

Analyzing IMDB movie ratings with respect to age group

Syed Asad Shah

22001127

Hamza Islam

22101350

In this project we are going to analyze IMDB movies ratings given by people falling in different age groups. We will try to find co-relation between the age group and the genre of movie, which might help in explaining different feelings and thoughts felt by that age group. We can also try to partition the data on other demographics like gender and country. We have obtained two IMDB datasets [1] from Kaggle i.e., movies.csv and rating.csv. Collectively, we will be analyzing around 170k+ rows of data. We will identify nodes as age-groups and genre and then create link between them using the ratings given by users to movie where of course we will use ratings value as edge-weight. The dataset obtained has the age groups defined as 0-18, 18-30, 30-45 and 45+ whereas movie genres are drama, comedy, action, adventure, fiction, romantic etc. and their combination as group.

We will perform the analysis such as degree distribution analysis, centrality analysis, page rank analysis, density analysis and path analysis etc. We can perform these analysis and others more that we might learn as the course progresses using timestamps also. This can help us find the change in relations or likeness or dislikeness over a course of time. We can visualize the trend over the period using Gephi or ORA. This analysis can also help us predict the popularity of an upcoming movie of a certain genre in different groups of people, which eventually will make filmmakers take better decisions when targeting the movie for a particular group or it can help marketing team to better focus on the right targeted audience for the promotion of a movie.

References

[1] Stefano Leone. IMDb movies extensive dataset. <https://www.kaggle.com/stefanoleone992/imdb-extensive-dataset>. Accessed: 2021-10-25