Influence of Actors in Movie Gross: A Network Analysis

Hilal Urun, AmirHossein Motameni, Seyed Ata Seyed Jafari

January 30, 2023

Introduction

The film industry is one of the largest and most dynamic markets globally. The success of a movie can be influenced by a variety of factors, including the cast, budget, marketing, and reviews. In this project, the focus is on the influence of actors on a movie's gross. The goal is to understand how actors' relationships, as measured by centrality measures in a network, can impact a movie's financial success.

Motivation

The motivation behind this project stems from the desire to gain insight into the factors that contribute to the success of a movie. By analyzing the network relationships between actors, it is possible to identify key players in the industry and understand their impact on movie success. This information can be useful for filmmakers, producers, and marketers who aim to make profitable movies.

In our proposal we wanted to use *GCN* and *GraphSAGE* to predict movie gross knowing the cast of the movie. Then we supposed to build a network in which we had two types of nodes. Actor nodes and Movie nodes. Also there should be an edge between nodes if the actor have played in the movie. Here we faced some problems. First of all, we had two different types of nodes which means we had to build two different types of feature vectors for our nodes. This when we have considered only one feature for movie nodes which was the *Gross*. So we decided to use node-embedding as feature vectors for the nodes. The main problem here again was the connectivity of two different types of nodes.

So we decided to change our point of view. We wanted to see the influence of actors in movie's financial success, and we realized that we can measure this influence using centralities. We are going to measure *Degree Centrality*. Then we are going to use this values to measure the correlation with *Gross*.

Dataset

The dataset used in this project is The Movies Dataset, which can be found on Kaggle here. The dataset includes information on movies, actors, and their connections. These files contain metadata for all 45,000 movies listed in the Full MovieLens Dataset. The dataset consists of movies released on or before July 2017. Data points include cast, crew, plot keywords, budget, revenue, posters, release dates, languages, production companies, countries, TMDB vote counts and vote averages. The information on the cast and gross of each movie was used to build the network and analyze the centrality measures. The dataset was selected because it provides comprehensive information on the relationships between actors, which is critical to understanding their impact on movie success.

We have build the network in the 'preprocessing_data.ipynb'. Our network has 46039 nodes and 115529 edges. Each node represent an actor and there is an edge between them if they played in a same movie together.

Method

To achieve the goal of this project, a network was built using the movie dataset from Kaggle. The actors were treated as nodes, and connections were established between actors who appeared in the

same movie. The centrality measures were then calculated for each actor to assess their importance in the network. We have used *Degree Centrality*. Finally, the relationship between these centrality measures and the movie's gross was analyzed to determine if there is a correlation between the centrality measures and the success of a movie in terms of its gross.

To see the results, I suggest you to look at the Notebook in our github repository. It is really well-commented and easy to follow.

Experiments

The results of the experiment showed that there is a positive correlation between degree centrality and movie's gross, indicating that actors with a high number of connections to other actors tend to be associated with movies with higher grosses. For better understanding, we suggest you to see this notebook.

Conclusion

Our findings indicate a strong relationship between the degree centralities of movie stars and the box office revenue of the films they star in. This correlation suggests that the popularity and influence of a movie's lead actors play a significant role in determining the financial success of a film. This highlights the importance of casting well-known and highly connected stars in the film industry. These results provide valuable insight for film producers and investors, as they can use this information to make informed decisions about casting and marketing strategies.Moreover, the results of this correlation analysis can be used to predict the potential box office revenue of a movie based on the degree centralities of its lead actors. By understanding the relationship between star power and box office success, film producers can make more strategic decisions about which actors to cast, and how to market their films to maximize revenue. Furthermore, these findings can also be used to guide talent management and representation, as they demonstrate the value of investing in and promoting actors with high degree centralities. In conclusion, the significant correlation between degree centralities of stars and movie box office revenue provides a valuable tool for decision making in the film industry.