Cindy Chen

DS210

Professor Leonidas Kontothanassis

Dataset: https://www.kaggle.com/datasets/maso0dahmed/netflix-movies-and-shows/data

○ **This dataset contains about 5752 entries along with various features such as title of movie, genre, runtime, imdb_scores, and many more.**

For my project, I used a Netflix Data set that contains different movies and their associated genres. My goal was to find the most popular genre, and get deeper insights into the genre trends and patterns. I started out by reading the csv file, and cleansing the data. I realized there were some missing values in the data set, therefore I just ignored it. I also created 3 branches in order to organize my codes. Next, I represented each movie as an individual node, and each genre as a separate node, with edges linking movies to their corresponding genres. This will allow me to analyze the interconnected nature between movies and genres.

To determine the most popular genre, I calculated the degree centrality for each genre node. This will allow me to measure the number of direct connections a node has, which helps indicate the most popular genre based on how many movies are associated with each genre. I printed out the degree centrality for each genre as shown below.

Output:

Genre:'crime', Degree Centrality:47
Genre:['horror', Degree Centrality:22
Genre:'european', Degree Centrality:21
Genre:['western'], Degree Centrality:0
Genre:['action'], Degree Centrality:0
Genre:['war', Degree Centrality:17
Genre:['western', Degree Centrality:20
Genre:['animation', Degree Centrality:29
Genre:'history'], Degree Centrality:27
Genre:['reality'], Degree Centrality:0
Genre:['action', Degree Centrality:34
Genre:['thriller'], Degree Centrality:0
Genre:'comedy', Degree Centrality:51
Genre:['horror'], Degree Centrality:0
Genre:['documentation', Degree Centrality:30
Genre:'horror', Degree Centrality:42
Genre:['animation'], Degree Centrality:0
Genre:['comedy', Degree Centrality:34

Genre:'reality', Degree Centrality:15
Genre:['history', Degree Centrality:6
Genre:['sport', Degree Centrality:3
Genre:['romance'], Degree Centrality:0
Genre:'family', Degree Centrality:46
Genre:'war', Degree Centrality:32
Genre:'history', Degree Centrality:42
Genre:'action'], Degree Centrality:30
Genre:['reality', Degree Centrality:18
Genre:'documentation', Degree Centrality:43
Genre:'war'], Degree Centrality:28
Genre:['fantasy'], Degree Centrality:0
Genre:['documentation'], Degree Centrality:0
Genre:'crime'], Degree Centrality:27
Genre:['family'], Degree Centrality:21
Genre:'sport'], Degree Centrality:24
Genre:['crime'], Degree Centrality:0
Genre:'western', Degree Centrality:28

Genre:['family'], Degree Centrality:0

Genre:['crime', Degree Centrality:28

Genre:'thriller', Degree Centrality:47

Genre:'horror'], Degree Centrality:22

Genre:'action', Degree Centrality:45

Genre:'western'], Degree Centrality:10

Genre:['fantasy', Degree Centrality:24

Genre:[], Degree Centrality:0

Genre:['scifi'], Degree Centrality:0

Genre:'romance', Degree Centrality:48

Genre:'reality', Degree Centrality:15

Genre:['history', Degree Centrality:6

Genre:['sport', Degree Centrality:3

Genre:['romance'], Degree Centrality:0

Genre:'family'], Degree Centrality:31

Genre:'scifi'], Degree Centrality:24

Genre:['music'], Degree Centrality:0

Genre:'fantasy', Degree Centrality:45

Genre:'scifi', Degree Centrality:46

Genre:['romance', Degree Centrality:31

Genre:'fantasy'], Degree Centrality:25

Genre:'european'], Degree Centrality:34

Genre:['war'], Degree Centrality:0

Genre:'reality'], Degree Centrality:20

Genre:'drama'], Degree Centrality:34

Genre:'sport', Degree Centrality:32

Genre:['history'], Degree Centrality:0

Genre:['comedy'], Degree Centrality:0

Genre:['scifi', Degree Centrality:32

Genre:'music'], Degree Centrality:24

Genre:'animation'], Degree Centrality:24

Genre:['music', Degree Centrality:15

Genre:['drama', Degree Centrality:36

Genre:'music', Degree Centrality:38

Genre:'drama', Degree Centrality:50

Genre:'comedy'], Degree Centrality:28

Genre:'romance'], Degree Centrality:31

Genre:['thriller', Degree Centrality:35

Genre:'animation', Degree Centrality:46

Genre:'documentation'], Degree Centrality:31

Genre:'thriller'], Degree Centrality:31

Genre:['drama'], Degree Centrality:0

The most popular genre based on degree centrality is 'comedy' with a centrality score of 51

We can see that a higher degree centrality signifies how the genre is frequently chosen for various concepts for movies in the dataset. In this case, we can see that "comedy" has the highest degree centrality of 51. By analyzing the degree centrality of comedy, we can make several analyses/implications. For example, filmmakers can see the high success of comedy acts, therefore they might be more prompted to increase production of comedy films, since it aligns more closely with the audience preferences. This will definitely help the film industry to highlight more comedic elements and eventually blend it into other types of genre.

The second method I chose to analyze the types of genres was the Betweenness centrality. Betweenness centrality helps measure the influence within a network which helps determine the shortest path between two other nodes. In this case, using betweenness centrality will help provide insight on how genres mediate the relationship between different movies. Firstly, I started off initializing all the genre nodes to 0, ensuring that each genre is evaluated fairly based on its connectivity within the network. I initialized the short_paths, short_path_counts, short_distance. Next, I decided to use BFS (breadth first search) to

help find the shortest path within each connecting node. Using this algorithm will help us explore the network one by one, ensuring that all paths are evaluated in the shortest manner. I tracked the shortest paths using a hashmap to count and store the paths and distances. I used a queue to manage the nodes being explored, starting out with the source node. As the BFS progresses, it updates "short_paths" and "short_path_counts", based on the nodes it visits. For each node visited, if a neighboring node is found that it can be reached shorter through the current node, the path count, and path details will be updated accordingly. By using the betweenness centrality through BFS, we can see which genres hold significant bridging roles within the network. Down below is the output of each genre and their centrality.

Output:

Genre: 'action': betweenness centrality: 0.02670961842786313
Genre: ['war']: betweenness centrality: 0
Genre: 'fantasy']: betweenness centrality: 0.08302525499627564
Genre: 'war']: betweenness centrality: 0.003995433789954338
Genre: 'action']: betweenness centrality: 0.05399897912992192
Genre: 'european': betweenness centrality: 0.0015220700152207
Genre: ['sport': betweenness centrality: 0
Genre: ['comedy']: betweenness centrality: 0
Genre: 'thriller': betweenness centrality: 0.0044771196618332091
Genre: 'animation']: betweenness centrality: 0.0003932014205986809
Genre: ['horror']: betweenness centrality: 0
Genre: ['history']: betweenness centrality: 0
Genre: ['fantasy']: betweenness centrality: 0
Genre: ['western': betweenness centrality: 0.01029173094477343
Genre: 'comedy']: betweenness centrality: 0.02446183953033268
Genre: 'family']: betweenness centrality: 0.10937796429503385
Genre: []: betweenness centrality: 0
Genre: ['music': betweenness centrality: 0.000380517503805175
Genre: ['action']: betweenness centrality: 0
Genre: 'sport': betweenness centrality: 0.0057077762557077625
Genre: ['reality': betweenness centrality: 0.0023510397743472375
Genre: ['music']: betweenness centrality: 0

Genre: 'music': betweenness centrality: 0.020502608194492133
Genre: ['documentation']: betweenness centrality: 0
Genre: ['romance']: betweenness centrality: 0.04773685839948457
Genre: 'animation': betweenness centrality: 0.00002061136478944698
Genre: 'drama': betweenness centrality: 0.06804734387413007
Genre: ['family']: betweenness centrality: 0
Genre: 'sport']: betweenness centrality: 0
Genre: 'reality': betweenness centrality: 0.003932014205986809
Genre: ['western']: betweenness centrality: 0
Genre: ['history': betweenness centrality: 0.000380517503805175
Genre: ['action': betweenness centrality: 0.005027878995268764
Genre: ['animation']: betweenness centrality: 0
Genre: ['war': betweenness centrality: 0.0025684931506849314
Genre: ['crime': betweenness centrality: 0.024674475807491356
Genre: 'family': betweenness centrality: 0.000012683916793505835
Genre: ['crime']: betweenness centrality: 0
Genre: 'european']: betweenness centrality: 0.06013871242937184
Genre: 'music']: betweenness centrality: 0.0007927447995941147
Genre: 'horror']: betweenness centrality: 0.011066717402333842
Genre: 'romance': betweenness centrality: 0.020000951293759513
Genre: ['drama': betweenness centrality:

Genre: ['documentation': betweenness centrality: 0.00210942184401085
Genre: ['horror': betweenness centrality: 0.0071822678843226784
Genre: ['thriller']: betweenness centrality: 0
Genre: 'fantasy': betweenness centrality: 0.0000111916912883875
Genre: ['comedy': betweenness centrality: 0.062284711255553314
Genre: ['family': betweenness centrality: 0.015220700152207
Genre: ['animation': betweenness centrality: 0.012398528665651954
Genre: 'horror': betweenness centrality: 0.00026636225266636225
Genre: ['reality']: betweenness centrality: 0
Genre: 'western': betweenness centrality: 0.0005707762557077625
Genre: 'thriller']: betweenness centrality: 0.017123287671232876
Genre: 'crime']: betweenness centrality: 0.007039573820395738
Genre: 'history']: betweenness centrality: 0.03019225193882728
Genre: 'scifi': betweenness centrality: 0.008800072938851097
Genre: ['romance']: betweenness centrality: 0
Genre: 'history': betweenness centrality: 0.0019377761800500202
Genre: 'documentation': betweenness centrality: 0.008339675291730086
Genre: 'romance']: betweenness centrality: 0.01902587519025875

0.054296915443016826
Genre: ['thriller': betweenness centrality: 0.0025126650252600313
Genre: 'crime': betweenness centrality: 0.012964717647820647
Genre: 'war': betweenness centrality: 0.004870624048706241
Genre: ['fantasy': betweenness centrality: 0.005136986301369863
Genre: ['drama']: betweenness centrality: 0
Genre: 'comedy': betweenness centrality: 0.0079839027715581359
Genre: ['scifi': betweenness centrality: 0.01573284251829935
Genre: 'documentation']: betweenness centrality: 0.03158523408881684
Genre: 'western']: betweenness centrality: 0
Genre: 'drama']: betweenness centrality: 0.06330657091561427
Genre: ['scifi']: betweenness centrality: 0
Genre: 'scifi']: betweenness centrality: 0.014123914405945025
Genre: 'reality']: betweenness centrality: 0.015125570776255707

The genre with highest betweenness centrality is ["'family']' with a score of 0.1094

In our case, we can see that the genre with the highest betweenness centrality is [family] with a score of 0.1094. This high score indicates that the "Family" genre frequently acts as a bridge on the shortest paths between other nodes in the network, thus connecting a diverse array of movies and genres. This centrality suggests that filmmakers often integrate the "Family" genre with other genres, serving as a crossover point in various cinematic combinations. For example, a family movie can be commonly associated with other genres such as comedy, adventure, fantasy and more. We can see that the "Family" genre reflects its broad appeals across different demographic groups, because family movies are typically designed to be inclusive for all ages. This inclusivity not only ensures a wide-reaching appeal to family films, but also enhances the cohesiveness of the movie genre network. Knowing that "Family" has a significant role in bridging different genres, filmmakers can use this as a leverage to their advantages by creating more family entertaining movies. By doing so, filmmakers can attract a larger and more diverse audience,

which can help enhance viewers' engagement and potentially increase the commercial success of their productions.

Sources:

- https://docs.rs/petgraph/latest/petgraph/visit/struct.Bfs.html
- https://doc.rust-lang.org/book/title-page.html
- https://cambridge-intelligence.com/keylines-faqs-social-network-analysis/
- https://drive.google.com/file/d/1stBcSKQiz1IlY3ZwGpsIoXAiDGOuMnaK/view
- https://en.wikipedia.org/wiki/Centrality