An Analysis of Twitter Users' Political Views using Cross-Account Data Mining

Shivram Ramkumar[1], Alexander Sosnkowski[1], David Coffman[2], Carol Fung[1], and Jason Levy[3]

[1] Virginia Commonwealth University, Richmond, VA, USA
cfung@vcu.edu,
[2] Duke University, Durham, NC, USA
[3] University of Hawaii, Honolulu, HI, USA

Abstract. During the 2016 US presidential election campaign, we witnessed a polarized population and the election outcome defied predictions from many media sources. In this study, we conducted a follow-up on the political view migration through tracking Twitter users' account activity. The study was conducted by following a set of Twitter users over a three year period. Each year, Twitter user activities were gathered and analyzed by a cross-account data mining algorithm designed by our team. This algorithm computed a numerical political score for each user to match their connections to other users. We observed that users became more moderate each year after the 2016 presidential election on both sides but division still exists.

1 Introduction

The 2016 US Presidential election resulted in a polarized country split between two conflicting and unorthodox political views. Traditional polling methods failed to predict the outcome of the election due to the massive political affiliation change among voters. Few studies have been conducted to follow up the political view migration of a large scale population after the election.

In the United States, more than 70% of the population are social media users [1] and as one of the most popular social network websites, Twitter has exceeded a hundred million daily users and nearly one billion total users [1]. This diverse range on one universal platform leads to the representation of almost any political view. Garimella et. al.[2] studied the behaviors of US Twitter users and found that a trend of increased political polarization has been formed in the past many years. Therefore, as a more accurate predictor for the upcoming election year and possible future elections, it could be beneficial to look at the change in the opinions of users across a relatively long period of time. In this paper, we conducted a study on the political view migration through following up a selected group of Twitter users over the period of three years.

Our contribution of this paper can be summarized as follows: (1) this is the first work of the same kind to analyze political views of Twitter users through political score computation over a period of time. (2) We developed an iterative political score computation algorithm through cross account data mining on Twitter accounts. (3) We conducted our analysis based on real data on Twitter user behaviors.

2 Related Work

Twitter has become a rising social media platform and is used in political campaigns. Much research work [3,4,5,6] has been done to study political activities on Twitter. For example, Ott [7] used Tweets from Donald Trump as an example to study the impact of social media on conventional journalism. Enli [4] discussed the Twitter strategies of polititians during the US 2016 Presidential election campaigns, and pointed out that amateurish yet authentic style in social media became a counter-trend in political communication. Garimella et. al.[2] studied the behaviors of US Twitter users and found that a trend of increased political polarization has been formed in the past many years. Guess et. al [8] surveyed Twitter users and found worrying amount of individual-level discrepancy that shows the division on political view among population.

In our work, we will ascertain the political score of Twitter users based on their Twitter activities and classify them using their score three times between 2017 to 2019. We will then search for changes and developments in the data to see how scores across a large user base

3 Analytical Model

3.1 Users' political activities in Twitter

Users can take several actions to express their political views in Twitter:

1. *Tweets:* Users can create their own Tweet, visible to all their followers, to express their political views.
2. *Hashtags:* Users can hashtag their Tweets so that it will be visible to other users who browse Tweets with that hashtag.
3. *Retweets:* Users can Retweet someone else's to further disseminate the message. When a user Retweets, it is generally a sign that the user agrees with the view carried in the Tweet.
4. *Comments:* Users can add comments to someone else's Tweets and both the Tweet and the comment. The comments can either agree or disagree with the Tweet.
5. *Likes:* Users can click *like* to show they agree with the content of the tweets.

On Twitter, a key feature for users is to follow any person with a public account which will provide them with updates. It is a common practice that Twitter users follow both people they agree with and ones they disagree with in order to keep up to date react to the Tweets. When a user posts a Tweet, they can choose to address it to a specific user by adding the mention symbol

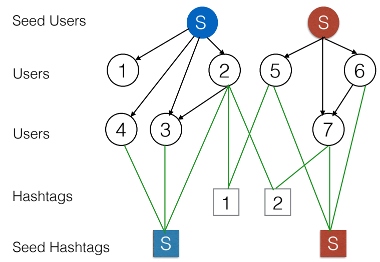

Fig. 1: An example relationship graph of the Twitter users and hashtags

@ followed by the target user's username into the Tweet. For example, after a user send the message "CNN is fake news @realdonaldtrump" by mentioning Donald Trump, the message will be notified to the account *realdonaldtrump*. Another symbol that is frequently used by Twitter users is the hashtag symbol #. Hashtags are often used to label Tweets into certain popular topics areas. For example, $\#NeverTrump$ was commonly used to by liberal users to label their Tweets against Donald Trump. Hashtags make it easy for users to find all posts that have been thus tagged. Finally, a commonly used feature by many Twitter users is the Retweet function which allows a user to take a pre-existing Tweet from a specific user and post it on their own account. In the majority of cases, users what they strongly agree with and would like to share with their own followers. As a result, looking a user's Retweets is a good metric at rating that user's personal views as the Retweeted message usually comes from a much larger account public account with constant ideals.

To study the political view (liberal or conservative) of Twitter users, we investigated a crowd of randomly selected Twitter users by studying their Tweets connected to political biased users and hashtags. More specifically, we investigated their Retweeting and hashtagging behaviors. The political score of each user and hashtag was computed based on a system of accumulated liberal and conservative credits. Through an iterative cross-count data mining methodology, the political scores were updated until the scores converged. We bootstrapped the algorithm using a set of seed users and hashtags that consisted of prominent politicians and heavily political hashtags.

3.2 Political View Computation Model

We illustrate the Twitter activities relationship in Figure 1. Circles are users and rectangles are hashtags. The edges between them represent Retweeting or hashtagging. The top round notes with red or blue colors are the labeled seed users. Red color represents conservative view and blue color represents liberal view. In our case, only prominent political figures such as Donald Trump and Barack Obama are used as seed users. We use a score from 0 to 1 to represent the users' political views. Zero represents complete liberal view and one represents complete conservative view. Seed users are labelled manually. We use a credit system to estimate the political score: the credit is accumulated through users' Retweeting and hashtagging behaviors. Let $U = \{u_1, u_2, ..., u_n\}$ represent the set of political scores of n users, where $0 \leq u_i \leq 1$ is the score of user i. We use $H = \{h_1, h_2, ..., h_m\}$ to denote the political score of hashtags, where $0 \leq h_i \leq 1$ is the score of hashtag i.

Liberal credits and Conservative credits When a user Retweets another user, commonly it is a sign that he/she agrees with the content of the Retweeted message. Therefore we use L_i and R_i to represent the liberal credits and conservative credits of user i, respectively. If a user Retweeted from a liberal user, then liberal credit will be accumulated. Similarly if a user Retweeted from a conservative user then conservative credit is increased.

When a user tags their Tweet using a political hashtag, it is a sign that the user's tweet conveys the same political view as the hashtag. For example, Tweets tagged with #neverTrump likely convey conservative content. Therefore, we accumulate the liberal credit and conservative credit based on hashtagging behavior as well.

User Political score We compute the political score of a user as follows:

$$u_i = \frac{L_i + w_u}{L_i + R_i + 2w_u} \tag{1}$$

where $w_u > 0$ is a parameter to put an initial weight on the computed results to normalize the political scores of users with small credits.

Hashtag political score We compute the political score of a hashtag based on the number of tweets from liberals and conservatives and as follows:

$$h_j = \frac{\sum_{\forall i \leq n} T_{ij} u_i^{t-1} + w_h}{\sum_{\forall i \leq n} T_{ij} + 2w_h} \tag{2}$$

where h_j is the political score of hashtag j; T_{ij} is the number of times user i mentioned hashtag j and u_i is the political score of user i. w_h is an initial weight on the computed results to prevent a large swing of political scores for hashtags with small number of posts from users.

Iterative Algorithm Based on the above method we derive an iterative algorithm to compute the scores of users and hashtags through iterative computation. The initial scores of users or hashtags at round 0 are set neutral (0.5) except seed users or seed hashtags (they are 0 or 1 if they are conservatives or liberals, respectively). The computation stops when the scores converge.

4 Experiments and Results

4.1 Data Collection

We crawled Twitter information such as users' Retweets and hashtags of 60,491 Twitter users with political interest over the three year period (2017 - 2019).

The seed users included prominent politicians such as Hillary Clinton and Donald Trump. These users are foundation for the data collection as they are vocal politicians with strong political beliefs. We crawled their followers as our study user base. We also labeled a set of hashtags as seed hashtags under the condition that they are predominantly political slogans with a one sided meaning. For instance, #neverTrump is considered a liberal hashtag and #MAGA is considered a conservative hashtag. We downloaded the most recent Tweet/Retweet activity for each user in our user base in each year. This way we collected our data set for the same collection of users in the year of 2017, 2018, and 2019, respectively.

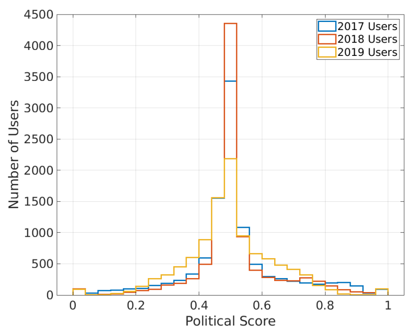

Fig. 2: Political User Scores Across All Three Years (2017 - 2019)

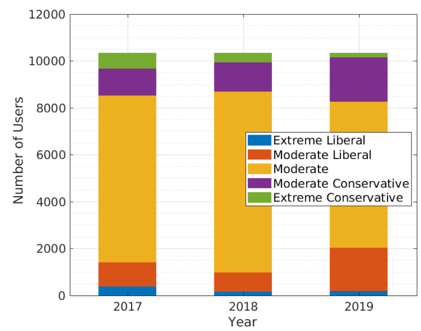

Fig. 3: Political User Scores Across All Three Years (2017 - 2019)

2017-2018	Switched to:				
	Extreme Liberal	Moderate Liberal	Moderate	Moderate Conservative	Extreme Conservative
Extreme Liberal	5.053%	9.043%	72.340%	13.564%	0.000%
Moderate Liberal	2.216%	11.464%	72.447%	13.873%	0.000%
Moderate	1.727%	7.553%	78.689%	12.031%	0.000%
Moderate Conservative	1.401%	8.231%	75.657%	14.711%	0.000%
Extreme Conservative	1.205%	9.187%	75.301%	11.295%	3.012%

Fig. 4: Political view migration 2017 - 2018

2018-2019	Switched to:				
	Extreme Liberal	Moderate Liberal	Moderate	Moderate Conservative	Extreme Conservative
Extreme Liberal	1.136%	17.614%	63.636%	17.614%	0.000%
Moderate Liberal	2.628%	20.401%	56.571%	20.401%	0.000%
Moderate	1.621%	17.626%	62.529%	18.223%	0.000%
Moderate Conservative	2.092%	19.710%	60.097%	18.101%	0.000%
Extreme Conservative	2.169%	15.422%	59.759%	18.554%	4.096%

Fig. 5: Political view migration 2018 - 2019

4.2 Data Analysis

We plotted the political score distribution of a set of users who were able to be scored in all three years (10,343 of them) in Figure 2. We can see that from 2017 to 2018, there was a notable decline of users with scores of [0, 0.2) (Extreme liberal) and (0.8, 1] (Extreme conservative) from 2017 to 2018. Many users with scores in the range of [0.2, 0.4) (moderate left) and (0.6,0.8] (moderate conservative) also shifted closer toward the median; however, in 2019 far more users with moderate liberal and moderate conservative scores exist when compared to either 2017 or 2018. The same year, 2019, also had the largest shift away from neutral values and across all three years. These general observations are also supported in Figures 3, 4, and 5. Figure 3, a migration graph representing changes in political mindset across users, has a rapid decline in users with extreme political beliefs as seen in the loss of blue or green colors on the graph from 2017 to 2018. Furthermore, Figures 4 and 5 provide insight into the migration of users across the political spectrum. Notably, from 2017 to 2018, the majority of the extreme liberal users (72.34%) and extreme conservative users (75.30%) migrated to more moderate positions; while from 2018 to 2019, a sizeable percentage of neutral (moderate) users became either liberal users (17.626%) or conservative users (18.223%). This can be interpreted as a loss of relevancy of political ideals across Twitter from 2017 to 2018 after the 2016 election and the revival of said relevance in 2019 as a new election season begins.

5 Conclusion

In this study, we designed a novel data mining method to identify US political views by analyzing Twitter data. After analysis, we perceived a pattern that users were gradually shifting from more extreme political positions to moderate ones. The shift toward more moderate stances may be tied to the relevancy of the Presidential election (which has the highest turnout of all political elections). As the social importance decreased after 2016, users generally devoted less attention to politics and political users; however, as a new election season approaches, users are adopting more polarized stances once again.

References

1. S. Aslam, "Percentage of u.s. population with a social media profile from 2008 to 2018."
2. K. Garimella and I. Weber, "A long-term analysis of polarization on twitter," *arXiv preprint arXiv:1703.02769*, 2017.
3. M. Ramos-Serrano, J. D. Fernández Gómez, and A. Pineda, "'follow the closing of the campaign on streaming': The use of twitter by spanish political parties during the 2014 european elections," *new media & society*, vol. 20, no. 1, pp. 122–140, 2018.
4. G. Enli, "Twitter as arena for the authentic outsider: exploring the social media campaigns of trump and clinton in the 2016 us presidential election," *European Journal of Communication*, vol. 32, no. 1, pp. 50–61, 2017.
5. A. D. Segesten and M. Bossetta, "A typology of political participation online: How citizens used twitter to mobilize during the 2015 british general elections," *Information, Communication & Society*, vol. 20, no. 11, pp. 1625–1643, 2017.
6. S. Waisbord and A. Amado, "Populist communication by digital means: Presidential twitter in latin america," *Information, Communication & Society*, vol. 20, no. 9, pp. 1330–1346, 2017.
7. B. L. Ott, "The age of twitter: Donald j. trump and the politics of debasement," *Critical Studies in Media Communication*, vol. 34, no. 1, pp. 59–68, 2017.
8. A. Guess, K. Munger, J. Nagler, and J. Tucker, "How accurate are survey responses on social media and politics?" *Political Communication*, vol. 36, no. 2, pp. 241–258, 2019.