An Analysis of Twitter Users' Political Views

DERI DEAN'S EARLY RESEARCH INITIATIVE

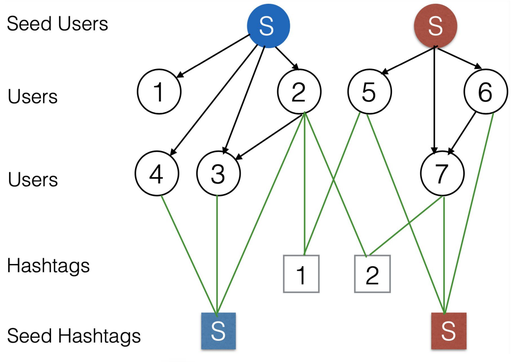
Shivram Ramkumar and Alexander Sosnkowski | Faculty Mentor: Dr. Carol Fung

Introduction

Since its launch in 2006, Twitter has become an eminent platform for rapid communication. More recently, its potential for mass contact with the public has increasingly attracted politicians. Such popularity among politically active figures provides a unique lens to view the contemporary American political climate and incentivizes the development of a system to analyze these digital networks. Our objective was therefore to examine these relationships and predict the affiliations of their members, especially in light of the 2016 presidential election upset that defied numerous public polls, to verify the ability of social media trends to supplement traditional polling.

Analysis Model

The algorithm begins with seed users who have a predetermined score of either 0 or 1 (0 for liberal, 1 for conservative). The Tweets and Retweets of other users establish a network of connections between users and seed users, through which users and hashtags receive a numerical score.



Algorithm

First, we gathered a set of seed users who were clearly identifiable as political extremes, such as certain party officials in Democratic and Republicans groups. Then, a larger randomly sampled population was chosen to assign each member a political score from 0 to 1 based on their Retweets of the seed users by the following algorithms:

$$L_i += \sum_{\forall j | u_j > \theta_l} R_{ij} | u_j - \theta_l |$$

$$R_i += \sum_{\forall j | u_j < \theta_c} R_{ij} | u_j - \theta_c |$$

$$L_i += \sum_{\forall j | h_j > \tau_l} T_{ij} | h_j - \tau_l |$$

$$R_i += \sum_{\forall j | h_j < \tau_c} T_{ij} | h_j - \tau_c |$$

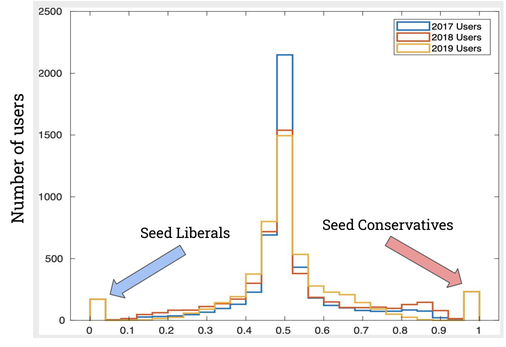
$$u_i = \frac{L_i + w_u}{L_i + R_i + 2w_u}$$

Notation	Description
$U = \{u_1, u_2, ..., u_n\}$	set of political scores of n users
$H = \{h_1, h_2, ..., h_m\}$	set of political scores of m hashtags
$R = \{R_{ij}\}$	retweeting matrix where $R_{ij} = k$ means users i retweeted user j k times.
$T = \{T_{ij}\}$	hashtagging matrix where $T_{ij} = k$ means user i taged hashtag j k times in their tweets
w_u	weight parameter for the user score
w_h	weight parameter for the hashtag score
L_i	liberal credit for user i
R_i	conservative credit for user i
θ_l	threshold to identify literal users
θ_c	threshold to identify conservatives users
τ_l	threshold to identify literal Hashtags
τ_c	threshold to identify conservatives Hashtags

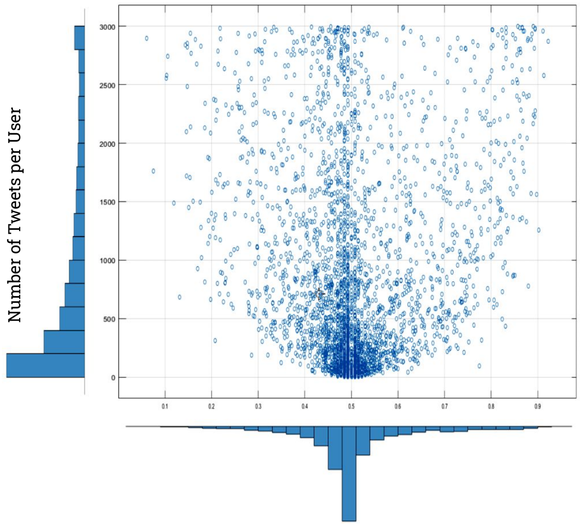
The algorithm accumulates liberal (L) and conservative (C) credit, with an initial 0.5 score via the constant W / 2W. After applying the algorithm iteratively, a score was recorded for each of the sample users.

Results

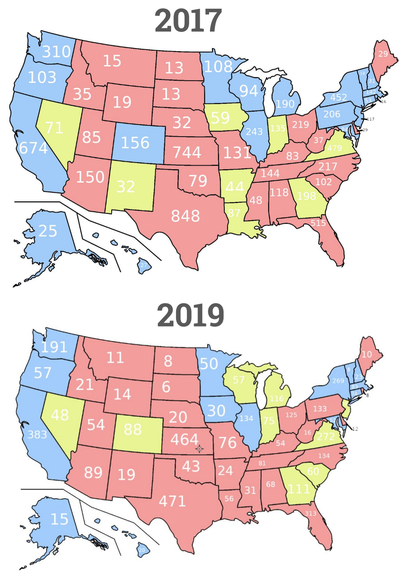
An initial sample of 88,057 users was followed from 2017 – present, with a shrinking sample (inactive, suspended, and deleted accounts):



Political Score (0 Liberal, 1 Conservative)



Political Score (0 Liberal, 1 Conservative)



The graph above shows user score on the x-axis and the number of collected Tweets from said user. Two qualities can be observed:

1. User activity is has greater density along the middle (increased moderate activity)
2. An increase in activity is correlated with greater polarization and political bias

The distribution graph to the left shows:

1. Greater non-political users in 2017 as opposed to 2018 and 2019
2. Increase in polarization in 2018 in extreme conservatives and extreme liberals (n > 0.8 and n < 0.2, respectively)
3. Decreased polarization in 2019, but greater and increasing presence of moderate conservatives and moderate liberals

State-wide distribution maps above resemble the presidential electoral map of 2016.

Conclusion

Our verification process shows that our algorithm can correctly identify users with strong liberal or conservative views with an accuracy of 92.5%. It provides insight into holistic, nationwide trends, as well as potential shifts. Furthermore, the moderate polarization resurgence in 2019 is indicative of the upcoming general election, and the trends on social media of waxing and waning partisanship are likely cyclical in nature, aligned with this 4-year pattern. In the context of the approaching 2020 cycle, it can be noted that user political scores have migrated further from neutrality (0.5) and instead toward political extremas (0 and 1).

Logical next steps involve sampling different user sets, developing more extensive accuracy vetting procedures, and investigating the addition of a linguistic component in the algorithm, e.g. natural language processing and keywords.

Ramkumar S., Sosnkowski A., Coffman D., Fung C., Levy J. (2020) An Analysis of Twitter Users' Political Views Using Cross-Account Data Mining. In: Fiaidhi J., Bhattacharyya D., Rao N. (eds) Smart Technologies in Data Science and Communication. Lecture Notes in Networks and Systems, vol 105. Springer, Singapore

VCU College of Engineering