Tia Hannah
DS210
12/15/22

Ds210 Project Report

The dataset I used was a Kaggle dataset about the top 1000 movies and tv shows on IMDB. In the dataset there are 16 columns which include data for the Link of the poster that imdb using, the Name of the movie, the Year when the movie released, certificates earned by the movie, the runtime,gener, IMDB rating,summary of the movie,the metascore,director, number of votes for the movie, the money earned by the movie, and 3 columns dedicated to the star actors of the movie.This dataset has around 1000 entries of different shows and movies from a wide range of years. I chose this dataset because I felt like it would give an accurate view of the director and score pair that was most connected to the other data entries in the graph. The features I used were the director of the movie and the metascore. I wanted to discover based on the data which nodes were the most connected to other nodes.The algorithm I used was the degrees of centrality algorithm. I created a petgraph where each node includes the metascore and the director of the movie. The edges of the nodes or the factor the nodes were connected by was the metascore. I found the connections between the nodes for each node in the graph and divide the answer by the total number of nodes to standardize the result. Based on my first run of the code, Node 588 is the most connected with a centrality score of .03924 and Node 539 is the least connected with a centrality score of 0. Node 588 is the node with director Jim Sheridan with a Metascore of 76. Node 539 is the node with the director Abhishek Kapoor with a meta score of 40. Within the dataset, many of the movies have the same metascore. This leads to the maximum and minimum centrality score directors changing despite the scores staying the same. This means that based on this dataset, the most common metascore for a movie to receive is 76 and the least common score is 40.