ISI KOLKATA

MASTERS THESIS

Effects of Social Media in Indian Election

Author:
Dipayan DAS

Supervisor:
Dr Kuntal Ghosh

A Dissertation submitted in partial fulfillment of the requirements for the degree of M.Tech CS

in the

Department:MIU and CSCR

Declaration of Authorship

I, Dipayan DAS, declare that this thesis titled, "Effects of Social Media in Indian Election" and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:

____

Date:

____

ISI KOLKATA
Department of Computer Science

CERTIFICATE

This is certify that the Dissertation entitled **"Effects of Social Media in Indian Election",** submitted by **Dipayan Das** is a record of bonafide work carried out by him, in the partial fulfilment of the requirement for the award of Degree of Masters of Technology (Computer Science) at ISI Kolkata This work is done during year 2019-2020, under our guidance.

Dr.Kuntal Ghosh

Associate Professor

ISI-KOLKATA

Abstract

Social media has a great influence on our social life and extant literature finds that Twitter trends can capture the electoral sentiment. Twitter is a micro-blogging app, is the latest gift of globalization for social networking. We have gathered around 0.3 million tweets from 11 February 2019 to 31 March 2019 from the twitter accounts of about 200 political leaders belonging to the ruling as well as opponent parties. The research mainly aims to focus on social media data, data on political landscape, political inclination of different political groups, data mining of politically relevant tweets, data extraction with the help of Java and Twitter API. Several Graphs over different time periods have been generated based on a sense of their common contexts in the tweets during such periods. These contexts, in turn, identify the important events occurring during the same period. We have made two significant observations based on the above analysis. First, we found that the degree distributions of the generated graphs, each of whose nodes represent the political members only, during different periods corresponding to the events, follow power-law distribution. This result closely resembles real world complex network characteristics. Second, based on the data, it has been observed that with respect to such influential incidents like Pulwama explosion and killing or the Balakot airstrike when the ruling party was using, in their tweets, certain type of words related to the sensational events, the opposition parties' tweet accounts show that such words were mostly absent and other words were more prevalent, a fact which may have crucial political significance in the context of the Indian General Election in determining which type of views on these issues may have played determining role in the final voting.

Acknowledgements

I would first like to thank my thesis advisor, Dr Kuntal Ghosh, for the door to his office was always open when I needed him and had doubts and whenever I was down he has pulled me up from the depth of the pit, he consistently allowed this thesis to be my work but steered me from a distance whenever I needed it. I would also like to thank with my fullest heart and gratitude Swarup Chattopadhyay from MIU for his sincere, caring and enjoyable guidance. Without his participation, it would be incomplete. Also, there was Anjan Chowdhury from CSCR for his constant backup which kept me going. Finally, I must express my deepest gratitude towards my family my beloved one and the Almighty for keeping their constant blessings upon me. …

Contents

List of Figures

List of Tables

0.1 Motivation

The present era of globalisation has bought in the advent and use of, as well as dependence on, the fields of communication, publicity, blogging, advertisement and electronic interaction via social media sites like Twitter, Facebook, Instagram, YouTube, TikTok etc. These platforms encourage and facilitate the user groups of varying ages to create personalized contents. These personalized contents by different users are again accessible by different other user groups following or using the same related apps. As a consequence, socially sharing certain content which may be in the form of videos, short notes, memes, voice texts, pictographs, can identify situations of mass sentiment on a macro scale which can be prominent indicators of events on an even bigger scale. This magnanimous volume of data which is also a prominent indicator of various socio-economic strata is currently the prime focus area of researchers and experimenters. In this context, Political Tweets which are user content data from various political leaders and parties on a social platform like Twitter, for the prediction of the situation of the ruling party with respect to the masses, is ambivalent without much research and study. The research, whatever has been done, however does not indicate an easy solution to the problem. From literature studies, it has been noted that mere amount of political tweets could on one hand meticulously predict election results in Germany [12] whereas following similar collective trend in analyzing social sentiment from twitter, failed to envisage the Presidential elections of United States in 2008 [11]. Our study in the present thesis is limited to the analysis of this user content in the context of India for various political personalities and whether these may identify the effect of simple yet significant discrete events in political field during a politically important period in study. We have noted that in spite of ambiguities, the prediction of election results from the usage of social media content is gaining importance amongst researchers. To publicize and create a stronghold over a user group, politicians make strategic moves and decisions to manage and post on their social media sites to influence and strengthen their campaigning phase. This is the point of our present research and as per data is concerned, it is observed that Twitter has approximately thirty-three to thirty-four million users registered in India which cannot be termed negligible when compared among developing countries in the world.

0.2 Review of Literature Study and Work Plan

Extant literature finds that twitter trends can capture electoral sentiment. One such work in the context of India [9] has demonstrated that for a particular general election study in the Indian context of 2014, twitter data may be used to understand the analyzing power of social media in this largely political diversified country. The paper analyses approximately 0.4 million tweets of three months user content data which after the analysis was able to envision results of electoral sentiment along

with envisaging variations in vote sharing by different political groups from the sentiment scales. This study has designed a template for data collection and cleaning which has mined relevant twitter data in the process, and shown that modulation of context learning is essential in the process to discard junk data and to conduct relevant domain data mining. Incorporation of a potential control variable Nationality Dummy (Nat_Dum) has been used to keep in command regional dynamics in such a political diverse country like India. Moreover, empirical results provide a strong background to the fact that subjects belonging from a particular political group share similar sentiment score and are loyal to the party in social media sites. This highlights the prominence in indicating vote swing in the political landscape of India, and shows that a structured analysis of electoral sentiment will fine-tune the prediction of vote change and vote sharing. Certain hypothetical examples have been studied which confirms a certain degree of misleading in envisioning election results. Phase wise data collection models have been studied to perform better with precision in politically diverse large counties like India, where election procedures and tenures exceed or near to almost a month. Our study on 2019 Indian General Election is to some extent influenced by this study of Khatua et al. [9] in 2014, though we have chosen a different route and a different research problem too. This I explain subsequently.
Our work plan was aided by a study available on net that talked about storing and data management of Twitter employing an optimized Neo4j database graph. This was used to solve communication on political grounds for 50 U.S states for the 2016 Presidential election [1]. The use of the database graph enables ease of complex modeling and visualizing communications among related subjects of interest and also change in user requirements was observed to be flexible to generate the database along with a user-friendly query interface. The study also developed a Python-based application using Twitter API to collect tweet from Twitter users. These methods were implied to differentiate the impact factor of the query optimization technique. All data were collected for eleven months using the same political database of queries in terms of response times. The study confirms and validates the usage of best practice design guidelines and that the designed database improved in extracting query response times.

Among the different type of literature survey that we conducted, one involved those research which show that from political tweets it is not easy to find the Political Orientation, and some researchers showed that methods which previously reported greater than 90% inference accuracy, actually achieve barely 65% accuracy on the normal Twitter users [6].

Again certain literature studies analyzed sentiment score based on twitter data by introducing POS specific characteristics. Tree kernel application of machine learning has been used in order to reduce complexity. For calculating the similarity between two trees a Partial Tree (PT) ker- nel was first proposed by Moschitti [10]. A PT ker- nel calculates the similarity between two trees by comparing all possible sub-trees. This tree kernel is an instance of a general class of convolution[2] kernels. The background of the study lines back to atomized detection and classifying tweets of sentiment linked with responses of users on different topics, including current issues, products. This could enable organizations or customer care departments to easily revert back to customers or other clients on micro blogs. Classification of sentiment data has been done using a binary task and a 3- way task is incorporated to further group the sentiments into positive, negative and neutral classes. Different models such as tree kernel and unigram have been employed both singly and in combination for representation and analysis of tweets. it is observed that the combinations outperform the unigram models. It is drawn from the analysis that polarity of tweets with specific speech parts is the most crucial for the classification of tweets. On the contrary, special characters like emojis and hash tags, contribute marginally to the classification of tweets. In our case, the political tweets also require such polarity analysis for addressing the research question here. Hand annotated emoticon dictionary, for example, been introduced which helps to map polarity of emoticons; along with an acronym dictionary with five thousand frequently used English translations. To simplify, acronyms used has been expanded in English like BRB- Be right back or BTS- behind the scenes whereas emoticons have been grouped with respect to different degrees of polarity like positive, negative, neutral, etc. Statically the data used in the experiments have been hierarchically classified. The different models which have been used in the study have been subjected to comparative appraisal and feature analysis has been done which depicts the average accuracy of different models in terms of percentage and standard deviation. on a whole, the study states that sentimental score analysis in Twitter is more or less similar to sentiment analysis for other genres. Another optimization study to address the complexity and improve the user interface in a real political platform of Twitter has been referred where different models and graph database has been developed for understanding the effect of query optimization. It is noted that query response time on an average is inversely proportional to an increased relationship in a graph database. the main contribution of the study introduces guideline design modulation of a graph database like Neo4j weekly database graph which eases the availability of fresh tweets to the participators of the political domain [1].

From an extensive survey it is inferred that users of twitter on a political landscape, help in strengthening political campaigns. It is also noted that following these politically relevant tweets acts as an important tool for political journalism and press interests. Less number of hops in between queries and attribute indexing are the prime indicators to achieve low query response time. The study elaborates previous work on similar databases, discusses proposed graph data models, experiments to design an efficient graph data model which focuses on query optimization and indexing mechanism. Property of edges has been modeled by the relationship quality of tweets. If an existence of a third entity is prominent in a two-way relationship, an intermediate node is formed to show the coherence and link all the entity nodes. The

chronology of nodes includes children nodes and root nodes forming a tree module where root nodes constitute children nodes as a representation of individual years which in turn has month node and again day nodes. In this branched tree module all the children nodes are interlinked to show coherence between the relationship pattern of twitter data. The validation of social network data concerning time-varying has been introduced by a graph data model. Cattuto et al. in their study collected data from social network using participants wearing badges with active Radio Frequency Identification devices during intervals of 20 seconds[4]. A combination of different social network topology with rich queries is got from the model. The study also showed a similar tree graph model to validate queries concerning time-based ranges and was tested in Neo4J graph to check if it performed well. Goonetilleke et al. stored micro-blogging queries in the most widely used graph databases: Neo4j and Sparksee [8]. In this study, different nodes with response characters like symbols, emoticons, and tweet relationships are studied. Neo4j graph is used for the nodes and edges and stated neighborhood of a node from the classification of queries and ordering them by counts.
For analysing our results we made a survey of the possible statistical methodologies used like community detection using Louvain [3] algorithm and used it in our study. Another important structural characteristics in the study of large-scale real-world complex networks is the degree distribution of the network. It has been empirically revealed that for real-world networks degree distributions follow, in general, power law though there are subtle issues associated with the same[5]. We have tested the applicability of this structural property in the graphs that have been generated from our data on tweets from Indian General Election 2019.
Finally, one particular tutorial inspired this work not only methodologically, but also instilling the idea that instead of crawling over the tweets of the millions of twitter users in India as done by Khatua et al. [9] for 2014 Indian election, it can be a good idea to concentrate for 2019 on important political personalities only in creating the graphs. This tutorial [7] dealt with the generation of twitter graphs of popular NBA (US professional basketball league) players to understand things like: which player in the NBA is the most central or which team has the highest density (of edges) etc. We borrowed this idea to some extent while incorporating the political contexts especially some crucial incidents occurring prior to the elections in our work.

0.3 Indian General Election 2019

India is a vast politically diverse country with a wide group of political parties. In this country, General elections are held in every five years for 543 Indian constituencies. The general period for elections is conducted by an independent body, the Election Commission of India. As compared to other nations, the election procedure in India is longer which reaches about a month to complete for all states. The longest recorded general election in India took place in nine phases from April 11, 2019, to

Leading Party List		
Name	Symbol	Party Supremo
All India Trinamool Congress	AITC	Mamata Banerjee
Bharatiya Janata Party	BJP	Amit Shah (National President) J P Nadda (Working BJP President)
Bahujan Samaj Party	BSP	Mayawati
Communist Party of India	CPI	Suravaram Sudhakar Reddy
Communist Party of India (Marxist)	CPIM	Sitaram Yechury
Indian National Congress	INC	Rahul Gandhi
Nationalist Congress Party	NCP	Sharad Pawar
National People's Party	NPP	Conrad Sangma

TABLE 1: Leading political Parties in 2019 Indian Election

May 19, 2019. The two major alliances in the 2019 general election were of Bharatiya Janata Party (BJP) which is the national democratic alliance and the other was Indian National Congress (INC) which is the United Progressive alliance.

In addition to these major parties, many smaller registered unrecognized parties also participated in this election. 2293 political parties contested in 2019 parliamentary election. Interestingly 300-odd parties. In fact, 149 political parties were registered with the poll panel between February and March on the eve of the announcement of the poll schedule. However, leading parties like BJP contested from 437 seats. Another interesting aspect of the Indian general election was the presence of independent candidates.

Several important and significant incidents of different socio-political dimensions occurred during the period preceeding the general elections. One such incident was the Pulwama masaccre and another that followed it after about two weeks was the Balakot airstrike Incident. There were some other events too which also were significant occurrences during the period. Such events may have played significant role in the electoral outcome. Our findings over the collected data during this period also justify such phenomenon of the influence of such political turn of events prior to the elections. The detailed analysis has been described in section 5.

0.4 Data Collection And Data Cleaning

In our analysis, Data Collection is one of the most important aspects of the analysis. Data is collected regularly from selected twitter accounts to be 203. such narendramodi, ArvindKejriwal, Arunjaitly, smritirani, etc and few popular political hashtags.

0.4.1 Collecting Tweets

The data is collected over a period of time prior to elections through twitter API which is a Java-based application. For collecting the tweets we have pre-selected 203 accounts based on political relevance and twitter popularity, such as narendramodi,ArvindKejriwal,arunjaitley,AmitShah SushmaSwaraj etc. We had a few popular hashtags #bjp , #congress,#hindu etc. Based on the above hashtags and the account names I have created the application to download the data regularly. This application needs to have either a hashtag list or a political accounts list in a file next line separated manner.Following this process over 2 months period of time we have collected over 252891 total tweets.While Collecting the tweets I had faced problems like every month I had to change the API keys . There is restriction of downloading the tweeter API , thats why I had made 3 sets of keys to download the data. We have chosen the political parties based on the very specific few reasons. In the Indian general election, 2019 it was BJP against rest of the political parties. That's why we have chosen the accounts in this such way that half of our users or the twitter holders were from BJP, rest half of the accounts were distributed among the rest of the parties such as Indian National Congress, Aam Aadmi Party, All India Majlis-e-IttehadulMuslimeen etc the details of which is provided in Table 2 and Table 3.

0.5 Data Analysis and Findings

For analysing the tweets, we resorted to Graph based analysis. We have generated the graphs in two ways: a. On the basis of Common words in their tweets of individual account holders where the nodes represent the individual users and b. On the basis of the common context in tweets among different account holders where the nodes represent the common words as well as the individuals. The detailed procedures of graph generation for the above two types is described below.

0.5.1 Graph Generated On the Basis of common Words

In this scenario, I have divided the entire data over a time period of tweets by users. Then I have created the graphs based on the common words. If during the period, two users have certain words in common depending on a context relevant to that

Political Accounts Aggregate	
Political party	User Count
BJP	86
Indian National Congress	45
Aam Aadmi Party	27
All India Majlis-e-Ittehadul Muslimeen	1
LokSatta,Surajya movement	1
All India Samathuva Makkal Katchi	1
Dravida Munnetra Kazhagam	2
Jan Adhikar Party Loktantrik	1
Janata Dal (Secular)	1
Jannayak Janta Party	2
Lok Janshakti Party	1
Nationalist Congress Party	2
Samajwadi Party	14
Shiv Sena	3
Swaraj Abhiyan	2
Telangana Rashtra Samithi	3
Telugu Desam Party	3
YSR CONGRESS	4
All India Trinamool Congress	2
Maharashtra Navnirman Sena	2

TABLE 2: Selection of Political Parties

Account Details		
ID	OriginalName	PoliticalParty
AjitPawarSpeaks	Ajit Pawar	Indian National Congress
AshokChandnaINC	Ashok Chandna	Indian National Congress
bhupeshbaghel	Bhupesh Baghel	Indian National Congress
bishnoikuldeep	Kuldeep Bishnoi	Indian National Congress
chalil	Dilip Chalil	Indian National Congress
DeependerSHooda	Deepender S Hooda	Indian National Congress
Dr_Uditraj	Dr. Udit Raj	Indian National Congress
DrParameshwara	Dr. G Parameshwara	Indian National Congress
gsurya	Suryanarayan Ganesh	Indian National Congress
INCChhattisgarh	INC Chhattisgarh	Indian National Congress
JP_LOKSATTA	Jayaprakash Narayan	LokSatta,Surajya movement
AAPHaryana	AAP Haryana	Aam Aadmi Party
janlokpal	Voice of the People	Aam Aadmi Party
meerasanyal	Meera Sanyal	Aam Aadmi Party
AamAadmiParty	AAP	Aam Aadmi Party
ArvindKejriwal	Arvind Kejriwal	Aam Aadmi Party
msisodia	Manish Sisodia	Aam Aadmi Party
raghav_chadha	Raghav Chadha	Aam Aadmi Party
SanjayAzadSln	Sanjay Singh	Aam Aadmi Party
asadowaisi	Asaduddin Owaisi	All India MajliseIttehadul Muslimeen
realsarathkumar	R Sarath Kumar	All India Samathuva Makkal Katchi
AmitShah	Amit Shah	BJP
AmitShahOffice	Office of Amit Shah	BJP
anandibenpatel	Anandiben Patel	BJP
AnanthKumar_BJP	Ananthkumar	BJP
aniljaindr	Dr. Anil Jain	BJP
ashishsood_bjp	Ashish Sood	BJP
AUThackeray	Aaditya Thackeray	BJP
BJP4Andhra	BJP ANDHRA PRADESH	BJP
BJP4Gujarat	Bharatiya Janata Party (BJP) Gujarat Pradesh	BJP
BJP4Haryana	Haryana BJP	BJP
BJP4Himachal	BJP Himachal Pradesh	BJP
BJP4Karnataka	BJP Karnataka	BJP

TABLE 3: Political Account Details with respective Twitter ids

Sample Tweets	
Account Info	Tweet
narendramodi	Addressing exservicemen before the dedication of #NationalWarMemorial to the nation.
Pawar_Sharad	My health is better now. Thank you all for the concern.
smritiirani	RT HMOIndia: The Union Cabinet today approved the Constitution (Application to Jammu & Kashmir) Amendment Order, 2019. This order will pav...
AmitShah	RT BJP4India: LIVE : PM Modi's mega interaction with Volunteers, Supporters and party Karyakartas from across the country. #MeraBoothSabse...
narendramodi	I found it surprising that the Congress would oppose Aadhaar but the reasons are now clear. Know why...

TABLE 4: Sample tweets

period, then they are connected. In our findings, it has been observed that such set of common words may vary relevant to different incidents over different periods. We have created a bag of words over a period of time through consultation with event expertise. The common words for generating a graph in a period is nothing but a subset of the bag of words which changes depending upon the changes in context over different periods. Let us consider two users say, user1 has n1 words in his tweets over a week and user2 has n2 words in his tweets over the same period, and let say that the length of our bag of words is k1, then the two users are connected only if they have a certain number of common words from those bag of words. The graph displayed in Fig1 and Fig2, is based on the common words related to the highly impactful event of Balakot air strike that occurred on 26th February.

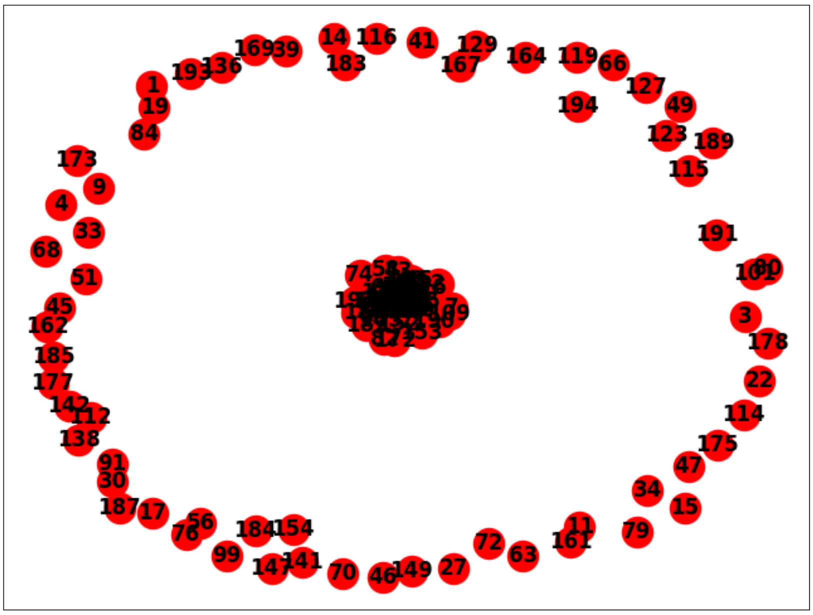

FIGURE 1: Figure showing a graph generated through common words related to Balakot air strike incident

In this scenario or to be specific data or tweets are used from date February 27, 2019, to March 2 2019, the graph is generated on the basis of the relevant common words corresponding to the Balakot strike (Fig1). We can observe from the graph that most of the nodes are not connected in the outer side, and the central part is highly connected. Next we have displayed the magnified graph of central part from Fig1.

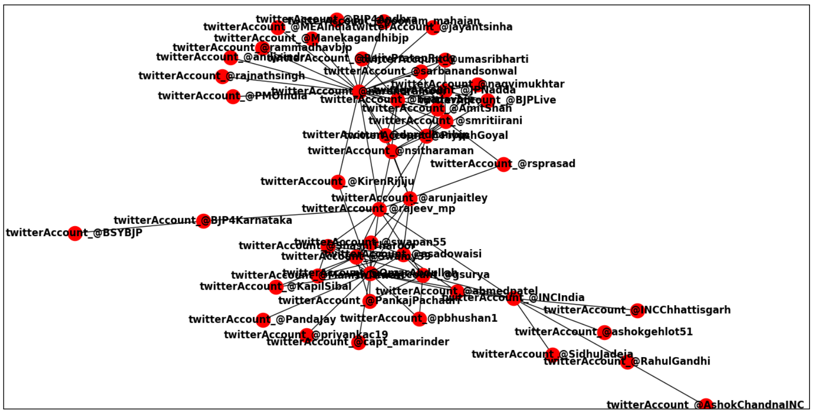

FIGURE 2: Magnified graph of central part of Figure 1

From Fig. 2, it is clear that the graph consists of two different clusters representing different individuals from the political parties. A closer inspection reveals that the upper cluster mostly links the BJP members based on their tweets with words which possibly bears implication of their political views wrt Balakot incident. Similar implication hold for the non-BJP parties in the lower cluster, only in their case it is the other political view expressed through the same incident, but still varying from the upper cluster, despite the fact that everybody was talking from the point of nationalism with respect to this incident.

Another important structural property in the context of real world networks, viz. degree distribution is studied from Fig. 2. The degree distribution of the network (Fig. 2) follow standard power law with exponent 2.73 after discarding some of the lower degree nodes shown in Fig. 4. This validates the real world characteristics of the generated network. Several such network degree distributions also have been shown over different incidents and periods, both short and long, from Fig 5 to 9.

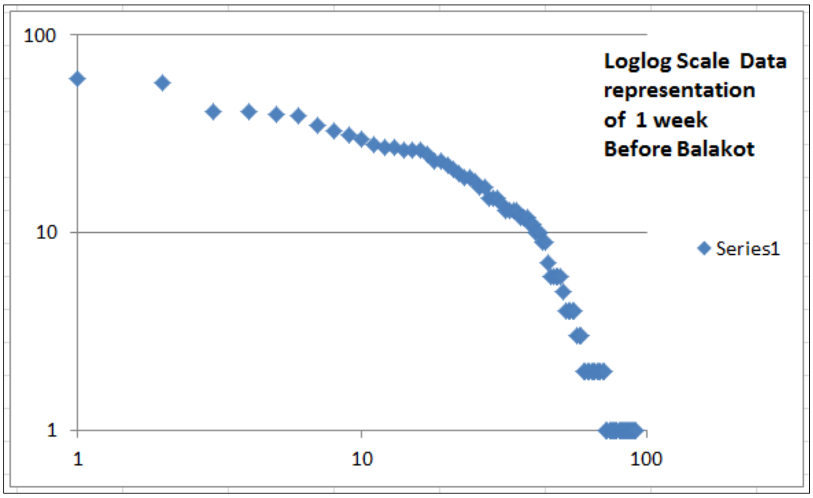

FIGURE 3: log-log degree distribution of the graph in Figure 2

After the study of the graphs, we plotted the degree distribution of the Log Log graph, and we found that it seems to follow the power law of the degree distribution if we remove some of the lower degree edges. We also observed that the value of the Gama in this power-law degree distribution was 2.78. Similar kind of results also was observed in each of the weeks, and in the entire data. Hence we can conclude that it may have some physical significance of this study.

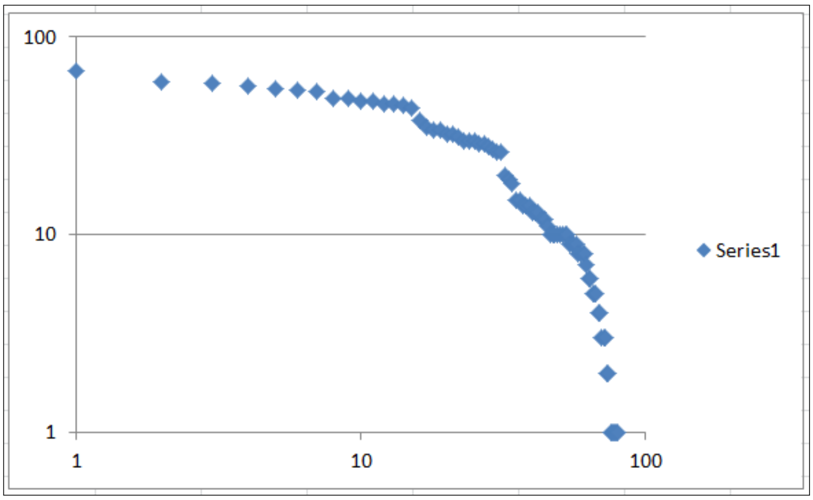

FIGURE 4: log-log degree distribution of the graph in the first half of the entire time period

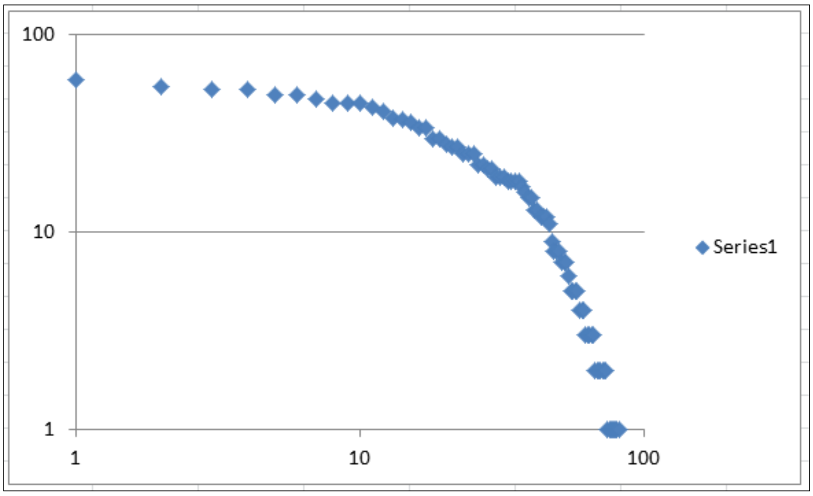

FIGURE 5: log-log degree distribution of the graph of entire data

0.5.2 Graph between relevant words and users in different contexts

On a weekly and monthly basis, different tweets have been meticulously observed in terms of political relevance and frequency of occurrence. By observation and consultation, an extensive bag of politically relevant words from tweets has been sorted from various political groups. Instances have been taken to check the probability of existence and frequency score of a similar set of words between two or more user groups. On analysis of the same, it has been noted that the selected bag of words when common between two or more user groups, stands as an indicator of users belonging to the same political groups, sharing the same line of interest and topics of conversation. The Bag of words are collected on the basis of their frequencies in each week and the relevance of those words in each week. Once the user groups and the bag of words show coherence, they are connected, thereby helping to generate small graphs. Figure 6 shows the example of such generated graph corresponding to a relevant word. Thereafter, Figure 7 is generated by accumulating all the relevant words described in Figure 6.

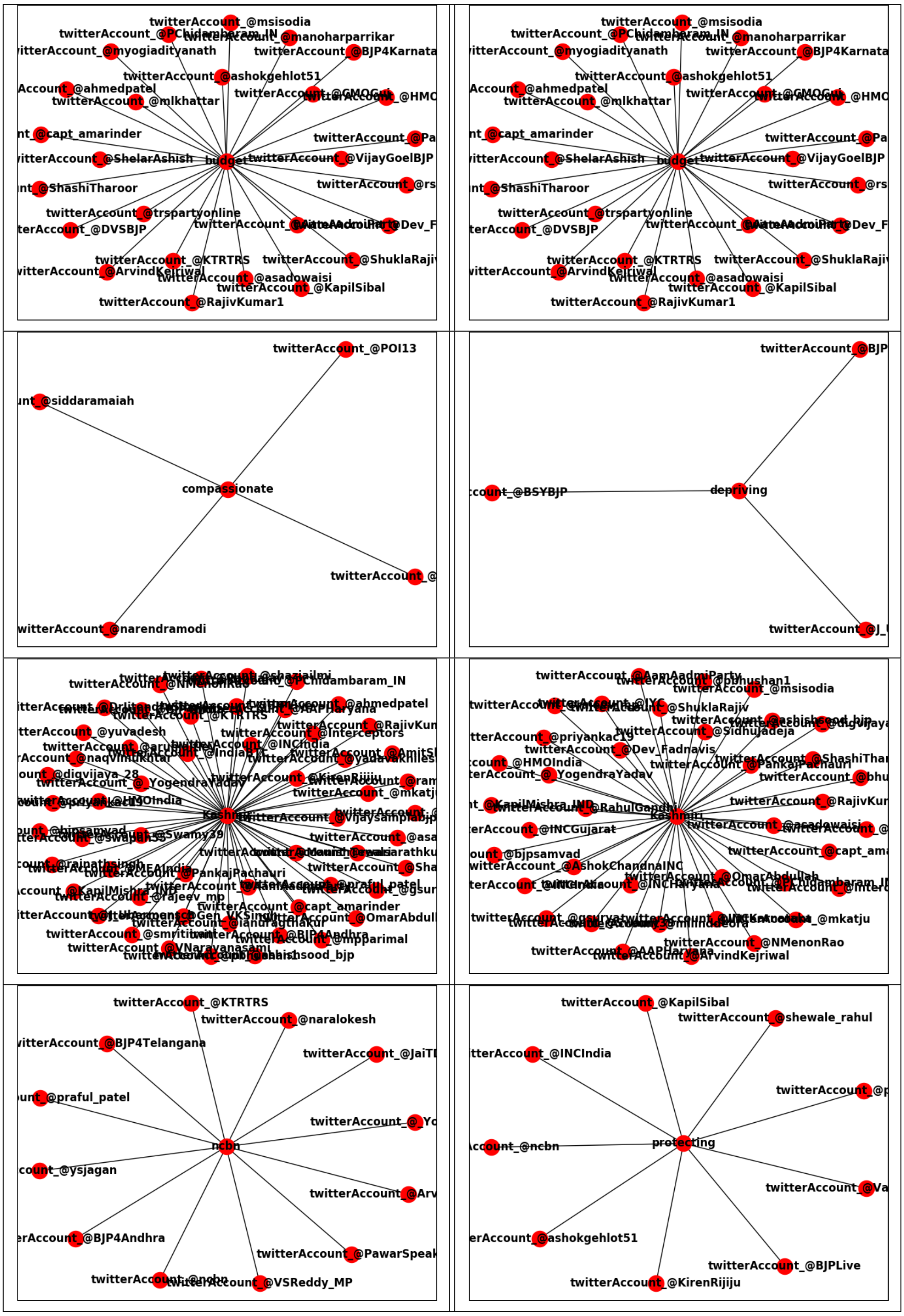

FIGURE 6: small graphs generated using a relevant words

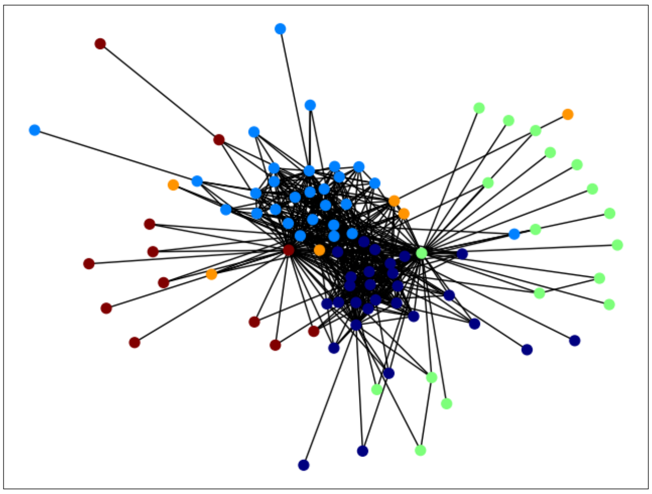

FIGURE 7: clusters identified by applying Louvain method. Different colors represent different clusters.

Community detection is also an important topological analysis in the study of real-world complex network. Here we also try find out the clusters and the contextual analysis of the respective clusters in the context of Indian General Election 2019. We have applied standard community detection method viz. Louvain method [3] over the graph in Figure 7 to detect the clusters present in the network. We have identified six clusters, represents through different colors in Figure 7, and the share of their political members in the respective clusters as shown in Table 5. From this table, our primary observation was in each of the group where ever BJP was in more in count the other parties were pretty less in the count and in the opposite scenario where ever the count of the oppositional party was hi we found the BJP count was less. we experimented the same thing in the rest of the data and found more or less similar kind of results. From this, we can say that it may be possible that in 2019 twitter share whatever BJP was communicating via social media, the oppositional parties were silent on those topics and whatever oppositio1nal parties were communicating the ruling party or BJP was silent on that. The detailed analysis of the clusters described in Table 5 have been done through the pie-chart as shown in Figures 8-14.

Details of groups					
Community	BJP	INC	AAP	AITMC	Words
Group0	10	2	3	0	Compassionate, Loksabha, Neighbor , Voter ,Woman
Group1	4	3	1	0	Allegations, Strike, Surgical Strike
Group2	8	8	0	0	Nation Building ,Surgical
Group3	21	10	0	0	Nadeeminc , NCBN , Pulwama, Suffered , Terrorist Attack , Undermining
Group4	7	1	0	0	Depriving ,Protecting ,Protection ,Reform ,Torture
Group5	4	7	0	0	Budgam , Kashmir ,Kashmiri, Kathiawari ,Srinagar
Group6	8	2	1	0	Budget ,Technologies

TABLE 5: Details of groups

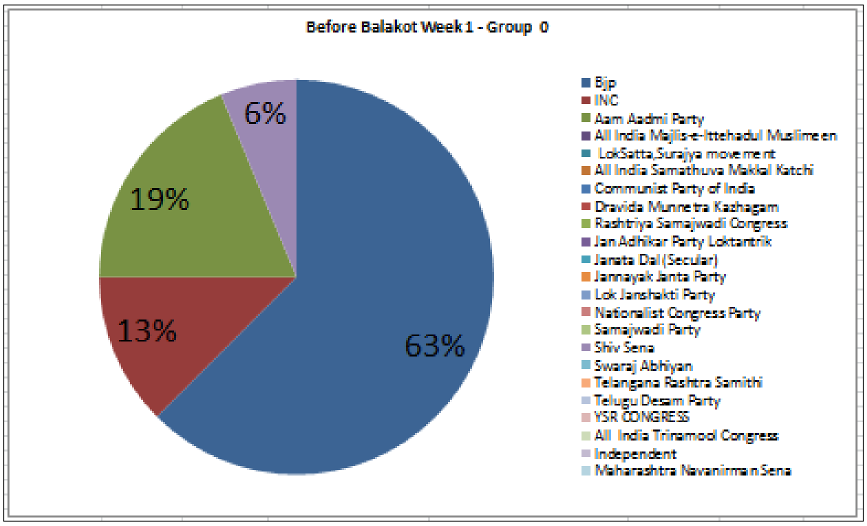

FIGURE 8: Pie chart of Group 0 User count

From the Pie Charts it is observed prominently that in cases where BJP has a strong count of user, the other parties' count of users is comparatively low, which included parties like Rashtriya Samajwadi Party, YSR Congress and Indian National Congress. This further puts light on the fact that majority people who tweeted based on similar bag of words, belonged from the same political group- BJP.

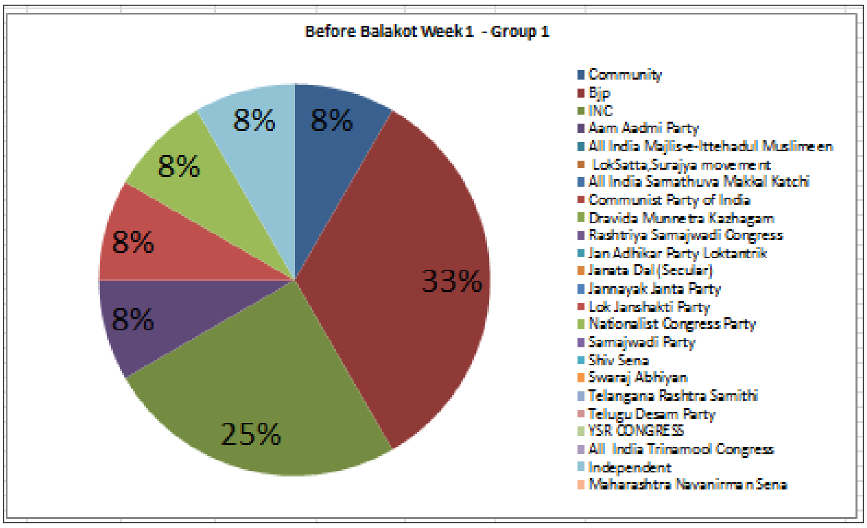

FIGURE 9: Pie chart of Group 1 User count

In this particular pie it is observed that after a strong percentage of similar bag of words from BJP, the Nationalist Congress Party too holds a strong user count in conversation but lower than BJP, and the rest parties to name like Shiv Sena, Communist party of India holds equal user count, all less than BJP and INC.

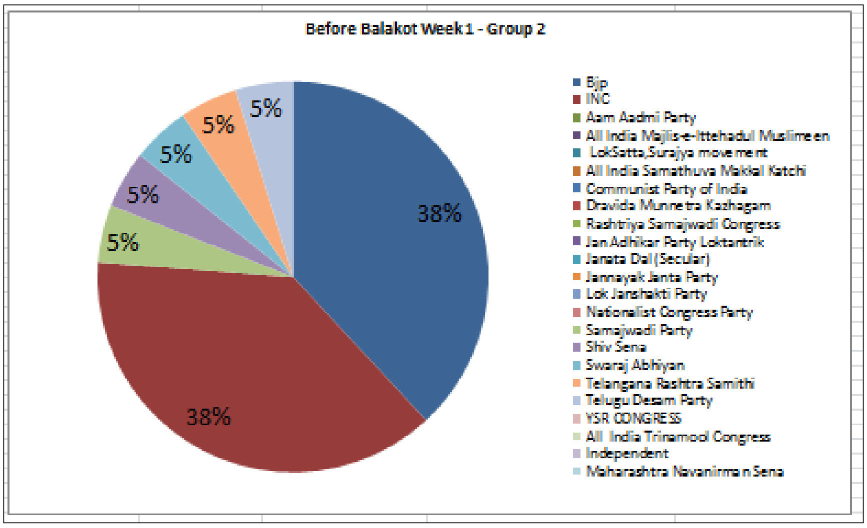

FIGURE 10: Pie chart of Group 2 User count

In this particular scenario, it is observed that, both BJP and INC share equal percentage of user count whereas other parties have relatively low score in terms of user count.

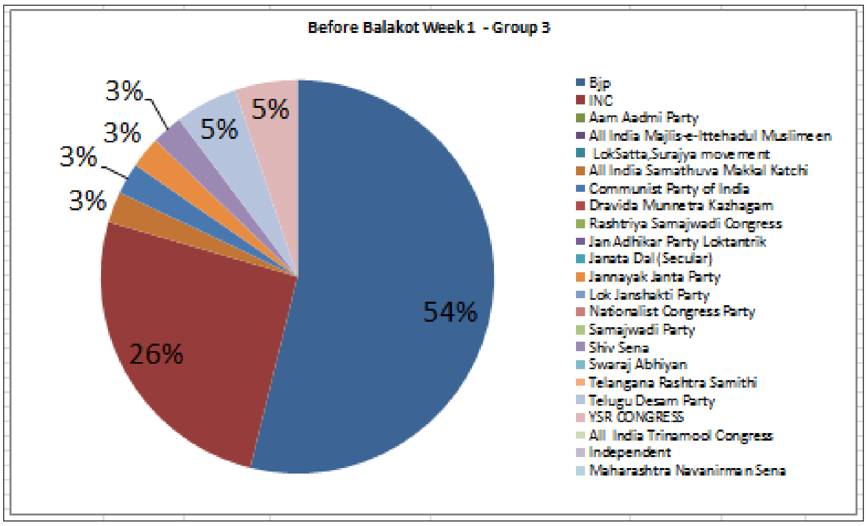

FIGURE 11: Pie chart of Group 3 User count

In this case too, it is evident that when the percentage of user count is high in case of BJP, the other parties have a comparatively weak user count.

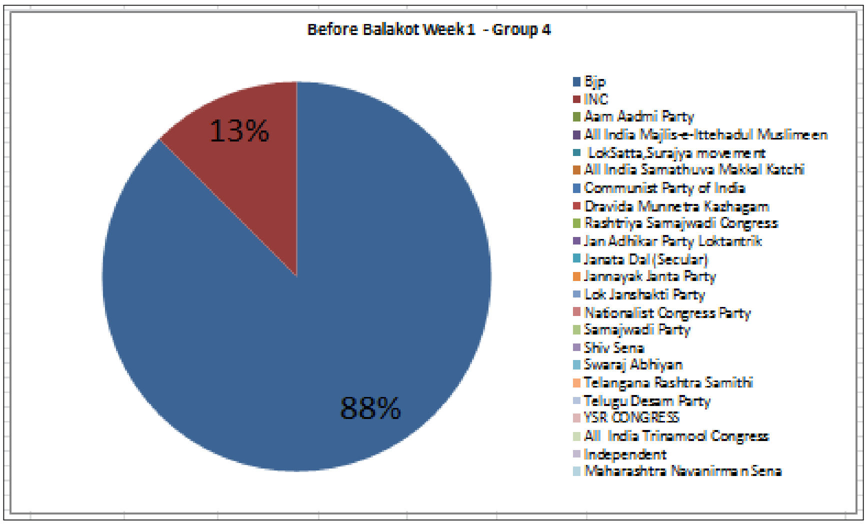

FIGURE 12: Pie chart of Group 4 User count

Here it is observed that only two parties share percentage of user count, BJP having the higher percentage as compared to INC.

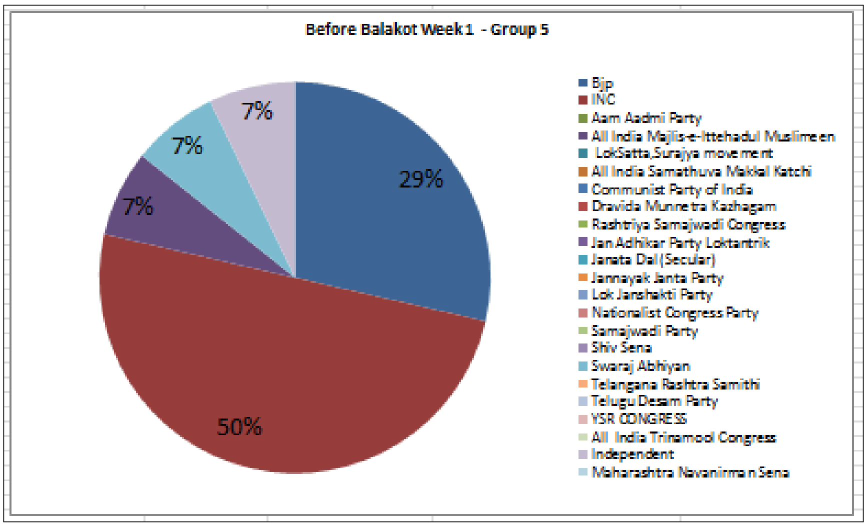

FIGURE 13: Pie chart of Group 5 User count

In this scenario, it is observed that the count of users is more in INC as compared to BJP and other political parties.

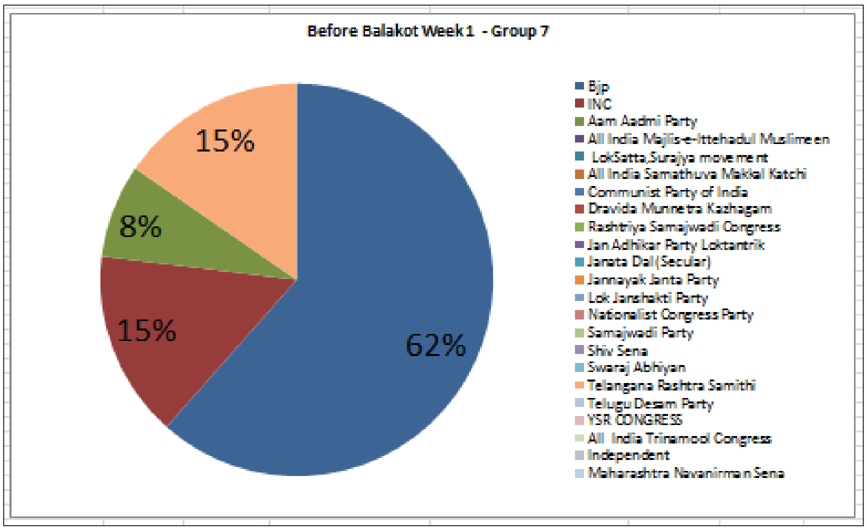

FIGURE 14: Pie chart of Group 7 User count

From the Pie Charts it is observed prominently that in cases where BJP has a strong count of tweets, the other parties' response is comparatively low, which included parties like Rashtriya Samajwadi Party, YSR Congress and Indian National Congress. This further puts light on the fact that majority people who tweeted based on similar bag of words, belonged from the same political group- BJP.

0.6 Conclusions

This study has made an attempt to find the patterns/trends in the political tweets during a period leading upto the Indian Genral Election 2019. Here we observed that in a period of certain striking incident like Pulwama, Balakot there was a clear political division in their tweet patterns of the 200 odd political personalities choosen. These also matched with their actual political positions/parties. This may have a role in the actual electoral verdict based on the influences of such political personalities. The method adopted here for generating graphs is also vindicated in the reflection of real world network characteristic. Another interesting finding is about a division in topics between the BJP and non-BJP parties in general which again may have important implication in the final outcome of the election results. Such division is reflected in the second type of graphs where nodes belong the relevant words as well the users who are discussing those.

This study has made an attempt to answer whether twitter trends can predict similar kind of Tweet shares which leads to political candidates belonging to the same party or party, and It has been shown that candidates of the same party usually share similar kind of tweets and hence as a group a party focuses on a particular set of political beliefs. Our research setting in Indian context offers a complex political landscape. For example, taking the case of Pulwama attack in Feb 14 2019, most of the political parties tried to get the political stronghold which has been distinctively observed from political tweets and its graph. Thus, it is crucial to analyse the explanatory power of twitter trends in predicting coherence in similar political mindset of different parties. Our findings broadly confirm existing studies. However, it also focuses on a few potential the shortcoming of using simplistic kind of analysis like connecting graph edges when a certain number of words are matched. Our contributions are in the number of graphs. First, our study has made an attempt to develop a kind of template for data collection and for storing indexing and searching words using Lucene jar extensively, which can be used by researchers for similar kind of work. Our political domain is driven data mining model has significantly improved our data collection and relevant tweet identification process utilizing certain key bag of words. As we have used Lucene,it hardly affected the unwanted tweets to interfere with our mining process.

Till now we have classified political accounts based on the common bag of words, that refers to that if two or more users use similar words in their tweets at a particular frequency, they generally belong to the same political party leaving aside certain outlier cases. On the contrary, if two or more users belong to the same political parties or classes, it is seen that they convey similar kind of tweets or political beliefs, except certain outlier cases.

However taking reference from this thesis, the kind of words a particular political party uses in a frame of time and the graphs generated will assist the future researcher to find interdependency of political beliefs, political tweets and political candidates. Future scope of research might focus on classifying the business of new Twitter accounts and determining their political inclination by applying machine learning algorithms.

Bibliography

[1] Tech. rep. URL: http://guides.neo4j.com/sandbox/us-elections-2016.

[2] Apoorv Agarwal et al. "Sentiment analysis of twitter data". In: *Proceedings of the Workshop on Language in Social Media (LSM 2011)*. 2011, pp. 30–38.

[3] Vincent D. Blondel et al. "" Fast unfolding of communities in large networks". In: *Journal of statistical mechanics: theory and experiment* (2008).

[4] Ciro Cattuto et al. "Time-varying social networks in a graph database: a Neo4j use case". In: *First international workshop on graph data management experiences and systems*. ACM. 2013, p. 11.

[5] Swarup Chattopadhyay, Asit K Das, and Kuntal Ghosh. "Finding patterns in the degree distribution of real-world complex networks: going beyond power law". In: *Pattern Analysis and Applications* (2019), pp. 1–20.

[6] Raviv Cohen and Derek Ruths. "Classifying political orientation on Twitter: It's not easy!" In: *Seventh International AAAI Conference on Weblogs and Social Media*. 2013.

[7] *Creating Beautiful Twitter Graphs with Python*. Tech. rep. URL: https://towardsdatascience.com/creating-beautiful-twitter-graphs-with-python-c9b73bd6f887.

[8] Oshini Goonetilleke et al. "Microblogging queries on graph databases: An introspection". In: *Proceedings of the GRADES'15*. ACM. 2015, p. 5.

[9] Aparup Khatua et al. "Can# Twitter_trends predict election results? Evidence from 2014 Indian general election". In: *2015 48th Hawaii international conference on system sciences*. IEEE. 2015, pp. 1676–1685.

[10] A. Moschitti. "Making tree kernels practical for natural language learning". In: *Proceedings of the Eleventh International Conference on European Association for Computational Linguistics, Trento, Italy*. 2006.

[11] B. O'Connor et al. "From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series". In: *Proceedings of ICWSM, 11*. 2010, pp. 122–129.

[12] A. Tumasjan et al. "Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment". In: *Proceedings of ICWSM, 10*. 2010, pp. 178–185.