**DEPARTMENT OF COMPUTER SCIENCE AND**

**ENGINEERING**

A report

Submitted for the partial fulfilment of the course

**19OH01-Social And Economic Network Analysis**

A report on

**“VISUALIZATION AND ANALYSIS FROM IMDB”**

A report submitted by

**P.Ajay Vikram (16z204)**

**S.Mohamed Farhan (18z231)**

**R.Sonaa (18z439)**

**S.Nitin (19z432)**

**V.Naveen (19z462)**



PSG COLLEGE OF TECHNOLOGY

**TABLE OF CONTENTS**

**Problem Statement 3**

**Dataset Description 3**

**Tools Used 4**

**Challenges Faced 4**

**Contribution of Team Members 4**

**Annexure 2 Output Snapshots 5**

**References 8**

**Problem Statement**

Currently with the rapid development, Movies has been a entertainment part of people’s life. A lot of sociology, biology, and information systems can use the network to describe, in which nodes represent individual and edges represent the relationships between individuals or the interaction between individuals. Therefore, the study of complex networks has been the important branch of many scientific fields. Analysing data from IMDB is to predict whether there will be links between two nodes based on the attribute information and the observed existing link information. Data analysis not only can be used in the field of social network but can also be applied in other fields. Scrapping data from the IMDB and analyzing the result from the nodes and edges. Data Scraping from a website is one of the way to get valuable data about present trends especially because most of the data in this age really comes from the datas coming from different website especially from Media Sites Now, if one wants to analyze what's the trend movie currently and use that data for personal or business reasons, scraping data from the popular movie website like IMDb is the way to go.

**Dataset Description**

This dataset consists of rating and reviews from IMDB.Data was collected from survey in which the reviewver use to review the movies using this movie app. The dataset includes node and edges.

|  |  |
| --- | --- |
| **Dataset Statistics** | |
| **Nodes** | 200 |
| **Edges** | 991 |
| **Network Diameter** | 9 |
| **Graph Density** | 0.025 |
| **Modularity** | 0.272 |
| **Average clustering coefficient** | 0.021 |
| **Strongly connected components** | 200 |

**Tools Used**

* Gephi

Gephi is the leading visualization and exploration software for all kinds of graphs and networks.

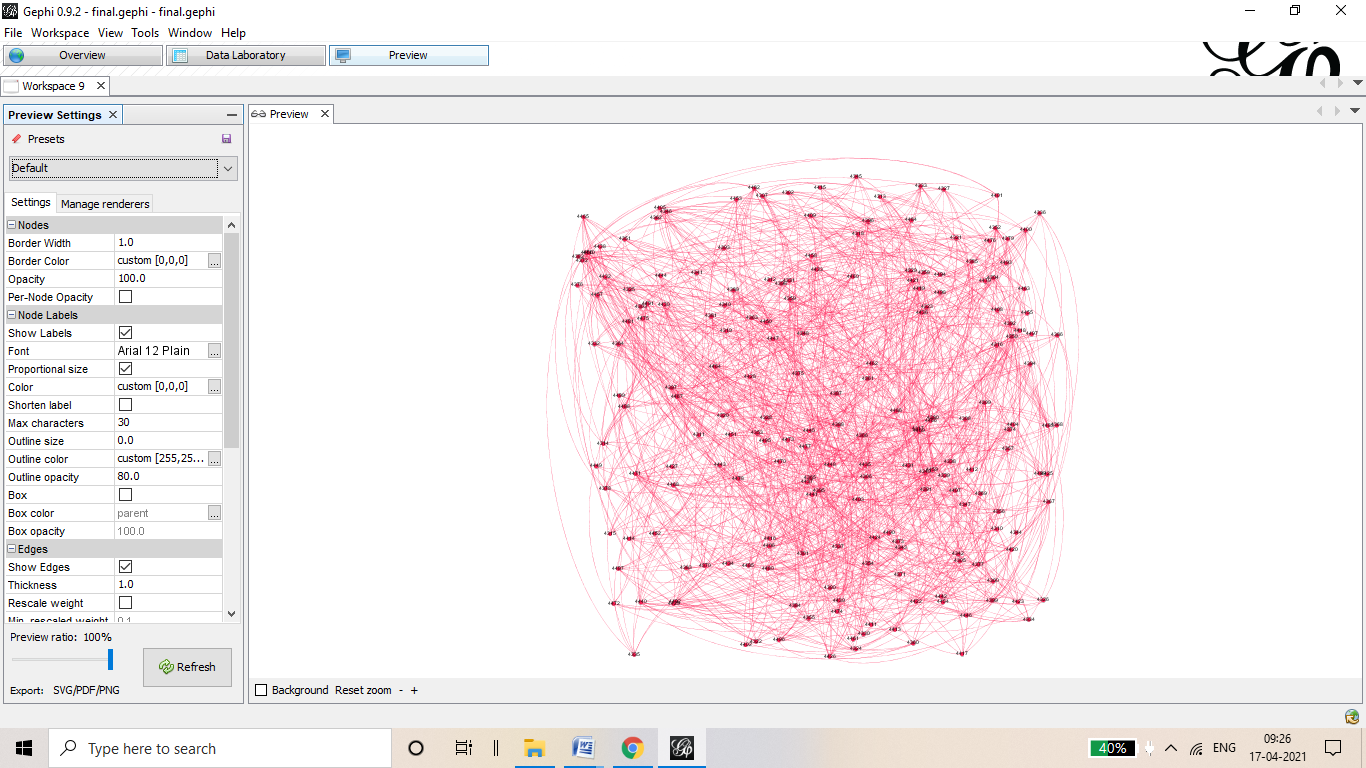
**Challenges Faced**

1. We cant able to find the dataset for the IMDB
2. Importing spreadsheet in gephi was difficult because it reached or exceeded its storage
3. The created data for was unable to fit within the model we created, after normalizing the data, the data fitted into the model with good accuracy.

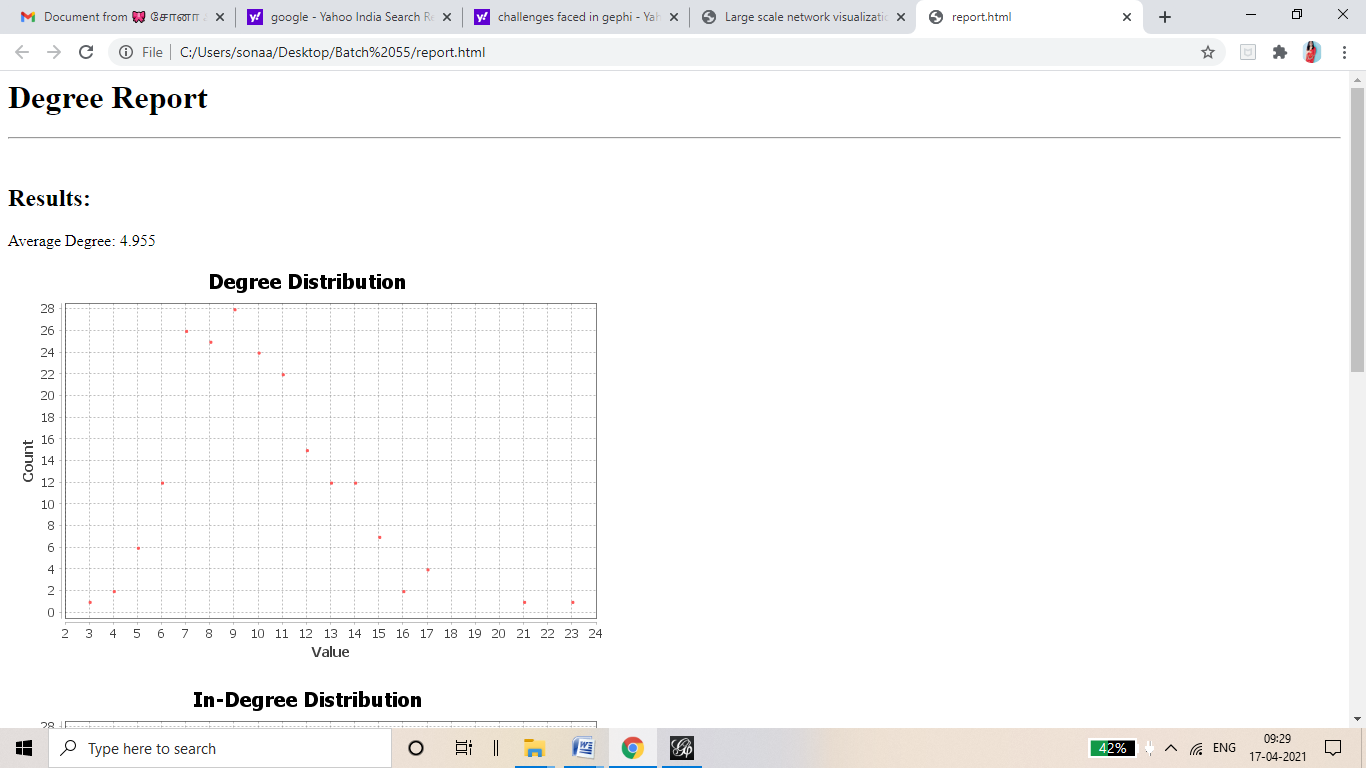
**Contribution of Team Members**

|  |  |  |
| --- | --- | --- |
| **Roll No** | **Name** | **Contribution** |
| 16z204 | Ajay Vikram.P | **Analysing average degree** |
| 18z231 | Mohamed Farhan.S | **Average weight degree** |
| 18z439 | Sonaa.R | **Model Training and report** |
| 19z432 | Nitin.S | **Clustering coefficient & report** |
| 19z462 | Naveen.V | **Average weight degree** |

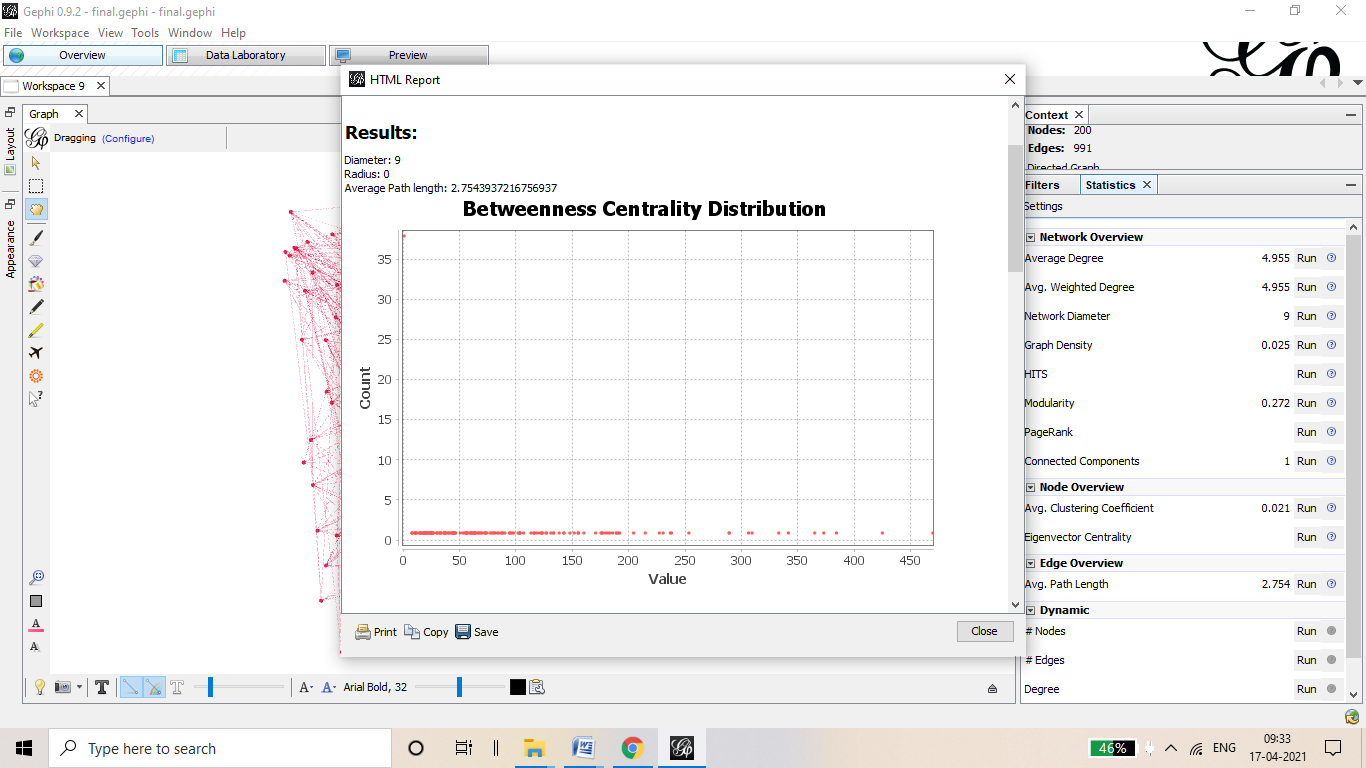
**Annexure 2 Output Snapshots:**

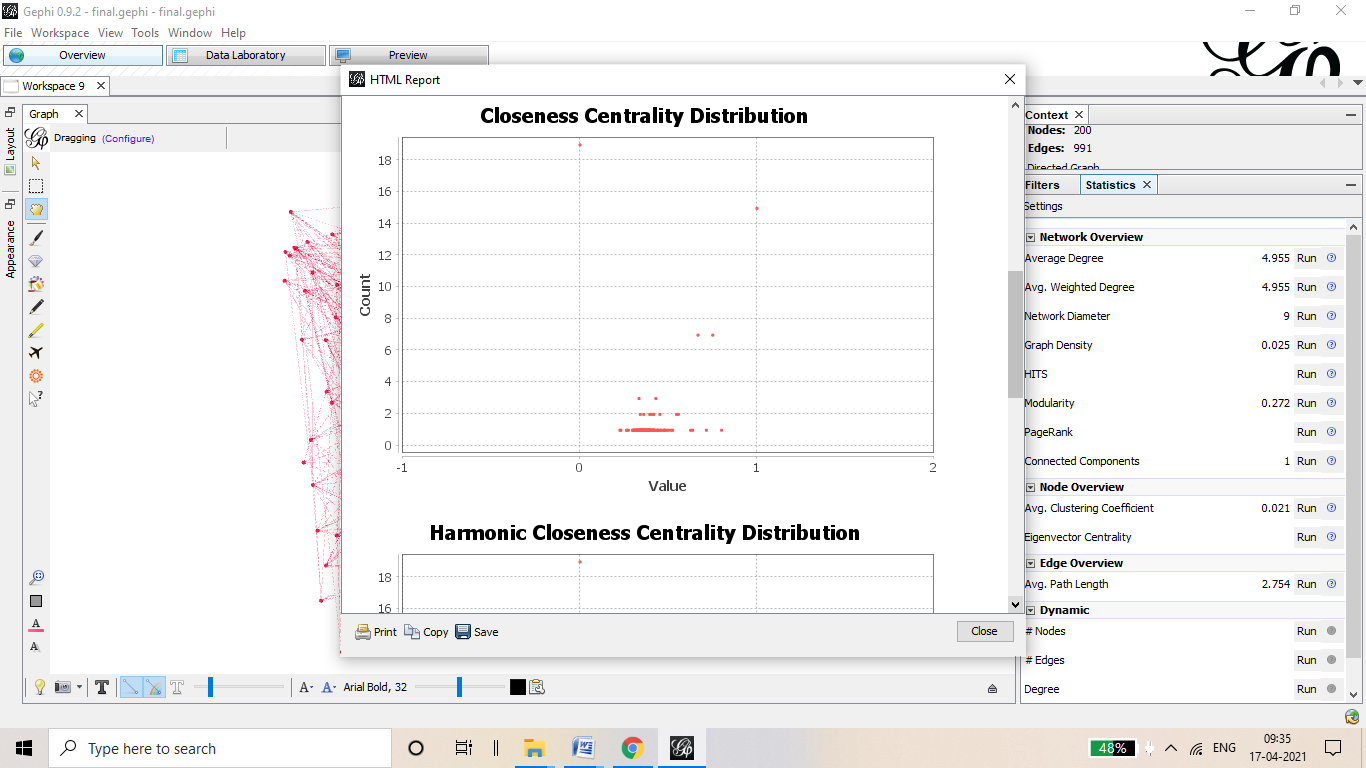


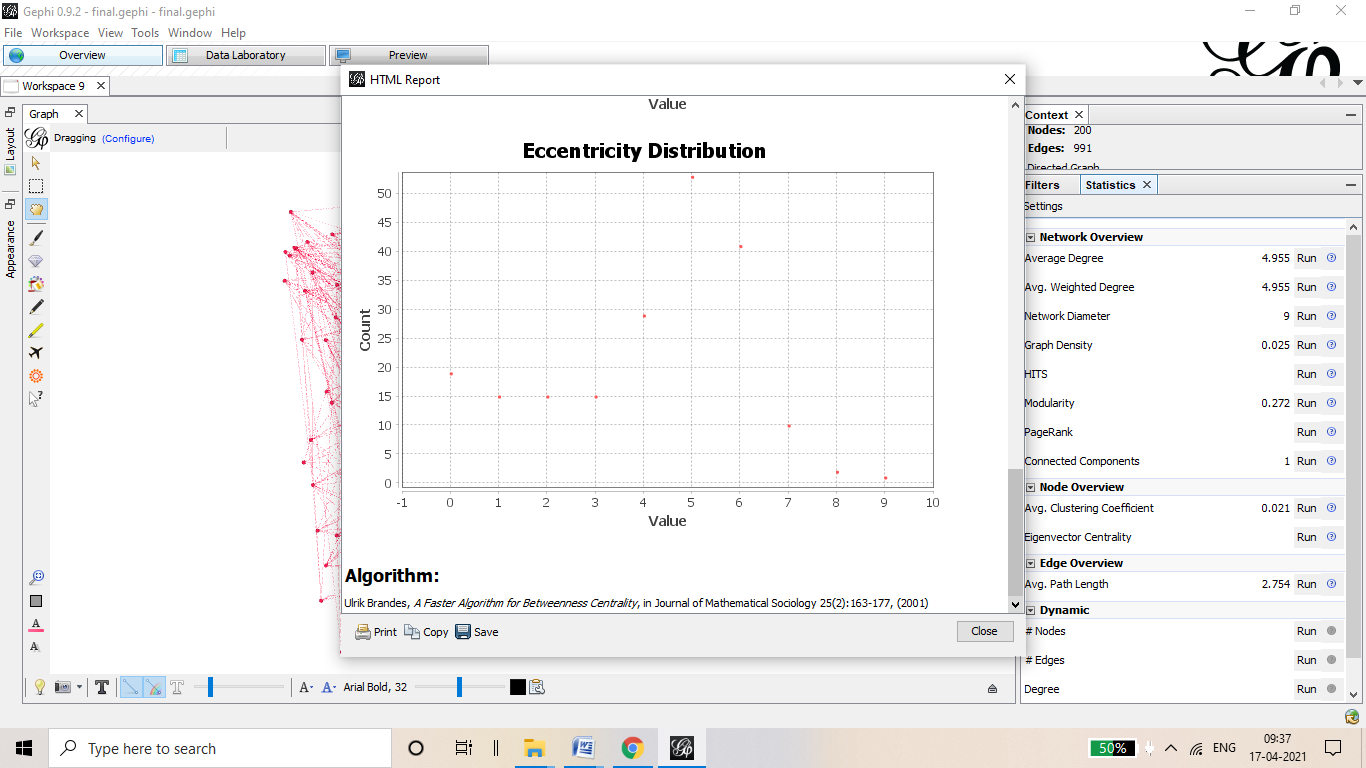
**Degree Distribution**



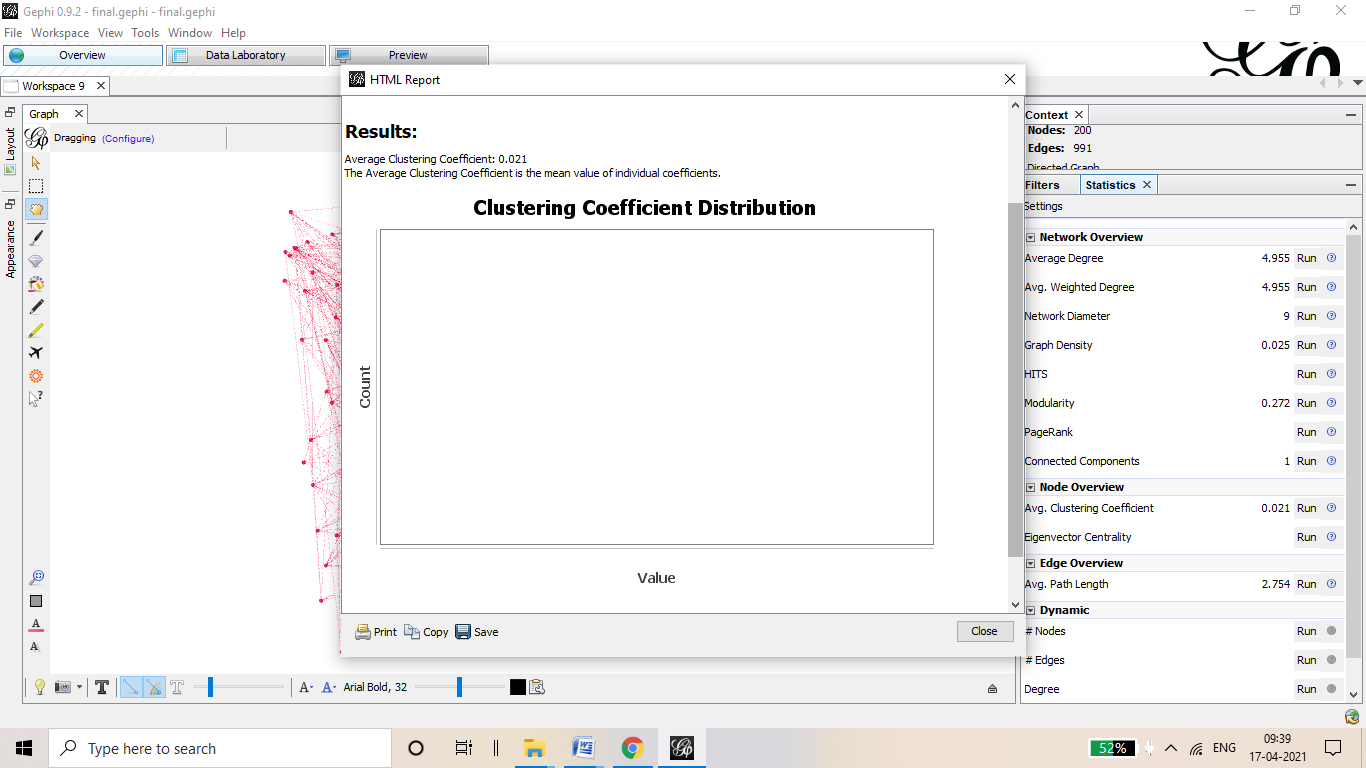
**Network Diameter**







**Clustering Coefficient**

****

**References**

[**https://www.promptcloud.com/blog/web-scraping-imdb-for-the-best-movies-and-shows/**](https://www.promptcloud.com/blog/web-scraping-imdb-for-the-best-movies-and-shows/)

[**https://github.com/Reljod/Python-Data-Scraping-IMDb-Movie-site-using-BeautifulSoup-Series-1-**](https://github.com/Reljod/Python-Data-Scraping-IMDb-Movie-site-using-BeautifulSoup-Series-1-)

[**https://towardsdatascience.com/data-analysis-and-visualization-of-scraped-data-from-imdb-with-r-5d75e8191fc0**](https://towardsdatascience.com/data-analysis-and-visualization-of-scraped-data-from-imdb-with-r-5d75e8191fc0)

[**https://github.com/nidaguler/Python-Data-Scraping-IMDb-Movie-site-using-BeautifulSoup-Series-DataFrame**](https://github.com/nidaguler/Python-Data-Scraping-IMDb-Movie-site-using-BeautifulSoup-Series-DataFrame)

[**https://dev.to/magesh236/scrape-imdb-movie-rating-and-details-3a7c**](https://dev.to/magesh236/scrape-imdb-movie-rating-and-details-3a7c)

[**https://github.com/topics/imdb-webscrapping**](https://github.com/topics/imdb-webscrapping)