ORIGINAL ARTICLE

Analysis and mining of an election-based network using large-scale twitter data: a retrospective study

Amartya Chakraborty[1] · Nandini Mukherjee[2]

Abstract

The user-generated Twitter data are a rich source of study and research that reflects the various social, economic, political, and other issues affecting people across the world. Analysis of the social interactions among users, who express themselves online, reveals different internal dynamics and provides detailed insights into real-world phenomena. In this paper, the structure and dynamics of the state assembly election-based tweet-reply network have been studied, as generated by Twitter users across the country of India for a period of 6-weeks. We study the flow of Twitter activity pertaining to the West Bengal assembly elections, along with the identification of the hashtags used by the three main political contenders. This information is used to identify the cluster-level dominance in the Twitter network over the 6-weeks of study. It is observed that this *cluster dominance* information is representative of the actual outcome of the elections, and can be effectively used as a forecasting tool. The collected tweets are used for lexicon-based emotion detection and further analysis. This highlights the reaction of the social media users in response to the events related to the election. It is observed that fear is the dominant emotion, while happiness is scarce in the opinions expressed during the studied duration. Next, the study and analysis of the complete reply-based social networks during weeks 1, 4, and 6 are undertaken. Important political and media actors are identified with standard network-level measures toward determining the efforts put in by the different clusters and individual actors involved in the election to control the network dominance.

Keywords Network analysis · Indian elections · Election network · Twitter network · Social networks · Cluster dominance

1 Introduction

The Online Social Media (OSM) platforms are a major source of the public opinion. These sources are rich in data about various social, geographical, economic, political, and other factors that drive the common people around the world to express themselves online. Such opinion is also representative of the people's perception and belief, and also molds the perception of other people with shared interests. The exchange of this discourse on online platforms can be visualized as a graph, where the nodes reflect users and the connecting edges represent their online interactions as per preset social network parameters. These networks correspond to different types of global events, ranging from natural or man-made disasters (Kim and Hastak 2018), social events (Korolov et al. 2016), political events (Perliger and Pedahzur 2011), to the ongoing pandemic (Naseem et al. 2021), etc.

In India, the largest democracy in the world, a federal structure of governance is followed. Consequently, separate elections are organized for choosing the national-level and state-level representatives. The cabinet of ministers of the central government is chosen from the political party or alliance that secures a majority of votes in the *Lok Sabha* elections. Similarly, for each of the constituent *States* or *Union Territories* of the country, the regional *State Assembly Elections* are conducted. This work is focused on the social network activity corresponding to one such assembly election for a particular state of the country. The analysis of the scraped Twitter data reveals that there are three major

✉ Amartya Chakraborty
amartya3@gmail.com

Nandini Mukherjee
nandini.mukhopadhyay@jadavpuruniversity.in

1 Department of Computer Science and Engineering, University of Engineering and Management, Kolkata, West Bengal 700160, India

2 Department of Computer Science and Engineering, Jadavpur University, Kolkata, West Bengal 700032, India

political alliances, which we identify as Cluster 1, Cluster 2, and Cluster 3 in this paper. The studied period - up to the elections and during the elections, witnessed heated campaigns, social unrest, and major political activities even in the presence of the ongoing COVID-19 pandemic.

The final efforts for the 8-phase election witnessed a tense political scenario amidst the shift of a number of influential Cluster 2 leaders to the Cluster 1 (Banerjee 2021). The people of the state took to different forms of media to voice their opinions and support for the different clusters or their leaders. Historically, Twitter has emerged as a relevant source of information-rich social opinion, and this election in the fourth-most populous state of the country is no exception to that phenomenon. The network of Twitter users who contributed to the social media platform by expressing themselves on this topic, is an enlightening research problem involving social network analysis and study. An attempt is made to discuss the analysis with relevance to the observations from hashtag-based *cluster dominance* estimation, toward determining the overall network influence of different actors and political clusters in the given scenario. The current article contributes as follow:

- We extract the Twitter data pertaining to 'bengal elections' on the 1st and 3rd weeks of the months of February, March, and April, 2021, i.e., for a total time period of 6 weeks,
- We determine and analyze daily and weekly Twitter activity during the period of study with respect to the topic of 'bengal elections',
- We identify the hashtags corresponding to each cluster, and propose the *cluster dominance* method for the weekly networks based on the period of study,
- Consequently we study how the *cluster dominance* information is critical for forecasting the election outcome with reference to the actual results,
- We determine and study the proportion of different emotions expressed in tweets by each political cluster,
- Finally we identify the different roles played by important actors in the studied election network using graph-based metrics.

The rest of the paper is organized as follows: Sect. 2 discusses the state-of-the-art research in the analysis of election-related online social networks; Sect. 3 describes the scraped data and identifies the relationship between the election dates and the variations in daily and weekly tweet count. With the help of the proposed *cluster dominance* technique, the user-level affinity has been determined for each of the individual clusters. This has been performed with the help of the hashtags extracted from the tweets, as discussed in Sect. 4. This section further studies the observed cluster-level social network dominance, and the measure of different

emotions expressed in the tweets for each of the 3 clusters using the extracted hashtag information. In Sect. 5, a comprehensive analysis of the social networks formed during the 1st, 4th, and 6th weeks of the study is undertaken with different network-level centrality measures to obtain in-depth inferences about the behavior and role of different network actors. Finally, concluding remarks are offered in Sect. 6.

2 Related works

The explosion of public expression on OSM platforms has inspired their analysis and study, and this has become a prime feature of the inter-disciplinary research in computer science, social science, etc. The concepts of network analysis, sentiment analysis and classification, emotion recognition and classification, etc., are undertaken by state-of-the-art research on related topics of election networks, as gathered from social media platforms like Twitter. The use of network analysis metrics in association with sentiment or emotion recognition, has been largely adapted to analyze the social media election networks across the different continents. This section undertakes a study of some of the recent literature in this field.

From the most populous democracy of the world in the Asian continent, India, tweets about the state assembly elections for the country's capital city Delhi had been analyzed using sentiment analysis tools and classification, for predicting the election results by Srivastava et al. (2015). By estimating the diffusion of sentiments about different contesting parties, Kagan et al. (2015) predicted the outcome of Indian and Pakistani elections of 2014 and 2013, respectively. From the same continent, the polarization during the Presidential Elections in the world's largest island country Indonesia, in 2019, was analyzed using social network analysis metrics as well as sentiment analysis techniques toward inferring valuable insights that reflected the ground truth (Habibi et al. 2019). The social network data pertaining to the same elections was used for research by Paskarina and Nuraeni (2021) and Rakhmawati and Mufidah (2020) who performed community detection with the help of network-level metrics, based on different candidate information. The highlighted trends and important communities pertaining to the Iranian Presidential election in 2009, were identified using Twitter conversations by Tabatabaei and Asadpour (2014).

Pertaining to the continent of North America, the Twitter data related to elections in the United States of America (USA) have been analyzed frequently using social network analysis methods. The 2012 Presidential Elections in the USA was subject to emotion and network analysis based on a pre-determined set of controversial political topics and related Twitter conversations in the work by Himelboim et al. (2016). The diffusion of rumor was determined using

network statistics on Twitter data for the same event by Shin et al. (2017). In a similar work, the 2016 Presidential election prediction was undertaken by analysis of tweet volume and sentiment pertaining to the different candidates' Twitter activity (Heredia et al. 2018). Social Network data from the protests after the same 2016 election were analyzed to identify the relation between network structure and the occurrence of protests by Renaud et al. (2019). In a recent work, Chen and Wang (2022) studied how the political incivility was propagated through a misleading election campaign video on YouTube during the 2020 U.S. Presidential elections.

In the case of the European continent, we find the mining and analysis of tweets about the French elections in 2012 was performed by Wegrzyn-Wolska and Bougueroua (2012). Corresponding to the same year, the German Bundestag elections were analyzed in terms of network sentiment and community structure by Plotkowiak and Stanoevska-Slabeva (2013). On the other hand, the Twitter activity regarding the Presidential elections in the African country of Nigeria was utilized for detecting sentiment polarity using lexicons as well as classifiers, related to the participating candidates, in order to predict the election's outcome (Oyebode and Orji 2019).

From the study of these state-of-the-art research works, it can be observed that the different network analysis and election outcome prediction techniques are mostly based on existing network-level measures. There has not been any research that compares the election outcome prediction capability of network analysis measures with the efficacy of contextual information that can be gathered from the network.

3 Data description

For the proposed work, Twitter data have been acquired using full-archive search stream for a pre-determined period of time, in keeping with the election in a particular state of India. The phase wise election dates are of significance to the study, and are listed in Table 1 (HT 2021). In order to capture the essence of the temporal Twitter activity starting with the initial stages of the election preparation upto the final round of polls, tweets containing the keywords 'bengal election' were gathered during the 1st and 3rd weeks for the months of February, March, and April (Chakraborty et al. 2022). The choice of these keywords for consideration was made after careful analysis of relevant n-grams (Chakraborty and Bose 2020), as described in the Appendix A. A total of 29462 tweets were gathered for the said duration, spanning over 6 weeks.

Table 1 West Bengal Assembly Election Dates, 2021

Phase	Date
1	27th March, 2021
2	1st April, 2021
3	6th April, 2021
4	10th April, 2021
5	17th April, 2021
6	22nd April, 2021
7	26th April, 2021
8	29th April, 2021

The acquired Twitter data were initially checked and was found to contain a large number of duplicate tweets. On removing the duplicates, the tweet corpus was reduced to contain 26180 tweets. The daily and weekly distributions of these tweets are illustrated in Figs. 1 and 2. The importance of duplicate removal and unique tweet identification is discussed in details in Sect. 4.2.

Observations:

- Fig. 2 highlights that the number of tweets about elections in the selected state keeps increasing from February to April. The studied periods of February and March precede the elections, and thus stand witness to the long-term and last-minute preparations by the contesting clusters. It is seen that the first weeks of March and April witnessed greater Twitter activity than the third weeks in each case.

- Local peaks in the weekly tweet distribution are observed on 1st, 6th, 7th, and 18th February. In the two weeks of study in this month, the first week shows more Twitter activity on weekends (6th and 7th February), while the third week's Twitter activity represents a bell curve with the peak on a Thursday, and minimal activity on weekends.

- In a striking similarity to the month of February, the first week of March also shows more activity on weekends (7th March), and the third week follows a bell curve in terms of the daily number of tweets (18th March).

- April is the only month where the studied period coincides with actual election dates - 1st, 6th, and 17th April. Consequently, peaks in tweet frequency are noted on these same dates from Fig. 1. The previously noted trend of tweeting more on weekends during the first week and more on weekdays during the third week in the months of February and March is disrupted here. We find that more people have taken to Twitter on the days of the election, irrespective of whether they are weekdays (1st and 6th April) or a weekend (17th April).

4 Tweet data analysis

4.1 Hashtag clustering and cluster dominance

From the tweet corpus, we have extracted the set of all hashtags that have been used. A total of 537 unique hashtags were identified manually. Hashtags are extensively used in tweets to give context to the tweet text. As already discussed, the corpus consists of tweets that pertain to the selected keywords. On inspection, it is seen that most of the hashtags correspond to some particular political party, or leader, or ideology, and are either in favor or against one of the parties, i.e., they can be easily mapped to a particular cluster. We have used the three pre-determined clusters, namely Cluster 1, Cluster 2, and Cluster 3, for segregating all such hashtags. Some representative hashtags for each cluster are listed below in Table 2.

It is important to note here that for a dataset with millions of tweets, the number of hashtags will also increase largely. In such cases, we propose to utilize a semi-supervised approach where some part of the hashtags can be manually assigned to specific political clusters. These hashtags, when vectorized, will enable a semi-supervised learning algorithm such as Label Propagation or Label Spreading to determine the labels for the remaining hashtags. The manual labeling approach is necessary for annotating any such data because the hashtags are not all in English but consist of Bengali words or phrases expressed in English alphabets. Only a knowledge of the actual political scenario will guarantee appropriate identification that the words or phrases are referring to particular political parties or alliances.

Using this hashtag cluster information, we have determined the weekly hashtag count made in favor of each particular cluster. Each week's network is built with a set
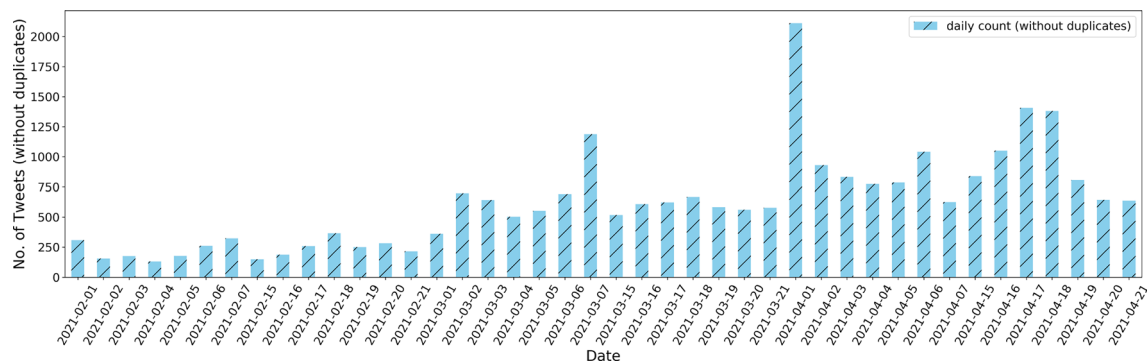

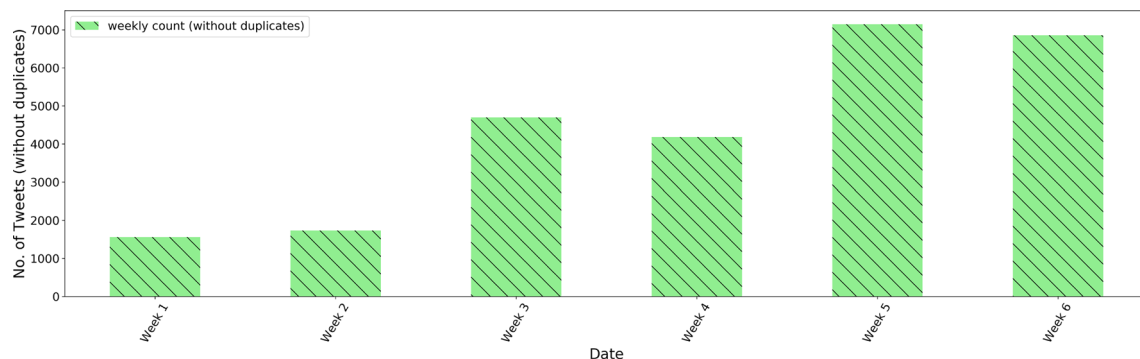**Fig. 1** Daily distribution of tweets for the studied period

Fig. 2 Weekly distribution of tweets for the studied period

Table 2 Hashtag Cluster-wise weekly number of tweets	Cluster	Sample tweets
	Cluster 1	bengalwithbjp, tmchataobanglabachao, banglarmanushbjpersathe
	Cluster 2	banglargorbomamata, bengalwithdidi, votefortmc
	Cluster 3	bengalneedsleft, savebengalfromtmcbjp, vote4sanjuktamorcha

of i nodes $(n_1, n_2, ..., n_i)$ and j edges $(e_1, e_2, ..., e_j)$. For each actor or node in the network there are y number of tweets $(t_1, t_2, ..., t_y)$, where each tweet consists of one or more hashtags. All such hashtags in the tweets of each node are appended to the set x defined as $(h_1, h_2, ..., h_x)$. Using these alphanumeric hashtag texts, the tweets of each node are classified as belonging to one of the three clusters on the basis of ground truth information. This helps in identifying the particular political cluster favored by the node. In this manner, the complete set of all the tweets made by all the nodes for each week's network is processed to identify the number of tweets made in favor of each of the three clusters.

This count has been converted to weekly relative percentage by taking the number of hashtags for each cluster over the total count of hashtags for all the three clusters, for that particular week. The obtained measure has been termed as the *cluster dominance*, represented by the equation:

$$\text{dominance of cluster}_i = \frac{\text{no. of hashtags of cluster i}}{\text{Total no. of hashtags that week}}$$

(1)

This process has been repeated for all the six weeks of tweet data that we have considered for this work. The result of this analysis is illustrated in Fig. 3. For establishing the validity of our proposed *cluster dominance* technique, we have tallied the different opinion poll and exit poll statistics produced by several polling agencies that are compiled at a wikipedia article (Wikipedia 2021), with our observations from the network characteristics.

Observations:

- From the figure, it is seen that the tweets with hashtags in favor of the Cluster 1 are maximum only in the first week. Thereafter, the hashtag count for the Cluster 1 has decreased, with the lowest value in week 4.
- In contrast, the hashtag count for Cluster 2 has increased every week with a peak in week 4, overtaking Cluster 1

after the first week, and then maintaining a consistent majority.
- Cluster 3 has the lowest presence overall, with maximum occurrence in week 2, a decrease till week 4, and then increasing till the end of the studied period.
- The results using *cluster dominance* reveal that during the 1st week of the study, Cluster 1 was dominating the online election network, and the same is observed in some of the opinion polls published using early stage data from around this time (Wikipedia 2021). However, the remaining weeks 2 and 3 before the polls observed Cluster 2 in a leading position, and this is also corroborated by the opinion polls conducted in this time (Wikipedia 2021). Thus, it is validated that the conclusions made from our proposed approach are in keeping with the ground reality of the election outcome.
- The weeks 4, 5 and 6 of our study coincide with the actual phases of the election, specifically, the 1st, 2nd and 3rd, and 6th phases respectively. The determined *cluster dominance* clearly identifies Cluster 2 as leading the scoreboard in terms of online presence and support, which can be translated into total votes and State Assembly seats gained as a result of public support. The actual election results (Wikipedia 2021) reveal the same picture, where cluster 2 leads the scoreboard in all of these phases.
- It is also observed that there is a dip in the online dominance of Cluster 1 during the 6th week of our study (corresponding to the 6th phase of the election), in comparison with the 5th week (2nd and 3rd phases of the election). This decrease in influence of Cluster 1 is also reflected in the actual poll results. While in the combined 2nd and 3rd phases they won almost 25% of the seats, about 19% of the seats went to their share in the 6th phase (Wikipedia 2021). The proposed *cluster dominance* method is thus able to capture the minute changes in the ground reality.
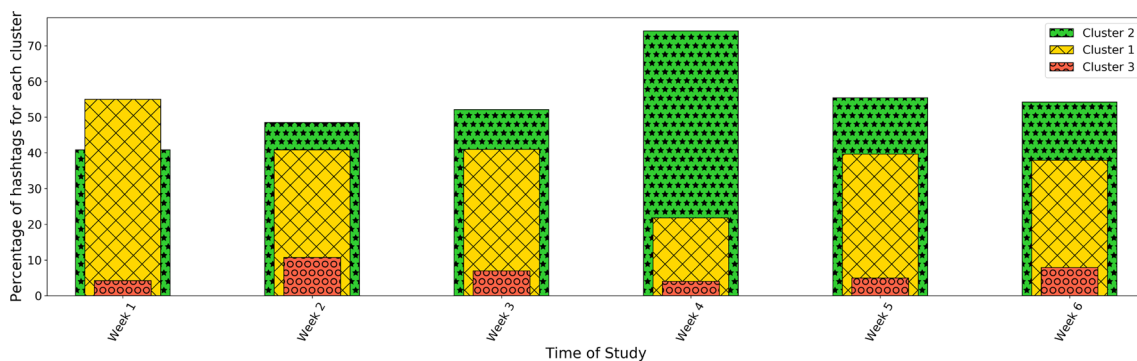

Fig. 3 Hashtag-based cluster dominance of each cluster over 6 weeks of study

- The statistical distribution of hashtag-based dominance of the three clusters, is illustrated in Fig. 4. On observation, it can be stated that Cluster 2 projects a clear majority in terms of establishing dominance in 'bengal election' related conversations on Twitter, followed closely by Cluster 1. Cluster 3 emerges as the contender with the lowest dominance in Twitter activity who also failed to secure any seats in the State Assembly. These observations are in keeping with the actual election outcomes, and thus, the proposed technique of *cluster dominance* is effective in forecasting the outcomes of elections based on temporal observations made in topical networks on social media platforms such as Twitter.

4.2 Robustness of proposed approach

The aforementioned cluster dominance method based on the extraction of hashtags from tweets faces some relevant challenges. The most important challenge is that of auto-generated tweets or tweets by bots or fake profiles, which may affect the whole analysis by introducing some unwanted bias in the dataset. From a study of relevant literature, it is observed that researchers have explored different means of identifying auto-generated or bot-generated content on social network platforms like Twitter.

Broadly, such measures either focus on manual verification and labeling of tweets (Kudugunta and Ferrara 2018; Lundberg et al. 2019) and extracting features related to the user accounts, followed by training and testing of supervised learning (classification) algorithms, or utilize available frameworks such as Botometer (Davis et al. 2016; Yang et al. 2022), Bot Sentinel (Bouzy 2018), etc., which provide an estimation of whether the tweets were auto-generated or posted from fake profiles or not. For networks dealing with the analysis and mining actual event (elections, disasters, etc.), the presence of fake profiles is the more relevant threat, as evident from state-of-the-art works dealing with fake profile identification (Ramalingam and Chinnaiah 2018; Roy and Chahar 2020; Wanda and Jie 2020; Boshmaf et al. 2015).

However, the main focus of the current work is not the identification of bots or auto-generated tweets in the network. Of course, in our current election data a well-coordinated effort by members of Information Technology (IT) cell (Jaffrelot and Verniers 2020; Jakesch et al. 2021; Ahmed et al. 2017) affiliated to one of the participating parties or coalitions will drastically increase the count of certain hashtags in the dataset. This may be misconstrued as an expression of support toward a particular cluster during cluster dominance analysis. Thus, the main challenge is to identify the legitimate tweets made by users from the total set of all tweets.

In the current work, the authors have overcome this challenge by eliminating all the 3282 duplicate tweets, as discussed in Sect. 3. This is a straightforward mechanism employed by various other researchers dealing with real-time social network data. For example, Imran et al. in their work on identifying mass emergencies from social network activity, have explicitly highlighted the 'de-duplication' process to reduce unwanted data (Imran et al. 2015). Similarly, Preotiuc-Pietro et al. (2019) removed the duplicate tweets while developing a complaint detection framework based on Twitter data. Habibi et al. (2019) performed an analysis of political polarization in Indonesia before the Presidential election where the data was processed initially by removing the duplicate tweets. In a similar manner, Oyebode and Orji (2019) removed the duplicate tweets in order to overcome any bias introduced in the online platform by reposting the same content. The work on election-oriented rumor detection and correction by Shin et al. (2017) also involves the removal of duplicate tweets for constructing the final dataset. Taking a cue from these relevant works, our proposed approach discussed in Sect. 3 ensures that even if numerous fake profiles or bot profiles post the same content, our dataset will include only the first entry of the unique tweet - all remaining duplicates are detected and removed. Also, from Eq. 1 it is ascertained that the effect of hashtag(s) included in each unique tweet is negligible in the determination of overall cluster dominance considering all unique hashtags. Thus, our proposed approach is robust enough to handle the manipulative posting of tweets by auto-generating bots or fake profiles on Twitter.
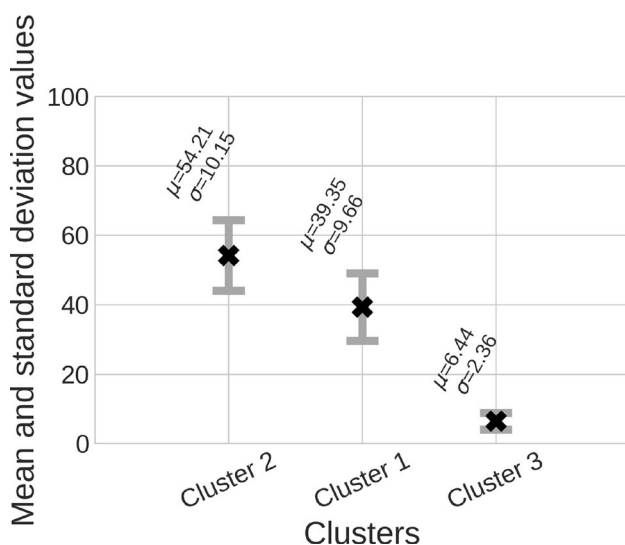

Fig. 4 Statistical distribution of cluster dominance based on hashtags

4.3 Cluster-based tweet emotion identification

In this part of the experiment, the tweets corresponding to the three hashtag-based clusters from Sect. 4.1, are extracted. These tweets are then processed and cleaned using a set of standard steps, namely *normalization*, *lowercasing*, *cleaning*, *splitting*, *lemmatization*, *stop-word removal* (Kowsari et al. 2019). The cleaned tweets are now processed using the *text2emotion* library (text2emotion Team 2020) which attempts to determine the emotions present in a text. This module assigns scores to the words in the text, based on the type and intensity of emotions expressed by those words. The different types of emotions detected are *happy*, *angry*, *surprise*, *sad* and *fear*. For all the tweets belonging to each cluster, these emotion scores have been determined. In this way, the emotion scores for all tweets pertaining to each cluster have been extracted.

Finally, these scores have been converted to weighted emotion scores in order to accommodate the inequality in the number of tweets pertaining to each cluster. The overall statistical distribution of different emotions in the tweets made by each cluster, are depicted in Figs. 5a, b, and c. On the other hand, Fig. 6a, b, c, d and e illustrate the cluster-wise weekly intensity of the respective emotions as detected in the tweets.

Observations:

- The *happy* emotion is detected most minimally in tweets for all the three clusters - happiness for Cluster 3 is fractionally more than that of Cluster 2, and Cluster 1 projects the most happiness in their tweets.
- *Fear* is found to be the dominant emotion in all the clusters, however, the amount of fear projected by Cluster 3

(a) Cluster 1   (b) Cluster 2   (c) Cluster 3
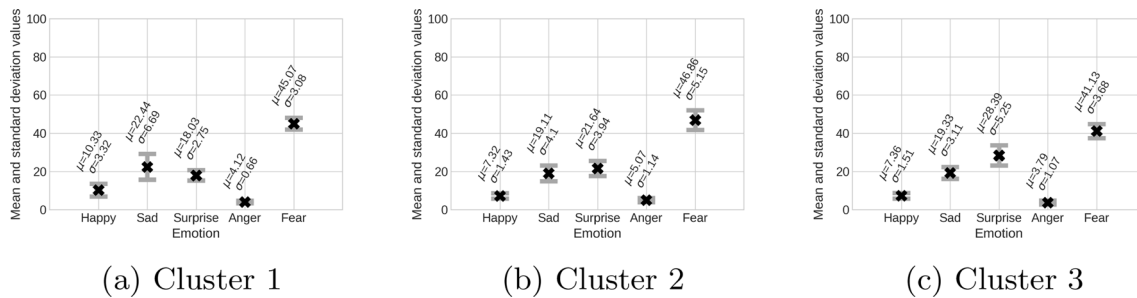
Fig. 5 Statistical Distribution of Emotions in tweets for the three Clusters

(a) happy emotion   (b) angry emotion   (c) sad emotion

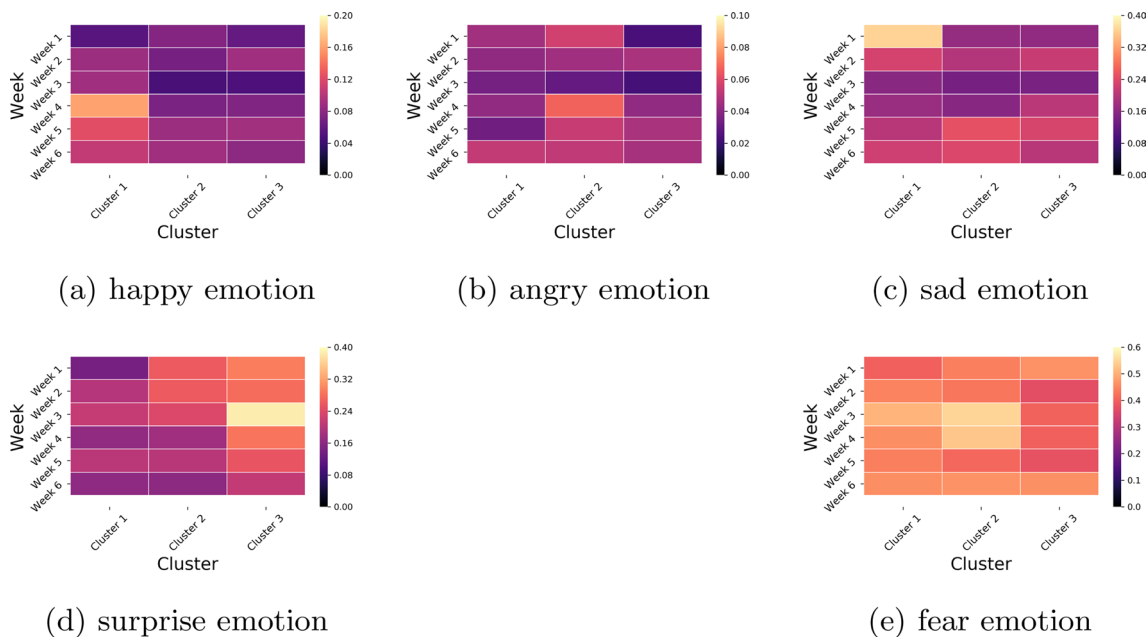(d) surprise emotion   (e) fear emotion

Fig. 6 Heatmaps of weekly cluster-wise emotions

is minimum, and that of the Cluster 2 is the maximum (during the 3rd and 4th weeks of study).

- Similar to the previous case, the emotion of *anger* is also maximum in case of the Cluster 2 and minimum in the tweets by Cluster 3.
- The emotion of *surprise* is consistently highest in case of the Cluster 3 (with a spike in week 3), followed by Cluster 2 and Cluster 1.
- The amount of *sadness* projected by the tweets of Cluster 1 is highest, followed by Cluster 3 and Cluster 2 respectively.
- The 4th week witnesses a simultaneous peak in *happiness* for Cluster 1, and in *anger* for Cluster 2. This may be indicative of the mentality of these cluster leaders and supporters before the commencement of phase-1 of the elections.
- A persistent spike in *fear* is detected for Cluster 2 during weeks 3 and 4, while, the *fear* peaks in week 3 and subsides in week 4 in the case of Cluster 1.

5 Twitter network analysis

There are different types of interactions possible between actors on online social media platforms, such as, reacting to the post (liking, etc.), replying to the post, sharing the post, etc. For the proposed work, the 'reply-to' network has been created using the scraped Twitter data for further network analysis. This network is generated based on the activity of tweets, made in reply to the pre-existing tweets by other users. If user A replies to the tweet made by user B, then a directed edge is formed from node A to node B. In this manner, the network is generated for the 'reply-to' relationships between all the users identified in the scraped data. As observed in Sect. 4.1, Cluster 1 was dominating the Twitter posts related to the state assembly elections, in terms of hashtag usage, during the 1st week of February, and Cluster 2 beats them to a position of dominance during the 2nd week of March, 2021. This trend could be better analyzed with the help of network-level measures for obtaining novel insights about the network actors, and their respective roles. Although our gathered dataset consists of data for all the 6-weeks of study as mentioned in Sect. 3, due to space constraints, we have put up the network-level analysis for only the 1st, 4th, and 6th weeks. As this covers the initial, medial, and final periods of our study (and the elections), the variation in observed network-level and actor-level statistics can be used for identifying the trend in the online dominance of individual clusters. For each observed network, the actors

have been identified as belonging to one of the three groups - Politicians, Media, and Users, which helps to better realize their specific roles in the network.

5.1 Standard network analysis metrics

Some of the standard metrics have been used for in-depth analysis of the network structure for each selected week of Twitter data, such as Giant Component (Janson et al. 1993), Degree Centrality, Betweenness Centrality (Hage and Harary 1995), Page Rank (Page et al. 1999), Eigenvector Centrality (Bonacich 2007), Authority and Hub Centrality (León 2013). The various centrality measures have been selected due to their ability to identify the different key roles played by the actors in the network.

5.2 Analysis of weekly reply networks

5.2.1 Week 1

The network during the first week of February consists of 1048 nodes and only 961 edges, so there is fewer than one edge per node on average. The Graph Density is identified as 0.001, which is a remarkably low value indicating the infrequent connections among actors in the network. The average clustering coefficient is determined as 0.008, with only 9 triadic closures detected in the network. On the other hand, the average degree of the network is calculated as 0.917, meaning a typical actor in this network collaborated with less than one edge in the network. There are 1045 strongly connected. and 181 weakly connected components in the network, so there are a good number of fragments in the network at this stage of the study. These measures may be reflective of the fact that the network of people engaged in election-related tweets is not really well-formed, and thus not well clustered or connected at this stage. Figure 7 illustrates the network structure during this studied period. The nodes are resized and colored based on respective degree. It is noteworthy how some small clusters and numerous individual actors, with one or two connections at most, lie at the fringe of the network, thus validating the observations from standard metrics. The Giant Component of the network includes 46.37% of the actors in the network, suggesting that the majority of the network is distributed in the form of isolated subgraphs. It is also observed that the network seems to be centered around one single actor with the highest degree, with the presence of other actors with relatively lesser degree values. The centrality measures, which have been listed previously, are used to identify the 10 actors who played the different types of important roles during this first week of the study.
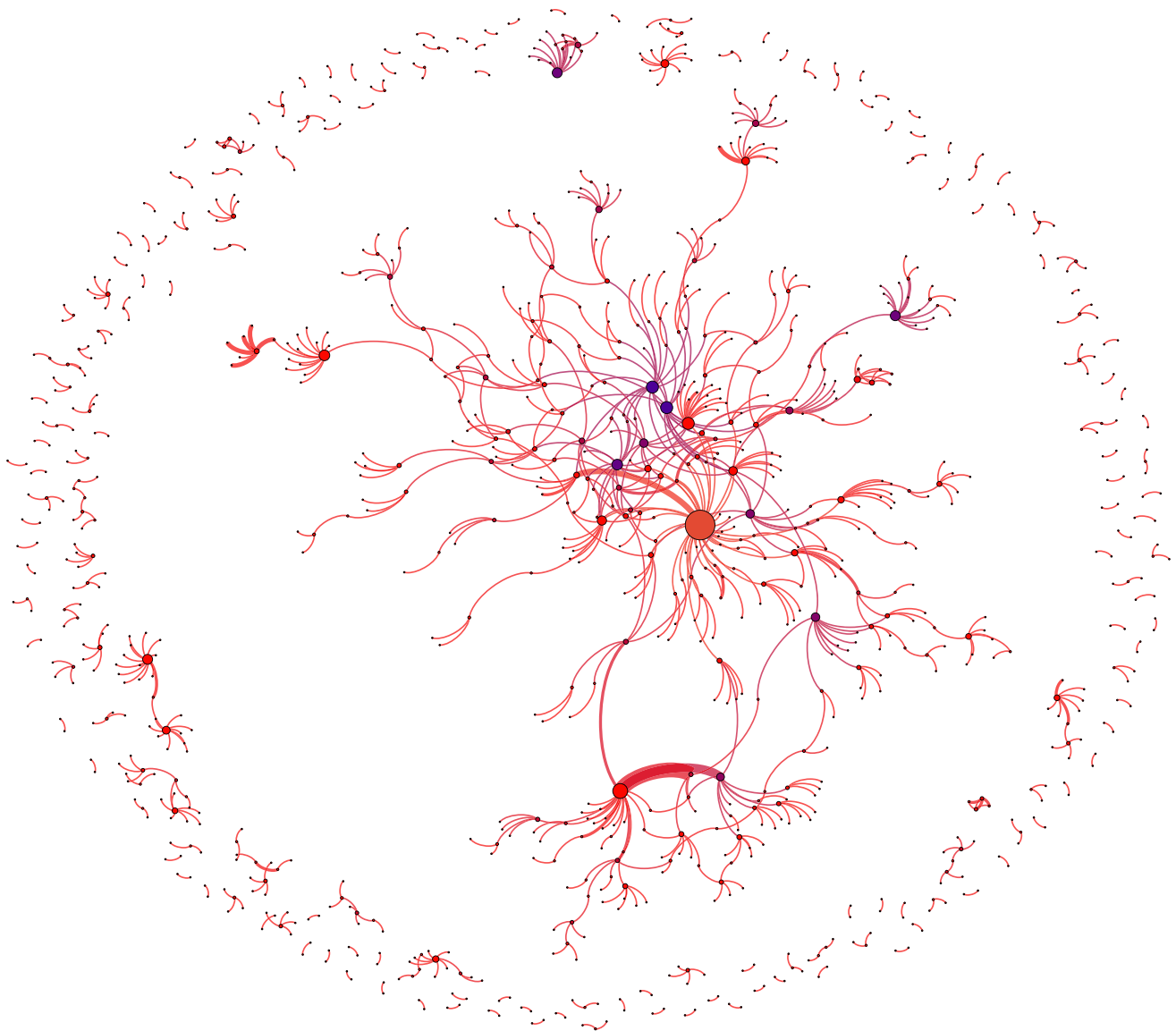
Fig. 7 Illustration of the studied network during week 1

Table 3 Top 10 most connected users and information spreaders during the first week, based on Degree and Betweenness Centrality

Rank	User	Degree centrality value	Type	User	Betweenness Value	Type
1	narendramodi	41	Politician	hatecommie	5	User
2	vk_bhatia2014	20	User	PoliticsMedi	4	User
3	Devipra20614032	16	User	Dharma2X	2	User
4	BJP4India	16	Politician			
5	nsitharaman	16	Politician			
6	SunjayJK	14	User			
7	JPNadda	14	Politician			
8	venkata71976090	13	User			
9	ANI	13	Media			
10	ndtv	13	Media			

Table 4 Top 10 influential users during the first week, based on Pagerank and Eigen Vector Centrality

Rank	User	Page rank value	Type	User	Eigen vector value	Type
1	narendramodi	0.01175	Politician	narendramodi	1	Politician
2	ANI	0.007024	Media	bainjal	0.578839	User
3	ndtv	0.006724	Media	Abhinav53799991	0.530059	User
4	BJP4India	0.004779	Politician	ShahabJafri55	0.530059	User
5	MahuaMoitra	0.004738	Politician	the_hindu	0.530059	Media
6	RahulGandhi	0.004633	Politician	chitrapadhi	0.530059	User
7	PTI_News	0.004378	Media	MRVChennai	0.530059	User
8	Swamy39	0.003793	Politician	apar_chu	0.530059	User
9	nsitharaman	0.003598	Politician	DrlathaC	0.530059	User
10	TimesNow	0.00354	Media	Swamy39	0.460151	Politician

Table 5 Top 10 sources of information and actors with most connections to knowledge centers during the first week, based on Authority and Hub Centrality

Rank	User	Authority value	Type	User	Hub value	Type
1	narendramodi	0.818741	Politician	Devipra20614032	0.271422	User
2	AmitShah	0.251961	Politician	citizen08205897	0.253929	User
3	JPNadda	0.245504	Politician	KoalaTweetin	0.248413	User
4	nsitharaman	0.210185	Politician	A_G_A_S_K_yadav	0.239645	User
5	BJP4India	0.159441	Politician	Harish80635127	0.231257	User
6	KailashOnline	0.156801	Politician	Venganiyer2	0.22533	User
7	smritiirani	0.121262	Politician	ABHAYPKUMAR1	0.215277	User
8	Swamy39	0.09589	Politician	berwal_nikita	0.161385	User
9	BJP4Bengal	0.090632	Politician	KumarTL	0.152915	User
10	TajinderBagga	0.087781	Politician	ProfSudhir	0.152801	User

The results are illustrated in Tables 3, 4, and 5, and are discussed below. Each table contains the top 10 ranking actors, along with their corresponding centrality score and type of actor. The following observations are made from the tables:

- The leaders of Cluster 1 are most well-connected, as seen in the table of Degree Centrality 3, as all four politicians among the top 10 most-connected actors belong to this cluster. Among media or news-based actors, *ANI* and *NDTV* appear to be the most connected. The remaining four actors in the list are general users.
- The Eigenvector and Page Rank centrality measures reveal different pictures. The former reveals that the majority of influential actors are general Twitter users, with only two politicians and one media house on the list. Both the listed politicians belong to Cluster 1. The media actor which made its way to this list is not even among the most connected ones of the same category, but is of influence as per Eigenvector centrality. In contrast, the Page Rank centrality highlights that six out of the ten most influential actors are politicians. Four

of these six politicians correspond to Cluster 1, while Cluster 2 and Cluster 3 have one contender each in this list. Along with the most connected media actors *ANI* and *NDTV*, *PTI News* and *TimesNow* are the most influential in their category. Clearly, as per the Page Rank algorithm, the position of influence is not available to general users even though some of them are widely connected. This difference in the observations based on two centralities, is due to the fact that Page Rank takes into account the in-degree of actors in the directed network, while Eigenvector centrality generally determines influence in undirected ones.
- The Betweenness centrality in Table 3, corresponds to the actors who spread information by acting as the frequent bridges in the shortest paths between nodes. In this network, only three actors are found to have non-zero Betweenness centrality, and they are general users. Remarkably, the media actors in the network do not play as much of an important role in information dissipation as do the general users.

- The top 10 knowledge sources are all actors of Cluster 1, as revealed from the results of Authority centrality in Table 5. No actors of other clusters, or media actors have been able to make it to this list. The Hub centrality toppers are all general users who are the direct connections to the top knowledge sources of the network.

5.2.2 Week 4

During this week, the Twitter network is formed by 1795 actors with 1774 connections among them. The graph density remains at 0.001 similar to the observation in Sect. 5.2.1. Only 5 triadic closures are identified in the whole network with an average clustering coefficient of only 0.002. The average degree is 0.988, a fractional improvement over the value observed for the network from week 1. There is also a relative increase in the number
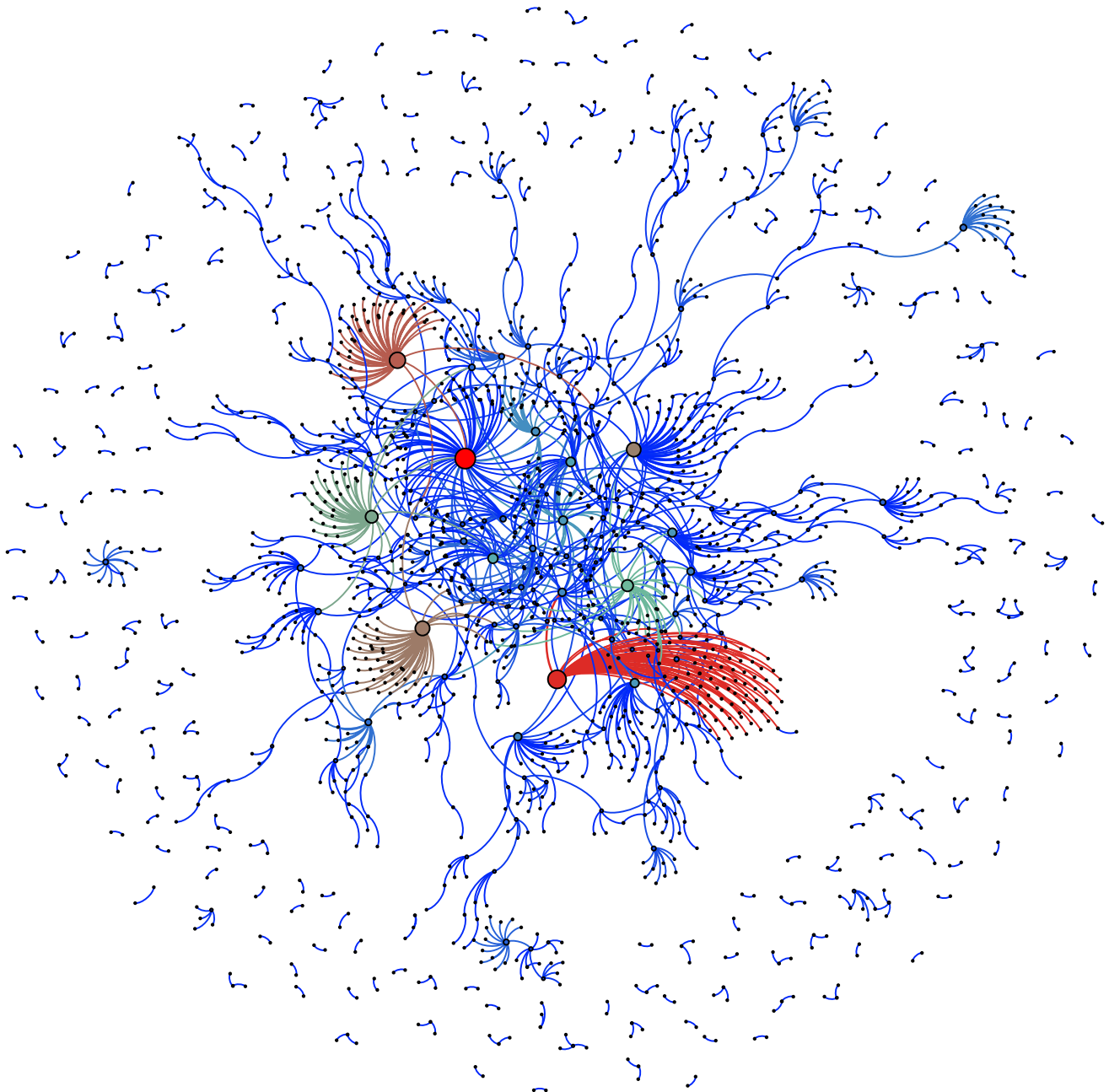

Fig. 8 Illustration of the studied network during week 4

Table 6 Top 10 most connected users and information spreaders during the fourth week, based on Degree and Betweenness Centrality

Rank	User	Degree centrality value	Type	User	Betweenness value	Type
1	iamComutpal	123	User	Baliramyadav007	63	User
2	narendramodi	70	Politician	LinySajan	33	User
3	Baliramyadav007	56	User	Alokmishra416	11	User
4	MamataOfficial	53	Politician	bengaluruboy48	10	User
5	sureshseshadri1	49	User	deep_chatterji	6	User
6	ANI	48	Media	radiantnavin	4	User
7	KolleoIsBack	44	User	rayroman_21	4	User
8	saxenaramesh	37	User	skmandin13	4	User
9	yag4121995	35	User	Skipper_93	4	User
10	AmitShah	33	Politician	MahmudOfGhazni	3	User

Table 7 Top 10 influential users during the fourth week, based on Pagerank and Eigenvector Centrality

Rank	User	Page rank value	Type	User	Eigen vector value	Type
1	narendramodi	0.010789	Politician	narendramodi	1	Politician
2	ANI	0.01068	Media	ANI	0.805565	Media
3	MahuaMoitra	0.00595	Politician	MahuaMoitra	0.441845	Politician
4	derekobrienmp	0.005724	Politician	AmitShah	0.4375	Politician
5	ndtv	0.005315	Media	derekobrienmp	0.417675	Politician
6	MeghUpdates	0.004637	User	BJP4India	0.40625	Politician
7	UnSubtleDesi	0.004416	User	MamataOfficial	0.3708	Politician
8	AmitShah	0.004147	Politician	RahulGandhi	0.36372	Politician
9	Swamy39	0.003947	Politician	ndtv	0.328125	Media
10	BJP4India	0.003656	Politician	MeghUpdates	0.234375	User

Table 8 Top 10 sources of information and actors with most connections to knowledge centers during the fourth week, based on Authority and Hub Centrality

Rank	User	Authority value	Type	User	Hub Value	Type
1	narendramodi	0.787012	Politician	vineetruia	0.267175	User
2	AmitShah	0.296569	Politician	saxenaramesh	0.241704	User
3	BJP4India	0.224993	Politician	Amit69468898	0.21387	User
4	PMOIndia	0.171528	Politician	sureshseshadri1	0.212054	User
5	JPNadda	0.162744	Politician	lokesh981849900	0.19199	User
6	MamataOfficial	0.160615	Politician	SanjibP44593972	0.176488	User
7	ANI	0.102026	Media	sremotes	0.169781	User
8	AITCofficial	0.085674	Politician	Harish80635127	0.149279	User
9	INCIndia	0.081157	Politician	SNCBly	0.149181	User
10	derekobrienmp	0.081002	Politician	saurabh9587	0.143853	User

of strongly connected components which stands at 1795, and smaller growth in the number of weakly connected components as well (224). So, the network structure is more well-formed than in the previous case. The Giant Component includes 67% of the actors or 1197 nodes, which also signifies the better-connected nature of this network. Unlike the previous case, here the network is well distributed with multiple actors of high degree and thus implies that the network dominance has changed at this time. This network is illustrated in Fig. 8, and the top 10 actors for the different network-level statistics, determined using standard measures, are highlighted in Tables 6, 7, and 8, and are discussed below.

Observations:

- The degree centrality measures in Table 6 reveal that during this week, the most connected network actor is a general user, who is preceded by the table leader from Week 1. Remarkably, the official Twitter handle of the leader of Cluster 2, has secured the 4th rank in this table - this same actor was in 86th position in terms of Degree centrality in the 1st week of our study! So, it appears that Cluster 2 has made fruitful efforts to take over the network by this point in time. Also, in contrast to the first week of study, in this week there are only 3 actors of Politician category, out of which two are from Cluster 1, and only one is from Cluster 2. *ANI* is the only media channel that has made its place in the top 10 most connected actors list. The remaining actors in this list are general Users.

- In the list of top 10 influencers based on Pagerank, the top 2 actors are the same as in Week 1 of the study. 2 out of 6 actors (Politicians) of Cluster 2 have made their way to this list too, which also highlights, that at this stage, the network dominance is not held by a single political cluster of actors, unlike week 1. Also, the sole actor from Cluster 3 in the Pagerank list of week 1, has been moved out of the list this week. Apart from the two media actors that have been retained, the two remaining actors in the list are general users, who, despite not being more connected, hold positions of influence. The Eigenvector centrality results also portray a changed situation, with 7 politicians, 2 media, and 1 user among the top 10 actors. Unlike the first week with only 2 politicians of Cluster 1 in the list, Table 7 shows 3 actors of Cluster 1, 3 from Cluster 2, and 1 from Cluster 3. There are 2 media actors and only 1 user, who have also made their way to this list. Unlike the previous case of week 1, here the network influence has clearly been taken over by the politicians, and Cluster 2 seems to have given an almost equal amount of effort to take over the Twitter network. The usage of only in-degree based information in Pag-

erank leads to a different assessment from Eigenvector centrality.

- In contrast to the first week, this week witnesses multiple users with non-zero Betweenness Centrality values, and the top 10 are listed in the table. Irrespective of the political inclination of these users, it is evident that they were motivated enough to have played an active role in information dissipation and connectivity in the network.

- The Authority centrality in Table 8 reveals, how 9 of the top 10 knowledge sources are Politicians, which is relevant, given that the network represents the actual election days. Out of these 9, 4 are from Cluster 1, 3 from Cluster 2, and 1 from Cluster 3, including the official Twitter handle of the Prime Minister of the country. *ANI* is the sole media actor in the list. Thus, Cluster 2 has indeed taken over much of the network dominance, given that the knowledge centers during week 1 were all from Cluster 1. Similar to the situation in week 1, we find that the highest hub centrality is for general users who act as pointers to the knowledge centers.

- The fourth week of our study actually ends on the first day of the state assembly elections, and is representative of the final effort put in by the political actors of different clusters. It is clear how Cluster 2 has made a huge mark in terms of network influence and connectivity just before the state elections commence. Clearly, the dominance of Cluster 1 has been challenged by Cluster 2, while Cluster 3 has not made much significant progress even during this time period. Among media actors, *ANI* has emerged as a robust contender across all the different network-level measures.

5.2.3 Week 6

This is the last week of our study, and is inclusive of 2 separate election days. The network here is more well-formed than in the previous two weeks, with 4935 actors and 5457 inter-connections in terms of replies on Twitter. A total of 50

Table 9 Top 10 most connected users and information spreaders during the sixth week, based on Degree and Betweenness Centrality

Rank	User	Degree centrality value	Type	User	Betweenness value	Type
1	narendramodi	337	Politician	AbhishBanerj	33	User
2	RahulGandhi	140	Politician	vipul2777	31	User
3	PMOIndia	134	User	PriaINC	25	User
4	AmitShah	117	Politician	Shivankkkkk	21	User
5	ANI	101	Media	PritishNandy	20	User
6	ECISVEEP	65	User	RajkumarkohliJ	18	User
7	ndtv	62	Media	taurusplay1	16	User
8	MahuaMoitra	59	Policitician	Tinniz1	16	User
9	BJP4India	55	Politician	PrudhviKrushna	13	User
10	ArvindKejriwal	49	Politician	Roguelost	10	User

Table 10 Top 10 influential users during the sixth week, based on Pagerank

Rank	User	Page rank value	Type	User	Eigen Vector Value	Type
1	narendramodi	0.018879	Politician	narendramodi	1	Politician
2	RahulGandhi	0.009321	Politician	RahulGandhi	0.429807	Politician
3	ANI	0.008524	Media	PMOIndia	0.399833	User
4	AmitShah	0.005741	Politician	AmitShah	0.349211	Politician
5	ndtv	0.005504	Media	ANI	0.31245	Media
6	PMOIndia	0.004669	User	ECISVEEP	0.19883	User
7	MahuaMoitra	0.003892	Politician	ndtv	0.187304	Media
8	PTI_News	0.003264	Media	MahuaMoitra	0.181165	Politician
9	timesofindia	0.002839	Media	BJP4India	0.173991	Politician
10	ECISVEEP	0.002834	User	ArvindKejriwal	0.155942	Politician

Table 11 Top 10 sources of information and actors with most connections to knowledge centers during the sixth week, based on Authority and Hub Centrality

Rank	User	Authority value	Type	User	Hub Value	Type
1	narendramodi	0.875861	Politician	AbidKha80673822	0.108309	User
2	PMOIndia	0.324114	Politician	deja_vu1992	0.108309	User
3	AmitShah	0.2061	Politician	AnshikaAggarwa8	0.108309	User
4	drharshvardhan	0.138417	Politician	jebnezer1	0.108309	User
5	BJP4India	0.125108	Politician	TheReal_Jassi	0.108309	User
6	MoHFW_INDIA	0.093456	Politician	GamingGurus4	0.108309	User
7	RahulGandhi	0.074615	Media	realderwal	0.108309	User
8	PIB_India	0.073373	Politician	Rahulsaha07	0.106869	User
9	JPNadda	0.069611	Politician	d_professional	0.106318	User
10	AyushmanNHA	0.061329	Politician	ajitsin30811844	0.100717	User

triadic closures are identified, in contrast to 5 in the previous week, which highlights the strong structure of the studied network. Similarly, the Average Degree for the network is 1.11, the highest among the 3 networks studied. The number of strongly connected components has also grown massively to 4924, with 556 weakly connected actors in the network. The Giant component includes 69.5% of the actors or 3433 nodes. This network is representative of the heat of the ongoing election, with the obvious involvement of more actors belonging to different categories, expressing themselves mostly in favor of some particular cluster. The network is also well distributed with many actors with important roles to play, as determined using network-level measures. The 10 most important actors corresponding to each measure are listed in Tables 9, 10 and 11. The dense network itself is illustrated in Fig. 9.

Observations:

- A massive difference is noted in the degree centrality-based analysis of the 6th week in comparison with the 4th week. The top 10 actors in terms of degree centrality are listed in Table 9. The leader and face of Cluster 1 has regained the top position in the list, and unlike the 4th week, no general user has made it to the top 10. There are 7 politicians on the list, 4 of which are from Cluster 1, 2 from Cluster 2, and only 1 from Cluster 3. The leader of Cluster 2 has moved down to the 17th position from the 4th position in the previous week of study. In terms of Degree Centrality, the network is dominated by politicians during this time period, with *ANI* and *ndtv* as the only 2 media actors in the list. The official Twitter handle of the *Election Commission of India* has also emerged as a well-connected actor in the present network. There is thus, a notable shift in network-level connections in favor of Cluster 1.

- The results of Pagerank analysis in Table 10, also reveal a similar scenario as our previous observation. Out of 5 politicians in the list, 3 are from Cluster 1, and there is only 1 actor from each of the other two clusters. The leader of Cluster 2 has been pushed back to the 17th position here as well. We observe 4 media actors in the list of influential users, highlighting the role that they have played during the final stages of the election. Last but not the least, the role of the *Election Commission of*
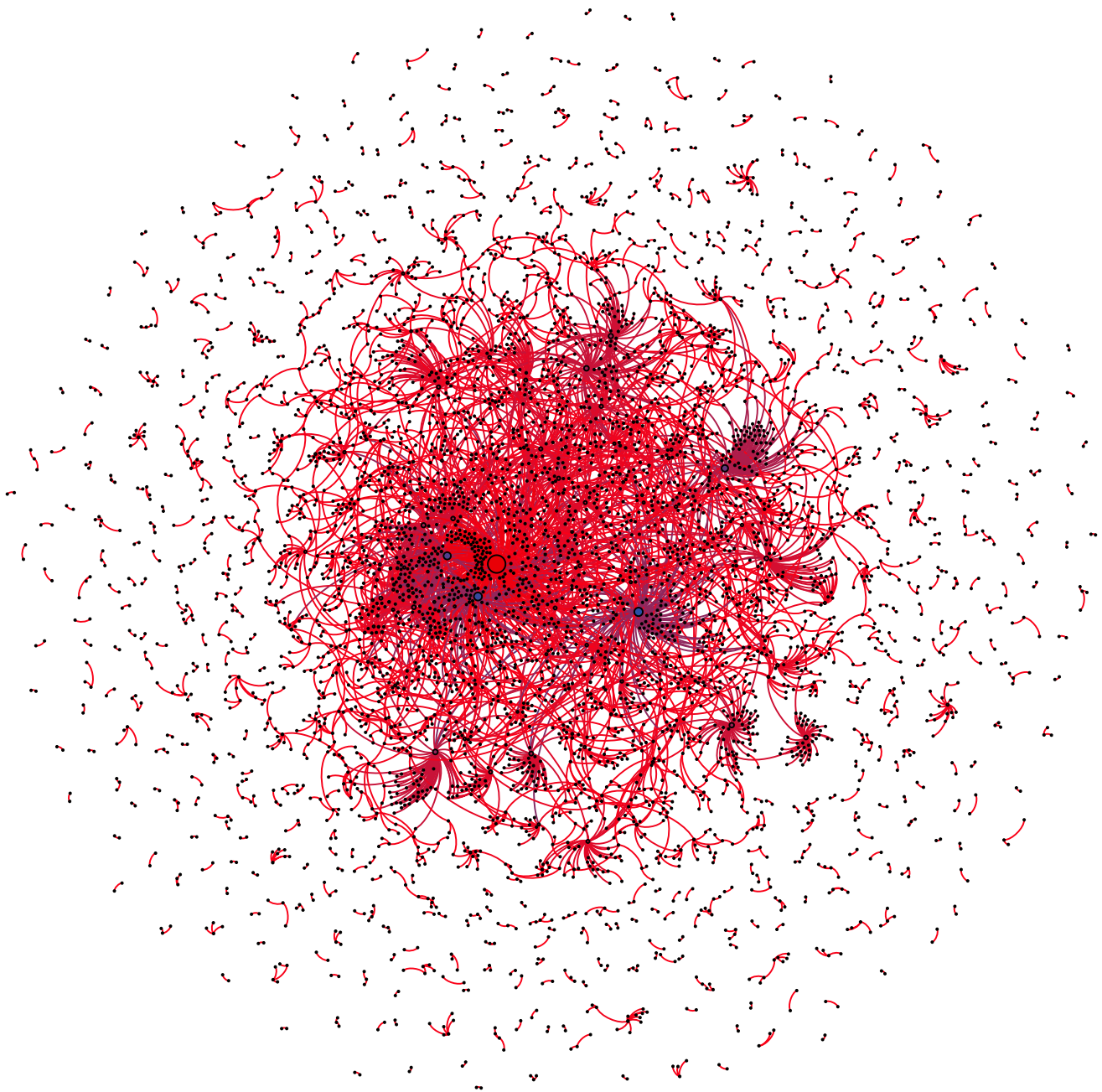
Fig. 9 Illustration of the studied network during week 6

India is also notable in terms of network influence. The consideration of an undirected network using Eigenvector Centrality, reveals an almost similar picture with the 7 most influential actors being politicians. 4 of the 7 politicians are from Cluster 1, 2 from Cluster 2, and only 1 from Cluster 3. The media actors observed in the list of the previous two weeks of study, are also part of this week's most influential actors. In terms of network influence, obviously, Cluster 1 has regained the dominating position while Cluster 2 is in second place. Unlike the previous week, there is no general user in either of the influence-based lists.

- In this week too we find many politically motivated common actors in the network who play a crucial role in the

network acting as bridges for information exchange. No political or media actor is there on this list.

- With only 1 media actor, the Authority centrality list in Table 11 is again dominated by politicians. Of the 9 actors of this category, 8 knowledge sources are from Cluster 1, including the official Twitter handles of the *National Health Authority*, *Ministry of Health*, and the Prime Minister's official Twitter account. Remarkably, the remaining actor in the list is from Cluster 3, and no actor from Cluster 2 is seen here. This is in contrast with the 4th week, when 3 actors from Cluster 2 made their way to the list. The only media actor in the list is the *(PIB)* or the Press Information Bureau, the media agency of the central government of the country, which is ruled by Cluster 1 itself. The top 10 users who act as pointers to the knowledge centers, are also mentioned in the list as per Hub centrality measures, and include no politician or media actor.
- The 6th week of our study corresponds to the final stages of the election, and pertains to the continuous struggle between the different clusters for gaining influence and attention in the network. Clearly, the dominance in terms of connectivity and influence, has been regained by Cluster 1 during this week while Cluster 2 closely follows suit. The inability of Cluster 2 actors who dominated the network in Week 4 to retain their positions is also notable. *ANI* and *ndtv* are the only two media actors who have maintained their presence in the network, as determined by the different measures.

From these observations, it is obvious that the network analysis hints that Cluster 1 can be easily predicted as the winner of the studied election. However, the discussions in Sect. 4 clearly reveal that the contextual information gathered from hashtags provides a true estimation of the election outcome.

6 Conclusion

This paper studies the reply-network on Twitter in reference to a state assembly election in India. The 6-weeks of data gathered for the study reveals that in the days leading up to elections, the tweet frequency was higher on weekends, where during the election period, the tweet frequency was higher on the election days. Identification of hashtag usage by social media users reveals how Cluster 1 started off with complete network dominance, but was slowly pushed back by the Cluster 2 which maintained this dominance all throughout. Cluster 3 has failed to make its mark in the network during the studied period of time. The proposed

cluster dominance technique is capable of forecasting the ground-level situation and influence of the different clusters as reflected in the different opinion and exit polls as well as the actual results. The identification of emotions expressed in cluster-wise tweets suggests, that happiness is the most scarce emotion, while fear is the dominant one. Different network-level measures are utilized to highlight the users who have played a pivotal role in the network during the 1st, 4th, and 6th weeks of study. These important actors have not only been instrumental in the network formation and cluster dominance on Twitter, but also in real life. We witness the network dominance shifting to Cluster 2 in week 4, and back to Cluster 1 in Week 6, using the different standard network analysis techniques. Thus, the current work performs network-level analysis and mining to forecast election results, and obtains novel insights regarding the dynamics of the network, and the role of actors therein. The authors also conclude that the contextual information is a better indicator of the election outcome than general network-level measures. This work can be further extended for election network mining and analysis by considering other sources of social network data, such as Facebook, YouTube, etc., which may offer better insights about the correlation between election outcomes, and human interactions on the social networks regarding the elections.

Appendix A: Bi-gram analysis of tweets

In the initial stages of data acquisition, bi-gram analysis was performed on the data for each week of our study. From the n-gram analysis of relevant weekly tweets, it was obvious that "bengal, elections" was the most frequent set of words used - almost 20 to 30 times more than any other bi-gram. The other n-grams that appeared in the list were either very abstract such as "bengal, assembly" or "election, 2021" which did not specifically refer to the topic of study. So, considering such bi-grams as keywords would result in the acquisition of more irrelevant tweets. Also, some of the relevant bi-grams containing bengali words in english, such as "poriborton, bengal" (poriborton means change in Bengali language), were only used in some hundred or lesser tweets. So, we had decided to go forward with the keywords "Bengal, election" for collecting the data for our work. The Figs. 10 and 11 are representative illustrations of relevant bi-grams corresponding to the first and last weeks of our study that highlight the reason behind our choice of keywords.
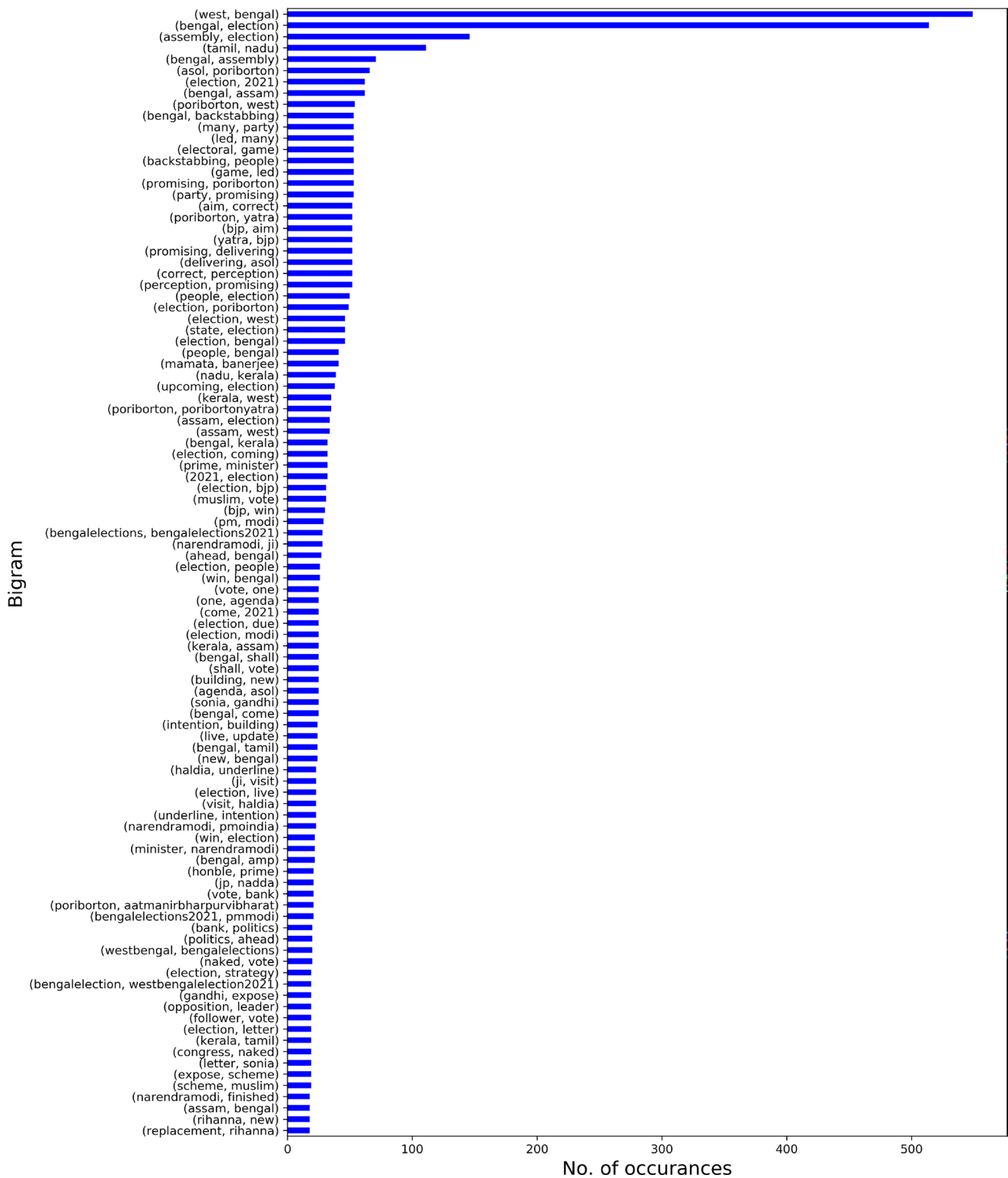
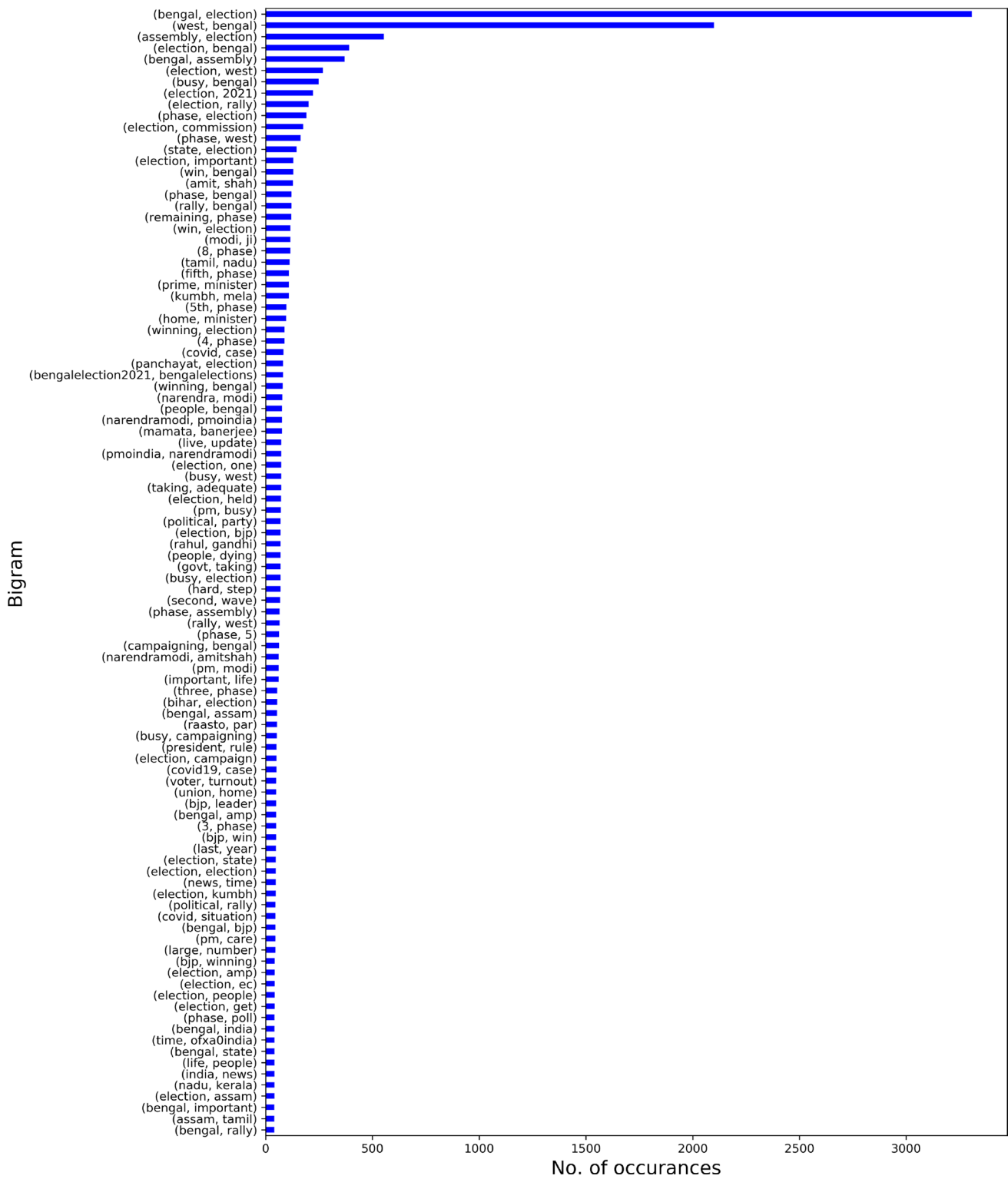**Fig. 10** Illustration of the relevant bi-grams during week 1

Fig. 11 Illustration of the relevant bi-grams during week 6

Declarations

Conflict of interest The authors declare no competing interests.

References

Ahmed S, Cho J, Jaidka K (2017) Leveling the playing field: the use of twitter by politicians during the 2014 indian general election campaign. Telemat Inf 34(7):1377–1386

Banerjee S (2021) In west bengal, bjp claims 41 tmc mlas ready to jump ship; can bring down mamata govt. https://www.republicworld.com/india-news/general-news/in-west-bengal-bjp-claims-41-tmc-mlas-ready-to-jump-ship-can-bring-down-mamata-govt.html, accessed: 2021-08-15

Bonacich P (2007) Some unique properties of eigenvector centrality. Soc Netw 29(4):555–564

Boshmaf Y, Ripeanu M, Beznosov K, Santos-Neto E (2015) Thwarting fake osn accounts by predicting their victims. In: Proceedings of the 8th ACM workshop on artificial intelligence and security, pp 81–89

Bouzy C (2018) Bot Sentinel. https://botsentinel.com/dashboard, [Online; Accessed 22-March-2023]

Chakraborty A, Bose S (2020) Around the world in 60 days: an exploratory study of impact of covid-19 on online global news sentiment. J Comput Soc Sci 3(2):367–400

Chakraborty A, Badyal N, Sharma A, Mukherjee N (2022) A novel centrality-based measure for election network analysis. In: 2022 IEEE silchar subsection conference (SILCON), IEEE, pp 1–6

Chen Y, Wang L (2022) Misleading political advertising fuels incivility online: A social network analysis of 2020 u.s. presidential election campaign video comments on youtube. Computers in Human Behavior 131:107,202. https://doi.org/10.1016/j.chb.2022.107202, https://www.sciencedirect.com/science/article/pii/S0747563222000243

Davis CA, Varol O, Ferrara E, Menczer F (2016) Botornot: a system to evaluate social bots. In: Proceedings of the 25th international conference companion on World Wide Web. International World Wide Web conferences steering committee, Republic and Canton of Geneva, CHE, WWW '16 Companion, pp 273–274, https://doi.org/10.1145/2872518.2889302

Habibi MN, Sunjana (2019) Analysis of indonesia politics polarization before 2019 president election using sentiment analysis and social network analysis. Int J Mod Educ Comput Sci 11(11)

Hage P, Harary F (1995) Eccentricity and centrality in networks. Soc Netw 17(1):57–63

Heredia B, Prusa JD, Khoshgoftaar TM (2018) Social media for polling and predicting united states election outcome. Soc Netw Anal Min 8(1):1–16

Himelboim I, Sweetser KD, Tinkham SF, Cameron K, Danelo M, West K (2016) Valence-based homophily on twitter: network analysis of emotions and political talk in the 2012 presidential election. New Media Soc 18(7):1382–1400

HT (2021) West bengal election 2021 schedule. https://www.hindustantimes.com/elections/west-bengal-assembly-election/schedule, Accessed: 2021-08-15

Imran M, Castillo C, Diaz F, Vieweg S (2015) Processing social media messages in mass emergency: a survey. ACM Comput Surv(CSUR) 47(4):1–38

Jaffrelot C, Verniers G (2020) The bjp's 2019 election campaign: not business as usual. Contemp South Asia 28(2):155–177

Jakesch M, Garimella K, Eckles D, Naaman M (2021) Trend alert: a cross-platform organization manipulated twitter trends in the indian general election. Proc ACM Human-Comput Inter 5(CSCW2):1–19

Janson S, Knuth DE, Łuczak T, Pittel B (1993) The birth of the giant component. Rand Struct Algor 4(3):233–358

Kagan V, Stevens A, Subrahmanian V (2015) Using twitter sentiment to forecast the 2013 pakistani election and the 2014 indian election. IEEE Intell Syst 30(1):2–5

Kim J, Hastak M (2018) Social network analysis: characteristics of online social networks after a disaster. Int J Inf Manag 38(1):86–96

Korolov R, Lu D, Wang J, Zhou G, Bonial C, Voss C, Kaplan L, Wallace W, Han J, Ji H (2016) On predicting social unrest using social media. In: 2016 IEEE/ACM international conference on advances in social networks analysis and mining (ASONAM), IEEE, pp 89–95

Kowsari K, Jafari Meimandi K, Heidarysafa M, Mendu S, Barnes L, Brown D (2019) Text classification algorithms: a survey. Information 10(4):150

Kudugunta S, Ferrara E (2018) Deep neural networks for bot detection. Inf Sci 467:312–322

León C (2013) Authority centrality and hub centrality as metrics of systemic importance of financial market infrastructures. Borradores de Economía; No 754

Lundberg J, Nordqvist J, Laitinen M (2019) Towards a language independent twitter bot detector. In: DHN, pp 308–319

Naseem U, Razzak I, Khushi M, Eklund PW, Kim J (2021) Covidsenti: a large-scale benchmark twitter data set for covid-19 sentiment analysis. IEEE Trans Comput Soc Syst

Oyebode O, Orji R (2019) Social media and sentiment analysis: The nigeria presidential election 2019. In: 2019 IEEE 10th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), pp 0140–0146, https://doi.org/10.1109/IEMCON.2019.8936139

Page L, Brin S (1999) The pagerank citation ranking: Bringing order to the web. Tech. rep, Stanford InfoLab

Paskarina C, Nuraeni RH (2021) Politics of hashtags: Social network analysis of online contestation in the 2019 indonesia presidential election. RIVISTA DI STUDI SULLA SOSTENIBILITA'

Perliger A, Pedahzur A (2011) Social network analysis in the study of terrorism and political violence. PS: Political Sci Polits 44(1):45–50

Plotkowiak T, Stanoevska-Slabeva K (2013) German politicians and their twitter networks in the bundestag election 2009. First Monday

Preotiuc-Pietro D, Gaman M, Aletras N (2019) Automatically identifying complaints in social media. arXiv preprint arXiv:1906.03890

Rakhmawati NA, Mufidah K (2020) Social network analysis of legislative candidates in indonesia general election 2019 using community detection. In: 2020 3rd International Conference on Computer and Informatics Engineering (IC2IE), pp 306–310, https://doi.org/10.1109/IC2IE50715.2020.9274669

Ramalingam D, Chinnaiah V (2018) Fake profile detection techniques in large-scale online social networks: a comprehensive review. Comput Elect Eng 65:165–177

Renaud M, Korolov R, Mendonça D, Wallace W (2019) Social network structure as a predictor of social behavior: the case of protest in the 2016 us presidential election. In: Chertov O, Mylovanov T, Kondratenko Y et al (eds) Recent developments in data science and intelligent analysis of information. Springer International Publishing, Cham, pp 267–278

Roy PK, Chahar S (2020) Fake profile detection on social networking websites: a comprehensive review. IEEE Trans Artif Intell 1(3):271–285

Shin J, Jian L, Driscoll K, Bar F (2017) Political rumoring on twitter during the 2012 us presidential election: Rumor diffusion and correction. New Media Soc 19(8):1214–1235

Srivastava R, Kumar H, Bhatia MS, Jain S (2015) Analyzing delhi assembly election 2015 using textual content of social network. Proc Sixth Int Conf Comput Commun Technol 2015:78–85

Tabatabaei SA, Asadpour M (2014) Study of influential trends, communities, and websites on the post-election events of iranian presidential election in twitter. In: Social network analysis-community detection and evolution. Springer, p 71–87

text2emotion Team (2020) text2emotion. https://pypi.org/project/text2emotion/#description, accessed: 2021-08-15

Wanda P, Jie HJ (2020) Deepprofile: Finding fake profile in online social network using dynamic cnn. J Inf Secur Appl 52(102):465

Wegrzyn-Wolska K, Bougueroua L (2012) Tweets mining for french presidential election. In: 2012 Fourth International Conference on Computational Aspects of Social Networks (CASoN), pp 138–143, https://doi.org/10.1109/CASoN.2012.6412392

Wikipedia (2021) 2021 west bengal legislative assembly election. https://en.wikipedia.org/wiki/2021_West_Bengal_Legislative_Assembly_election, accessed: 2021-12-06

Yang KC, Ferrara E, Menczer F (2022) Botometer 101: Social bot practicum for computational social scientists. J Computat Soc Sci pp 1–18