**Social World Connectivity among the Indian Celebrities**

Harshit Bhatt

[*hbhatt2014@my.fit.edu*](mailto:hbhatt2014@my.fit.edu)

*PhD. Computer Science*

*Florida Institute of Technology*

Research Paper on Twitter Data Mapping Project

Complex Networks (CSE 5656)

Fall 2014

**ABSTRACT**

In this work I have developed a social world (twitter) relationship among the Indian celebrities based on their mention in tweets. Using *Twitter API* I have collected tweets of majority of *verified* Indian celebrities. The data collected through *Twitter API* has been passed through various algorithms to deduce some meaningful data.

Nodes in the network are various celebrities whereas the link (edges) between them is the common users. The link strength is measured by the number of common entities between nodes whereas the Node size is the number of unique user mentions for each celebrity.

Another network which forms a relationship among the celebrities on the basis of their mention in same tweets is made which helps to gauge the cohesiveness among the celebrities.

1. **INTRODUCTION**

The main purpose of devising this network was to get a gauge of the celebrities’ fame and their internetwork through the Social Media Data. The data here is the tweets, and only those tweets are included which mention or are related to, any of these Celebrities. Only those celebrities are included who are verified from Twitter.

Celebrities in the list are from the field of Indian Cinema, Cricket and Politics. Due to the popularity of the celebrities among their fan base I was able to collect a large amount of data with the help of Twitter Streaming API.

The JSON data which includes a wide range of information about a tweet helped in defining the base table for the network, through which I was able to deduce various other meaningful sub-tables and views which modeled the various graphs in my result.

Tables which mapped the various graphs were on:

1. List of number of unique users for each celebrity.
2. List of common users between each pair of celebrities.
3. List of total number of each celebrity mentioned in tweets.
4. List of each pair of celebrities which are mentioned in the same tweet and their count.
5. List of every user and their mention counts to the celebrity (this list is for each celebrity).

The first two tables make a network which proposes the common interest among users between celebrities, in this network:

Nodes -> Celebrities

Node Size -> Celebrities unique users count

Edges -> Common Users between celebrities

Edge Strength -> Common Users count

The third and fourth table define a network which proposes how commonly two celebrities are mentioned in the same tweets, in this network:

Nodes -> Celebrities

Node Size -> Celebrities mention count

Edges -> Celebrities common in a tweet

Edge Strength -> Celebrities common in a tweet count

The last table could just be used to look upon which users tweet more about a celebrity or which particular user is common in the network if combined with the above two networks.

1. **METHODOLOGY AND SIMULATION**
2. The first step was to set up Twitter Streaming API which was processed by getting my special token and secret key from twitter for this application.
3. The python library Tweepy was used as a bridge to collect tweets from the Twitter Inc.
4. The whole program is set up in the Python language, which helped to collect tweets and its other relevant information, which was further saved into MySQL database tables so as to map the networks.
5. The main table for the network is named *Celebrity\_Tweets* which includes the first data that is collected directly from the tweets.
6. The fields are ->
7. user\_twitter\_id -> it has the twitter\_id of the person who has send that tweet.
8. user\_name -> user twitter handle.
9. tweet -> the actual tweet which has been sent by the user.
10. users\_mention\_id -> twitter\_id of the celebrities that are mentioned in the tweet.
11. hashtags -> hashtags that are in the tweet and are related to those celebrities.
12. To get all (most of) the celebrities in India I first choose three most famous Indian celebrities (based on general prospective) which by their twitter handle are @SrBachchan, @iamsrk, @msdhoni.
13. With the help of *Twitter API* I was able to collect all the *Verified* friends and friends of friends of these celebrities.
14. Each celebrity friends details are saved by their respective table name having fields as ->
15. user\_id -> celebrity twiiter id
16. user\_name -> celebrity twitter handle
17. The same was done for the first three celebrities for their friends list and their own details.
18. Now a list was made to remove the duplicates from these tables and get a unique list of celebrities which was saved in the table named *celebrities*.
19. There were total of 467 unique celebrities and verified groups which I finally got as the input for data collection.
20. The program is now executed with the input of these unique celebrities.
21. The data collected is saved in the *celebrity\_tweets* table as discussed above.
22. The data was collected within a range of 23 days and total tweets collected were around 352,000.
23. After the data collection some relevant tables and views were deduced from the main table, which were for the specific networks which we have mapped to show the final results.
24. Two views namely :
25. celebrity\_mentioned\_count -> this view stores the total mentions which a celebrity got during the time period. It has fields as :

celebrity\_twitter\_id -> celebrity twitter id

celebrity\_twitter \_name -> celebrity twitter handle

count -> total user mentions

1. celebrity\_distinct\_hashtags\_count -> this view gives the total number of different hashtags that are connected to the celebrity. It has fields as :

celebrity\_twitter\_id -> celebrity twitter id

celebrity\_twitter \_name -> celebrity twitter handle

distinct\_celebrity\_hashtags -> total distinct hashtags connected with the celebrity

1. Four different tables namely :
2. users\_common -> this table gives the number of users that are common between each pair of celebrities. It has fields as :

celebrity1\_name -> one of the celebrities twitter handle in the pair

celebrity1\_id -> first celebrity twitter id

celebrity2\_name -> the other celebrity twitter handle in the pair

celebrity2\_id -> second celebrity twitter id

common\_users -> number of users that are common between them

1. hashtags\_common -> this table gives the number of hashtags that are common between each pair of celebrities. It has fields as :

celebrity1\_name -> one of the celebrities twitter handle in the pair

celebrity1\_id -> first celebrity twitter id

celebrity2\_name -> the other celebrity twitter handle in the pair

celebrity2\_id -> second celebrity twitter id

common\_hashtags -> number of hashtags that are common between them

1. celebrity\_common\_mention\_tweets -> this table gives the number of tweets in which the celebrities are commonly mentioned. It has fields as :

celebrity1\_name -> one of the celebrities twitter handle in the pair

celebrity1\_id -> first celebrity twitter id

celebrity2\_name -> the other celebrity twitter handle in the pair

celebrity2\_id -> second celebrity twitter id

common\_mention\_tweets -> number of tweets which have the celebrities common between them

1. number\_unique\_users -> this table gives the total unique users that a celebrity got mentioned by. It has fields as :

celebrity\_name -> celebrity twitter handle

celebrity\_id -> celebrity twitter id

count -> number of unique users for the celebrity

1. I have also saved the users details who have mentioned the celebrity in another database with tables name respective of the celebrity name containing the fields as :

user\_id -> user twitter id who has mentioned that celebrity

user\_name -> users twitter handle

count -> number of times that celebrity has been mentioned by the user

1. Two networks were constructed with the help of these tables. Gephi Software was used to map these networks. The networks were :
2. The first network uses users\_common and number\_unique\_users tables to generate a relationship which proposes - the common interest among users between each pair of celebrities, in this network:

Nodes -> Celebrities

Node Size -> Celebrities unique users count

Edges -> Common Users between celebrities

Edge Strength -> Common Users count

**For example =>**

Common Users = 159

**Unique Users = 6353 Unique Users = 852**

Common Users = 278

**Unique Users = 6353 Unique Users = 797**

1. The second network uses celebrity\_mentioned\_count and celebrity\_common\_mention\_tweets tables to generate a relationship which proposes how commonly two celebrities are mentioned in the same tweets, in this network:

Nodes -> Celebrities

Node Size -> Celebrities mention count

Edges -> Celebrities common in a tweet

Edge Strength -> Celebrities common in a tweet count

**For example =>**

Common Mentions = 30

**Total Mentions = 23443 Total Mentions = 1693**

Common Mentions = 148

**Total Mentions = 23443 Total Mentions = 1087**

1. **RESULTS**
2. **CONCLUSION**

After analyzing the two networks we get to know about, that how much does people mention about the celebrities and who are the celebrities which they consider most related to each other and to themselves.

We also get to know the users who tweet most about celebrities and whether or not they are loyal to one celebrity or many.

My study shows that as in accordance to the general perspective the social media also shows the same relationship among the celebrities, making the celebrities from the same community (politics, cinema and cricket) to be related more to each other than with other community and the users being more common between those community celebrities. Surely there are some exceptions but they are overshadowed by the strong between the celebrities.

Some celebrities or organizations have a very large number of mentions than others which may be due to those celebrities (politicians, cricketers and movie stars) being trending in India for some cause (movie release, an event being taking place or due to their certain initiatives) examples of which could be given with @iamsrk, @deepikaspeak movie stars mentioned a lot due to their movie released within this time period or @IndSupLeague an organization which is conducting a football event within India during this time or @narendramodi India’s Prime Minister for starting a cleanliness drive campaign within the country.

The common mention between some celebrities also has some great numbers like for @iamsrk, @deepikaspeak, @juniorbachchan, @farahkhan are mentioned a great number of times with each other due to a movie release which had all of them working together. The same was for the teams like @FCGoa, @atlidkolkata, @PuneFC and others due to their involvement in a league.

This same relationship could also be attributed with the other two networks which shows the relationship between the celebrities from the prospective of hashtags.

Finally I could say that if popularity could be gauged by the mention of celebrities it would hold true to what general perspective is.