Jessica Cannon

Professor Kontothanassis

DS210

7 May 2023

Final Project Write-Up

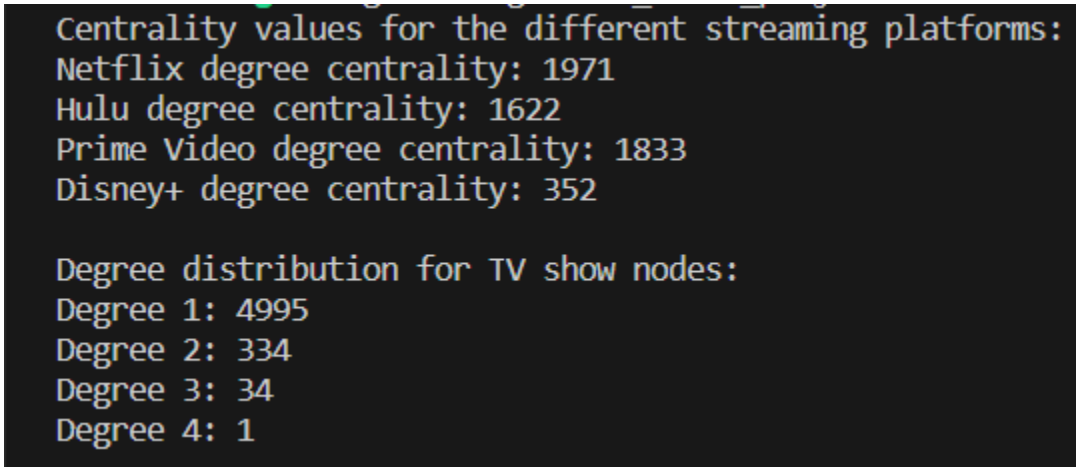
Collaborators: None

Resources: Lecture 16, 18, 20-28, 30, 34 Notes, https://docs.rs/csv/latest/csv/struct.Reader.html , https://docs.rs/petgraph/latest/petgraph/graph/struct.NodeIndex.html , https://depth-first.com/articles/2020/02/03/graphs-in-rust-an-introduction-to-petgraph/#:~:text=The%20most%20broadly-useful%20graph%20implementation%20is%20Graph%2C%20which,%5D%29%3B%20assert_eq%21%20%28graph.node_count%28%29%2C%202%29%3B%20assert_eq%21%20%28graph.edge_count%28%29%2C%201%29%3B%20%7D , https://www.sciencedirect.com/topics/computer-science/degree-centrality#:~:text=Degree%20centrality%20is%20one%20of%20the%20easiest%20to,degrees%20also%20have%20high%20centrality%20by%20other%20measures.

For my final project, I used a dataset called “TV shows on Netflix, Prime Video, Hulu, and Disney+”:(https://www.kaggle.com/datasets/ruchi798/tv-shows-on-netflix-prime-video-hulu-and-disney). Using all of this data (contained in a CSV file called tv_shows.csv, which has 5000+ entries), I created an undirected network graph consisting of vertices and edges, where the individual TV shows are the vertices and the streaming platforms are the edges connecting them - they represent the presence of a particular show on a specific streaming platform. From this graph, I then created a variety of functions that I used to effectively make correlations between

TV shows across certain streaming services. For example, I have a function called "calculate_degree_centrality" that calculates and prints the degree centralities for the platform nodes. I also have one called "calculate_degree_distributions" that finds and prints the degree distributions for the tv show nodes, as well as a "calculate_exclusive content" function, a "calculate_platform_overlap" function, and an "age_rating_sort" function (this one sorts the tv shows by different age categories and prints how many per age category are available on each platform).

In order to run this code, simply type "cargo run" into the terminal while in the project folder directory "\ds210_final_project". This will also run the tests module, which ensures that each function in my code works as intended. By running my code to analyze the contents of tv_shows.csv, a variety of information can be discovered.



The centrality function tells us how many shows each platform in the dataset contains. From the results, we see that Netflix has the highest number of shows (1971), while Disney+ has the lowest (352). Meanwhile, the degree distribution allows us to see the amount of shows that are only on 1 streaming service vs 2 streaming services, 3, and 4 respectively. From the results above, we see that a vast majority of shows can only be found on 1 streaming platform; however,

there are still several shows that can be found on more than one platform. These results were taken into consideration when implementing the next two functions, which determine the exclusive content of each platform as well as how much each platform overlaps with one another:

```
Percentage of TV shows exclusive to each platform:
Netflix: 36%
Hulu: 30%
Prime Video: 34%
Disney+: 6%

Overlap counts between streaming platforms:
Netflix-Hulu overlap count: 129
Netflix-Prime Video overlap count: 108
Netflix-Disney+ overlap count: 12
Hulu-Prime Video overlap count: 161
Hulu-Disney+ overlap count: 36
Prime Video-Disney+ overlap count: 4
```

From these results, we see that Netflix has the most exclusive content on its platform at 36%. However, it also has a large amount of overlap with both Hulu and Prime video. That being said, Hulu and Prime video have the highest overlap overall with one another, sharing 161 shows. Disney+ does not have a strong amount of overlap with other platforms; however, this may also be due to the fact that this dataset does not include all of the shows present on its platform.

```
Age Category: Adult
Netflix: 482
Prime Video: 164
Disney+: 2
Hulu: 261

Age Category: All
Disney+: 129
Netflix: 176
Prime Video: 173
Hulu: 124
```

Finally, we have the results of the age_rating_sort function (the output above only shows results for 2 of the age categories - there are also results for Children, Teen, Mature, and Unknown content). We can conclude from this that if you are looking for content suitable for all ages, each platform has a strong amount to offer (the numbers of shows are relatively close to each other). However, if you are looking for adult content, Netflix has a much larger variety to choose from, while Disney+ does not have nearly as much in this regard. Using these results, as well as for the other age categories, is very effective in helping to select a streaming platform that is suitable for the kind of content you are looking to watch.

Overall, using this code provides a variety of insights about the relationships between different streaming platforms and the content they contain.