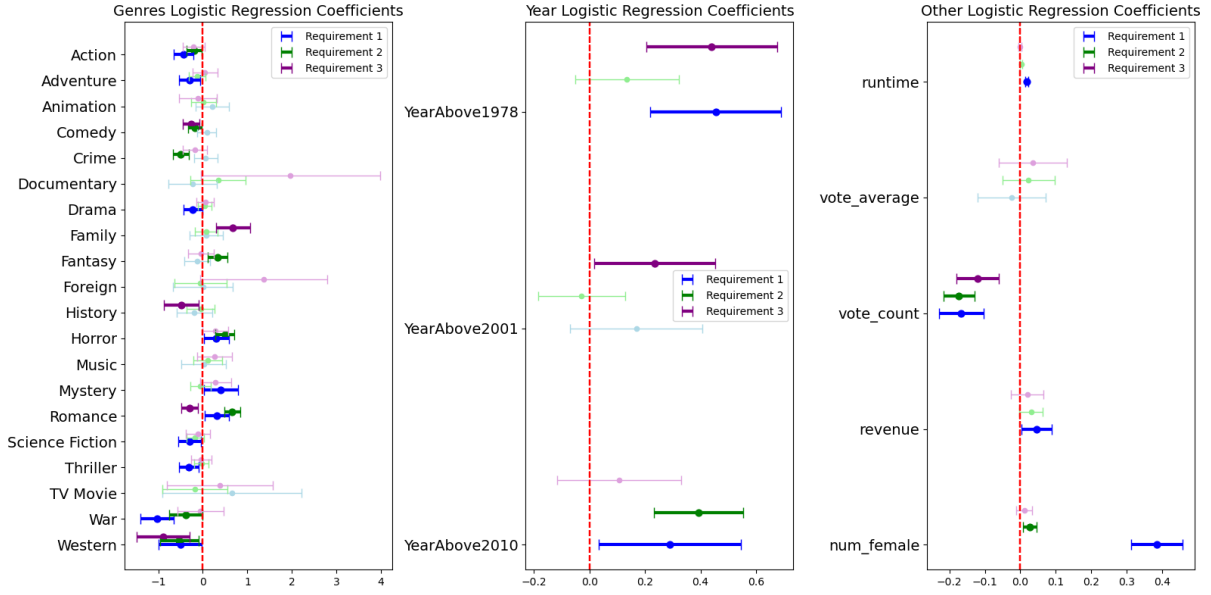


Figure 2: Logistic Regression fitted coefficients over movie-level factors. Female ratio has been omitted (for better visible scale) but has a statistically significant coefficient of over 3 for each requirement.

### 5.3 MovieGalaxies Labelling Gender

We used a plain transformer encoder with four layers, hidden dimension of 256, and four heads. It uses a GPT2 (Radford et al., 2019) module to encode the character’s name and combine it with the rest of the node’s features. We also used weighted binary cross entropy loss with a higher weight on predicting female over male to account for the class imbalance, and train for 20 epochs. There are 28287 nodes, with 67% of characters genders unknown, 23% of characters male, and 10% of characters female. Due to the low amount of labelled data, we split the data into 10 partitions and train an ensemble of 10 transformers, where each transformer is trained on 9 of the 10 partitions and the last is used as evaluation. Predictions are made by averaging the predictions of the ten transformers. Overall, the ensemble model achieves a prediction accuracy of 88.5%, with 90.5% accuracy for females and 87.8% accuracy for males.

Since each node is labelled independently, the predictions are skewed by typically gendered names. For example, all ten models agree that a character named Tony is a male, a character named Ashley is female. Out of all of the nodes, only 21% of them had any disagreement between the models on what they should be labelled.

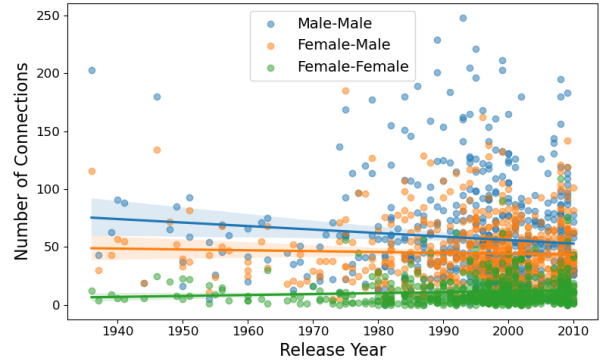


Figure 3: Number of connections vs release year, separated by the three types of connections.

### 5.4 MovieGalaxies Connection Analysis

We analyze how the graph connection features change over time through linear regression between the release date indicator variables and the connection features. Max weight has not significantly changed over time for any of the three kinds of max weight connection features. The model does predict a statistically significant decrease in number of connections between men. However, it does not predict a corresponding increase in connections between female characters, or female and male characters, as can be seen in Figure 3. This does correspond with previous analysis of how the number of male characters does seem to be slightly decreasing over time.

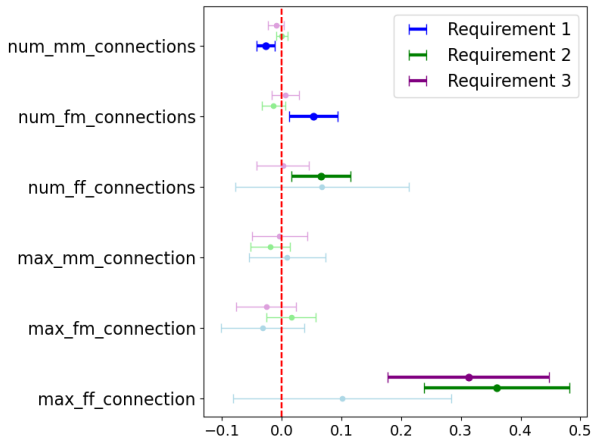


Figure 4: Coefficients for the connection variables in the fitted logistic regression models.

We also proceed with the same logistic regression procedure as with the entire Bechdel test dataset, including the six additional connection variables and removing the indicator variable for year above 2010, since the MovieGalaxies dataset only has one movie later than 2010. The three logistic regression models achieve an accuracy of 57.3% on this smaller dataset.

Figure 4 shows the fitted effects of the connection variables in the logistic regression models. Despite all variables (excluding year above 2010) from the previous models also being included, none of those are considered significant in these models. This may be due to the smaller dataset, but also possibly due to these connection variables being better predictors of Bechdel test. The first model also doesn't predict a positive effect for number of scenes with two women, which may be a flaw, since if two women appear in a scene together then the movie must have at least two women and pass requirement 1.

The effects noted by this model do reflect what we would logically expect. Movies with a high number of scenes with at least two female characters, and also if two female characters that appear in a large number of scenes together are positively related with the Bechdel test rating. These results suggest that MovieGalaxies graphs are potentially good predictors of Bechdel test rating, and creating a video model that takes in movies and generates more graphs of this structure may be helpful in further analysis of gender in movies.

## 5.5 Limitations

The Bechdel test and MovieGalaxies datasets are both crowdsourced datasets. As such, they may be skewed by the personal interests of the people who created them. These datasets also tend to only cover Western movies, so the results are most likely not representative of media in other countries and cultures. The release date of movies is also very left skewed towards recent movies. This could be due to personal interests again, or because due to the rapid technological development over the past few decades, more movies are being made. We have tried to account for this by using indicator variables for specific time periods when appropriate, but this is still a factor that may affect our analysis of changes over time.

The model is extremely skewed towards predicting a movie has a rating of 3, which is most likely due to how over 55% of the data has a rating of 3. Due to this major limitation, the results of the logistic regression should be taken as indicative of possible common stereotypes, rather than concrete trends that always occur.

## 6 Conclusion

We find that at a surface level, it does appear as though there have been improvements in gender representation over time. Movies do tend to pass the Bechdel test more often now, and the gender ratio has had slight improvements. However, deeper inspection reveals that there may have been a slight decrease in the number of male actors appearing in movies since the 1900s, possibly due to inflation and increase in salary for the average actor, which unintentionally has made the gender ratio for female characters increase. Additionally, the number of scenes featuring two female characters in movies has not gone through a significant increase, suggesting that screentime for women has in fact, not increased. This coupled with the increases in Bechdel test ratings suggest that perhaps many movie studios are trying to hit diversity quotas nowadays, by hiring some female actors to do some female-centric scenes to garner good faith in terms of diversity, but do not go the extra step of actually balancing the gender ratio and spotlight between male and female characters. Follow-up work into the exact conversations men and women tend to have in movies would shed further light into what kind of roles women tend to have now, and if those are also improving over time or not.

## References

Bechdel test movie list.

Rounak Banik. 2017. The movies dataset. <https://www.kaggle.com/datasets/rounakbanik/the-movies-dataset>. Accessed: 2022-10-30.

Lauren Dundes and Madeline Streiff. 2016. Reel royal diversity? the glass ceiling in disney's mulan and princess and the frog. *Societies*, 6(4).

Benjamin Hine, Dawn England, Katie Lopreore, Elizabeth Skora Horgan, and Lisa Hartwell. 2018a. The rise of the androgynous princess: Examining representations of gender in prince and princess characters of disney movies released 2009–2016. *Social Sciences*, 7(12).

Benjamin Hine, Katarina Ivanovic, and Dawn England. 2018b. From the sleeping princess to the world-saving daughter of the chief: Examining young children's perceptions of 'old' versus 'new' disney princess characters. *Social Sciences*, 7(9).

Jermain Kaminski, Michael Schober, Raymond Albaladejo, Oleksandr Zastupailo, and César Hidalgo. 2018. Moviegalaxies - Social Networks in Movies.

Arjun M. Kumar, Jasmine Y. Q. Goh, Tiffany H. H. Tan, and Cynthia S. Q. Siew. 2022. Gender stereotypes in hollywood movies and their evolution over time: Insights from network analysis. *Big Data and Cognitive Computing*, 6(2).

Kelly Lawler. 2017. 'beauty and the beast's 'gay moment' may have been much ado about nothing.

Alice Payne. 2022. Leigh whannell's the invisible man discussing narratives of domestic abuse and gaslighting through the cassandra myth. *Humanities*, 11(1).

Cassandra Primo. 2018. Balancing gender and power: How disney's hercules fails to go the distance. *Social Sciences*, 7(11).

Alec Radford, Jeff Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. 2019. Language models are unsupervised multitask learners.

Madeline Streiff and Lauren Dundes. 2017. From shapeshifter to lava monster: Gender stereotypes in disney's moana. *Social Sciences*, 6(3).