Characterizing the 2016 U.S. Presidential Campaign using Twitter Data

Ignasi Vegas, Tina Tian
Department of Computer Science
Manhattan College
New York, USA

Wei Xiong
Department of Information Systems
Iona College
New Rochelle, USA

Abstract—This paper models the 2016 U.S. presidential campaign in the context of Twitter. The study analyzes the presidential candidates' Twitter activity by crawling their real-time tweets. More than 16,000 tweets were observed in this work. We study the interactions between the politicians and their Twitter followers in the retweet and favorite networks. The most frequently mentioned unigrams are presented, which serve the best featuring the political focuses of a candidate. The mention network among the politicians was constructed by parsing the content of their tweets. In this paper, we also study the Twitter profile of the users who follow the presidential candidates. The gender ratio among the Twitter subscribers is examined using the government's census data. We also investigate the geography of Twitter supporters for each candidate.

Keywords—Twitter; social networks; data mining

I. INTRODUCTION

Online social networks, such as Facebook, Twitter and LinkedIn, have gained increasingly popularity over the decade [1]. Vast quantities of real-time, fine grained data are available on social networking sites, including text, images and other multimedia records. The tremendous amount of data on social networks can be extracted using their Application Program Interfaces (APIs), thus be leveraged for analysis. For example, Sakaki et al. monitored real-time activities on Twitter to detect earthquakes [2]. Tian et al. extracted knowledge from Facebook to build an intelligent search system, which improved users' search experience on the web [3] [4]. Archambault and Grudin surveyed the Microsoft employees to study the usefulness of Twitter in organizational communication and information-gathering [5]. Bakhshi et al. studied a pool of images on Instagram and found that photos with faces are more likely to receive likes and comments [6]. Overall, the availability of massive quantities of data on social media has given a boost to the scientific study of the field of social networks [1].

Due to the rapidly growing number of users, online social media has become the ideal platform for politicians to engage and interact with their potential voters. Among all, Twitter, particularly, has become an integral part in the political campaign [7]. Twitter is an online microblogging service that enables users to share with the public short messages called "tweets". Currently, there are more than 310 million monthly active users on Twitter worldwide [8]. Users may subscribe to other users' tweets, in which case subscribers are called "followers". Users may also rebroadcast other people's tweets

to their own feed, a process known as a "retweet". It allows posts to propagate throughout the network and thus raise their visibility [9]. Moreover, individual tweets can be marked as "favorites" by other users. The content of a tweet may contain text, hyperlinks, images and video clips. Messages regarding the same topic can be grouped using hashtags, a form of metadata consisting of words or phrases preceded by a hash symbol ("#"). Similarly, the "@" sign followed by a username is used to refer to a specific user [10].

A. Background on the 2016 U.S. Presidential Election

In this study, we analyze the presidential candidates' Twitter activities by collecting their real-time tweets. We started extracting data from September 26, 2015 and we will keep gathering and monitoring the Twitter data until the Election Day, which will occur on November 8, 2016. At the beginning of the study, there were five active presidential candidates. They are Hilary Clinton and Bernie Sanders from the Democratic Party and Donald Trump, Ted Cruz and John Kasich from the Republican Party. As of the time of writing, only two candidates remain in the presidential campaign: Hilary Clinton as the nominee of the Democratic Party and Donald Trump being the nominee of the Republican Party.

B. Related Work

The exponential growth of Twitter has made it a popular subject for research in multiple disciplines [11]. One stream of research studied the influence and passivity of users. For example, Romero et al. revealed that users with many followers may not necessarily be influential to the community [12]. Another stream of research investigated the commercial and marketing usage of Twitter. For instance, Jansen et al. examined the use of Twitter for sharing consumer opinions to targeting products and brands [13].

With the successful campaign of Barack Obama in the 2008 U.S. Presidential election, the importance of Twitter in politics has become clear [14] [15]. Twitter, being a platform for political deliberation, has attracted attention of many researchers [7]. Tumasjan et al. investigated whether online tweets can reflect offline political sentiment in the context of a German election [7]. Conover et al. examined the retweet network and the mention network in pushing the political communication on Twitter during the 2010 U.S. congressional midterm elections [16]. A proof-of-concept model was developed by Livne et al. to predict candidate's victory using data in the same context [17].

The previous studies of Twitter in politics, however, focused on voters instead of the candidates themselves. Livne et al. analyzed the differences between candidates [17], but had an emphasis on each political party as a whole. In this paper, we concentrate our attention on individual candidates.

Our study analyzes the U.S. presidential candidates' Twitter activity by collecting their real-time tweets using Twitter's REST API [18]. The system monitors the interactions between the politicians and their followers by studying the patterns emergent from the retweet networks. The paper also investigates and compares the gender ratio and geographical distribution of the candidates' Twitter followers.

The rest of the paper is organized as follows. Section II describes the data set used in the experiment and explains the methods and algorithms adapted in analyzing the data. In Section III, we present the results and visualize them in the form of charts and maps. Section IV concludes the paper and proposes future directions.

II. DATA SET AND METHODOLOGY

The study leverages data crawled from Twitter's REST API between September 26 of 2015 and the time of writing. During the one year of data collection, we observed approximately 16,805 tweets. Tweets were extracted from the candidates' verified Twitter accounts. A verified account is a validation mechanism on Twitter that ensures the identity of the user. One of the candidates, Senator Bernie Sanders, has two verified and highly active Twitter accounts (@BernieSanders and @SenSanders). Therefore, tweets from both accounts are stored. Analysis of Bernie Sanders in this paper is based on the combined data from his two Twitter accounts. Table I shows the total number of tweets and the average number of posts tweeted per day by each candidate. As one can see in the table, the candidates are grouped by their political party. Politicians in the same party are sorted alphabetically by their last name. The same listing order is used for the tables in the rest of this paper.

TABLE I.   TOTAL AND DAILY VOLUME OF TWEETS BY CANDIDATES

Party	Candidates	Total number of tweets	Average number of tweets per day
Democratic	Clinton	3075	12
	Sanders	3824	15
Republican	Cruz	2437	10
	Kasich	2309	9
	Trump	5160	21

In this paper, we analyze three aspects of the data – top favorited and retweeted posts by candidates, most frequently mentioned terms by candidates, and profile analysis of followers, where we study their gender ratio and geographical distribution. Figure 1 shows the architectural overview of the system. The rest of this section elaborates each module.
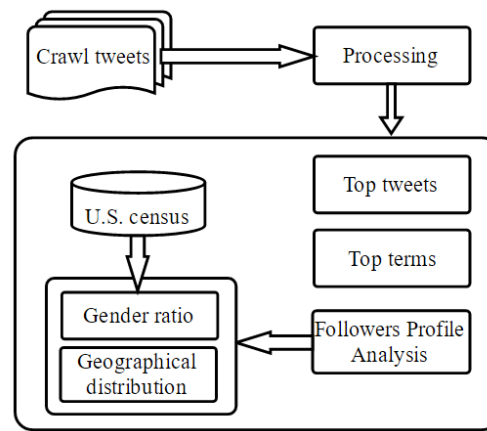

Fig. 1.   System architectural overview

A. Top Tweets

As mentioned in Section I, individual tweets can be labeled as "favorites" by other Twitter users. They can also be retweeted, thus to be shared and rebroadcasted. The volume of favorites and the rate of retweets of a post indicate its influence on the Twitter network [7]. To evaluate the relationship between the two measures, we extracted the number of favorites and the number of retweets of all the tweets posted by the candidates and calculated their correlation. As seen in Table II, the amount of favorites and the level of retweets show very high correlation. The overall correlation of all tweets is as high as 0.95. In our data set, the number of favorites of a post is in general larger than the number of retweets. Therefore, the former was chosen as the criterion for top tweets. Our system extracts daily tweets published by the candidates and selects the most influential message from the pool with the highest number of favorites, which we refer as top tweets.

TABLE II.   CORRELATION BETWEEN THE NUMBER OF FAVORITES AND THE NUMBER OF RETWEETS

Candidates	Correlation between favorites and retweets
Clinton	0.99
Sanders	0.95
Cruz	0.92
Kasich	0.98
Trump	0.96
Overall	**0.95**

One factor that needs to be taken into account is the proper interval between the time a tweet is posted and when the volume of favorites is measured, since the latter is a constantly changing variable that accumulates through time. To investigate the scope of this problem, we used a sample of tweets and monitored the revolution of favorites in the next 72 hours after the day those tweets are broadcasted. The level of favorites was observed every 12 hours. We collected the first record at the end of the day (11:59PM PST), which is denoted as $count_0$ in Table III. The next measure was examined after 12 hours, which is represented as $count_1$ in the table, and so on. In other words, for every tweet in our sample pool, a series of numbers was recorded, ranging from $count_0$ to $count_6$.

TABLE III.   OBSERVATION OF THE NUMBER OF FAVORITES

Time	Start of observation	After 12 hours	After 24 hours	After 36 hours	After 48 hours	After 60 hours	After 72 hours
Favorites	$count_0$	$count_1$	$count_2$	$count_3$	$count_4$	$count_5$	$count_6$

To analyze the degree of augmentation of favorites, we calculated E_i, the percentage of increment of the number of favorites compared to the previous record retrieved 12 hours ago:

$$E_i = \frac{count_i - count_{i-1}}{count_{i-1}} \times 100 \ (\%), \ where \ i = 1, 2, …, 6 \quad (1)$$

Table IV shows the average increment of the volume of favorites from after 12 hours of the day of post to after 72 hours. From the table, one can see that the degree of increment begins with 6.16% after 12 hours and drops significantly after the first 36 hours. After 48 hours from the day of post, the change of favorites reduces to below 1% and becomes relatively stable. Therefore, we adopted 48 hours as the waiting interval. In our experiment, volume of favorites was obtained 48 hours after a tweet is published.

TABLE IV.   AVERAGE INCREMENT OF THE NUMBER OF FAVORITES WITH TIME

Time	After 12 hours	After 24 hours	After 36 hours	After 48 hours	After 60 hours	After 72 hours
Percentage of increment (E_i)	6.16%	3.36%	1.48%	0.94%	0.40%	0.25%

B. Top Terms

In this work, we investigate the content of the tweets by extracting unique unigrams from the candidates' accounts. Table V shows the number of unique terms mentioned by candidates in their tweets. Terms are stored in a knowledge base and sorted in the order of their appearances. The top terms play an important role in identifying content produced by each candidate. These keywords serve the best as features to reflect a candidate's political beliefs.

TABLE V.   NUMBER OF UNIQUE TERMS IN TWEETS

Candidates	Number of unique terms
Clinton	11456
Sanders	12738
Cruz	6565
Kasich	8135
Trump	12090

Stop words were filtered out from the list of terms. In this study, we considered three types of stop words. They are 1) common functional terms, such as "the", "but", "and", etc., 2) frequently occurred words in a political campaign without individual characteristics, for instance, "America", "people", "president" and so on, and 3) stop words targeting only certain candidates. For example, Bernie Sanders regularly includes his username @BernieSanders in the retweets. While it reveals nothing about the content of Sanders' tweets, it can be an important indicator when mentioned by other candidates.

C. Followers Profile Analysis

Previous research suggested that the number of supporters on social media can be successful acting as a sign for electoral success [19]. Some candidate, such as Donald Trump, has more than 9 million followers on Twitter as of the time of writing. In this study, we extract the followers' Twitter profiles and examine them in two facets. We are interested in learning the gender ratio among the followers for each candidate, as well as their geographical distribution within the U.S..

Twitter does not ask users to share their gender. However, registered users are required to provide their full name when signing up. To determine the gender of a user, our approach utilizes his profile name. We trimmed the name by removing non-English characters and checked it against a list of 4275 female first names and a list of 1219 male given names provided by the U.S. Census Bureau [20]. We were able to identify gender of 55% followers in total. Figure 2 shows the number of followers with gender identified and the number of followers with unknown gender. The gender of a user cannot be determined based on the profile name in the following situations: 1) the name is not in English, 2) it is an unusual name written in English, for example, a foreign name, 3) the name provided in the user profile is a screenname or nickname that does not exist in the lists in [20].

Another aspect we investigated is the geography of the followers. More specifically, we are interested in learning which U.S. state a follower is based in. Twitter allows its users to list their geographical location in the profiles. In most cases, this information is manually entered by the user. Thus, the geographical data for some users may be missing or incorrect [21].

To analyze the location data, first, they were passed to a list of U.S. states, which contains the full name of each state and its abbreviation. If the state of the location cannot be determined, we then queried it in another list, which is constructed with all the cities and towns in the U.S. together with their mapping states [22]. Using this approach, we were able to identify the location of approximately 47% followers. Table VI lists the size of overall followership for each candidate and the number of followers with location identified.
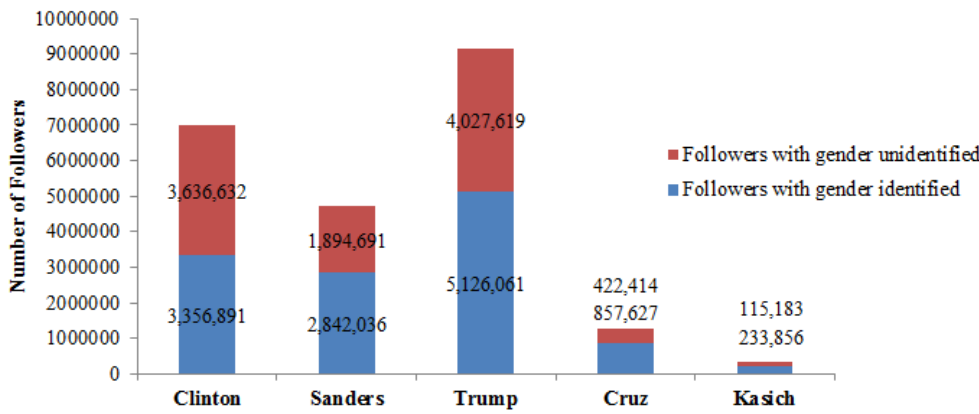
Fig. 2.   Number of followers with gender identified

TABLE VI.   NUMBER OF FOLLOWERS WITH GEOGRAPHICAL LOCATION IDENTIFIED

Candidates	Total number of followers	Followers with location identified
Clinton	6993523	3007215
Sanders	4736727	2652567
Cruz	1280041	627220
Kasich	349039	209423
Trump	9153680	4119156

The location of a user cannot be analyzed if the data is missing or incorrect. Another scenario is that a city along can be interpreted as different places in the U.S.. Two cities from different parts of the country may share the same name. For example, there are Manhattan in New York State, Manhattan in Kansas, Manhattan in the state of Illinois, and so on. If a user specifies his location as Manhattan, we cannot know which Manhattan he is referring to based on this information.

In Section III, we will see that some states have more followers than others for a particular candidate. It is likely that these states are more supportive of the politician. However, there is another possible situation that these states have larger population than others regardless of the public opinions. Therefore, the state population [23] has been considered in the analysis. We used P_i to represent the number of followers per ten thousand among the overall population of a U.S. state:

$$P_i = \frac{follower\_count_i}{population_i} \times 10000 \ (^0/_{000}), \ i \in \{U.S.\ states\} \quad (2)$$

To further investigate the geography of the followers, we applied Jenks natural breaks classification method [24] to all the P_i values of each candidate. Jenks natural break classification method is a clustering algorithm designed for one dimensional data to arrange values into different groups. In our experiment, we split the P_i data into six classes. The classes with higher P_i values represents the more supportive (positive) U.S. states, while the other groups with lower P_i values indicates the less supportive (negative) states. Results of the method will be demonstrated in Section III.

Our system also examines the percentage of high-impact followers of each candidate. It is anticipated that users with a large number of followers also have strong influence in the real world. These users may include popular artists, politicians, and so on.

III.   RESULTS

This section presents the results of our work using the methodology described in Section II. We divide the section into three subtopics: top tweets, top terms and follower profile analysis.

A. Top Tweets

As previously seen in Section II, our system selects the daily top tweets for each candidate based on the volume of favorites received. Table VII shows the average number of favorites collected per tweet for each candidate. The standard deviation reveals that the amount of variation is large among tweets. In the table, we also list the highest number of favorites a candidate has received during the period of observation.

TABLE VII.   AVERAGE NUMBER OF FAVORITES AND HIGHEST NUMBER OF FAVORITES

Candidates	Average favorites	Standard deviation	Highest favorites
Clinton	3881	12751	551388
Sanders	4395	3427	56412
Cruz	901	846	16417
Kasich	409	510	5138
Trump	12911	12830	262457

Figure 3 provides an example of how the system visualizes the change of top tweets. Each data point in the chart represents the number of favorites gained by the top tweet of the day. The line chart is available for demonstration on our website Tweetlitics.net. It provides interaction with the users by displaying the time and content of the top tweet when a user hover the mouse over a data point. For example, the chart in Figure 3 shows the top tweet of July 20, 2016.

The system monitors the trend of evolution in the number of favorites. A burst in the volume is often caused by emerging events or news. For instance, we can see in Figure 3 that Donald Trump's tweet on July 20 collected 221105 favorites, which is more than twice as many as the favorites of other top tweets in the month. The tweet was posted shortly after the wife of Donald Trump, Melania Trump, delivered her speech at the Republican National Convention.
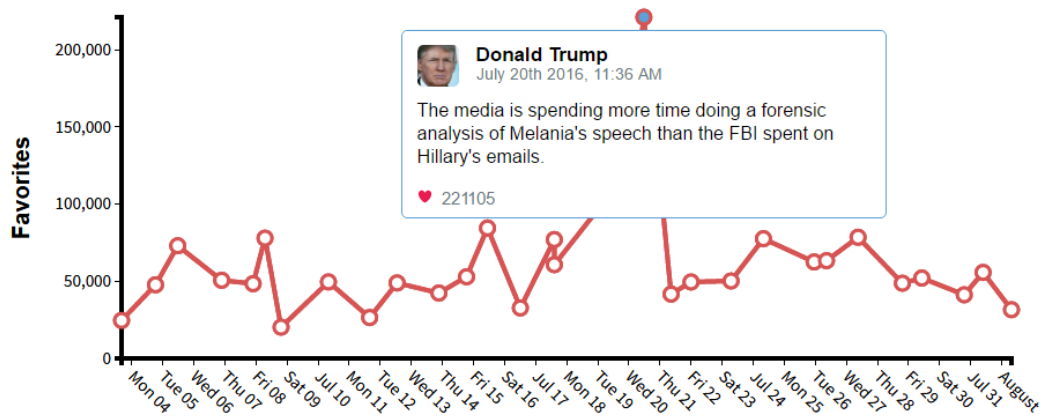
Fig. 3.   Example of top tweets by Donald Trump

B. Top Terms

Our system shows the top 20 terms that are regularly tweeted by each candidate. Results are updated periodically. Table VIII gives a glance of some of the top terms in each candidate's profile.

TABLE VIII.   TOP TERMS BY CANDIDATE

Clinton	Sanders	Cruz	Kasich	Trump
women	health	#gopdebate	ohio	@foxnews
health	wall	obamacare	#gopdebate	@cnn
trump	wage	#iacaucus	clinton	clinton
gun	jobs	iowa	hilary	cruz
rights	working	tax	work	hilary
families	climate	donald	jobs	crooked
#gopdebate	security	trump	hampshire	ted

We found Donald Trump frequently mentioning other candidates, such as Hilary Clinton and Ted Cruz. His name, on the other hand, is also regularly referred by Hilary Clinton and Ted Cruz. Besides, John Kasich often includes Clinton in his tweets. Figure 4 demonstrates the relationship of mentions among the candidates. An arrow pointing from figure *A* to figure *B* represents the mentioning of *B* in *A*'s tweets.
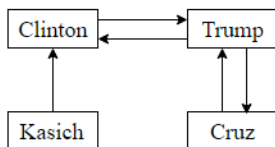

Fig. 4.   Relationship of mentions among candidates

Terms relating to health are addressed by a few candidates, including Clinton, Sanders and Cruz. Words about economics (e.g., wage, jobs, work, tax) are frequently mentioned by Sanders, Cruz and Kasich. Compared to the Republican records, the Democratic profile covers a wider range of topics, such as economics, health, security (e.g., gun, security), rights (e.g., women, rights) and climate.

C. Followers Profile Analysis

As mentioned in Section II, we extracted Twitter profile of each candidate's followers and analyzed their gender ratio. Figure 5 shows the number of male and female followers of each candidate. Interestingly, the four male candidates all have

more male followers than female subscribers. Donald Trump, especially, has 66.7% supporters being male. In contrast, Hilary Clinton, who is the only female candidate, has slightly more female followers (50.4%) than male followers (49.6%).
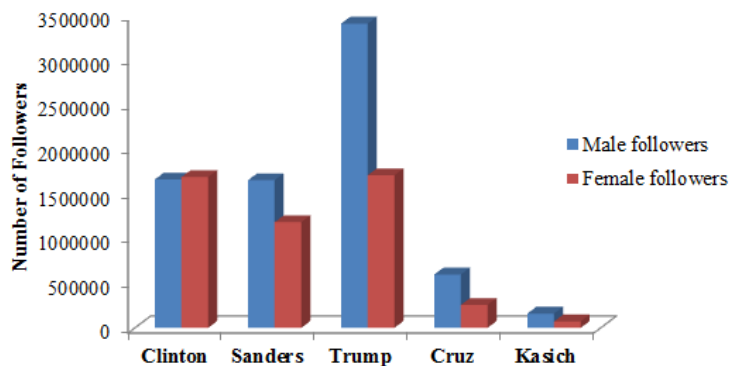

Fig. 5.   Gender ratio of followers

Besides gender ratio, we also parsed the geographical data on the followers' records. As previously discussed in Section II, the number of Twitter followers in each state was examined. The size of followership was then compared with the permyriad (one ten-thousandth) of the total population of a state, and the proportion was calculated. Figure 6 shows the results of all U.S. states for Hilary Clinton.

Table IX gives a summary of proportion of followers for each candidate, including the average proportion among the 50 U.S. states and its standard deviation, the highest proportion and the lowest proportion.

TABLE IX.   STATISTICS OF PROPORTION OF FOLLOWERS

Candidates	Average proportion	Standard deviation	Highest proportion	Lowest proportion
Clinton	24	28	171	3
Sanders	14	17	105	2
Cruz	7	7	38	1
Kasich	3	3	16	1
Trump	33	37	233	5

This paper also compares the proportion of Twitter followers in each U.S. state among the presidential candidates. As seen in Figure 7, each state is marked with a value, which

specifies the highest proportion of followers of that state. In order to reveal the "winner" of each state, we use different colors to represent the candidates. In Figure 7, states with

Hilary Clinton having the highest proportion of followers are marked with blue, while states that follow Donald Trump the most are illustrated in red.
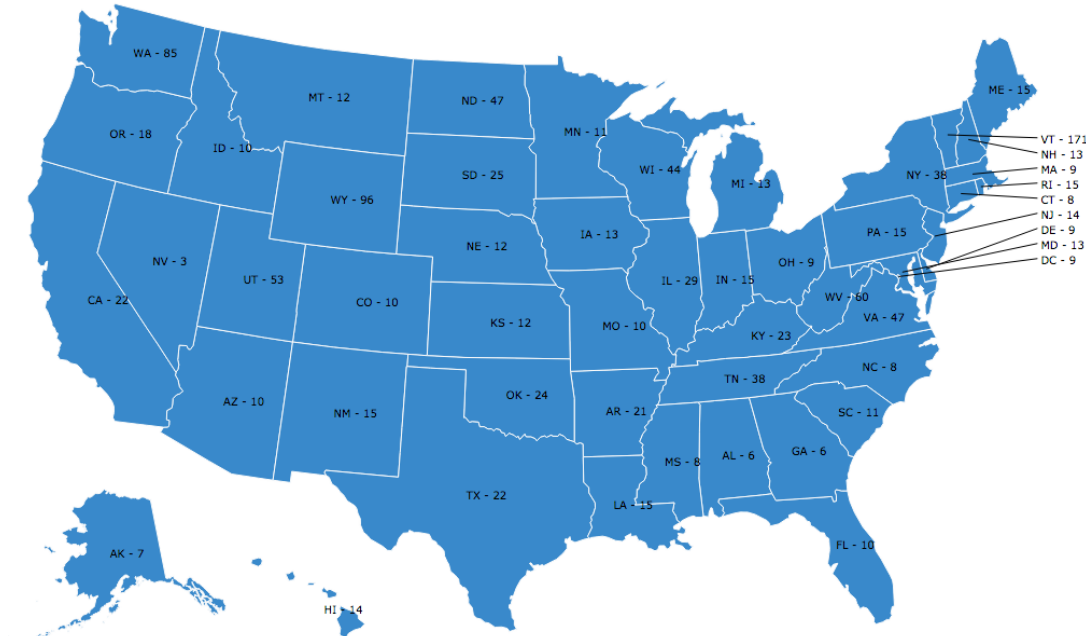

Fig. 6.   Proportion of followers of Hilary Clinton

Fig. 7.   Highest proportion of followers in each U.S. state

To better understand the geography of supporters for each candidate, we clustered the proportion of followers of each state into different classes by applying Jenks natural breaks optimization [24]. Through experiments, we found that

splitting the states into six groups renders the best classification. Table X summaries the results in each class, including the range of proportion of followers and the number of states that fall in that range.

TABLE X.   RESULTS OF PROPORTION OF FOLLOWERS AFTER APPLYING JENKS NATURAL BREAKS METHOD

	Clinton		Sanders		Cruz		Kasich		Trump	
	Range	States	Range	States	States	States	Range	States	Range	States
Group 1	0 – 3	1	0 – 2	1	1	1	0 – 1	1	0 – 1	1
Group 2	3 – 15	32	2 – 7	15	30	30	1 – 5	27	1 – 2	23
Group 3	15 – 29	8	7 – 13	22	11	11	5 – 9	13	2 – 3	18
Group 4	29 – 47	5	13 – 25	6	5	5	9 – 15	5	3 – 5	4
Group 5	47 – 60	2	25 – 40	4	2	2	15 – 21	3	5 – 8	3
Group 6	60 – 71	3	40 – 105	3	2	2	21 – 38	2	8 – 16	2

Another aspect investigated in this work is the social influence each candidate's followers may have in the Twitter community. To study this matter, we traced the number of fans of every Twitter follower. Table XI lists the number of supporters that have a large social impact. Specifically, we examined the percentage of supporters that have more than 10000 fans, 100000 fans, and 1000000 fans respectively. As one can see in the table, despite the smaller overall number of followers (recall in Table VI), Ted Cruz and John Kasich have the largest percentage of influential subscribers. Another interesting finding is that between the two nominees of the Republican Party and the Democratic Party, Hilary Clinton has more affecting supporters than Donald Trump does, although the latter is followed the most on Twitter.

TABLE XI.   STATISTICS OF FOLLOWERS WITH A SOCIAL IMPACT

Candidates	Followers with over 10000 fans	Percentage	Followers with over 100000 fans	Percentage	Followers with over 1000000 fans	Percentage
Clinton	28396	0.41%	4245	0.06%	462	0.007%
Sanders	23313	0.49%	3028	0.06%	294	0.006%
Cruz	8676	0.68%	1288	0.10%	94	0.007%
Kasich	3370	0.97%	506	0.15%	47	0.013%
Trump	27243	0.30%	3670	0.04%	299	0.003%

We developed a website (Tweetlitics.net) to demonstrate the results of our study and the comparisons between candidates. The website is written in JavaScript. AngularJS was used as the framework for the client-side, while Node.JS was adapted for the server-side. We chose Node.JS mainly for its ability of parallel processing in order to deal with the massive amount of Twitter data. MongoDB was used as the knowledge base for data storage.

IV.   CONCLUSIONS AND FUTURE WORK

This paper closely monitors the Twitter activity of the candidates during the 2016 U.S. presidential campaign. We analyzed the interactions between the politicians and their Twitter followers in the retweet/favorite networks. The study collects the real-time tweets published by the candidates and keeps track of the daily top tweets. We found that a burst in the volume of favorites often corresponds to an emerging event. The study also gathers the top terms tweeted by each candidate. These keywords can feature the political focuses of a candidate or a political party. The Democratic Party seems to include a larger range of subjects in their tweets, such as economics, health, rights, security and climate. It is found that some candidates frequently mention others on Twitter. With the extracted top terms, we were able to construct the mention network among the politicians.

This paper also studies the user profiles of the candidates' Twitter supporters. Using the government census data, we examined the ratio of male followers and female subscribers for each candidate. We found that besides Hilary Clinton, the other candidates have the majority of their supporters being male. Moreover, we investigated the geographical distribution of the candidates' Twitter followers. It is found that Donald Trump has the highest number of supporters in most of the U.S. states. Lastly, we studied the proportion of influential supporters of each candidate. We found that despite the larger volume of Twitter followers, Donald Trump has a smaller number of impacting supporters compared to what Hilary Clinton does.

This study has several limitations. First, we have found in the study that the Twitter followers in the presidential election are a small part of the general voters. Comparing the size of followership with the overall population of a U.S. state, on average, only 0.24% of the population follows Hilary Clinton and 0.33% subscribes Donald Trump on Twitter. Second, the paper only considered unigrams as the top terms extracted from the candidates' tweets. In the future, we plan to include n-grams in the analysis.

Additionally, the results of this paper are based on the tweets broadcasted by the presidential candidates. It would be interesting to study the public opinions by steaming tweets published by the general public. Future work includes conducting a sentiment analysis regarding the election by mining the content on Twitter regarding the candidates and other political events.

REFERENCES

[1] C. C. Aggarwal, Social Network Data Analytics. Springer, 2011.

[2] T. Sakaki, M. Okazaki, and Y. Matsuo, "Earthquake shakes Twitter users: real-time event detection by social sensors," Proceedings of the 19th International Conference on World Wide Web, pp. 851-860, 2010.

[3] S. A. Chun, T. Tian, and J. Geller, "Enhancing the famous people ontology by mining a social network," Proceedings of the 2nd International Workshop on Semantic Search over the Web, article no. 5, 2012.

[4] S. A. Chun, and T. Tian, "Social Network-based Entity Extraction for People Ontology." Third International Conference on Advances in Information Mining and Management, 2013.

[5]   A. Archambault and J. Grudin, "A longitudinal study of Facebook, LinkedIn, & Twitter use," Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 2741-2750, 2012.

[6]   S. Bakhshi, D. A. Shamma, and E. Gilbert, "Faces engage us: photos with faces attract more likes and comments on Instagram," Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 965-974, 2014.

[7]   A. Tumasjan, T. O. Sprenger, P. G. Sandner, and I. M. Welpe, "Predicting elections with Twitter: what 140 characters reveal about political sentiment," Fourth International AAAI Conference on Weblogs and Social Media, pp. 178-185, 2010.

[8]   Twitter, https://about.twitter.com/company, retrieved on 09/21/2016.

[9]   D. Boyd, S. Golder and G. Lotan, "Tweet, tweet, retweet: conversational aspects of retweeting on twitter," System Sciences (HICSS), 43rd Hawaii International Conference, pp. 1-10, 2010.

[10]   Twitter, https://support.twitter.com/articles/166337, retrieved on 09/21/2016.

[11]   A. Java, X. Song, T. Finin, and B. Tseng, "Why we Twitter: understanding microblogging usage and communities. Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis, pp. 56–65, 2007.

[12]   D. M. Romero, W. Galuba, S. Asur, and B. A. Huberman, "Influence and passivity in social media," Proceedings of the 20th International Conference Companion on World Wide Web, pp. 113-114, 2011.

[13]   B. J. Jansen, M. Zhang, K. Sobel, and A. Chowdury, "Twitter power: tweets as electronic word of mouth," Journal of the American Society for Information Science and Technology, vol. 60, pp. 1-20, 2009.

[14]   M. J. Choy, M. L. F. Cheong, N. L. Ma, and P. S. Koo, "A sentiment analysis of Singapore presidential election 2011 using Twitter data with census correction, Research Collection School Of Information Systems, 2012.

[15]   J. Borondo, A. J. Morales, J. C. Losada, and R. M. Benito, "Characterizing and modeling an electoral campaign in the context of Twitter: 2011 Spanish presidential election as a case study," Chaos: An Interdisciplinary Journal of Nonlinear Science, vol. 22, no. 2, 2012.

[16]   M. D. Conover, J. Ratkiewicz, M. Francisco, B. Gonc¸alves, A. Flammini, and F. Menczer, "Political polarization on Twitter," Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media, 2011.

[17]   A. Livne, M. P. Simmons, E. Adar, and L. A. Adamic, "The party is over here: structure and content in the 2010 election," Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media, 2011.

[18]   Twitter's REST API, https://dev.twitter.com/rest/public, retrieved on 09/21/2016.

[19]   C. Williams, and G. Gulati, "What is a social network worth? Facebook and vote share in the 2008 presidential primaries," Annual Meeting of the American Political Science Association, pp. 1-17. 2008.

[20]   US Census data, http://www.census.gov/topics/population/genealogy/data/1990_census/1990_census_namefiles.html,retrieved on 09/21/2016.

[21]   S. Chandra, L. Khan, and F. B. Muhaya, "Estimating twitter user location using social interactions - A content based approach," Proceedings of the 2011 IEEE International Conference on Privacy, Security, Risk and Trust and IEEE International Conference on Social Computing, pp. 838-843, 2011.

[22]   List of cities and towns in the U.S., https://developers.google.com/adwords/api/docs/appendix/ geotargeting, retrieved on 09/21/2016.

[23]   Population estimates from the U.S. Cencus Bureau, http://www.census.gov/popest/data/national/ totals/2015/NST-EST2015-popchg2010-2015.html. retrieved on 09/21/2016.

[24]   G. F. Jenks, "The data model concept in statistical mapping", International Yearbook of Cartography, vol. 7, pp. 186-190, 1967.