STUDY OF BOLLYWOOD ACTORS NETWORK

BALAKAUSHAL DAMARAJU RAVI TANDON

1 INTRODUCTION

Movie Actors network is most difficult network to design, build and to analyse. This difficulty arises mainly due to the fact that in a film industry many actors come and retire. So no.of actors in network increase as time moves forward. So decade-wise analysis is important to see the trend of film industry. This trend can also be study the logistics in the industry.

2 DESIGN

The network contains actors as node and links between nodes exist if the nodes have co-acted in same film. The weight of the link depends on the no. of films they have acted together. We will create the network for 4 decades i.e. 70s, 80s, 90s, 2000s respectively. The number of nodes in each network are as follows:

- 1. 70s: 771
- 2.80s:880
- 3.90s:1555
- $4. \ 2000 {\rm s:} \ 1946$

The network was designed using the fact that the database we collected were structured such that the actors were listed in the film of their importance in the film, from main lead to supporting artists to junior artists. The following scores are assigned to a actors for each movie in following fashion:

- 15pts :Leading actors
- 12pts :Second lead
- 10/6pts : Main Supporting actors
- 3pts : All remaining actors.

The score of the actor is finally equals to the average score per movie. We can color the actor by applying threshold to his score. The colors are as follows

The colors are as follows

- 1. **Red** for star actors
- 2. Black for big actors
- 3. Green for supporting actors
- 4. Yellow for remaining actors

The network diagrams are given at end of this report.

3 ANALYSIS

We have to analyse the networks to find the different properties of network which may be helpful to us to understand some aspects of bollywood.We will study the following properties in the graphs. The coloring int graph is as follows.

1. Cluster coefficient:

The following figure will compare the cluster coefficients of the 4 decades the graph is cumulated cluster coefficient vs index number.

(a) 70s: Red



Figure 1: Cumulative Cluster Coefficient

- (b) 80s: Green
- (c) 90s: Blue
- (d) 00s: Pink

As we see that the graphs of cumulative cluster coefficient for 70s, 80s, 2000 overlap with each other. However Slope of 90s is less decreasing and also starts at higher value and reach 0 at the higher count of actors. This shows that at 90s, the cluster formation between actors were high, it means that there was no problem of actors from that decade to act with any other actor. So 90s era has less polarization compared to others.

2. Degree Distribution:

The following figure displays the log-log graph of degree distribution of nodes for 4 decades. The graph is logdegree vs logcount of the degree.

(a) 70s: Red



Figure 2: Degree Distribution

- (b) 80s: Green
- (c) 90s: Blue
- (d) 00s: Pink

We can infer that in 1970s and 1980s the degree distribution was same. But in 90s and 00s, the distribution is different. In 90s the count of lower degree was less than 00s but in higher degrees,90s overtake 00s in count. Thus 90s actors have acted/experimented with lot of new and different actors

3. Score:

It is the simple property. The graph will hep you to count the number of actors of particular score present in that era.

- (a) 00s: Red
- (b) 90s: Green
- (c) 80s: Blue



Figure 3: Score distribution

(d) 70s: Pink

The 2000 era has maximum no of star actors (i.e score greater than 10) compared to the other decades and 2000 era is not over so more good actors can be awaited. However above average good actors are present more in 90s which was era of budding actors. As expected, small actors are present in 90s-2000s era because no. of films has increased as no. of actors.However all actors don't become stars. and increase in low grade movies also gave rise to small actors.

4. Assortivity:

First, we have found the associate score of a node which is average of all the scores of the neighbours where the neighbours scores are weighted according to link weights. We can call this score as associate score for an actor per movie.

- (a) 00s: Red
- (b) 90s: Green



Figure 4: Assortivity

- (c) 80s: Blue
- (d) 70s: Pink

Now looking at the graph we find that the slope of curve of 2000s is maximum and the slope is minimum in 90s. This implies that the network in 2000s was most assortive in nature while in 90s it was its assortivity was least and remaining 2 networks were also sufficiently assortive compared to the 90s network. The high assortivity is due to increase in multistarrer films and big bufget family drams.

5. Multi-Starrer films:

Here we have counted the films for each value of i where i is the no of star actors present in the movies

- (a) 70s: Red
- (b) 80s: Green
- (c) 90s: Blue



Figure 5: 7Multi-starrers

(d) 00s: Pink

This property will help us to inference the actors relationship. 70s and 80s have highest no of movies with at least 1 big actor. This Big actor may be male or female. 90s and 2000s have highest no. of 2 film actors which we may consider as high rated hero heroine pairing. This means that since the score of heroines have increased the female oriented movies also started in this era. Otherwise we can say that 2 hero movies with low rated heroines were also seeing the trend in this era. 70s also have seen 2 actor movies which is just below the curve of 90s. The percentage of 3 big star movies is less in all decades considering the budget considerations of producer. The new inference we derived from graph is that 0 - star movies is less in 00s which are high in remaining decades. Thus giving the possibility of high amount parallel cinema or low grade cinema releasing in that decades.

6. Survivability:

It is a vague property but we have found the use of this property in calculating



Figure 6: Career Chart for Actors common in 4 decades

ups and downs in the actor's career whose span consists of four decades. we have considered the actors who have acted in all 4 decades. We have see the trend of their career by reading the corresponding data files. There are some actors who are at their peak in all decades, like Amitabh Bachchan. Some like Dharmendra who rule for 3 decades slided to secondary roles.

We can see the general trend in this graph for actors common in all decades.

4 Important considerations in design and analysis

We have taken the data for the network from movie database site. So reliability of our analysis depends on reliability of data provided by that site.

There is a possibility of noises in network due to possible errors like repetition of actors due to wrong spelling. The films of one particular decade making its presence in other decades. We have tried to remove most of the errors, however there may be some unknown errors like spelling mistakes which can be removed only if we know the spelling of all the possible actors in the bollywood (The impossible task).

5 Conclusion

We have seen the properties all the 4 networks of bollywood actors. We conclude by inferring that bollywood has changed its trend in decade 2000 with entry of new actors. The 2000 decade has not yet over but has maximum actors compared to other decades though 90s have produced more movies than 2000. We have see the trend of forming larger cluster or groups have in 90s but this trend disappeared in 2000 with 2000 trend being the same as 70s and 80s where there was concept on big actor and small actor. We say that bollywood movies have changed gradually.



Figure 7: 70s network

Paiek



Figure 8: 80s network



Figure 9: 90s network



Figure 10: 2000s network

References

- [1] Database:http://www.bollywood.de/
- [2] Ramasco, Jose J.; Dorogovtsev, S. N.; Pastor-Satorras, Romualdo -Self-organization of collaboration networks
- [3] M. E. J. Newman -Scientific collaboration networks- Network construction and fundamental results