Gender Representation in Star Wars

Degrees of separation between female characters, percentage of interactions that are between two women, and more

My project aimed to quantify gender representation in Star Wars by determining how often female characters interact with each other and comparing this to how often female characters interact with male characters, and how often male characters interact with other male characters. I also determined, for each female character in each movie, how many degrees of separation there were between her and another female character. The results, while not particularly surprising, illustrate the need for more diverse storytelling in the Star Wars franchise.

My project involved seven different databases, which each described a different movie in the sequence of films between episodes 1 and 7. Data for episodes 8 and 9 were not available.

The data was in the form of a json file. It contained two vectors - nodes and links. Each node contained the name of a character and the total number of scenes they appeared in. Each link had a target and a source, which were integer values that corresponded to the indices of the characters in question. This was accompanied by 'value,' which described the number of scenes the two characters appeared in together. So, each movie's social network is described by an undirected graph with weighted edges.

Unfortunately, biographical information - such as gender - was not included in the original data, so I created a csv file and manually wrote out each character in the entire series and their gender. I gave droid characters a unique gender so that I could include or exclude them as needed. I wrote a test function to make sure I wrote the names correctly - this function made sure that every name in the graph appeared in the gender HashMap.

I read the json files into a custom struct, which had a vector of nodes and a vector of links (each of which were also custom structs). I read my gender database into a HashMap. All of my functions and calculations required iterating through the links of the graph, using the indices to find each character's name, and getting the gender from the HashMap based on that key.

I had three modules for this project - graph.rs, which contained the structs for the graph and functions for reading in data and calculating adjacency lists; gendercounting.rs, which contained all of my functions for calculating gender statistics and degree of separation; and makedata.rs, which had a single function that randomly generated 1000 characters and created a graph and HashMap from those characters. The purpose of makedata.rs was to show that my code works on larger datasets, since the actual data for each movie only has a couple dozen characters each. main.rs is where I implemented my calculations and printed the results. It is also where I put my tests.

One issue I had while working on this project - specifically when I was calculating the degrees of separation - was the fact that the data was generated from the original scripts. In other words, there were characters from deleted scenes that were included in the data and resulted in incorrect results. To correct this, I had to go through each movie, remove cut characters, and adjust the indices in the links to reflect the change. There was also a cut scene between Padme and Beru in episode 2, which resulted in them having zero degrees of separation - something that confused me for a long time since I knew they never interacted with each other. I deleted this link from the data manually.

My results were very interesting. I've included the output below:

```
kmbidkhelp@Jarvis:~/ds210/project/draft$ cargo run --release
    Finished release [optimized] target(s) in 0.03s
     Running `target/release/draft`
EPISODE 1 - DEGREE OF SEPERATION BETWEEN WOMEN
        TEY HOW - 1
        PADME - 0
        JIRA - 0
        SHMI - 0
        RABE - 1

EPISODE 2 - DEGREE OF SEPERATION BETWEEN WOMEN
        PADME - 1
        TAUN WE - 1
        BERU - 1
        SHMI - 1

EPISODE 3 - DEGREE OF SEPERATION BETWEEN WOMEN
        PADME - 0
        MON MOTHMA - 0
```

```
PREQUEL TRILOGY STATS
        Percentage of female characters among all characters: 0.11353707
        Percentage of female characters among all non-droid characters: 0.1269417
        Percentage of interactions that are between two women among all interactions: 0.007620799
        Percentage of interactions that are between two women among female interactions: 0.032056194
        Percentage of interactions that are between two men among all interactions: 0.5944989
        Percentage of interactions that are between two men among male interactions: 0.63735527
```

```
EPISODE 4 - DEGREE OF SEPERATION BETWEEN WOMEN
        LEIA - 1
        BERU - 1

EPISODE 5 - DEGREE OF SEPERATION BETWEEN WOMEN
        LEIA - 7

EPISODE 6 - DEGREE OF SEPERATION BETWEEN WOMEN
        LEIA - 0
        MON MOTHMA - 0
```

```
ORIGINAL TRILOGY STATS
        Percentage of female characters among all characters: 0.08095238
        Percentage of female characters among all non-droid characters: 0.08966861
        Percentage of interactions that are between two women among all interactions: 0.0012391574
        Percentage of interactions that are between two women among female interactions: 0.0042194095
        Percentage of interactions that are between two men among all interactions: 0.3941653
        Percentage of interactions that are between two men among male interactions: 0.4551834
```

```
EPISODE 7 - DEGREE OF SEPERATION BETWEEN WOMEN
        CAPTAIN PHASMA - 1
        REY - 0
        MAZ - 0
        LEIA - 0
        JESS - 1
```

```
EPISODE 7 STATS
        Percent female characters: 0.18518518
        Percent female characters no droids: 0.20833333
        Percentage of female interactions that are between two women: 0.02112676
        Percentage of all interactions between two women0.008645534
        Percentage of male interactions that are between two men:0.44625407
        Percentage of all interactions that are between two men:0.39481267
```

```
GENERATED DATA
        Example of node name and getting gender: name - GPDBMZLC, gender - f
        Time to calculate degrees of seperation: 3.385843ms
        Percent female characters: 0.33
        Percent female characters no droids: 0.5030488
        Percentage of female interactions that are between two women: 0.20438558
        Percentage of all interactions that are between two women: 0.10858954
        Percentage of male interactions that are between two men:0.21751808
        Percentage of all interactions that are between two men:0.12028976
```

DEGREES OF SEPARATION

I used breadth-first search to calculate the degrees of separation between women.

Predictably, the worst results here come from the original trilogy, which has three named female characters in total across all three films. Leia is the only one that appears in more than one film, and is the only female character in episode 5 (hence seven degrees of separation). Of the three films, episode 6 is the only one where two female characters interact with each other directly.

Interestingly, in the prequel trilogy, the number of female characters decreases with each film, with episode 3 having only Padme and Mon Mothma. The most interesting result here is that in episode 2, there are four female characters, but none of them interact with each other.

Episode 7 has practically the same degrees of separation as episode 1. I think it's interesting that the characters who don't interact with other women have primarily combat roles. Captain Phasma is a stormtrooper and commander, and the only female character associated with the First Order. Jess is a pilot who appears in a couple aerial fights.

STATISTICS

In every trilogy (including episode 7, which was released in 2015), female-female interactions make up less than one percent of all interactions. In contrast, nearly 40% of all interactions in the original trilogy are between two men. This is about the same as episode 7. However, male-male interactions make up a staggering 60% of all interactions in the prequel trilogy. Even if you restrict the number of interactions in consideration to solely interactions involving women, female-female interactions make up less than 5% in every trilogy. Doing the same analysis for men produces 45% for the original trilogy and episode 7, and 63% for the prequel trilogy. I think this is an interesting comparison since the prequel trilogy was more recent, and has comparatively more female characters than any of the other trilogies. Despite this, it seems that the female characters in

the prequels have a less prominent role than the female characters in other trilogies.

Although the percentage of female-female interactions is abysmally low in every trilogy, it is evident that the number improves depending on when the movies were released. The original trilogy, which was released in the late seventies, has .12% female-female interactions. The prequel trilogy, released in the late nineties, has .76% female-female interactions. Episode 7 has .86%. The percentage of characters that are female also improved with time - the original trilogy's three female characters make up 8% of all characters, and 8.97% of non-droid characters. The prequel trilogy has 11.4% female characters among all characters, and 12.7% among non-droid characters. Episode 7 has 18.5% female characters, and 20.8% among non-droid characters. It's interesting to note that excluding droids increases the percentage by roughly one percentage point in every trilogy. Since droid characters are also a minority of characters, I think this shows how much of a difference just a few characters can make in representation.

CONCLUSIONS

Overall, this project has quantifiably shown the disparity between male and female representation in the Star Wars franchise. I think it speaks volumes that in every trilogy, more than 95% of interactions between women and other characters are between a woman and a man. Meanwhile, male characters have much more evenly distributed interactions - close to 50/50 in most cases. I also think it's important to note that, while I gave droids a unique gender, most droids in the Star Wars universe are portrayed as masculine, and are referred to with masculine pronouns. If I had counted droids as masculine when I was calculating male-male interactions, the disparity would be even higher.

The reason I think this is important is that the demographics of Star Wars characters doesn't match that of its audience. While it is true that Star Wars is more popular with men than women, a study by Statista found that women made up more than one-third of people who said they were avid Star Wars fans. Outside of diehard fans, the audience gender share of Star Wars movies in

recent years has consistently been roughly 60-40, with men making up the small majority. The lack of female characters who interact with each other ignores a significant demographic of Star Wars fans.