

A network analysis of

**NETFLIX**

**Actors/Directors and Movies**

Simrat Benipal, Jose Perales, Ishrat Naba  
Group # 11  
Netflix II

# Research Questions

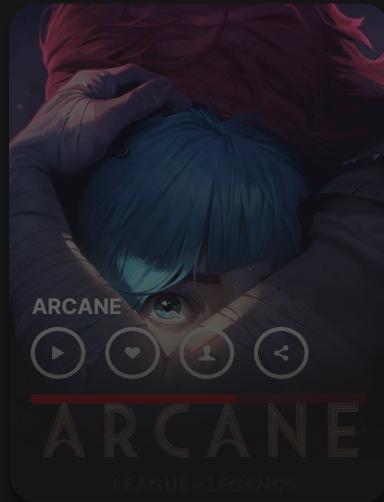
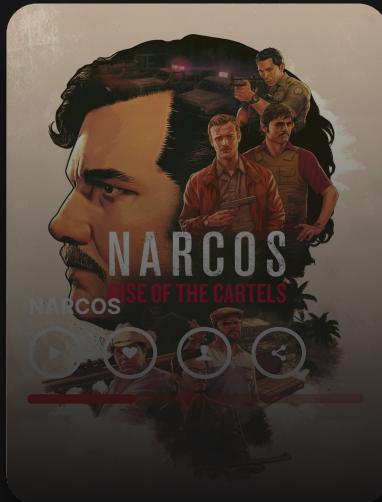
- Can distinct communities be identified within the Netflix movie network, and if so, what characteristics (e.g. country of origin, genre, release year) define these communities?
- Who are the key individuals within the Netflix movie network, and how do they facilitate connections between different communities?

We believe that by answering these questions, we will be able to assist large streaming platforms (more largely Netflix) create more cross-collaboration between actors across the network as well as help them offer more diverse content that is tailored to the tastes of audiences in different locations.

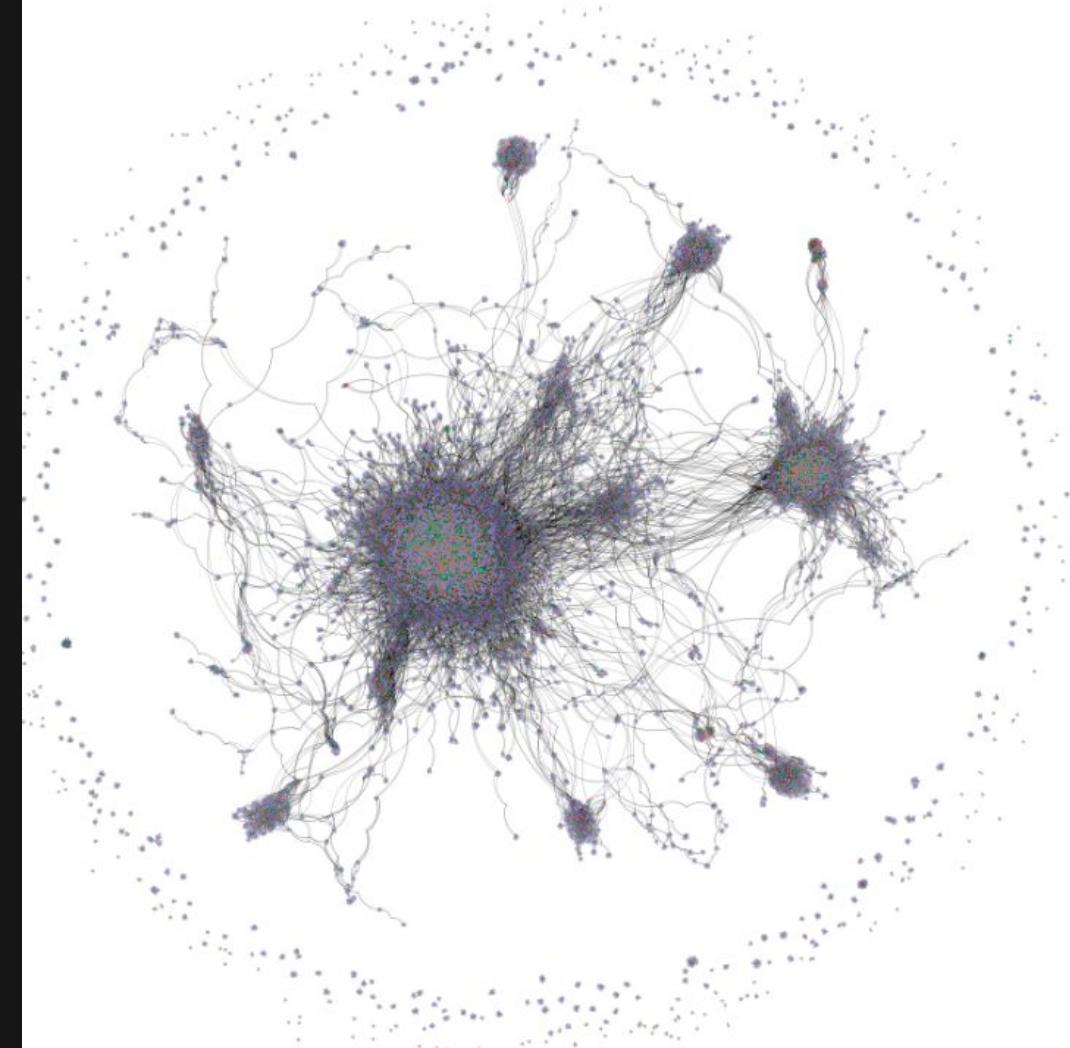
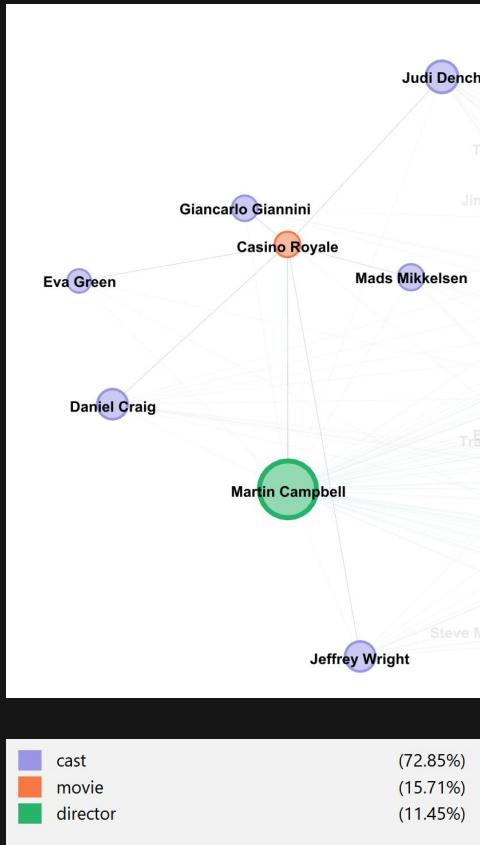


# Network

- **Nodes:** Movies and People (Cast, Directors) involved in the filmmaking.
- **Links/Edges** denote various interactions in the movie-making process.
  - Cast  $\leftrightarrow$  Movie, if cast members appeared in the Movie.
  - Directors  $\leftrightarrow$  Movies, if director directed the movie.
  - Directors  $\leftrightarrow$  Cast, if director and cast worked together.

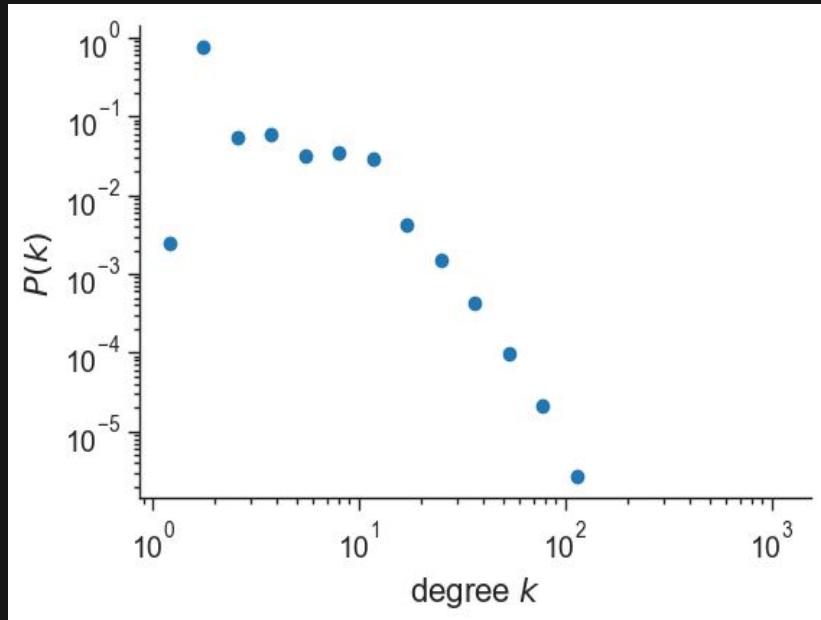


# Network Visualizations



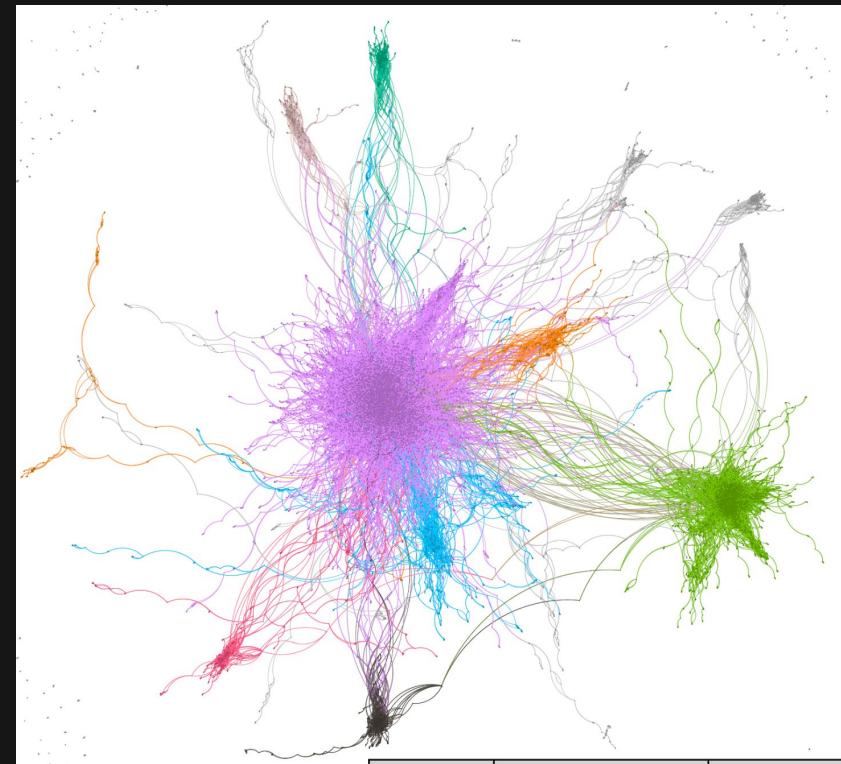
# Results

- Nodes = 35,133. Edges = 95,500.
- Connected components = 512
- Minimum Degree = 1, Maximum Degree = 133
- Average Clustering Coefficient = 0.6448
- Average Shortest path length = 8.65
- Average degree per type:
  - **Movies: Average degree of 8.28** (reflects collaborative nature due to 8 directors/actors per film)
  - **Directors: Average degree of 10.99** (high degree indicates central role in industry due to number of connections)
  - **Cast members: Average degree of 3.63** (lesser connected/well-known than directors due to lack of opportunity in the industry)



# Community Structure

- Analysis showed the existence of multiple communities.
- Community detection using Louvain Method (Avg Modularity among 100 tries = 0.8351)
- These communities were based on different film cinema industries.
  - E.g.: American Film Industry (Hollywood), Indian Film Industry (Bollywood), Nigerian Film Industry (Nollywood) etc.
- Analysis of key individuals (“Bridge Nodes”)
- Using Betweenness centrality measure
  - Om Puri, Anupam Kher



| Color       | Film Industry             | % of nodes |
|-------------|---------------------------|------------|
| Pink        | United States (Hollywood) | 51.46%     |
| Light Green | India (Bollywood)         | 14.50%     |
| Blue        | Latin America             | 5.19%      |
| Orange      | East Asia/Hong Kong       | 2.63%      |
| Black       | Nigeria (Nollywood)       | 2.63%      |
| Dark Green  | Egypt                     | 2.11%      |

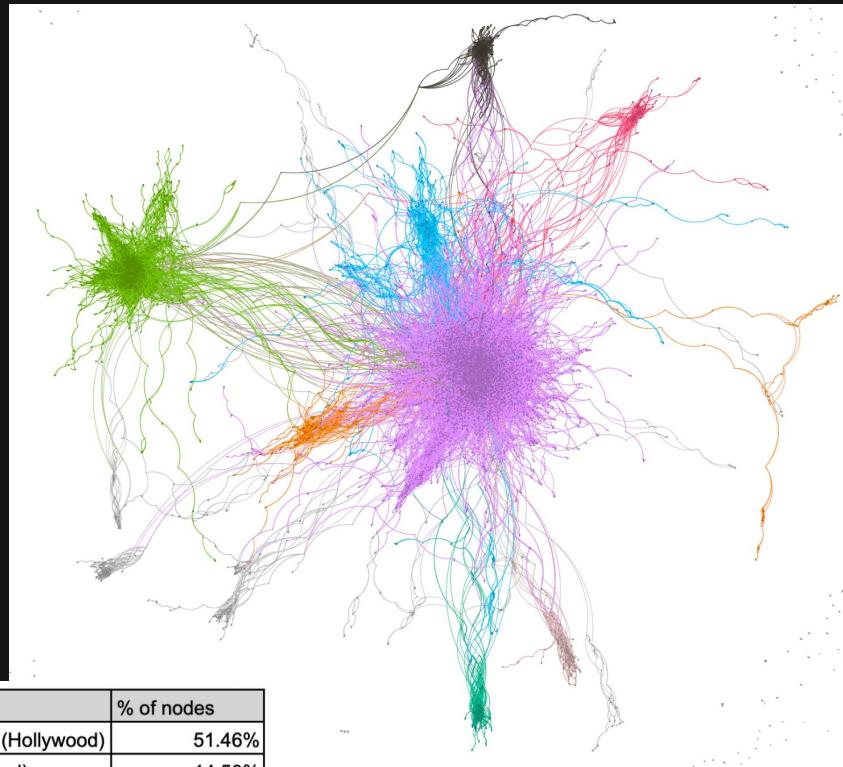
# Null Model Analysis

- Ten configuration null models, total run time equal 3+ hours.
- Lower clustering coefficient in null model signifies that there is less clustering in the network compared to the actual network
- Lower Average path length in null models signifies that the null model is more well-connected as it has fewer steps to reach certain nodes
  - Structural constraints (geographic distribution, language barriers) may contribute to this
  - Exclusive networks refer to limited access groups within the network, which can influence collaboration.
  - Preferential attachment suggests that nodes with higher connectivity attract more connections, leading to the formation of hubs/central nodes in the network.

| Network        | Clustering Coefficient | Avg. Path Length |
|----------------|------------------------|------------------|
| Netflix Movies | 0.6482                 | 8.65             |
| Null Model     | 0.00072                | 5.24             |

# Discussion

- Different community structures were identified using Louvain Method
- These different communities existed because of different cinema industries.
- “Bridge nodes” were identified using Betweenness Centrality measure.
  - Ex: Anupam Kher
  - Ex: Om Puri



| Color       | Film Industry             | % of nodes |
|-------------|---------------------------|------------|
| Pink        | United States (Hollywood) | 51.46%     |
| Light Green | India (Bollywood)         | 14.50%     |
| Blue        | Latin America             | 5.19%      |
| Orange      | East Asia/Hong Kong       | 2.63%      |
| Black       | Nigeria (Nollywood)       | 2.63%      |
| Dark Green  | Egypt                     | 2.11%      |

# Discussion cont.

- Based off our findings we can observe many things, some examples including: for future directors, they should cast the actors with high betweenness centrality as they have more community connections → more professional relationships, thus there's more of a guarantee in how well their films will do.
- We can also see that India is the second most connected in Netflix when it comes to actors/actresses, so some more collaboration between them and those in the US would do well with the audiences in both locations

| Person            | Role     | Degree | Centrality |
|-------------------|----------|--------|------------|
| Steven Spielberg  | Director | 126    | 0.0578     |
| Robert Rodriguez  | Director | 82     | 0.3551     |
| Lasse Hallström   | Director | 74     | 0.0315     |
| Don Micheal Paul  | Director | 88     | 0.0235     |
| Quentin Tarantino | Director | 80     | 0.0214     |

| Person          | Role     | Degree | Centrality |
|-----------------|----------|--------|------------|
| Anupam Kher     | Cast     | 71     | 0.0375     |
| Om Puri         | Cast     | 55     | 0.0338     |
| Priyanka Chopra | Cast     | 31     | 0.0287     |
| Amrish Puri     | Cast     | 32     | 0.0287     |
| Leena Yadav     | Director | 28     | 0.0148     |

# Future Work

- Extend the analysis to include TV shows and other Netflix content as well.
- Extend the analysis to other streaming platforms.
- Contrast and compare analysis of different streaming platforms.



# THANKS

# REFERENCES

Dataset: [https://www.kaggle.com/datasets/  
shivamb/netflix-shows/data](https://www.kaggle.com/datasets/shivamb/netflix-shows/data)

