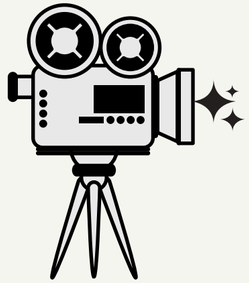
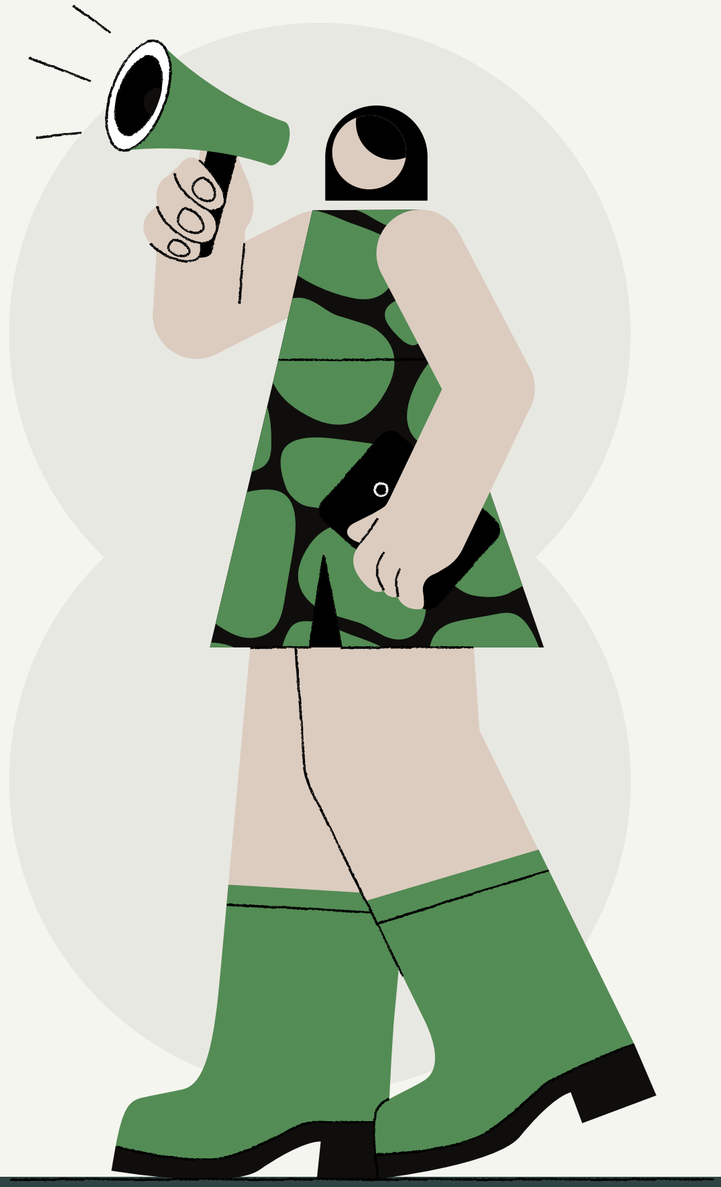
Analysis of Bollywood Movie Community and Nepotism

Presenters:
Apoorv Jain - 2023EET2190
Madhur Sharma - 2023EET2193

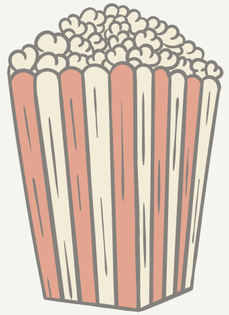


Report Outline

- Introduction
- Background
- Motivation
- Problem Definition
- Related Work
- Results
- Conclusion & Future Work



Introduction



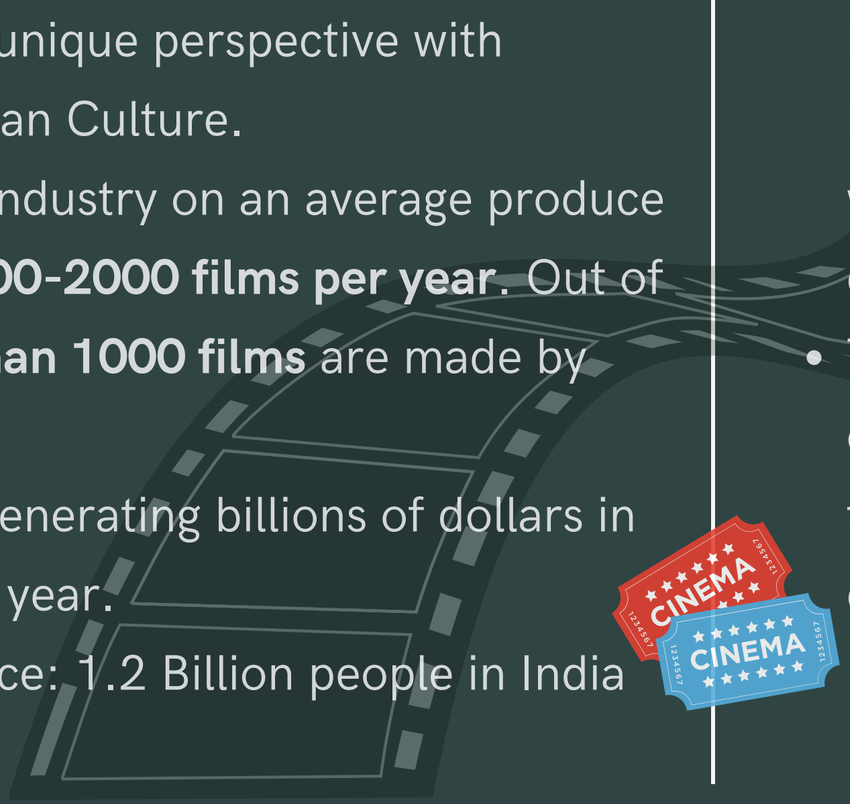
Welcome to our exploration of Bollywood Movie Industry, a global cinematic powerhouse, with captivating storytelling and larger-than-life productions.

Today, we will try to unveil the role of social connections in shaping careers and projects in this influential industry.

Background

Overview of Bollywood

- Bollywood, the heart of the global film industry, is renowned for its cinematic diversity, and unique perspective with respect to Indian Culture.
- Indian Movie Industry on an average produce more than **1500-2000 films per year**. Out of which **more than 1000 films** are made by bollywood.
- Bollywood is generating billions of dollars in revenue every year.
- Target Audience: 1.2 Billion people in India only.



Importance of Relationships in the Industry

- Beyond the glamour of the silver screen, Bollywood thrives on relationships. Historical successes of stars shows trend where personal ties shaped casting decisions.
- The depth of these relationships often extends beyond the screen, influencing the public's perception and industry dynamics.

Meet the Oscar Triumphs

A Brief History of Indian Oscar Triumphs

Indian winners at the Academy Awards

Best Costume Design
Bhanu Athaiya
"Gandhi"
1983

Best Documentary (Short Subject)
Guneet Monga
"Period. End of Sentence."
2019

Best Sound Mixing
Resul Pookutty
"Slumdog Millionaire"
2009

Best Documentary (Short Subject)
Kartiki Gonsalves, G
"The Elephant Whispe
2023

Best Original Song (Jai Ho)
A.R. Rahman, Gulzar
"Slumdog Millionaire"
2009

Best Original Song (Na
M.M. Keeravani, Cha
"RRR"
2023

Best Original Score
A.R. Rahman
"Slumdog Millionaire"
2009

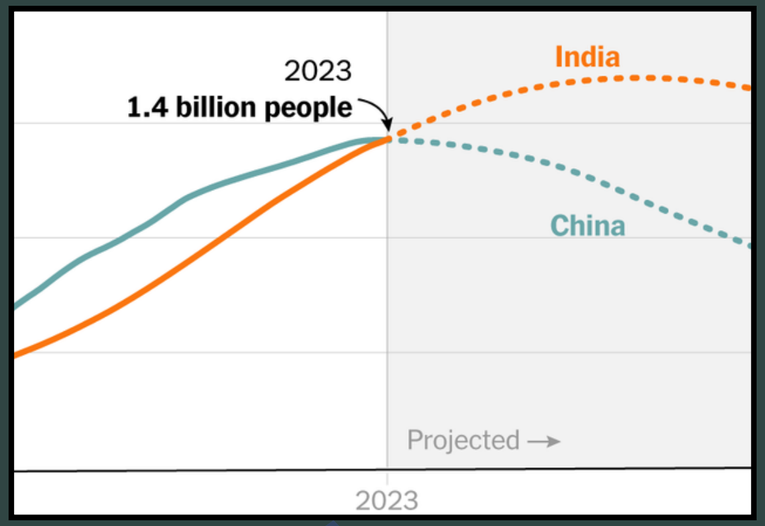
Excluding honorary awards and awards for technical achievements
Source: The Academy of Motion Picture Arts and Sciences

sta

Motivation

Total 10 Oscars with 20 nominations.

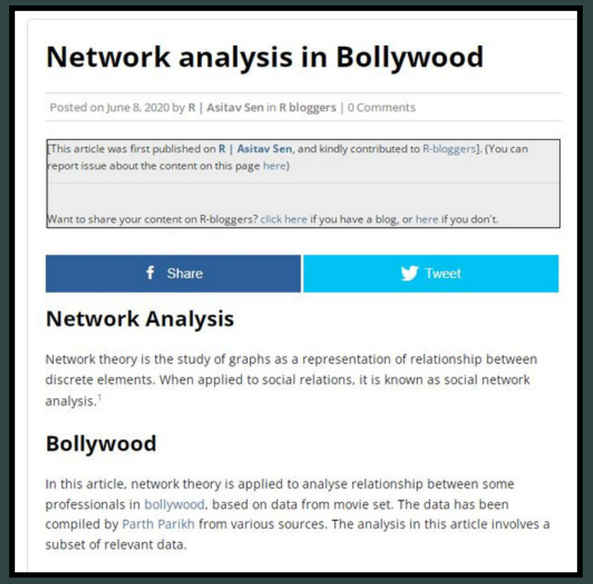
- **Gandhi**(1983) and **Slumdog Millionaire**(2009), were not even directed by an Indian.
- Other Oscars were awarded for Short Documentaries.
- Ray of Hope: RRR won Oscar for ‘Nattu Nattu’ last year, but still no award for Indian Original Film.



Huge Population

- As per projection of various reports of past years, India has overtaken China in population with 1.46 Billion people walking in the country.
- We have a huge population, still we are not able to get a **single Oscar**, for an **original Hindi Film**.

Related Work



Network analysis in Bollywood

Posted on June 8, 2020 by R | Asitav Sen in R bloggers | 0 Comments

[This article was first published on R | Asitav Sen, and kindly contributed to R-bloggers]. (You can report issue about the content on this page here)

Want to share your content on R-bloggers? click here if you have a blog, or here if you don't.

Share Tweet

Network Analysis

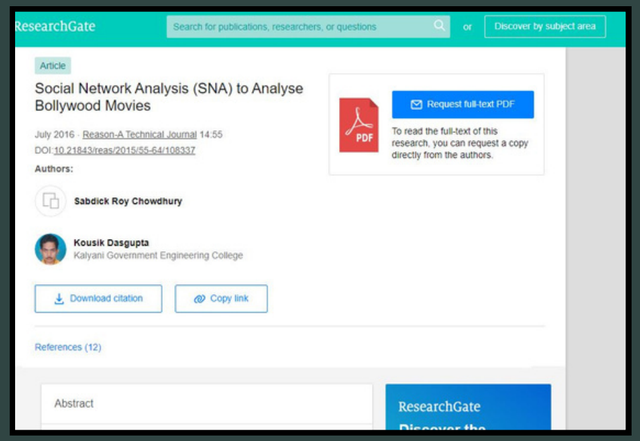
Network theory is the study of graphs as a representation of relationship between discrete elements. When applied to social relations, it is known as social network analysis.[1]

Bollywood

In this article, network theory is applied to analyse relationship between some professionals in bollywood, based on data from movie set. The data has been compiled by Parth Parikh from various sources. The analysis in this article involves a subset of relevant data.

1) Analysis of node importance and clusters in Bollywood https://www.r-bloggers.com/2020/06/network-analysis-in-bollywood/

2) Predicting futute trends of revenue for different stakeholders of Bollywood https://www.researchgate.net/publication/305785018_Social_Network_Analysis_SNA_to_Analyse_Bollywood_Movies



ResearchGate Search for publications, researchers, or questions or Discover by subject area

Article

Social Network Analysis (SNA) to Analyse Bollywood Movies

July 2016 · Reason-A Technical Journal 14:55

DOI: 10.21843/reas/2015/55-64/108337

Authors:

Sabdick Roy Chowdhury

Kousik Dasgupta
Kalyani Government Engineering College

Request full-text PDF

To read the full-text of this research, you can request a copy directly from the authors.

Download citation Copy link

References (12)

Abstract

ResearchGate

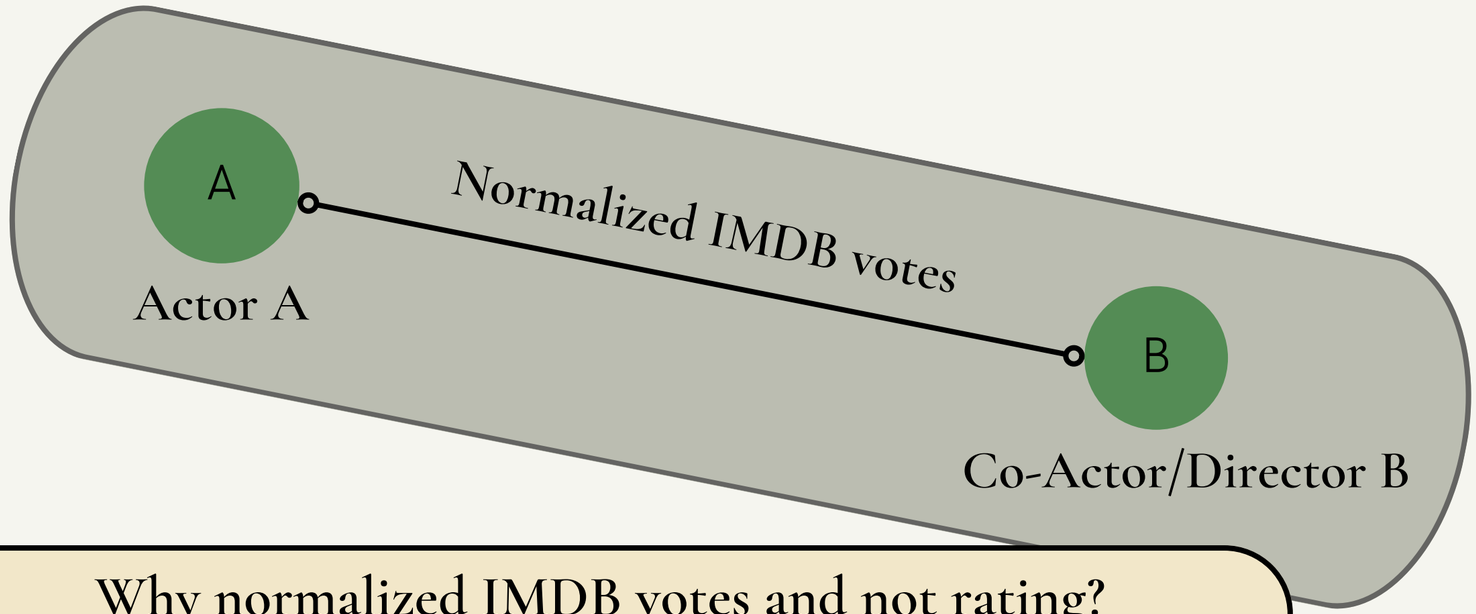
Dataset

We have used Kaggle dataset -
https://www.kaggle.com/datasets/thedevastator/timdb-bollywood-films

- It consists of all bollywood films from 1960 to 2019.
- It is retrieved from the TMDB(The Movie Database) API.
- We are using data from 1990 to 2015 for making the graph network.

```
Index(['Unnamed: 0', 'title_x', 'imdb_id', 'poster_path', 'wiki_link',
       'title_y', 'original_title', 'is_adult', 'year_of_release', 'runtime',
       'genres', 'imdb_rating', 'imdb_votes', 'story', 'summary', 'tagline',
       'actors', 'wins_nominations', 'release_date', 'directors', 'writers',
       'budget', 'revenue', 'popularity'],
      dtype='object')
```

Graph Representation



Why normalized IMDB votes and not rating?

- Number of votes give more insight on the popularity of the film.
- Less rating isn't gonna stop a movie from earning money in India.

NOTE:
Multiple Edges Between two nodes is Replaced by One edge with Average of all edge weights.

Data Preprocessing

Handling Missing Values

- Missing IMDB votes is imputed with median of non-missing entries.
- Dropping film rows with no actors information.

Normalization

- Limiting number of actors to 10 for each film.
- Normalization of IMDB votes feature.

Outlier Removal

- Dropping of the FIlm : 3 Idiots with very high IMDB votes in comparison to others.

Nepotism Data

We used OpenAI API for collecting dataset of insiders and outsiders who debuted in the time period of 2015 to 2019.

Prompt - “You are a Bollywood Relations Expert, i will give you a list of names of bollywood stars, you need to tell me if they are related to someone in the bollywood(by blood) or not by giving 1 or 0 as well as name of the related star. Return the Output for each star in JSON format like this- {"star_name":" ", "is related":1 or 0, "related_star_name": " " }. Please make the JSON compatible.”

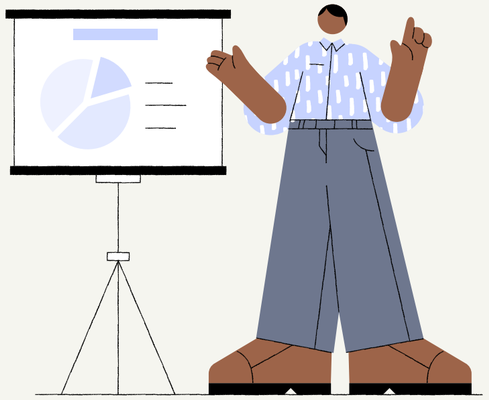
The aggregated output had to be preprocessed to generate the correct JSON formatted file.

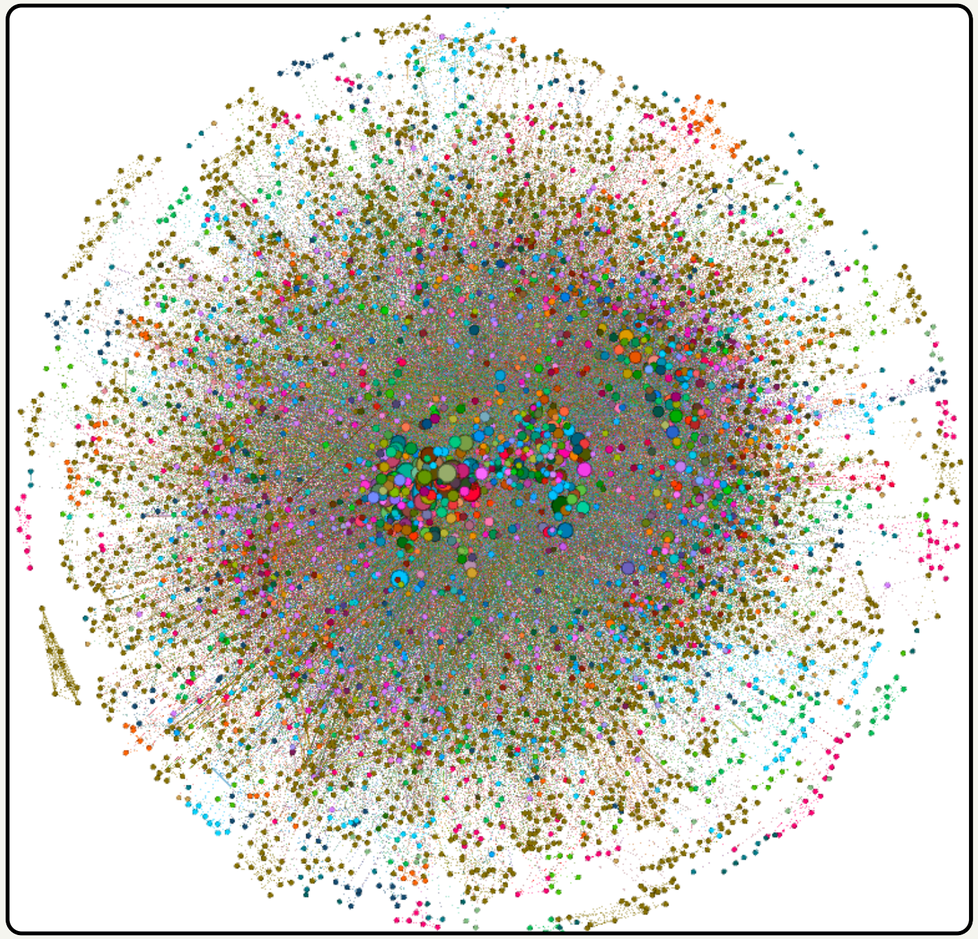
4%

```
ChatCompletionMessage(content='{"star_name": "Vicky Kaushal", "is_related": 0, "related_star_name": ""}\n
ChatCompletionMessage(content='{"star_name": "Preeti Sood", "is_related": "0", "related_star_name": "N/A"
ChatCompletionMessage(content='{"star_name": "Brinda Trivedi", "is_related": 0, "related_star_name": ""}\
ChatCompletionMessage(content='{"star_name":"Kumar Kanchan Ghosh", "is_related": 0, "related_star_name":
ChatCompletionMessage(content='{"star_name":"Mukesh Hariawala", "is_related":0, "related_star_name": ""}\
ChatCompletionMessage(content='{"star_name":"Karishma Sharma", "is_related": 0, "related_star_name": ""}\
ChatCompletionMessage(content='The output is in the following format:\n\n[\n  {"star_name":"Vaishnavi Dha
ChatCompletionMessage(content='{"star_name": "Payal Rajput", "is_related": 0, "related_star_name": ""}\n{
```

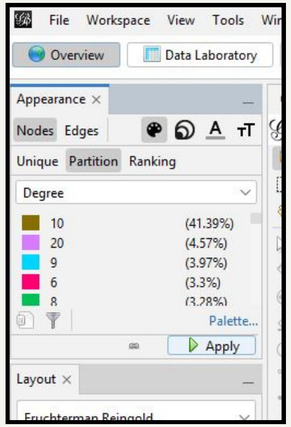
- Out of total 1000 debutants, around 40 were insiders.
- This is only 4 % of total debutants.
- Remark: 1000 debutants include all lead roles as well as supporting roles.

Network Using Gephi





Color Palette



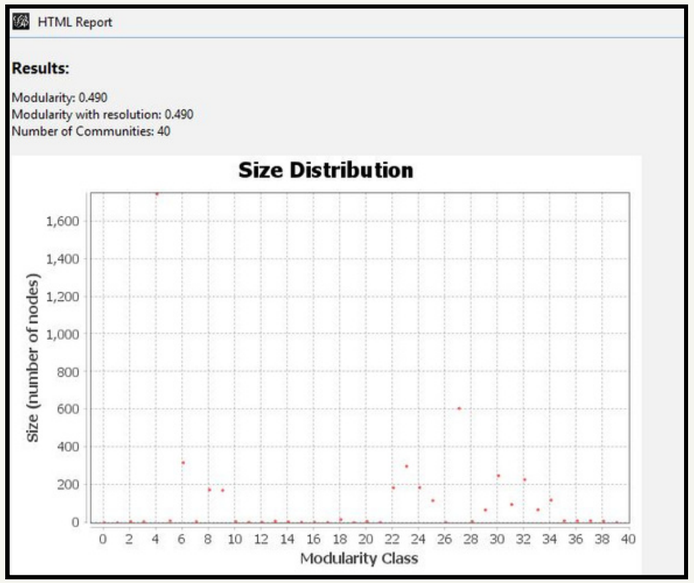
Nodes: 4802

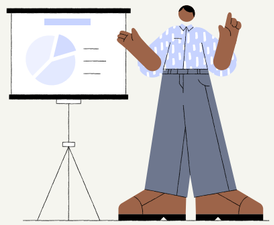
Edges: 67970

#Conn. comp = 22

Modularity: 0.49

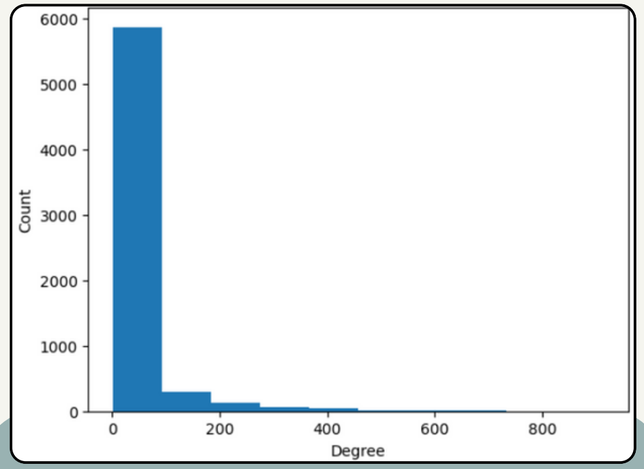
Diameter = 7





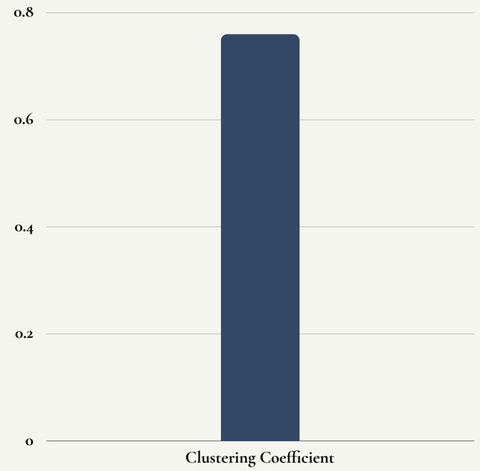
Results

Real World Properties



1) Power Law

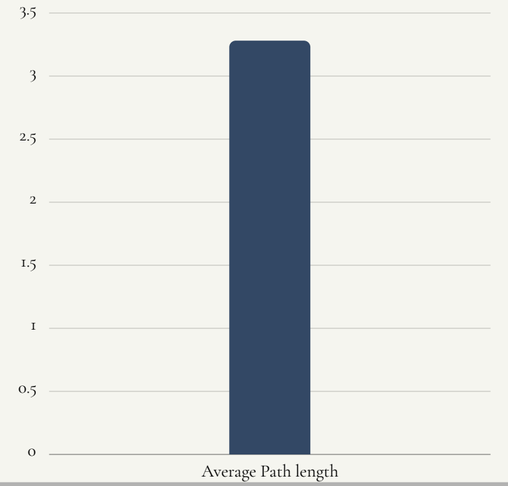
The degree distribution is following the Power Law where nodes with higher degree are less and nodes with lower degree are more.



0.76

2) Clustering Coefficient

Real world data have high clustering coefficient.



3) Small World Property

Average path length = 3.2808811035636363

Every node is at most 3 hops away from other, following the small world property.

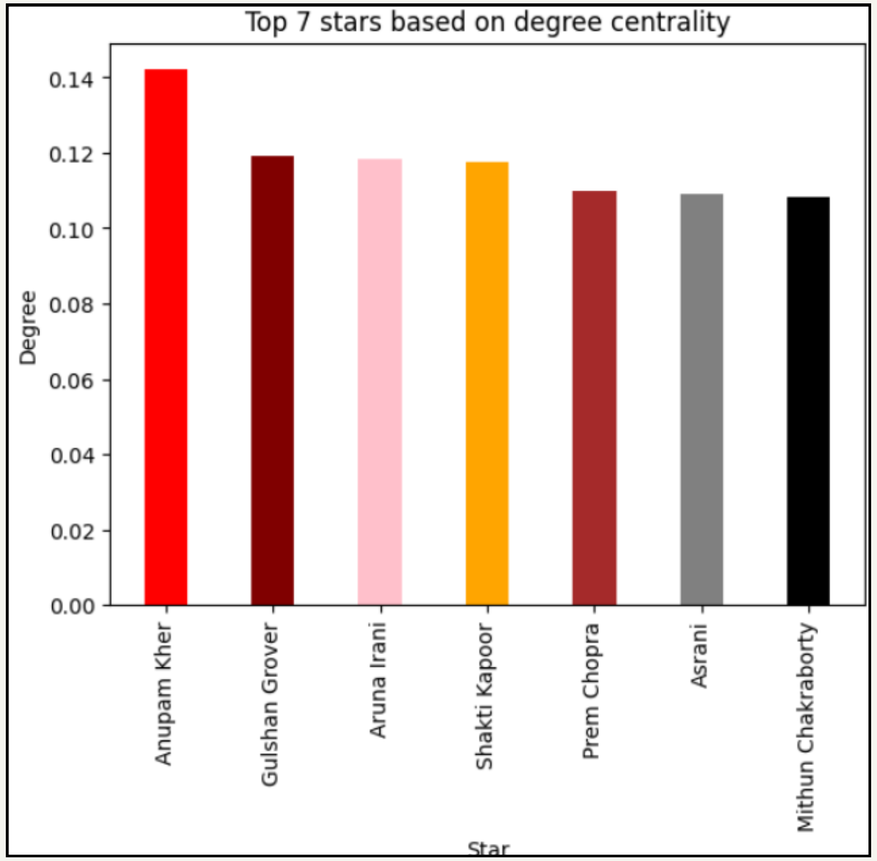
Centrality Measures

Degree Centrality

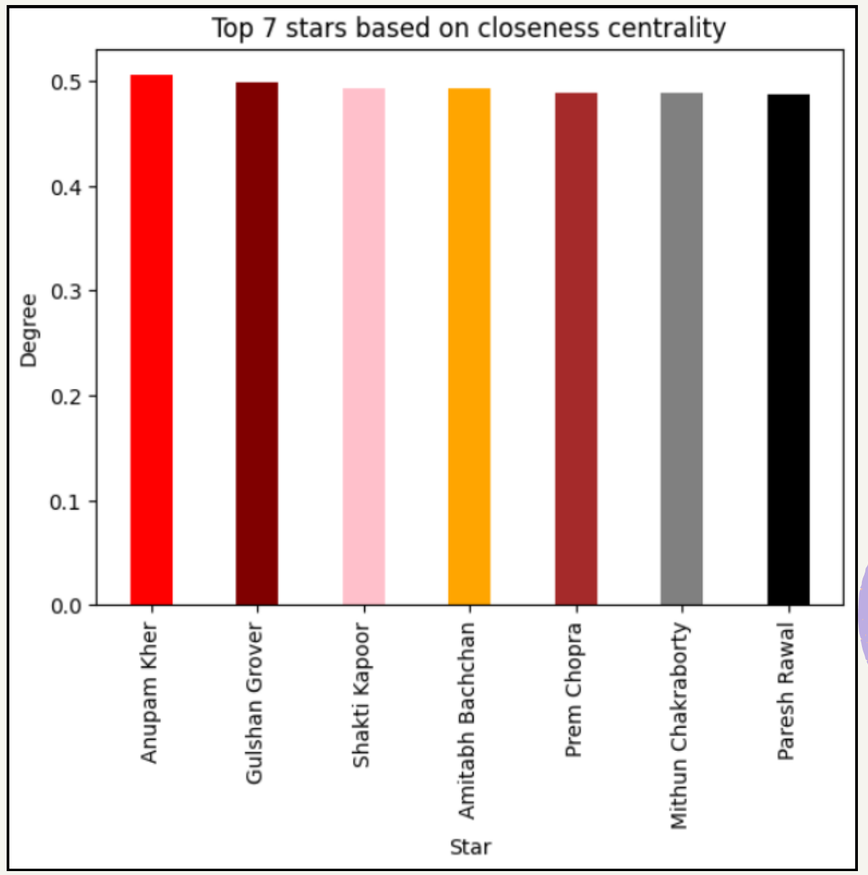
We are finding stars **importance based on film success.**We know that **Aamir Khan is a more prominent figure than Anupam Kher or Shakti Kapoor**. This proves that Degree centrality is not a practical metric for node importance.

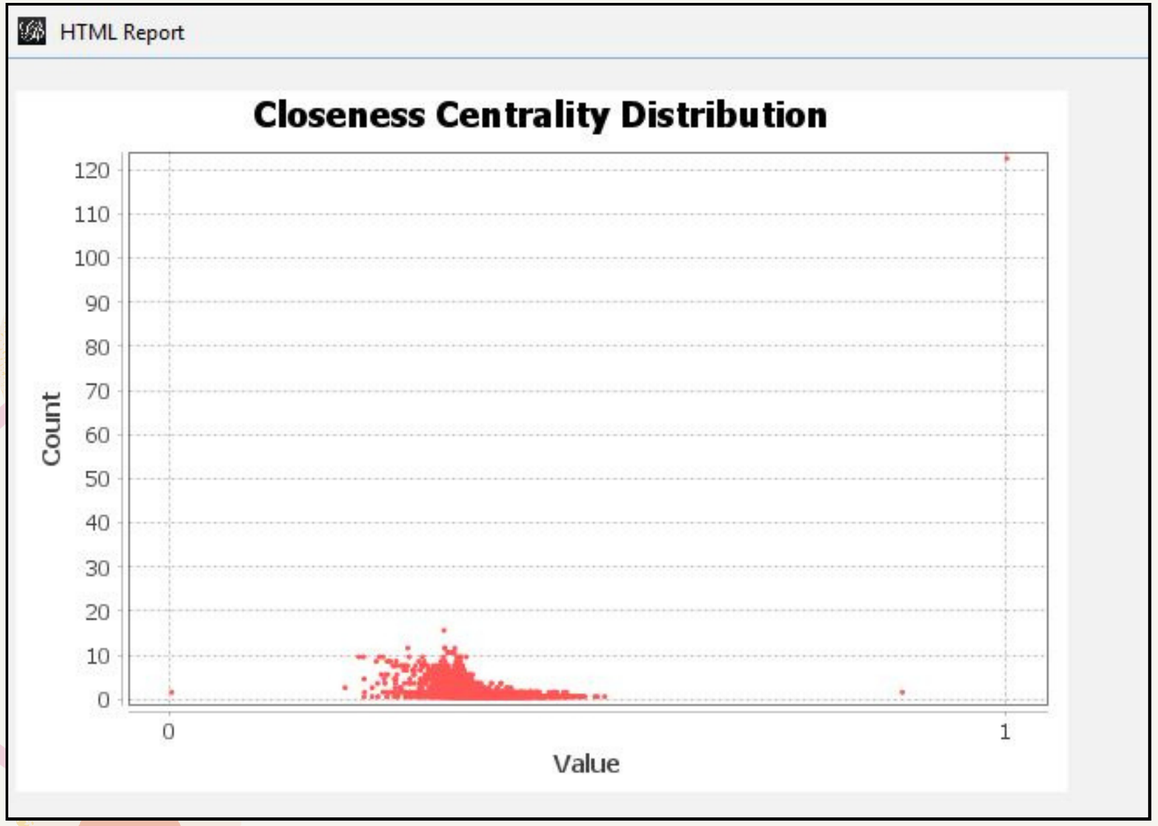
Most IMDB votes Films-

1) Taare Zameen Par
2) PK
3) Dangal
4) My Name Is Khan

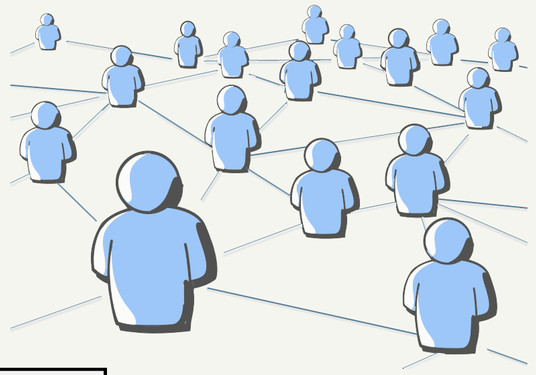


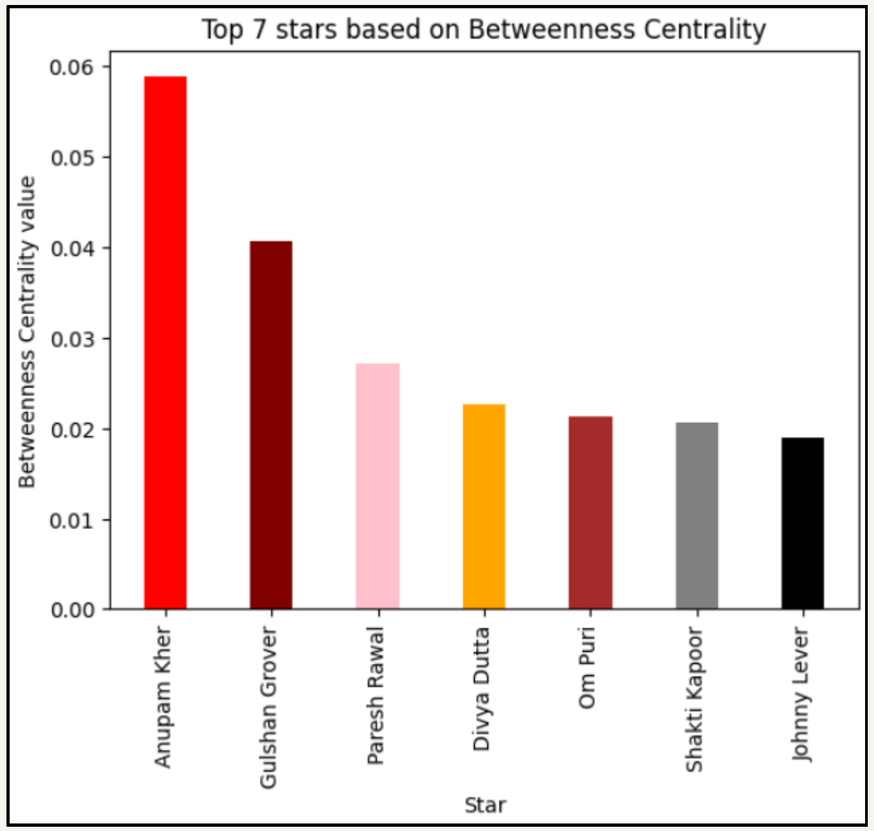
Closeness Centrality

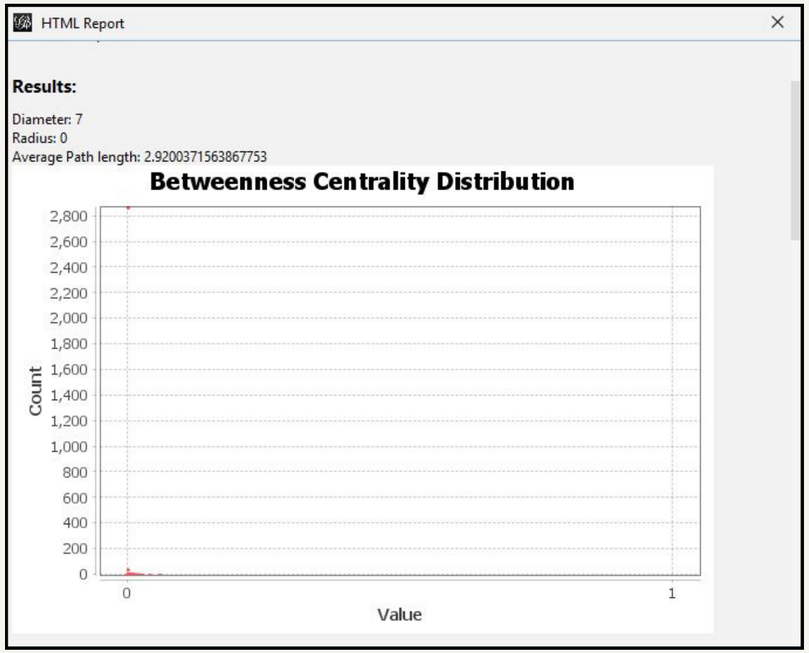




Betweeness Centrality







Personalized Pagerank Algorithm

Different Weightage to Actors of a film-

Weightage(ith actor of the film j) = (Imdb votes of the film j)*(alpha)^(i)
where alpha < 1, (importance attenuation factor)

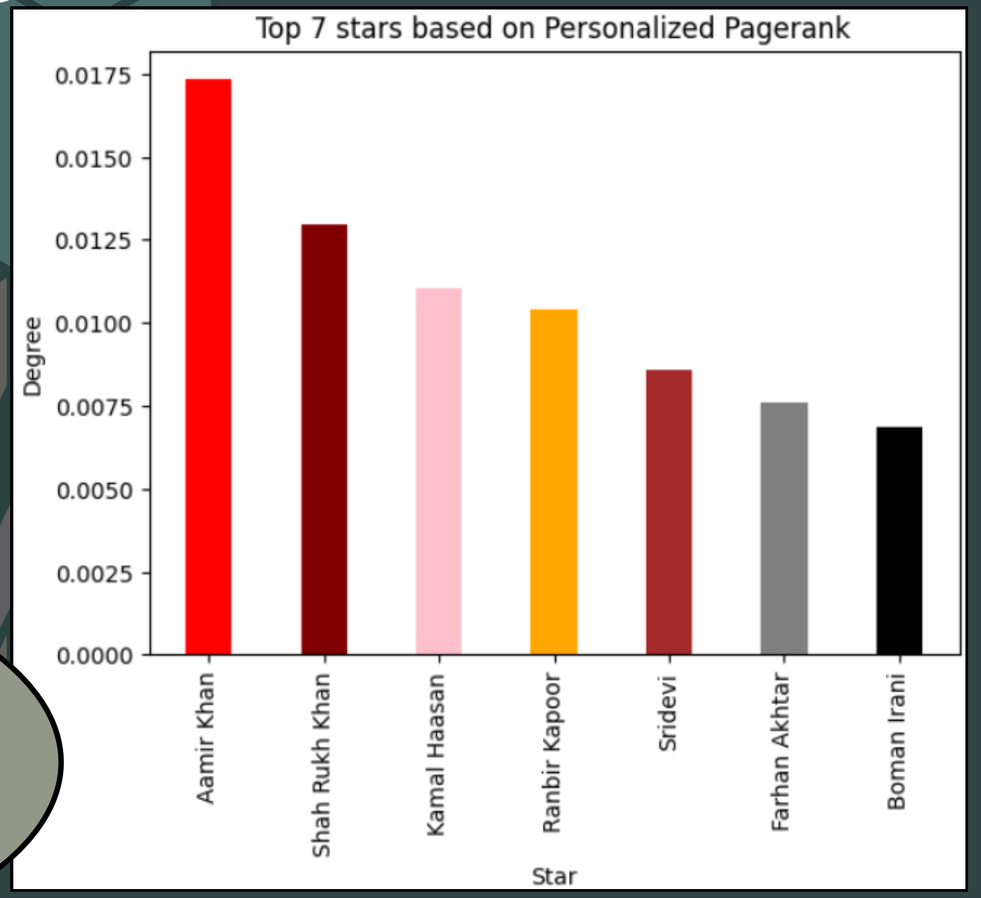
Example

Dhoom 3
['Aamir Khan', 'Abhishek Bachchan', 'Katrina Kaif', 'Uday Chopra', 'Jackie Shroff']

Personalized weights for pagerank -

Weightage(ith actor) = Average of weighted importance of all his films

Personalized Pagerank



This ordering looks more correct as **Aamir Khan** and **Shahrukh Khan** has the most collection films during 1990 to 2015.

We will use this prominence of each actor in our further analysis.

Analysis of Insider Debut Films

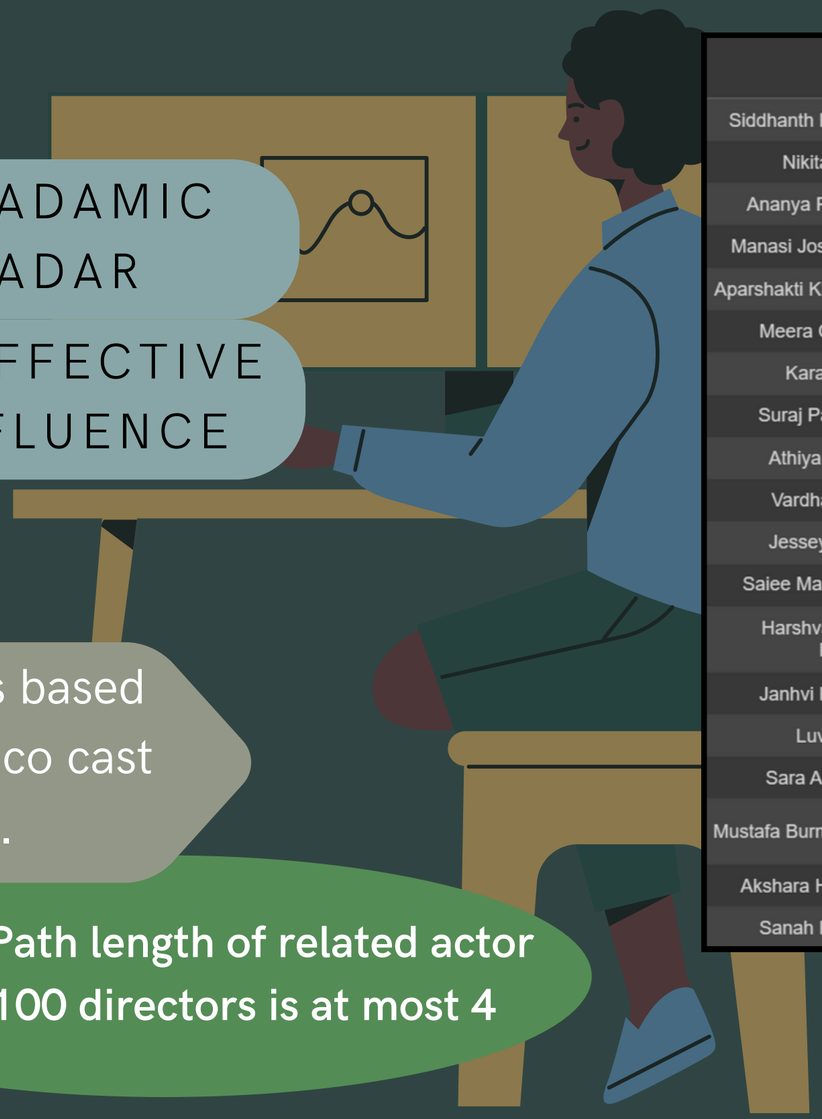
Techniques Used -

1) JACCARD COEFFICIENT

2) ADAMIC ADAR

3) PERSONALIZED PAGERANK

4) EFFECTIVE INFLUENCE

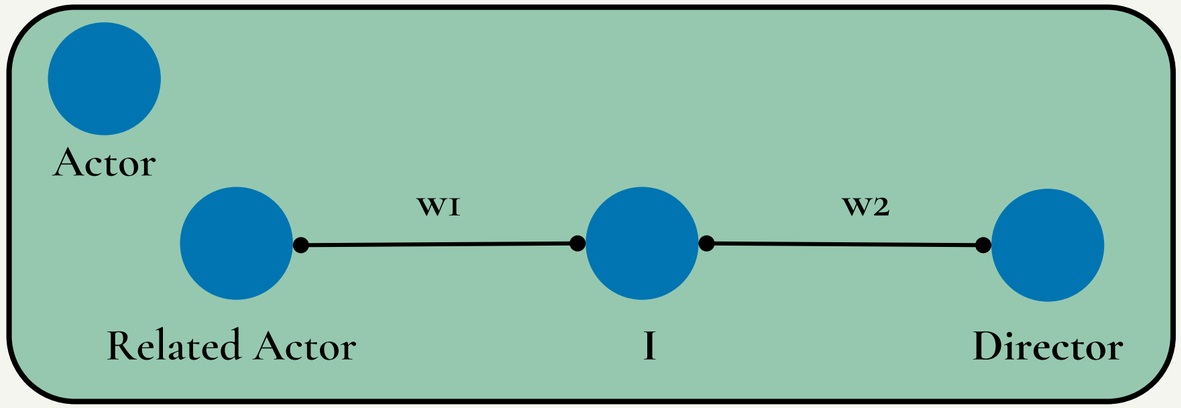


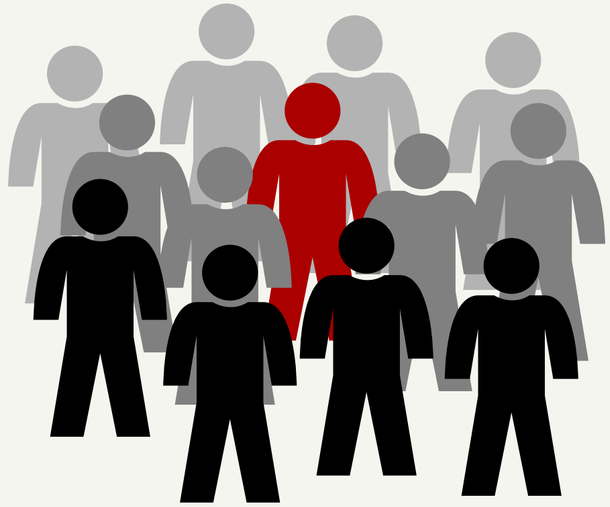
We will evaluate these methods based on the correlation with average co cast prominence of the actors.

Average Path length of related actor with top 100 directors is at most 4

Debut	Related to	Average path with top 100 directors
Siddhanth Kapoor	Shraddha Kapoor	3.347368
Nikita Dutta	Tanushree Dutta	3.221053
Ananya Panday	Karan Johar	3.168421
Manasi Joshi Roy	Rohit Roy	3.284211
Aparshakti Khurana	Ayushmann Khurrana	3.705263
Meera Chopra	Priyanka Chopra	2.926316
Karan Deol	Sunny Deol	3.031579
Suraj Pancholi	Aditya Pancholi	3.094737
Athiya Shetty	Sunil Shetty	3.000000
Vardhan Puri	Amrish Puri	3.010526
Jessey Lever	Johnny Lever	2.978947
Saiee Manjrekar	Mahesh Manjrekar	3.073684
Harshvardhan Kapoor	Anil Kapoor	2.936842
Janhvi Kapoor	Sridevi	3.210526
Luv Sinha	Sonakshi Sinha	3.189474
Sara Ali Khan	Saif Ali Khan	2.894737
Mustafa Burmawala	Abbas Alibhai Burmawalla	3.221053
Akshara Haasan	Kamal Haasan	3.421053
Sanah Kapoor	Shahid Kapoor	3.052632

Effective Influence



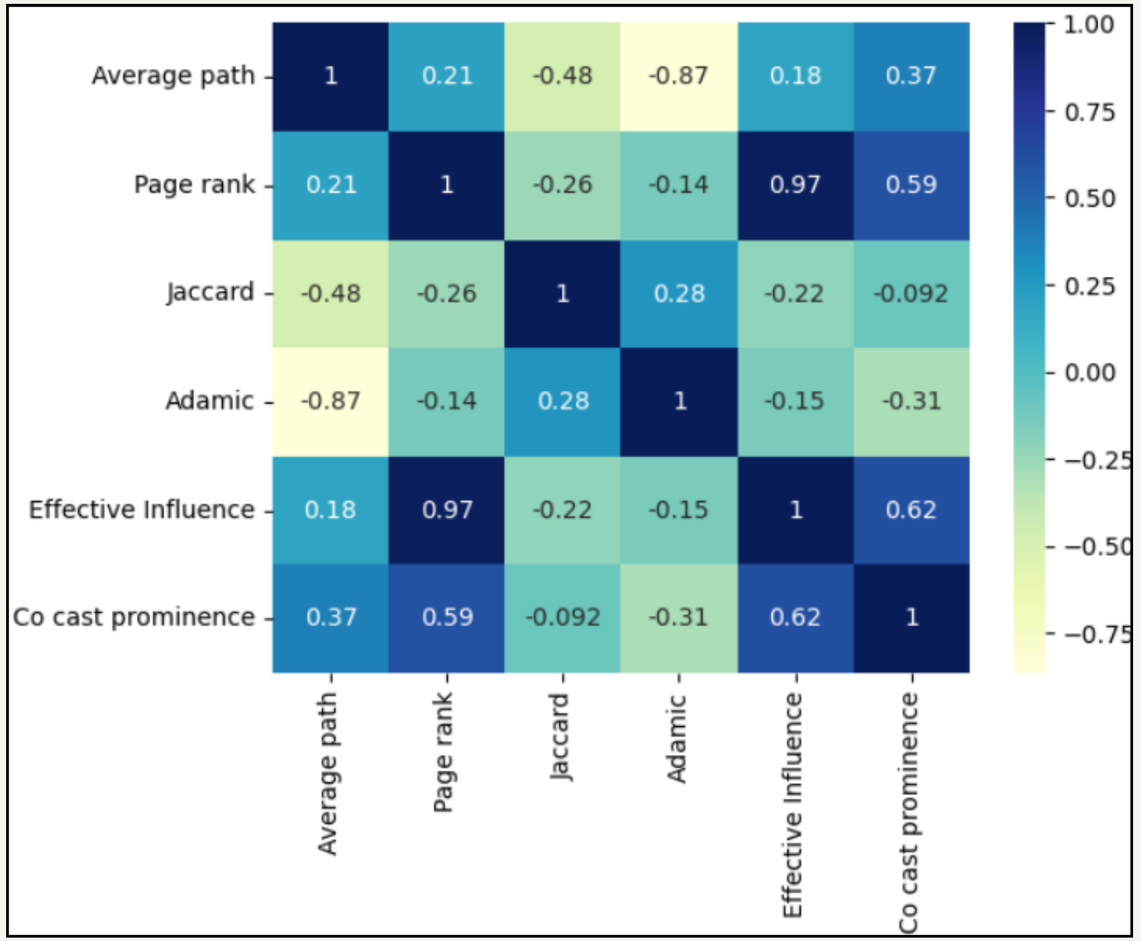


Effective Influence = PR(Actor) + PR(I)*w1*(alpha)^1 + PR(Director)*w2*(alpha)^2
where alpha < 1, attenuation factor, PR (PageRank value)

The effective influence of the node depends on the actors in between and past film successes between those weights, represented by the weights.

Correlation Matrix

Pagerank and Effective Influence shows the related actor 's role in getting film with great co-actor/director for the insider actor.



Insights

Avg insider actor number: 3.80952380952380 93
Avg outsider actor number: 5.664014146772767

Avg insider rating: 5.221475582102 76
Avg outsider rating: 6.019464320372654

Avg insider co-cast prominence : 0.0010549132386782907
Avg outsider co-cast prominence : 0.0010213362099271474

1) Average rating of insiders first movie is about 1.0 rating less than an outsiders which maybe the reason of overall mediocre work.

2) The actor number of insider is 2 less than an outsider's actor number.

3) The average prominence of co cast of an insiders is slightly more than outsiders.

4) The average path length of related actor with top 100 directors is at most 4 which is inherently present from the small world property.

Conclusion

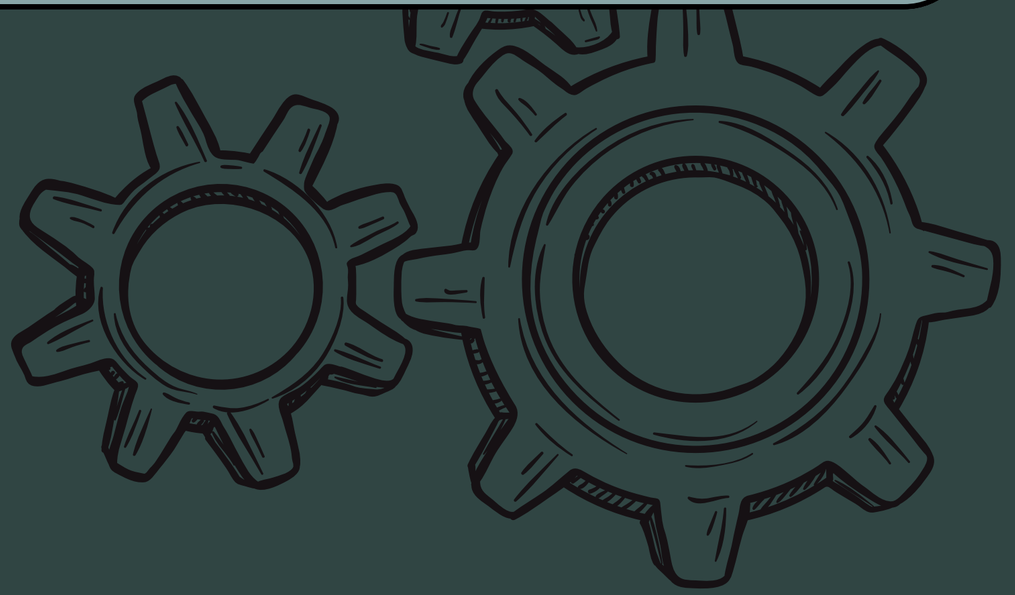
With the limited dataset, the slight advantage of having an established inside connection is evident from the results that we have got, which helps in getting good roles with already established cast.

We have also shown the quality of films in the debut film of an insider in comparison to an outsider. But it is not so prevalent in the current era as more outsiders are getting chances in the industry.

All in all it boils down to the talent of the individual to stay in the industry.

Future Work

1) Considering the type of relation between actors.
2) Analysis on a Bigger dataset with different time period.
3) More Centrality measures and embedding methods can also be used.



Thank You

“

In order to succeed, we must first believe that we can."

— Nikos Kazantzakis