

Research Article

Predicting Politician's Supporters' Network on Twitter Using Social Network Analysis and Semantic Analysis

Asif Khan ¹, Huaping Zhang ¹, Jianyun Shang,¹ Nada Boudjellal ¹,
Arshad Ahmad ², Asmat Ali,^{1,3} and Lin Dai¹

¹School of Computer Science and Technology, Beijing Institute of Technology, Beijing, 100081, China

²Department of Computer Science, University of Swabi, Anbar 25000, Swabi, Pakistan

³Department of Computer Science, University of Peshawar, Peshawar 25000, Pakistan

Correspondence should be addressed to Huaping Zhang; kevinzhang@bit.edu.cn

Received 16 December 2019; Accepted 1 August 2020; Published 1 September 2020

Academic Editor: Muhammad Sajjad

Copyright © 2020 Asif Khan et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Politics is one of the hottest and most commonly mentioned and viewed topics on social media networks nowadays. Micro-blogging platforms like Twitter and Weibo are widely used by many politicians who have a huge number of followers and supporters on those platforms. It is essential to study the supporters' network of political leaders because it can help in decision making when predicting their political futures. This study focuses on the supporters' network of three famous political leaders of Pakistan, namely, Imran Khan (IK), Maryam Nawaz Sharif (MNS), and Bilawal Bhutto Zardari (BBZ). This is done using social network analysis and semantic analysis. The proposed method (1) detects and removes fake supporter(s), (2) mines communities in the politicians' social network(s), (3) investigates the supporters' reply network for conversations between supporters about each leader, and, finally, (4) analyses the retweet network for information diffusion of each political leader. Furthermore, sentiment analysis of the supporters of politicians is done using machine learning techniques, which ultimately predicted and revealed the strongest supporter network(s) among the three political leaders. Analysis of this data reveals that as of October 2017 (1) IK was the most renowned of the three politicians and had the strongest supporter's community while using Twitter in a very controlled manner, (2) BBZ had the weakest supporters' network on Twitter, and (3) the supporters of the political leaders in Pakistan are flexible on Twitter, communicating with each other, and that any group of supporters has a low level of isolation.

1. Introduction

Social media technologies have been utilized effectively in various domains, for instance, software engineering [1–5]. Likewise, social media networks are preemptive platforms for politicians, political parties, political institutions, and foundations to connect to citizens. Hundreds of thousands of political parties and politicians are using social media nowadays. They can get the attention of people easily and rapidly when compared to other traditional ways of communication. Every politician and political party has millions of followers on these social media networks, and the politicians try interesting and advanced ways to encourage the citizens to engage in politics. Besides this, social media helps the politicians in various decision making processes by indicating recommendations, for instance, devising policies/

strategies based on the past experiences, recommending and selecting suitable candidates for a particular constituency, recommending a suitable person for a particular position in the party, and launching a political campaign according to the sentiments of citizens on different issues and controversies among others [6–15].

Many researchers have analyzed Twitter in different ways: the importance of Twitter in journalism and political news coverage [6], the use and importance of Twitter in medical education [16, 17], prediction in e-commerce using sentiment analysis [18], sentiments of people towards terrorist events [19], privacy and security issues [20], use of Twitter by the government to engage citizens [21, 22], manipulation, and misinformation propagation during elections [23, 24], among numerous others. Analyzing politics on social media is a hot topic in the past few years.

Twitter is considered to be on top as it is commonly used among politicians. The style of Donald Trump's tweets and its change over time has been analyzed [25]. The work of Tromble in [26] has studied politician-citizen engagement. In political prediction, researchers mostly analyzed Twitter for political campaigns and elections predictions. Election prediction using sentiment analysis [27, 28], social network analysis [29], topic modeling and sentiment analysis [30], and aggregating online and offline data [31] were studied, but all these studies used election time data. To the best of our knowledge, no research had drawn attention to the supporters' network of political leaders in normal time (long before the election), which motivated us to investigate specifically the supporters' network of political leaders of Pakistan on Twitter in normal time that can lead to the prediction of a political leader's fate in the election.

Every politician has supporters and adversaries that can change their futures. These supporters can help a politician to be more influential and win an election. A question still lies there, if the followers of a politician are supporters or adversaries? This study focuses on the supporters' network of three famous political leaders of Pakistan, namely, Imran Khan (IK), Maryam Nawaz Sharif (MNS), and Bilawal Bhutto Zardari (BBZ). This study is mainly based on the following research questions:

RQ1: which political leader has the strongest supporter community based on the analysis of network structure and community detection techniques?

RQ2: which political leader has the strongest supporters' community based on analyzing the content of supporters on Twitter?

To address the aforementioned research questions, this study has utilized multicriteria decision making and machine learning algorithms by (1) studying the politician's supporter network using social network analysis techniques and (2) analyzing the politician's supporter networking using semantic analysis.

Initially, fake supporters were identified and removed. All the tweets, comments, and retweets of the identified fake supporters were discarded. The approach used in the first part predicts the strongest supporters' network by analyzing network structure and running community detection on the supporters of three political leaders on Twitter. Additionally, the supporters' reply network and retweet network of each political leader have been investigated in this study. The retweet behavior of supporters shows us the information diffusion among the supporters, and the replying behavior shows the communication between supporters of the leader(s).

The second approach predicts the strongest supporters' network by analyzing the sentiments of supporters of every political leader using supervised machine learning techniques such as Support Vector Machines (SVM), Naïve Bayesian (NB), and K-Nearest Neighbour (KNN). These sentiments are classified into Positive, Neutral, and Negative. Next, a word cloud is built for every political leader from the content used by supporters.

The main contributions of this study are

- (1) identifying and removing the fake supporters in a supporter's network of political leaders
- (2) identifying the existing supporters' communities of political leaders on Twitter
- (3) determining the strongest supporter's network of the three political leaders using network analysis
- (4) determining and categorizing the strongest supporter's network of the three political leaders using semantic analysis.

This study can particularly assist in understanding the supporters of the political leaders of Pakistan. In general, it helps political leaders to understand their supporters in a better way; they can make decisions based on the supporter communities' sentiments, information diffusion patterns, and communication patterns. Besides this, the political leaders can adapt their party agenda and slogans according to the sentiments and needs of the people, which will ultimately assist in gaining more support in their political campaigns. Nevertheless, this study can also help other countries in policymaking while enhancing bilateral relations.

The rest of this paper is organized as follows: Section 2 provides all the related work to this study; the research methodology is discussed in Section 3; the results and discussion are in Section 4; and, finally, Section 5 provides the conclusions and the possible future work of this study.

2. Related Work

Many researchers have analyzed different political leaders and parties on twitter. Verweij [32] analyzed the Twitter network of Dutch journalists and politicians using some basic network metrics like in-degree and out-degree and they concluded that the subgroups do not influence the contact between political leaders and journalists. In 2015, Dokoohaki et al. [29] implemented link analysis techniques to predict Swedish elections. The data of the political party was collected during election days. They presented evidence that link analysis can play an important role in prediction. The work of [33, 34] investigated the aspect of the Greek MPs Twitter network. They studied MPs (members of parliament) who belonged to the two major parties of Greek. They also studied the common followers and unique followers of MPs.

In [35], the authors studied the political alignment during parliamentary election 2015 in Venezuela using social network analysis and unsupervised machine learning methods. The data was taken a week before elections. Japan's 48th general election for the lower house has been analyzed in [36]; they investigated the ruling party and opposition party's retweet network and found that they are almost similar. The works of [37, 38] predicted the 2017 French elections using term weight and sentiment analysis. In 2016, Xie et al. [31] studied the three leaders of Taiwan during the election, and they predicted the popularity of leader based on the number of mentioning a leader, proportion of Amazon

web data of politicians, daily average Facebook likes, and the ratio of search indexes.

Furthermore, in 2018 in [39], they predicted Taiwan's election based on politician's popularity. They predicted popularity based on Facebook likes count, comment count, and event detection. They used the data during the election campaign.

This study focuses on the political leaders of Pakistan on Twitter. Some researchers have work on the politics of Pakistan on Twitter, but their works were limited to the prediction of elections using the data collected during election campaign time. The work of [27, 40] analyzed the 2013 General Election of Pakistan, and they have analyzed the four significant parties, Pakistan Tehreek-i-Insaf (PTI), Muttahida Qauomi Movement (MQM), and Pakistan People's Party using sentiment analysis. Furthermore, in the work of [41], they investigated political parties on Twitter and predicted the 2018 general election of Pakistan using sentiment analysis. They trained a deep neural network on 2013 elections Twitter data and tested on 2018 Twitter data. Aragón et al. [42] analyzed communication dynamics by investigating reply and retweet network of Spanish leaders during the 2011 election campaign, but their work did not study the supporters' reply and retweet networks.

This study is different from the work cited above. This study investigates the support of the three political leaders of Pakistan (MNS, IK, and BBZ). The supporters' reply and retweet networks of the three leaders have been studied, and the strongest support network of the political leader has been predicted using social network analysis and semantic analysis, sentiment analysis using machine learning techniques and word cloud.

3. Supporters' Network of Political Leaders

This section discusses the methodology used in this study. Due to the modular nature of the research questions of this study, the research methodology of each research question is explained separately in the following subsections.

3.1. Data Extraction and Finding Political Leaders' Demographics. A new application on Twitter [43] was created in order to obtain access tokens (consumer_key, consumer_secret, access_token, and access_secret) for the Twitter Search API. The tweets were collected using a python library "Tweepy" [44]. It is an open-source library for python, which facilitates access to the Twitter API [45]. It gives access to all objects as well as methods of the official Twitter API. Figure 1 shows the process of tweets collection. The data (Tweets) were in the form of JSON (JavaScript Object Notation), which is in a simple and lightweight data-interchange format. Humans can use (read and write) it quickly because it consists of key/value pairs, and it is easy for a machine to generate and parse it. The downloaded JSON file of tweets contained enormous amounts of data, such as user details, texts, retweets, replies, mentions, links, hashtags, and locations. The data was then parsed to the MySQL database while ignoring irrelevant information. Tweets that contained @ImraKhanPTI, #ImraKhanPTI

@MaryamNSharif, #MaryamNSharif, @BBhuttoZardar, and #BBhuttoZardar were collected. The tweets were between 14th September 2017 to 13th October 2017. The total number of supporters was 159,683, and the number of Tweets was approximately 456,990.

Table 1 shows the Twitter profile details of each political leader of Pakistan, which is considered in this research.

Table 1 shows that IK has more followers on Twitter than the other three leaders. MNS uses Twitter more than the other two leaders. IK uses Twitter very precisely as he only has 6.2 K tweets and follows just 18 people.

3.2. Fake Supporters Detection. Fake supporters in the politician's supporters' network were detected and removed. Every day, hundreds and thousands of people are registering into the social media network. Apart from legitimate users, many fake people get register into those websites too. Those fake people always tend to be real, they often spam the real/legitimate users, and they post illegal and inappropriate content. A fake user/supporter in a supporter network of a political leader can misrepresent the ideology of a political party, or they can create some fake content to create hate for other communities (supporters of other politicians) by posting inappropriate contents about other politicians or groups. There are some people (or bots) in a political network that have a special objective to manipulate the opinions of voters by promoting and spreading a specific agenda:

- (1) Supporting and overcoming a weak or negative reputation of a political leader
- (2) harming the reputation of a political leader (opponent)

In this study, a supporter is considered to be a fake supporter if he/she tweeted about a political leader but is not following him. The list of followers of political leaders was scrapped [46] and stored in MySQL; the supporters' list was matched with the followers' list of leaders, and unmatched supporters were considered fake and were removed along with their tweets. 82 users in the IK network, 207 in MNS network, and 25 in BBZ's network were identified and removed. Furthermore, fake supporters in the political network were detected and removed using (i) a traditional method, i.e., ratio of tweets, and (ii) using the ML technique. The first method uses supporters' metadata and basic activity feature: (a) friends count, (b) followers count, (c) statuses count, and (d) "verified," which is a binary field that shows whether the user account is verified by Twitter or not. In this research, a supporter is considered to be fake who has less than or equal to 10 followers and friends and more than 100 tweets. This method identified 18 fake supporters in IK's network and 11 in MNS's network.

The method using the ratio of tweets did not give the desired results. Similar to [24, 47–50], it is believed that there might be many bots or supporters in the network of the three political leaders that behave like a bot by copying, sharing, and retweeting content to distort and manipulate the real picture of a leader.

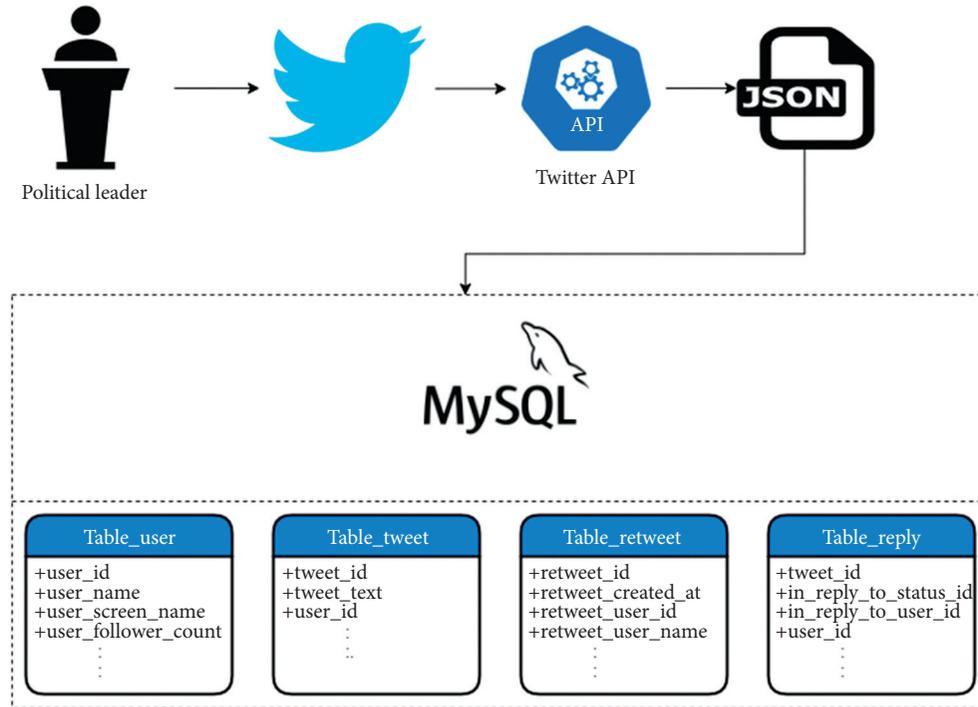


FIGURE 1: Data collection process.

TABLE 1: Twitter profile of the political leaders.

Name	Party name	Followers (millions)	Following (K)	User name	Tweets (since joining) (K)	Joined since
Imran Khan	Pakistan Tehreek-i-Insaf (PTI)	10.7	18	@ImranKhanPTI	6.2	March 2010
Maryam Nawaz Sharif	Pakistan Muslim League Noon (PMLN)	5.4	9.7	@MaryamNSharif	62	January 2012
Bilawal Bhutto Zardari	Pakistan People's Party (PPP)	3.6	1.9	@BBhuttoZardari	11.4	July 2011

The second method detects fake supporters using Botometer [51]. Botometer is publicly available and has an API. It detects fake users (bots) using machine learning techniques and has achieved a high accuracy (0.95). It classifies a user account into fake or real by using more than 1000 features of a user's metadata and information obtained by the content and network structure. Botometer grouped all those features, mainly in six classes: Network Features, Content Features, Sentiment Features, Friends Features, Temporal Features, and User Features. It gives output between 0 and 1. In this study, a supporter's account is considered to be a "fake" if the output score is 0.43 or higher. This technique identified about 36% of the total supporters of IK, 30.3% of MNS, and 20.7% in BBZ's to be fake. Those supporters and their tweets were removed. Finally, the data contains 54,058 supporters in IK's network, 36,942 in MNS's network, and 13,386 in BBZ's network.

3.3. Network Analysis of Supporters' Network. A community is a collection of individuals as a cluster to such an extent that there is high relatedness among individuals inside the community. In other words, a collection of people who tends

to have common interests, goals, habits, likes, dislikes, and choices leads to making virtual communities/groups and clusters in a social network. Detecting such clusters and communities in social media is an essential task in many aspects and applications of life, such as science, sociology, psychology, computer science, and marketing. The association of individuals into groups gives critical experiences into aggregate elements that can be utilized for information dissemination, online web marketing, terrorism control, medical fields, politics, sports, culture, clubs, and many more.

The virtual groups inside online networking content are the direct impact of the homophile guideline, and the influence these groups have over their individuals is the effect of influence. Structure-based Social Network Analysis (SNA) denotes connections between individuals in the form of a graph, and it uses graph-based algorithms to mine communities or subgroups [52]. Such structures of graphs can be of incredible significance for understanding how data engenders and spreads in a network of individuals.

To address RQ1, the experiment has been set up into two parts: supporters' reply network and supporters' retweet network.

3.3.1. *Reply Network Analysis of Supporters.* To analyze communication patterns between supporters of a political leader and predict the strongest community, the authors studied the reply networks of each political leader. Supporters who replied to political leaders in their tweets are grouped into one network.

$$G^{\text{rep}} = (V^{\text{rep}}, A^{\text{rep}}), \quad (1)$$

- (i) where V^{rep} denotes supporters, who reply to the political leaders in the network
- (ii) A^{rep} is the set of directed links of reply between supporters and the leaders

Equation (1) creates a retweet network. Self-loop, i.e., $v_i^{\text{rep}} \rightarrow v_j^{\text{rep}}$, is ignored, where v_i is a person in the network, who replies to his tweet and then analyzes the supporter's reply network of each leader and calculates the average degree, average weighted degree, network diameter, average path length, and nodes connectivity in the network.

Then, the Louvain method [53] is applied to mine communities in the reply network of each political leader, analyzing those communities, supporters existing in those communities, and analyzing whether the supporters directly took an interest in the leader by replying to him/her or through some other discussion by another group of people.

3.3.2. *Retweet Network Analysis of Supporters.* The supporters' retweet network of each political leader is studied. It contains those supporters who retweeted the leader or the leader retweeted someone else.

$$G^{\text{ret}} = (V^{\text{ret}}, A^{\text{ret}}), \quad (2)$$

- (i) where V^{ret} denotes the supporters who contributed to the leaders' support by retweeting him/her in the network and
- (ii) A^{ret} is the set of directed links of retweet links, i.e., V_i retweet V_j

Create a retweet network using (2). Self-loop here too, i.e., $v_i^{\text{ret}} \rightarrow v_j^{\text{ret}}$, is ignored. Then, characteristics of the supporters' retweet network for a single leader were calculated, i.e., average degree, average weighted degree, network diameter, average path length, and nodes connectivity in the network.

Then, the authors used the Louvain method [53] to extract communities in a retweet network of each political leader and analyze those communities and study whether the supporters directly took an interest in the leader by retweeting him/her or by some other group of people or leader.

To validate the strongest supporter (reply and retweet) network using social network analysis techniques, this study combines the network of the three political leaders using

$$G = (G_{\text{MNS}}(V, E) + G_{\text{IK}}(V, E) + G_{\text{BBZ}}(V, E)), \quad (3)$$

- (i) where $G_{\text{MNS}}(V, E)$ denotes supporters' (reply or retweet) network of MNS
- (ii) $G_{\text{IK}}(V, E)$ denotes supporters' (reply or retweet) network of IK
- (iii) $G_{\text{BBZ}}(V, E)$ denotes supporters' (reply or retweet) network of BBZ
- (iv) V denotes supporters
- (v) E denotes the links

Next, communities were extracted. The PageRank of each political leader in that network was calculated. The larger the size of the community of a leader and the higher the PageRank of a leader are, the more influential the leader is and the more support he/she has.

3.4. *Semantic Analysis of Supporters' Social Network.* RQ2 analyzes the tweets of supporters of every political leader by

- (i) sentiment analysis and
- (ii) semantic word cloud

Sentiment analysis, also known as opinion mining or emotional AI, refers to computational linguistics, Natural Language Processing, and text analysis to analytically extract, identify, and analyze affective states and subject information. People express their views and opinions on different topics over social network sites. Many government and private organizations need to know about the views, opinions, thoughts, feelings, behavior, and attitude of the public. This can help them to make new policies that can help them to get better output. Many people take an interest in politics, and they discuss politics. They try to take part in the discussion related to a political party or political leader. Citizens express their viewpoints about a leader or the campaign he/she is running. It is imperative to know what people think about a political leader.

As discussed earlier, fake supporters and their tweets were removed in this study. As a result, our data contained 213,080 tweets, 92,415 tweets from the supporters of IK, 86,721 tweets from the supporters of MNS, and 33,944 tweets from the supporters of BBZ. This study considered all those tweets written in the English language only. As a result, the final data contained 140,508 tweets in total, 73,846 tweets regarding IK, 45,197 tweets regarding MNS, and 21,465 regarding BBZ.

The extracted data from Twitter contains several syntactic features, which are not useful for this study. First of all, the tweets were preprocessed, as data filtering and cleaning techniques are required to remove irrelevant information. Many preprocessing steps are involved in this process: removing URLs, filtering, removing questions, removing special characters, removing stop-words, and removing emoticons.

Sentiment analysis can be achieved mainly using three approaches, lexicon based approach, machine learning based approach, and hybrid approach. This study used supervised machine learning techniques to predict the sentiments of supporters of political leaders. The machine learning based

TABLE 2: Supporters' reply network of political leaders.

Parameters	Imran Khan	Maryam Nawaz Sharif	Bilawal Bhutto Zardari
Nodes	21249	27933	5146
Edges	24753	31594	7294
Network diameter	5	17	13
Average path length	2.359	4.188	2.666
Number of weakly connected components	119	1977	217
Number of strongly connected components	2105	19138	6985

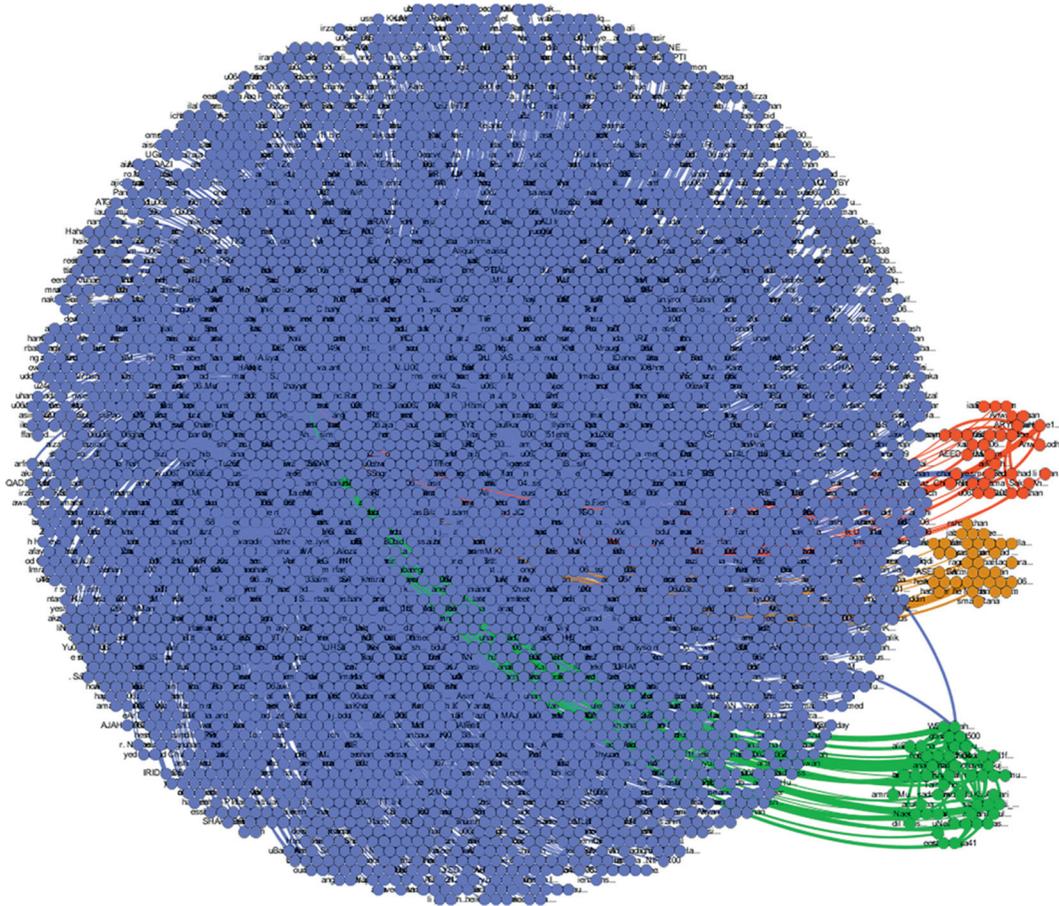


FIGURE 2: Communities in IK's supporter reply network. The blue colour community contains supporters of IK, green colour contains PTI's media group, orange colour contains BBZ, and red contains MNS.

approaches need labelled data to train and test the model. There are numerous labelled datasets available online; however, that did not fit the dataset used in this study. The authors manually labelled 21,000 tweets, which contained 7000 positive, 7000 negative, and 7000 neutral ones. In this study, SVM [54], NB [55], and KNN [56] have been considered. K -fold Cross-Validation was used to ensure that the dataset will not fall in overfitting during training and testing. $K = 10$ was taken in Cross-Validation.

At last, the best classifier was chosen for further use on unseen tweets. If the sentiments of people about a political leader are positive, the prediction about him/her will have a high probability of winning. By using sentimental analysis, a leader can know what his people think about him, what they

want from him, and how they react to various things. Such information and analyses can help the politicians in decision making and refining their policies for the well-being of the public and their interest.

Furthermore, semantic word cloud of supporters' content has been generated. It maps the cooccurrence and frequency of different terms that appear in the tweets of supporters. Word clouds are used in the representation of text visually. It adds clarity during the analysis of the texts. It also divulges patterns in the data that can guide the future investigation. Word frequencies can be spotted easily. All the tweets were first preprocessed. This can be used by researchers, political leaders, and political parties to understand the sentiments, topics, and discussion used by supporters.

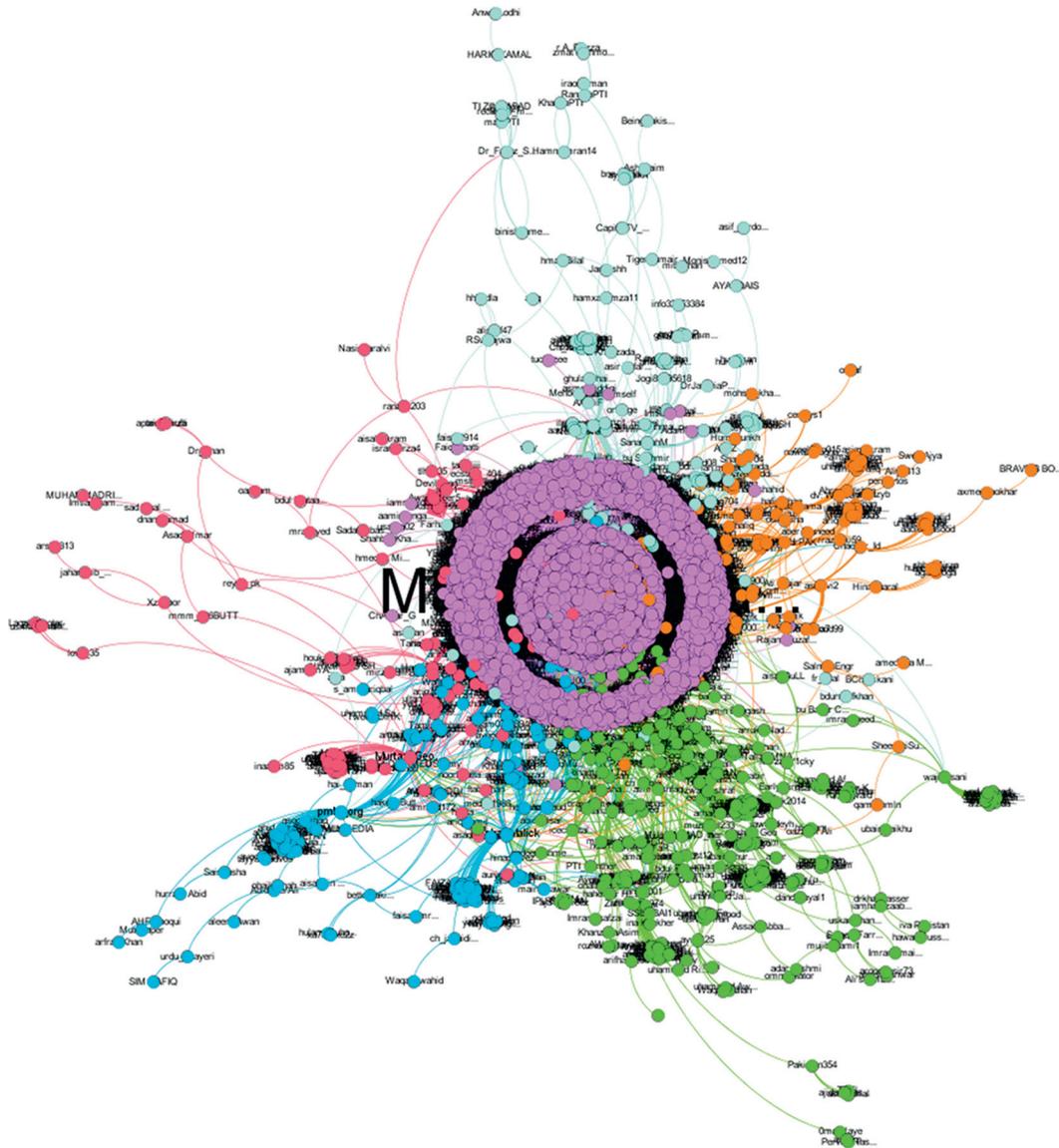


FIGURE 3: Communities in MNS’s supporters’ reply network. The lilac coloured community contains supporters of MNS, green colour contains IK, cerulean colour contains PMLN’s leader, arctic colour contains BBZ, and orange colour contains PMLN’s media cell.

4. Results and Discussion

To answer RQ1 and RQ2, the results of experiments are discussed mainly in two parts, social network analysis on supporters’ network and semantic analysis of supporters’ network.

4.1. Social Network Analysis on Supporters’ Network. To answer RQ1, the following experiments are conducted to mine communities and predict the strongest support of a political leader.

4.1.1. Reply Network. Reply network of each political leader has been constructed, and Table 2 contains the properties of the supporters’ reply network of each political leader.

The community detection method is applied to extract the communities for the supporter’s reply network of IK. Figure 2 shows the three significant subcommunities in this

network. This study shows a 60% view of final figures for clarity using Gephi [57]; the same method is followed for the rest of the experiments. The extracted communities in Figure 2 contain 86.26% of nodes and 87.37% of edges of the entire network. The other nodes in 13.74% consisting of insignificant communities containing few nodes were ignored in this analysis. The giant community that contains IK himself is 82.93%. The other three communities are having sizes 1.36%, 1.19%, and 0.79% of the total network. On manual study, it was found that the community with 1.36% relates to PTI’s media group, orange to BBZ and red to MNS. Reply links between the community of IK and the other two are visible.

The communities were extracted in MNS supporter’s reply network by implementing a community detection algorithm. Figure 3 shows six major communities, which consist of 78.33% of the whole network. The other 21.67% of the network consists of 438 insignificant communities. Those communities

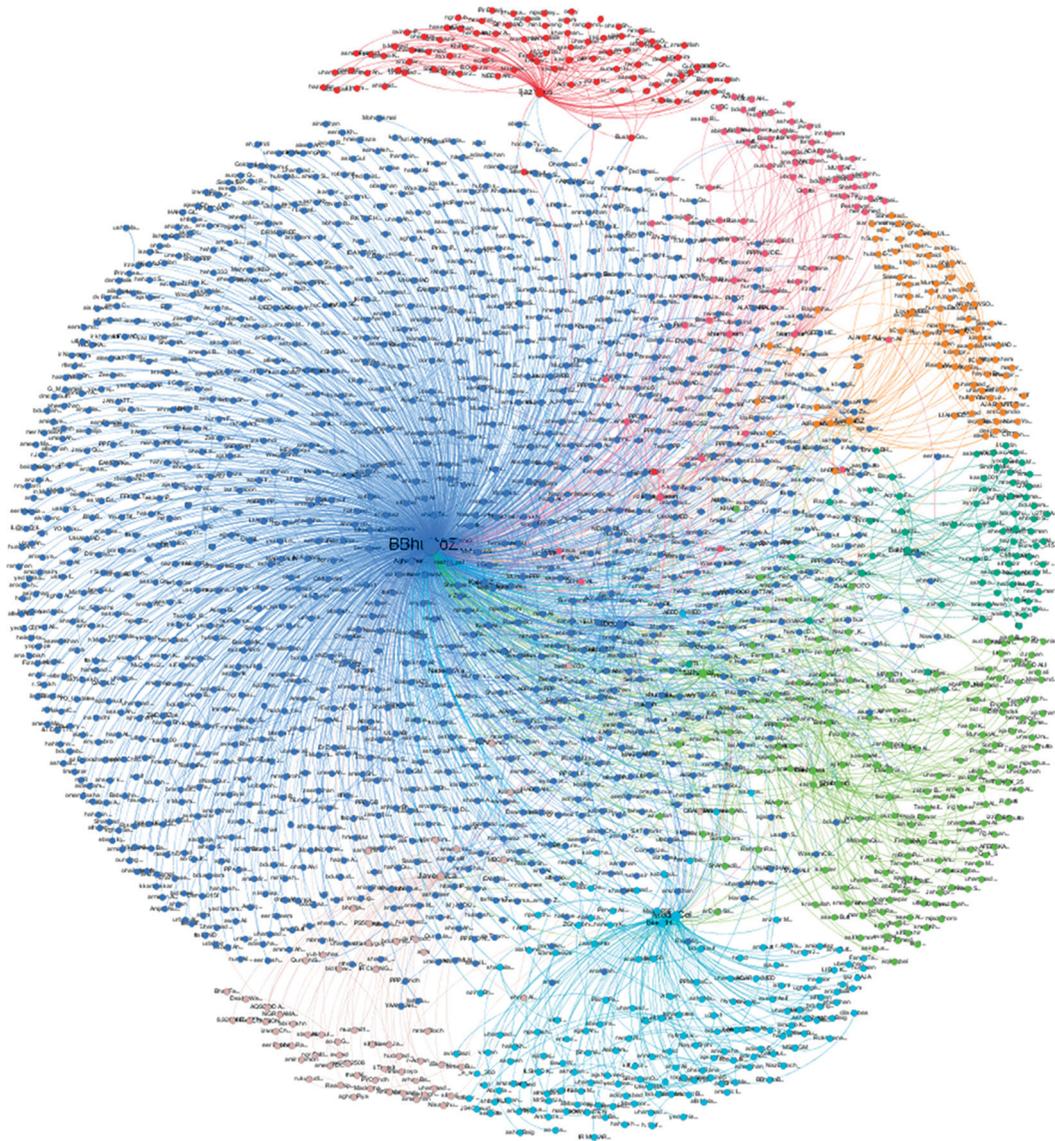


FIGURE 4: Communities in BBZ's supporters' reply network. The blue colour community contains supporters of BBZ, green colour contains PPP media cell, orange colour contains MNS, pink colour contains IK, and cerulean colour contains PPP's political leaders.

were ignored in this study as the sizes of those were between 0.001% and 0.94%. A giant community having lilac colour has a size of 52.14% and contains supporters who replied directly to MNS. On manual study, it was noticed that the people in the community having a green colour weigh 4.76% of the total network related to IK. The community having cerulean colour consists of 2.87% containing another political leader who is a member of the Pakistan Muslim League Noon (party of MNS). The other community of arctic colour whose size is 2.2% of the whole network contained some supporters from BBZ who replied to MNS tweets. The remaining two communities, weighing 3.8%, each contained some people, in-touched with the media cell accounts of PMLN (Pakistan Muslim League Noon).

Figure 4 shows eight major extracted communities consisting of 80.18% nodes and 85.7% of links of the total supporters' reply network of BBZ. The giant community

having the blue colour has a size of 50.08% consisting of supporters who lied in the BBZ community and replied to BBZ tweets. On a manual study, it was found that the green community which has the size of 7.47% contained the media cell of PPP (Pakistan People Party) and the cerulean colour community having sizes 6.72%, 3.82%, 3.53%, 3.02%, 2.90%, and 2.64% contained other political leaders who are members of BBZ's party. Orange community having a 3.82% size contains some people from MNS's party and the pink community having a size of 3.02% contains people from IK's party.

Table 2 shows that the number of nodes is high in MNS supporters' network, but Figures 3 and 5 show that she has less support. The number of weakly connected components is less in IK, followed by BBZ and MNS. The smaller the network diameter, the stronger the network. IK has a network diameter of 5 and then BBZ and MNS, respectively.

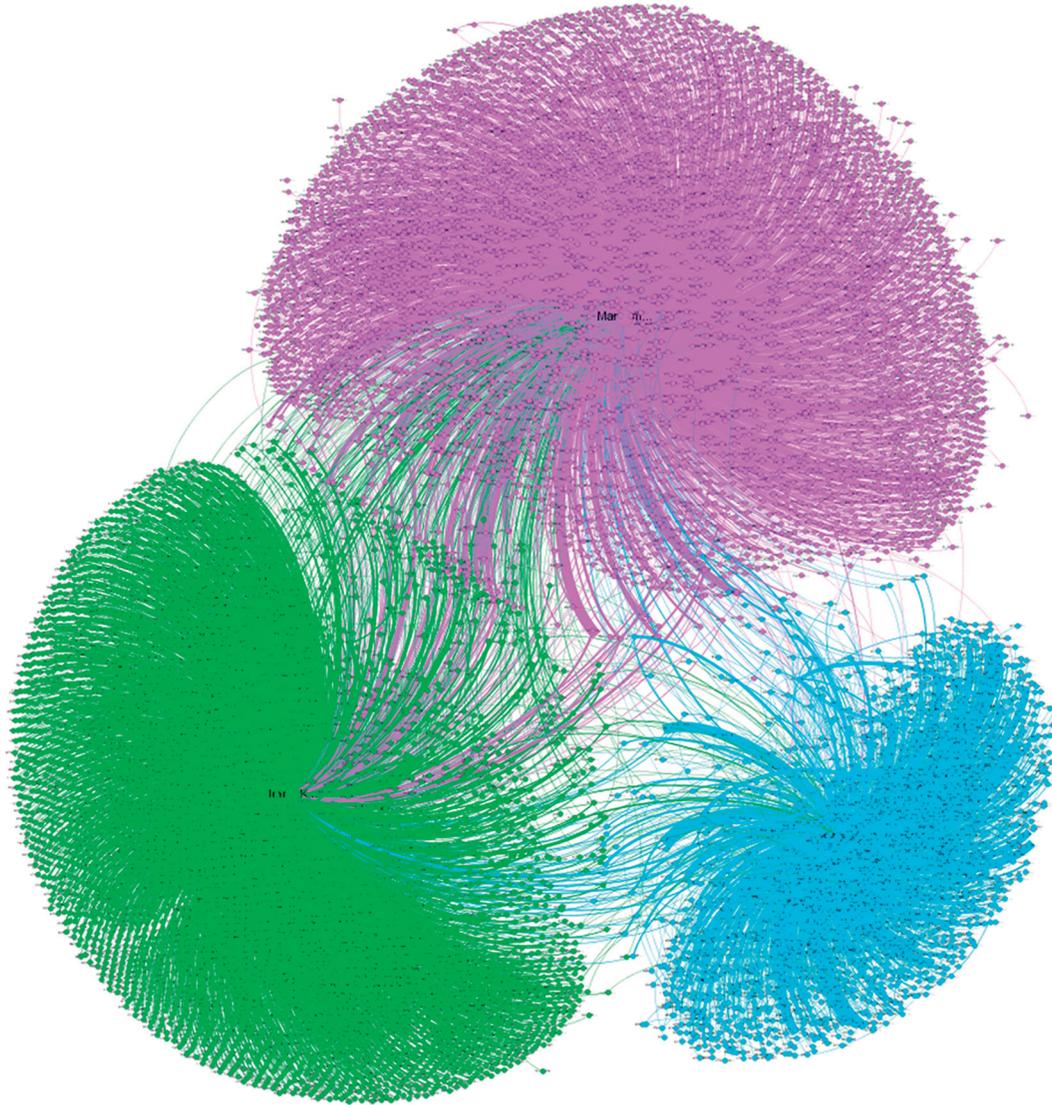


FIGURE 5: Supporters reply communities of IK, MNS, and BBZ. The green colour community contains supporters of IK, violet colour contains MNS, and blue colour contains BBZ.

TABLE 3: Supporters’ retweet network of political leaders.

Parameters	Imran Khan	Maryam Nawaz Sharif	Bilawal Bhutto Zardari
Nodes	28128	10800	6180
Edges	62586	37215	14806
Network diameter	8	10	10
Average path length	4.846	5.633	3.749
Number of weakly connected components	135	164	81
Number of strongly connected components	27750	10564	6084

The average path length of the IK supporter network is less followed by BBZ and then MNS. All these characteristics lead us to the fact that IK has the strongest supporter network.

The supporters’ reply network of the three leaders has been validated. The PageRank of IK is 0.177, MNS is 0.170, and BBZ is 0.051. Figure 5 shows that the three major extracted communities consist of 81.79% of the entire

network. The other communities belong to some other leaders and parties. Supporters’ reply community having green colour belongs to IK having size 34.70%, MNS 34.52%, and BBZ 12.57%. It concludes that IK is more influential and has strong support than MNS and BBZ, respectively. Figure 5 also shows the communication links between the supporters of the three leaders. It can be seen that the supporters of IK and BBZ are more flexible, and they have

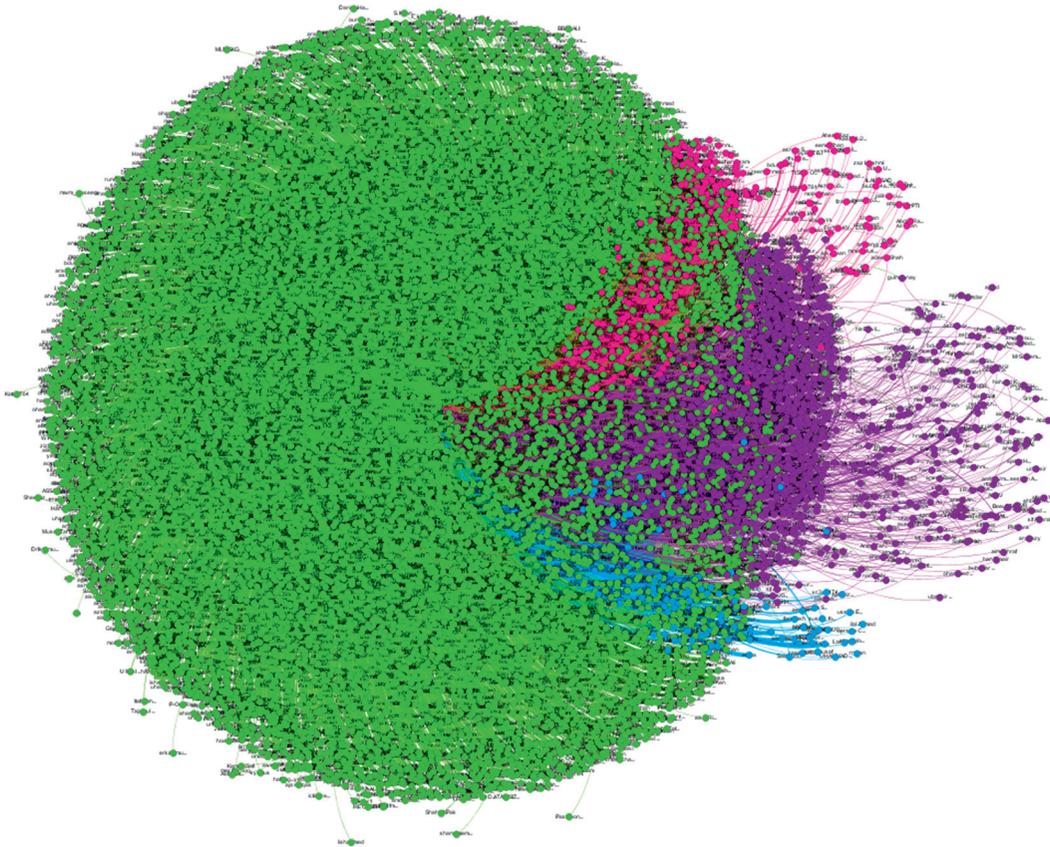


FIGURE 6: Communities in IK's supporters' retweet network. The green colour contains supporters of IK, violet colour contains PTI's account, and the rest are related to other leaders from PTI and other political parties.

more connections with the other communities. The supporters of MNS have more communication links with IK's supporters but few links with BBZ's supporters.

4.1.2. Retweet Network. The supporters' retweet networks of each political leader which shows information diffusion have been constructed. The properties of the supporters' retweet network of each leader are given in Table 3.

Figure 6 shows five major extracted communities in the supporters' retweet network of IK consisting of 94.90% of the entire network. The giant green community had the size of 80.16% IK and supporters who retweeted IK directly. The violet community is having the size of 10.43% consisting of the official page of PTI (Pakistan Tehreek-i-Insaf); the other communities with sizes 2.95%, 1.23%, and 1.13% consist of other leaders of PTI and some other political parties.

Figure 7 shows five significant communities in the retweet network of MNS consisting of 89.35% of the full network. 64.46% consists of MNS and supporters who retweeted her directly. The green community with the size of 11.02% consists of people that belong to the media cells of PMLN (Pakistan Muslim League Noon). The other communities had the sizes of 6.5%, 6.38%, and 0.99%, and the rest belong to other members of PMLN and other parties.

Figure 8 shows eight major communities in the supporters' retweet network consisting of 71.31% of the entire

network. The huge lilac community, which is only 24.90% of the whole network, consists of BBZ and supporters who directly retweeted him. The second-largest green community having the size 12.14% belongs to the media cell of PPP (Pakistan People's Party). The other communities belong to other parties and some prominent members of Bilawal's party who were supporting him by retweeting him and they have their followers.

Table 3 shows that the information diffusion rate of IK is higher as the number of supporters who retweeted IK is higher than in MNS and BBZ. The number of weakly connected components is less in MNS followed by IK and then BBZ, but IK has almost three times more nodes than MNS. IK has the smallest network diameter than that of the MNS and BBZ. The average path length of BBZ is 3.909, followed by IK and then MNS. On average, it can be concluded that IK has more influence and has strong support as compared to the other three political leaders.

The strongest retweet political network has been validated. PageRank of IK is 0.163, that of MNS is 0.044, and that of BBZ is 0.011. Figure 9 shows the three major extracted communities. The giant community having green colour is comprised of IK supporters, the purple community of MNS, and the blue community of BBZ. Communication links can be seen between the supporters of the three political leaders. The PageRank and the size of IK are higher than those of the other two leaders, which leads us to the strong support of

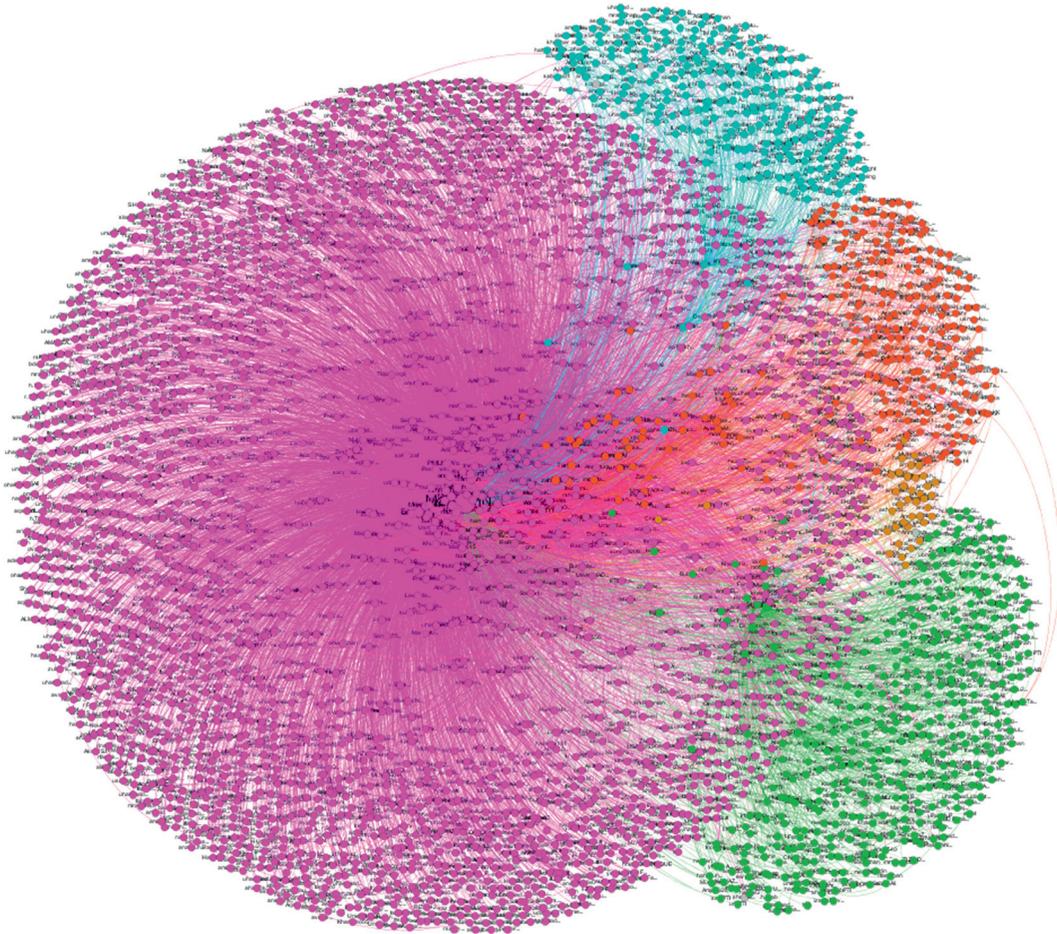


FIGURE 7: Communities in MNS's supporters' retweet network. The violet colour consists of MNS' supporters; green colour consists of PMLN media cell; the rest are related to other leaders from PMLN and other parties.

people to IK. The information diffusion in the network of IK is higher.

From the above experiments and discussion, the answer of RQ1 can be concluded that IK has strong support from his supporters as compared to MNS and BBZ, respectively.

4.2. Semantic Analysis of Supporters' Network. To address RQ2, experiments were conducted to extract sentiments of supporters towards the three political leaders using machine learning techniques and generating the word cloud from the content used by supporters.

4.2.1. Sentiment Analysis of Supporters. NB, K-NN, and SVM were trained. Accuracy (in percentage) of the NB classifier, SVM classifier, and KNN classifier is plotted in Figure 10. It was found that SVM performed better than the other two classifiers with an accuracy of 78.89%; NB has 75.40%; and K-NN has 65.35%. The classifier was considered to classify unseen and unlabeled tweets in this research. Unlabeled tweets about the three political leaders were classified into three classes, i.e., Positive, Neutral, and Negative.

Figure 11 shows the percentage of polarity of supporters' tweets about IK, MNS, and BBZ. The number of supporters' tweets for every political leader is different; it is because a specific time slice has been considered in this study. IK's followers are more than those of the other two politicians. The percentage of sentiment polarity was considered to have better and balanced results. It can be seen that the people having a positive sentiment about IK have a high rate as compared to the BBZ and Maryam Nawaz. People's sentiment about BBZ has a high negative rate as compared to the other two politicians. IK and MNS are popular, and people talk more about them on Twitter as compared to BBZ. It is stated that a political leader, whose supporters have a high percentage of positive sentiments, has a strong supporter community. From these results, it can be concluded that IK has a stronger supporter community followed by MNS and BBZ, respectively.

4.2.2. Semantic Word Cloud of Supporters. In this analysis, the frequency of words used by supporters of the three political leaders is visualized using semantic word clouds. Table 4 shows some abbreviations used by the supporters in their tweets.

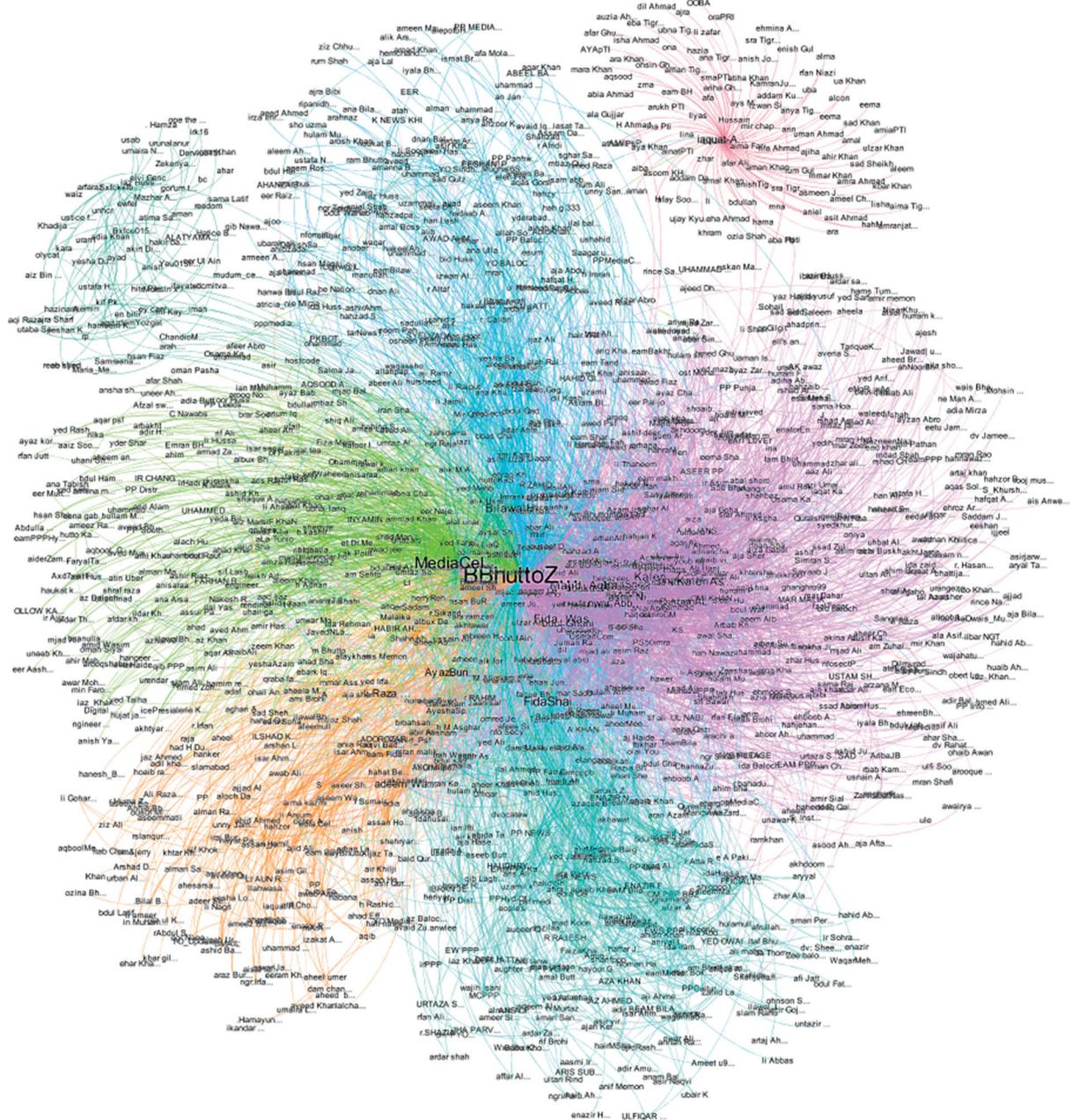


FIGURE 8: Communities in BBZ's supporter retweet network. The lilac colour stands for BBZ, green colour stands for PPP media group, and other communities belong to other parties and some prominent members of PPP.

Figure 12(a) shows word cloud for tweets of IK. It can be seen that the frequency of some words tends to have a positive sentiment like Great, Leader (great leader), Good, Allah, Leader, against (against corruption), More, Want, Support, Well, Right, Need, Save, and Nation. The frequency of these words is high, and it can be seen in the word cloud of IK. Figure 12(a) also shows some words which give negative sentiments, i.e., Corrupt, Shame, etc. It can be seen in the word cloud of IK that there are some words like PMLN, Go, NS, Nawaz, and Sharif. This is because IK initiated a slogan “Go Nawaz Go” which was a movement against the corruption of Nawaz Sharif (Ex-Prime Minister of Pakistan and father of MNS) and many supporters of IK raised this slogan.

In the manual study, it was found that the word “corruption” in the word cloud of IK was actually against Nawaz Sharif. As can be seen that “Go, Nawaz, NS, and Corrupt” words have almost the same frequency.

Figure 12(b) shows word cloud for MNS. It can be seen that MNS’s word cloud has many words having almost the same frequency as compared to IK’s word cloud where some words have a very high frequency. It also shows words having positive sentiment like, Great, Allah, Love, Right, Support, People, Sher (lion), Good, God, Health, work, Ameen, Help, Nation, Pray, Happy, Better, Nice, Wish, Need, soon (better health soon), Insha’Allah, Life, Mother, and Save. During the time of the data collection for this

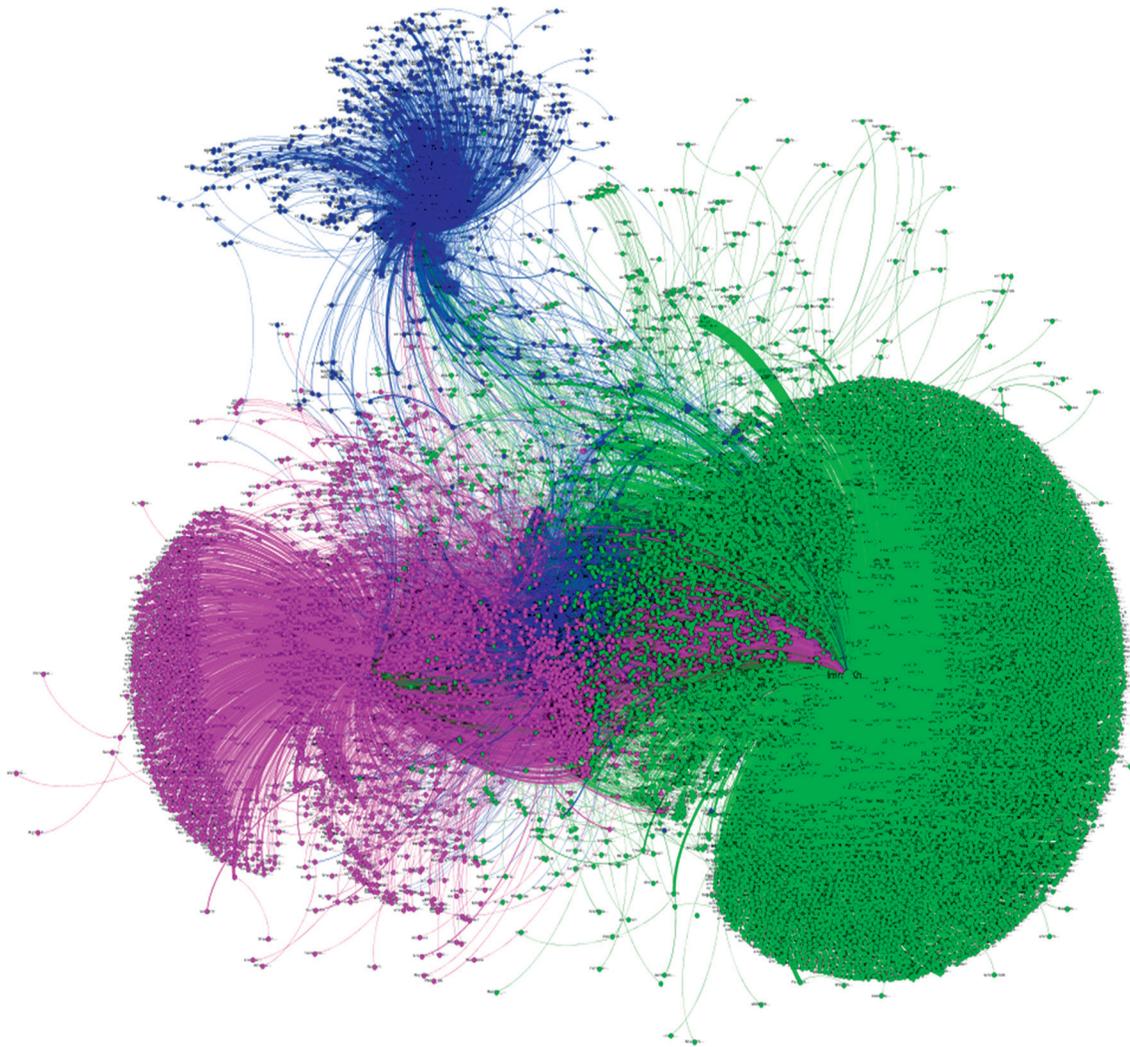


FIGURE 9: Communities of the supporters of IK, MNS, and BBZ. The green colour community contains supporters of IK, violet colour contains MNS, and blue colour contains BBZ.

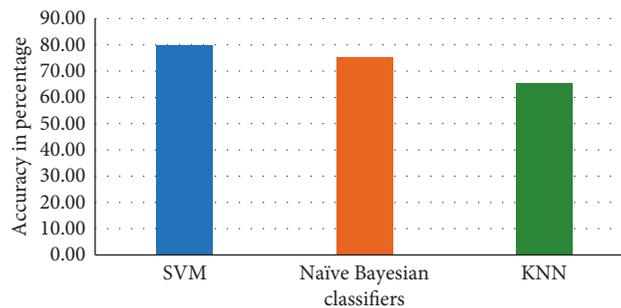


FIGURE 10: Classifiers accuracy.

research, MNS’s mother was at the hospital and was fighting against cancer. Most of the supporters were praying for health. It can be seen that supporters of MNS have a huge love for her; they pray for her and her family. The supporters call her a great leader too. Figure 12(b) also shows us words having negative sentiments like, Chor, Loot, Corrupt,

Shame, case (the case against her corruption scandal in the courts), Jail, Go, Stop, NAB, London (London flats—scandal), Liar, Never, Court, Fool, Hell, and False. Nawaz Sharif and his children (including MNS) were part of a corruption scandal like flats in London, for which they were facing some cases in NAB and courts by that time, and later they were

Karachi, Hyderabad, Mazar, etc. This gives a viewpoint that there are some people (leaders) from the party who strongly support BBZ and want to make him superior. It can be seen in Figure 12(c) that there are other renowned personalities who are supporting BBZ. Words like, Jeay (long live), Next (next Prime Minister), Peace, Young, Work, Wish, Thank, Need, Pak, Educated Nice, Best, Peace, More, against (Against corruption), Public, and Right show positive sentiments. Words like Shame, Corrupt, Kid, Terror, Sad, and Out show a negative sentiment of people. The result discussed previously answers RQ2 that IK has the strongest supporter network on the basis of content analysis.

5. Conclusion and Future Work

This study was conducted to assist in decision making and predicting the future of a political leader based on multicriteria and machine learning algorithms. The supporters' network of three political leaders of Pakistan on Twitter was analyzed using link analysis and semantic analysis. This study analyzes the communication based on reply network and information diffusion based on the retweet network and predicted the strongest supporter network. 21,000 tweets were manually labelled and used to train machine learning techniques, i.e., SVM, NB, and KNN, to predict the strongest support network of a leader based on the sentiments. The dataset analyzed in this study was taken prior to the election in October 2017. It was concluded that SVM works better and has an accuracy of 79.89% on the dataset used in this study, which is higher than NB (75.40%) and KNN (65.35%).

This study also concludes that IK has the strongest support network and is more influential on Twitter compared to MNS and BBZ. This can assist and lead us in predicting the results of the election. Besides, the supporters of every political leader have some flexibility and they use to communicate with other leader's supporters. It was also observed that IK is more vigilant in his content by using social media. The proposed method can help political leaders, political parties, and government organizations in decision making for devising policies and their future plans. The proposed method can be used to support government organizations in many ways, i.e., identifying some groups that secretly work against the government. Besides, it can also find out the links between followers of different religions, actors, gamers, and supporters of any products.

There are some limitations in this study. The communities having small sizes were not analyzed. The number of communication links between two supporters' networks and their weights was not studied. The topics discussed and the issues raised in these tweets were not focused on. The trends and the issues that arose in the tweets concerning these politicians will be studied in the future. Some famous slogans can be investigated, which is more famous among the people of Pakistan. Detection of communities using edge content will be investigated in the future. Besides this, the Tweets posted in the Urdu language are also worth exploring in the future.

Data Availability

The data used in this study are available on Twitter and can be accessed using Twitter API. The python code for the Twitter API can be provided on request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Authors' Contributions

Asif Khan, Huaping Zhang, Jianyun Shang, Lin Dai, and Arshad Ahmad were involved in the conceptualization. Asif Khan, Huaping Zhnag, Nada Boudjellal, Jianyun Shang, Lin Dai, and Asmat Ali were responsible for the investigation and analysis. Asif Khan, Hauping Zhang, Nada Boujdllela, Jianyun Shang, Lin Dai, and Arshad Ahmad were involved in the methodology. Asif Khan, Nada Boudjellal, and Asmat Ali were responsible for the data collection. Asif Khan, Arshad Ahmad, and Asmat Ali were involved in the data labeling and validation. Asif Khan and Nada Boudjellal were responsible for experimentation. Asif Khan was responsible for original drafting. Asif Khan, Huaping Zhang, Jianyun Shang, Lin Dai, and Asmat Ali were responsible for visualization. All authors were involved in the results validation. All authors were responsible for reviewing and editing. Huaping Zhang, Lin Dai, and Jianyun Shang were involved in the administration. Huaping Zhang, Jianyun Shang, and Lin Dai were involved in the supervision. Huaping Zhang, Jianyun Shang, and Lin Dai were responsible for funding acquisition.

Acknowledgments

The research work was funded by the National Key Research and Development Project of China (Grant No. 2018YFC0832304) and by the National Science Foundation of China (Grant No. 61772075). The authors are thankful to them for their financial support.

References

- [1] A. Ahmad, C. Feng, S. Ge, and A. Yousif, "A survey on mining stack overflow: question and answering (Q&A) community," *Data Technologies and Applications*, vol. 52, no. 2, pp. 190–247, 2018.
- [2] A. Ahmad, C. Feng, K. Li, S. M. Asim, and T. Sun, "Toward empirically investigating non-functional requirements of iOS developers on stack overflow," *IEEE Access*, vol. 7, pp. 61145–61169, 2019.
- [3] A. Ahmad, K. Li, C. Feng, and T. Sun, "An empirical study on how iOS developers report quality aspects on stack overflow," *International Journal of Machine Learning and Computing*, vol. 8, no. 5, pp. 501–506, 2018.
- [4] A. Ahmad, *Research on Comprehending Software Requirements on Social Media*, Beijing Institute of Technology, Beijing, China, 2018.
- [5] A. Ahmad, "An empirical evaluation of machine learning algorithms for identifying software requirements on stack

- overflow: initial results,” in *Proceedings of the IEEE International Conference on Software Engineering and Service Sciences (ICSESS)*, Beijing, China, October 2019.
- [6] B. J. Brands, T. Graham, and M. Broersma, “Social media sourcing practices: how Dutch newspapers use tweets in political news coverage,” *Managing Democracy in the Digital Age*, Springer, Berlin, Germany, pp. 159–178, 2017.
 - [7] G. Enli and L. T. Rosenberg, “Trust in the age of social media: populist politicians seem more authentic,” *Social Media + Society*, vol. 4, no. 1, 2018.
 - [8] M. A. Baum and P. B. K. Potter, “Media, public opinion, and foreign policy in the age of social media,” *The Journal of Politics*, vol. 81, no. 2, pp. 747–756, 2019.
 - [9] S. A. Eldridge, L. García-Carretero, and M. Broersma, “Disintermediation in social networks: conceptualizing political actors’ construction of publics on twitter,” *Media and Communication*, vol. 7, no. 1, pp. 271–285, 2019.
 - [10] M. Lalancette and V. Raynauld, “The power of political image: justin Trudeau, Instagram, and celebrity politics,” *American Behavioral Scientist*, vol. 63, no. 7, pp. 888–924, 2019.
 - [11] A. Hellweg, “Social media sites of politicians influence their perception by constituents,” *The Elon Journal of Undergraduate Research in Communications*, vol. 2, no. 1, pp. 22–36, 2011.
 - [12] A. Ceron, L. Curini, S. M. Iacus, and G. Porro, “Every tweet counts? How sentiment analysis of social media can improve our knowledge of citizens’ political preferences with an application to Italy and France,” *New Media & Society*, vol. 16, no. 2, pp. 340–358, 2014.
 - [13] M. Broersma and T. Graham, “Social media as beat,” *Journalism Practice*, vol. 6, no. 3, pp. 403–419, 2012.
 - [14] C. Vaccari and A. Valeriani, “Follow the leader! Direct and indirect flows of political communication during the 2013 Italian general election campaign,” *New Media & Society*, vol. 17, no. 7, pp. 1025–1042, 2015.
 - [15] W. J. Grant, B. Moon, and J. Busby Grant, “Digital dialogue? Australian politicians’ use of the social network tool twitter,” *Australian Journal of Political Science*, vol. 45, no. 4, pp. 579–604, 2010.
 - [16] A. L. Walsh, M. E. Peters, R. L. Saralkar, and M. S. Chisolm, “Psychiatry residents integrating social media (PRISM): using twitter in graduate medical education,” *Academic Psychiatry*, vol. 43, no. 3, pp. 319–323, 2019.
 - [17] B. N. Reames, K. H. Sheetz, M. J. Englesbe, and S. A. Waits, “Evaluating the use of twitter to enhance the educational experience of a medical school surgery clerkship,” *Journal of Surgical Education*, vol. 73, no. 1, pp. 73–78, 2016.
 - [18] A. Rasool, R. Tao, K. Marjan, and T. Naveed, “Twitter sentiment analysis: a case study for apparel brands,” *Journal of Physics: Conference Series*, vol. 1176, no. 2, Article ID 022015, 2019.
 - [19] J. G. D. Harb and K. Becker, “Comparing emotional reactions to terrorism events on twitter,” *Communications in Computer and Information Science*, vol. 926, pp. 107–122, 2019.
 - [20] S. Ali, N. Islam, A. Rauf, I. Din, M. Guizani, and J. Rodrigues, “Privacy and security issues in online social networks,” *Future Internet*, vol. 10, no. 12, p. 114, 2018.
 - [21] E. Bonsón, D. Perea, and M. Bednárová, “Twitter as a tool for citizen engagement: an empirical study of the Andalusian municipalities,” *Government Information Quarterly*, vol. 36, no. 3, pp. 480–489, 2019.
 - [22] A. Haro-de-Rosario, A. Sáez-Martín, and M. del Carmen Caba-Pérez, “Using social media to enhance citizen engagement with local government: twitter or facebook?” *New Media & Society*, vol. 20, no. 1, pp. 29–49, 2018.
 - [23] A. Oehmichen, K. Hua, J. Amador Diaz Lopez, M. Molina-Solana, J. Gomez-Romero, and Y.-k. Guo, “Not all lies are equal. A study into the engineering of political misinformation in the 2016 US presidential election,” *IEEE Access*, vol. 7, pp. 126305–126314, 2019.
 - [24] C. Featherstone, “South African bot behaviour post the July 2018 twitter account cull,” in *Proceedings of the ICABCD 2019—2nd International Conference on Advances in Big Data, Computing and Data Communication Systems*, Winterton, South Africa, 2019.
 - [25] I. Clarke and J. Grieve, “Stylistic variation on the Donald Trump twitter account: a linguistic analysis of tweets posted between 2009 and 2018,” *PLoS One*, vol. 14, no. 9, Article ID e0222062, 2019.
 - [26] R. Tromble, “The great leveler? comparing citizen-politician twitter engagement across three Western democracies,” *European Political Science*, vol. 17, no. 2, pp. 223–239, 2018.
 - [27] T. Mahmood, T. Iqbal, F. Amin, W. Lohanna, and A. Mustafa, “Mining twitter big data to predict 2013 Pakistan election winner,” in *Proceedings of the 2013 16th International Multi Topic Conference, INMIC 2013*, pp. 49–54, Lahore, Pakistan, 2013.
 - [28] P. Sharma and T. S. Moh, “Prediction of Indian election using sentiment analysis on Hindi twitter,” in *Proceedings of the 2016 IEEE International Conference on Big Data*, pp. 1966–1971, Washington, DC, USA, 2016.
 - [29] N. Dokoochaki, F. Zikou, D. Gillblad, and M. Matskin, “Predicting Swedish elections with Twitter: a case for stochastic link structure analysis,” in *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM 2015*, pp. 1269–1276, Paris, France, August 2015.
 - [30] B. Li, D. Guo, M. Chang, M. Li, and A. Bian, “The prediction on the election of representatives,” in *Proceedings of the 2017 International Conference on Security, Pattern Analysis, and Cybernetics, SPAC 2017*, Shenzhen, China, December 2017.
 - [31] Z. Xie, G. Liu, J. Wu, L. Wang, and C. Liu, “Wisdom of fusion: prediction of 2016 Taiwan election with heterogeneous big data,” in *Proceedings of the 2016 13th International Conference on Service Systems and Service Management, ICSSSM 2016*, Kunming, China, 2016.
 - [32] P. Verweij, “Twitter links between politicians and journalists,” *Journalism Practice*, vol. 6, no. 5–6, pp. 680–691, 2012.
 - [33] A. Klefodimos, G. Lappas, A. Triantafyllidou, and A. Yannacopoulou, “Investigating the Greek political twittersphere: Greek MPS and their twitter network,” in *Proceedings of the South-East Europe Design Automation, Computer Engineering, Computer Networks and Social Media Conference*, Kunming, China, 2018.
 - [34] G. Stamatelatos, S. Gyftopoulos, G. Drosatos, and P. S. Efraimidis, “Deriving the political affinity of twitter users from their followers,” in *Proceedings of the 16th IEEE International Symposium on Parallel and Distributed Processing with Applications*, pp. 1175–1182, Melbourne, Australia, 2019.
 - [35] R. Castro, L. Kuffó, and C. Vaca, “Back to #6D: predicting Venezuelan states political election results through twitter,” in *Proceedings of the 2017 4th International Conference on eDemocracy and eGovernment, ICEDEG 2017*, pp. 148–153, Quito, Ecuador, 2017.
 - [36] M. Yoshida and F. Toriumi, “Analysis of political party twitter accounts’ retweeters during Japan’s 2017 election,” in *Proceedings of the 2018 IEEE/WIC/ACM International Conference on Web Intelligence*, Santiago, Chile, 2018.
 - [37] L. Wang and J. Q. Gan, “Prediction of the 2017 French election based on twitter data analysis,” in *Proceedings of the*

- 2017 9th Computer Science and Electronic Engineering Conference, pp. 89–93, Colchester, 2017.
- [38] L. Wang and J. Q. Gan, “Prediction of the 2017 French election based on twitter data analysis using term weighting,” in *Proceedings of the 2017 10th Computer Science and Electronic Engineering Conference*, Colchester, UK, 2019.
- [39] Z. Xie, G. Liu, J. Wu, and Y. Tan, “Big data would not lie: prediction of the 2016 Taiwan election via online heterogeneous information,” *EPJ Data Science*, vol. 7, no. 1, p. 1, 2018.
- [40] S. Ahmed and M. M. Skoric, “My name is Khan: the use of twitter in the campaign for 2013 Pakistan general election,” in *Proceedings of the 2014 Annual Hawaii International Conference on System Sciences*, pp. 2242–2251, Waikoloa, HI, USA, 2014.
- [41] M. Bilal, S. Asif, S. Yousuf, and U. Afzal, “2018 Pakistan general election: understanding the predictive power of social media,” in *Proceedings of the 12th International Conference on Mathematics, Actuarial Science, Computer Science*, Karachi, Pakistan, 2019.
- [42] P. Aragón, K. E. Kappler, A. Kaltenbrunner, D. Laniado, and Y. Volkovich, “Communication dynamics in twitter during political campaigns: the case of the 2011 Spanish national election,” *Policy & Internet*, vol. 5, no. 2, pp. 183–206, 2013.
- [43] Twitter Apps, Twitter developers <https://apps.twitter.com/>.
- [44] Tweepy. [Online]. Available: <https://github.com/tweepy/tweepy>.
- [45] Twitter Developer. <https://developer.twitter.com/en/docs>.
- [46] No Title. [Online]. Available https://github.com/irfreitas/tweepy_mysql_follower_scrape.
- [47] J. Fernquist, L. Kaati, and R. Schroeder, “Political bots and the Swedish general election,” in *Proceedings of the 2018 IEEE International Conference on Intelligence and Security Informatics, ISI 2018*, pp. 124–129, Miami, FL, USA, November 2018.
- [48] J. A. Caetano, J. Almeida, and H. T. Marques-Neto, “Characterizing politically engaged users’ behavior during the 2016 us presidential campaign,” in *Proceedings of the 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM 2018*, pp. 523–530, Barcelona, Spain, 2018.
- [49] C. Shao, P.-M. Hui, P. Cui, X. Jiang, and Y. Peng, “Tracking and characterizing the competition of fact checking and misinformation: case studies,” *IEEE Access*, vol. 6, pp. 75327–75341, 2018.
- [50] F. Masood, G. Ammad, A. Almogren et al., “Spammer detection and fake user identification on social networks,” *IEEE Access*, vol. 7, pp. 68140–68152, 2019.
- [51] C. A. Davis, O. Varol, E. Ferrara, A. Flammini, and F. Menczer, “BotOrNot,” in *Proceedings of the 25th International Conference Companion on World Wide Web—WWW’16 Companion*, pp. 273–274, Montréal, Canada, 2016.
- [52] S. Fortunato, “Community detection in graphs,” *Physics Reports*, vol. 486, no. 3–5, pp. 75–174, 2010.
- [53] V. D. Blondel, J. L. Guillaume, R. Lambiotte, and E. Lefebvre, “Fast unfolding of communities in large networks,” *Journal of Statistical Mechanics: Theory and Experiment*, vol. 1, Article ID P10008, 2008.
- [54] W. S. Noble, “What is a support vector machine?” *Nature Biotechnology*, vol. 24, no. 12, pp. 1565–1567, 2006.
- [55] K. M. Leung, “Naive Bayesian classifier,” *Lecture Notes*, Springer, Berlin, Germany, 2007.
- [56] S. Singh, J. Haddon, and M. Markou, “Nearest-neighbour classifiers in natural scene analysis,” *Pattern Recognition*, vol. 34, no. 8, pp. 1601–1612, 2001.
- [57] M. Bastian, S. Heymann, and M. Jacomy, “GEPHI: an open source software for exploring and manipulating networks,” in *Proceedings of the 3rd International AAAI Conference Weblogs Social Media*, Ann Arbor, MI, USA, 2009.