CALTECH/MIT
VOTING TECHNOLOGY PROJECT

A multi-disciplinary, collaborative project of
the California Institute of Technology – Pasadena, California 91125 and
the Massachusetts Institute of Technology – Cambridge, Massachusetts 02139

ELECTION MONITORING USING TWITTER

Nicholas J. Adams-Cohen
Clare Hao
Cherie Jia
Nailen Matschke
and R. Michael Alvarez
Caltech

Key words*: Election monitoring, election observation, Twitter data*

VTP WORKING PAPER #129

Election Monitoring Using Twitter

Nicholas J. Adams-Cohen, Clare Hao , Cherie Jia, Nailen Matschke, and R. Michael Alvarez

Abstract

As the presence of independent election observation is essential for safeguarding the integrity of democratic processes, numerous tools have been developed for observing and monitoring of elections. While arguably the best way to detect election problems involves in-person monitoring, it is impossible to scale up this approach in large elections. Quantitative tools involving post-election statistical forensics are also useful, but require data unavailable until after an election. In this paper, we discuss the potential of utilizing social media data from Twitter to monitor elections, using the United States 2014 midterms and 2016 primary elections as a test of the methodology.

Keywords

Election monitoring; election observation; Twitter data

Author Note

Nicholas Adams-Cohen is the corresponding author and is a graduate student at the California Institute of Technology. Clare Hao is an undergraduate student at the California Institute of Technology. Cherie Jia, is an undergraduate student at the California Institute of Technology. Nailen Matschke graduated the California Institute of Technology in 2017 and is currently a software engineer for Microsoft. R. Michael Alvarez is a professor of political science at the California Institute of Technology, as well as the co-director of the Voting Technology Project, a joint Caltech-Massachusetts Institute of Technology initiative. The author email addresses are (in order): nadamsco@caltech.edu, chhao@caltech.edu, cjia@caltech.edu, nmatschk@caltech.edu, and rma@hss.caltech. The mailing addresses for all authors: Division of the Humanities and Social Sciences, California Institute of Technology, Pasadena, CA 91125, USA.

We thank colleagues from the Caltech/MIT Voting Technology Project for their discussions, in particular Charles Stewart III. We also wish to thank Pablo Barberá and Jonathan Nagler for their advice, and Eugene Vinitsky for his assistance. Pablo Barberá shared his data on ideological estimates of Twitter users, and we thank him for his willingness to provide that data. Alvarez also thanks the students in his PS 120 class in 2013 and 2014 for discussions that lead to the idea behind this project. Partial support for the research reported in this paper was provided from a grant to the Caltech/MIT Voting Technology Project, by the Carnegie Corporation of New York; the statements made and views expressed in this paper are the responsibility of the authors. The Caltech Summer Undergraduate Research Fellowship (SURF) program supported some of the work of Jia on this project. Hao's work was supported through the 2016 Associates SURF Fellowship.

1. Introduction

The 2000 United States Presidential election produced heightened interest among researchers for the study of election technology and administration. Back in the 2000 presidential election, concerns were raised about the reliability and accuracy of voting machines, as well as with long lines in polling places, voter registration problems, and issues with absentee ballots. Research conducted by the Caltech/MIT Voting Technology Project, launched in the immediate aftermath of the 2000 presidential election, estimated that as many as six million votes may have been lost in the 2000 presidential election because of these issues (Caltech/MIT, 2001). Since 2000, the research community has grappled with the important question of how to quantify the problems that arise in the course of an election, and with developing tools to detect and measure these problems as quickly as possible during and after an election (Alvarez, Ansolabehere, & Stewart, 2005).

Unfortunately, the existing techniques for monitoring elections are highly imperfect. Powerful quantitative tools allow researchers to find voter problems with ballot data, but these techniques only detect voter problems *after* an election. Similarly, survey data allows pollsters to ask citizens directly about their voting experiences, but again these questions can only detect problems that occurred in the past. Qualitatively monitoring elections, wherein individuals directly observe the voting, does allow for the immediate detection of voting issues, but only at polling places where observers are present. Given the cost and difficulty in training individuals to accurately observe polling places, this technique is difficult to scale, making it difficult to detect issues at a national level.

Overall, researchers face trade-offs studying election issues: either they focus on large samples of data gathered from past elections to more accurately pinpoint problems, precluding the detection of election issues in real time, or they gather qualitative data on the day of the election, necessarily resulting in a relatively small-sample of data which makes national-level inferences virtually impossible. It is necessary to develop new tools that can provide information that is quickly actionable with regard to an ongoing election, allowing problems to be solved as they arise.

The goal of our research is to develop and test a tool that utilizes Twitter data to identify, locate, and assess problems with election administration and voting technology as they occur. We believe Twitter data has the potential to balance the trade-off we just described: if voters discuss the problems and issues they run into during the voting process on Twitter, it represents a large source of real-time data from across the United States that is easily accessible. While this source of data has great potential, this method is only feasible if 1) users from across the country discuss election problems as they arise and 2) a tool exists that can actively track and gather these discussions. In this piece, we develop one such tool and assess whether or not the data collected has the potential to detect election issues in real time. This work thus represents a primary but necessary assessment of our methodology, with future work planned that further exploit and utilize these data. We discuss these issues later in the conclusion.

In the following sections, we first elaborate on past approaches to monitoring elections. We then discuss other research that uses Twitter data to study political attitudes and behavior before describing our own methodology. We proceed by presenting some results regarding the 2014 general elections and 2016 presidential primaries. We first demonstrate our approach has merit by demonstrating Twitter users do share many commonly noted election issues online, and then validate our approach by showing discussions of electoral issues are correlated with major events in the electoral cycle, both temporally and spatially.

2. Past Approaches to Monitoring Elections

When eligible citizens attempt to vote, many problems arise, as was documented in research in the immediate aftermath of the 2000 presidential election in the U.S., and has been the source of research in the years since (e.g., Caltech/MIT 2001, Stewart 2013). Voters may attempt to register for an election, only to find on Election day that their name is not in the voter roll. Absentee ballots may not be sent to a voter on time, or the voter may have trouble understanding their absentee ballot. Citizens may also show up to vote, only to find a long line at their polling place or a malfunctioning voting machine. All of these problems have continued to plague elections in the U.S., and abroad.

One of the early attempts to develop a simple-to-compute metric for voter problems with election technology was the residual vote (Caltech/MIT, 2001; Ansolabehere & Stewart, 2005). The residual vote concept is simple; one only needs to know the number of ballots that were cast for a particular race and the total number of ballots cast in that election. Dividing the first by the second gives the residual vote, which is an important diagnostic measure of the reliability of both in-person voting technologies (Ansolabehere & Stewart, 2005; Sinclair & Alvarez, 2004) and voting-by-mail (Alvarez, Beckett, & Stewart, 2012). But the residual vote measure is imperfect, and does not allow the researcher to study the many other problems that a voter might experience as they try to vote.

Others have tried using polls or surveys to collect more detailed information regarding the voting experience. Some available datasets, including the U.S. Census Bureau's Current Population Survey (CPS) voter supplement, have information that can be used to assess a few aspects of the voting experience. The CPS question on why voters do not vote has been used to estimate the number of ballots potentially lost due to issues with voter registration and polling places (Caltech/MIT, 2001). Others have used recent large-scale surveys, like the Cooperative Congressional Election Survey (CCES) or the Survey of the Performance of American Elections (SPAE) to study different aspects of the voting experience, like convenience voting (Alvarez, Levin, & Sinclair, 2012) or long lines at polling places (Stewart, 2013). Still other scholars have implemented comprehensive surveys in individual states, allowing for a detailed analysis of a voter's Cexperience in the context of a single state's administrative, legal, and procedural setting (Alvarez, Atkeson, & Hall, 2012; Atkeson et al., 2010; Atkeson et al., 2014). The state-level approach has yielded important studies about how administrative procedures can be implemented differently across a state's electoral jurisdiction, due to the discretion accorded to individual poll and election workers (Atkeson et al., 2010). These survey-based approaches for studying the voting experience are important, and they are clearly yielding useful and significant results that advance academic research and the state of policymaking in this area, but only identify the types of problems voters are *likely* to face, failing to detect when and where these problems actually occur on Election Day.

In addition to survey data, other researchers have used qualitative methods to study the voting experience by using in-person election monitoring and observation. These approaches were largely developed and deployed internationally (e.g., Bjornlund, 2004; Hyde, 2007, 2011), but they have been used increasingly in the U.S (Alvarez, Atkeson, & Hall, 2012). In-person election monitoring can generate a great deal of very useful qualitative data, and can be used to help detect and possibly deter election fraud (Hyde, 2007). However, in-person election monitoring can be very expensive to implement in a large election, as it is costly to train a large number of observers and deploy them during a large national election (like a U.S. presidential election). It is also logistically difficult, if not impossible, to cover a large national election; in particular, many U.S. states have laws and procedures in place which make direct, in-person, observation difficult to conduct. Thus, while qualitative approaches can yield a great deal of useful information about the administration of an election, they are costly and have coverage problems.

While these different approaches for studying the voting experience are producing a great deal of useful data and information, these methods have difficulties detecting and resolving problems in real time during an election. The toolkit that researchers and election administrators have developed and used in the past do not provide information that is quickly actionable with regard to an ongoing election in which problems arise. To solve this issue, we attempt to build and test a new methodology to monitor elections, leveraging social media data as a source of real-time data that allows one to find election problems as they occur.

3. Using Twitter Data to Study Public Opinion and Political Behavior

With Twitter's widespread use (as of October 2016, Twitter reports having 313 million active monthly users, with at least 500 million tweets per day), there have been a number studies that use Twitter data to analyze political issues.[1]

Of the studies that utilize Twitter data, many use social media text data as a proxy for public opinion surveys. One prominent study by O'Connor et al. (2010) collects tweets that mention either John McCain or Barack Obama and compare the sentiment expressed in these messages to conventional opinion polls. Their research reveals that analyzing Twitter data, even with relatively simple sentiment extraction algorithms, provides a good approximation for traditional opinion surveys. Beauchamp (2017) expands on this work by showing Twitter data can track state-level polling data. Other studies have demonstrated that there are measurable correlations between the public “mood” of Twitter messages and major economic/political events (Bollen, Mao, & Pepe, 2009) as well as the closing values of the stock market (Bollen, Mao, & Zeng, 2011). While the majority of this research relies on analyzing the content of the Twitter message itself to infer the user's sentiment, another line of research has shown that analyzing the network structure of a Twitter user can accurately reveal their political leaning (Barberá, 2015).

Other researchers have looked at how Twitter and other social media platforms encourage or discourage political behavior. A meta-analysis of 36 studies revealed that, while the vast majority of research demonstrates a positive relationship between social media use and civic/political participation, only half of the

[1] For up-to-date information on Twitter use, see https://about.twitter.com/company.

coefficients across these studies were statistically significant (Boulianne, 2015). Furthermore, social media use appears more strongly correlated with protest activities compared with election-campaign participation, and that these effects are strongest in studies including more younger users. Overall, social media use does not appear “transformative” with regard to political participation, with “metadata rais[ing] serious doubts about causal effects” (Boulianne, 2015, p. 534).

Any research using social media data to study political behavior must contend with the fact that Twitter users do not represent a random sample of the population at large. In the United States, public opinion as measured on Twitter tends to display a greater liberal and pro-Democratic bias when compared with more traditional polling data (Mitchell & Hitlin, 2013). Moreover, when analyzing Twitter data, researchers need to be wary that the tweets analyzed do not include “spam” sent to deliberately manipulate how popular a candidate appears on social media or otherwise disrupt political communication (Ratkiewicz et al., 2011; Thomas, Grier, & Paxon, 2012).

Our study, rather than using social media as a measure of public sentiment or finding the impact of social media on behavior, uses Twitter as a tool to *monitor* the election process. The previous studies mentioned above demonstrate that there are a large number of Twitter users using the service to, in one way or another, discuss political events. We hope to leverage this tendency and detect potential issues with the voting process.

4. Election Problems in 2014 and 2016, As Seen Through Twitter

We discuss in the Supplementary Materials the details about how we collect the tweets for our analysis, but the general approach involves collecting all tweets mentioning keywords related to specific voting problems, as well as information about the user sending the tweet. For an initial examination of the data we collected, we present in Table 1 the counts of tweets for each keyword, as well as their incidence rate from our total collection of election tweets during the April 2014 to February 2015 period in the second and third columns of the table, and for the April 2016 through June 2016 period in the fourth and fifth columns of the tables. The keywords in Table 1 are arrayed in descending order (based on their 2014-2015 incidence), with the most mentioned keyword (“Ballot”) at the top of the table and the least mentioned keyword (“Voting identification”) at the bottom of the table.

It is important to note that one keyword, “Ballot”, was by far the most frequently tweeted keyword in both sets of data, occurring in 66.19% of the tweets we collected during the first period and in 79.45% of the tweets in the later period. The next two keywords, in terms of frequency, were “Voter id” (appearing in 17.17% of tweets collected in 2014-2015 and 6.56% of tweets collected in the spring of 2016) and “Early voting” (appearing in 11.58% of the tweets collected in 2014-2015 and 6.11% of those from the spring of 2016). In 2014-2015, we collected a sizable number of tweets referencing “absentee ballot”, with 32,875 tweets (1.77% of all tweets), and “voting machine”, with 27,311 tweets (1.47% of all tweets collected). In 2016, the incidence of some of the less frequent tweeted keywords changed somewhat, with more tweets in that period about “vote by mail” (0.98%), “Provisional ballot” (1.99%), and “poll worker” (1.57%).

It is also important to see that some of the keywords we monitored did not occur with much frequency in the data we collected. For example, “voting identification” only appeared in 12 tweets in 2014-2015 and once in 2016, while “polling place line” only appeared in 60 tweets collected in 2014-2015 and four times in 2016. Tweets referencing “pollworker” or “precinct line” were also relatively rare in our data, occurring only in a few hundred instances.

In summary, this first pass through our collected data helps to demonstrate that this approach to collecting information about potential issues in the administration of an election has merit. During the 2014-2015 period of study, we collected and categorized 1,856,301 tweets, though most regarded a very small number of the keywords we used. In the shorter period we covered in 2016, we collected just over a million tweets (1,046,957). It is striking to us that some of the keywords that refer to specific administrative issues that have been widely discussed in both academic research and the media were not frequently mentioned in the tweets we collected during this period.

Next, we examine the tweets we collected in each period by the inferred ideological stance of the Twitter user. To do this, we use the ideological ideal point estimates from Barberá (2015). Barberá provided us a list full list of Twitter users with an estimated ideological score, as well as their Twitter user identification number. We matched Barberá's data to ours by Twitter user identification number, and we obtained relatively high match rates: 28% of users (44% of tweets) matched in our 2014-2015 sample, while 32% of users (54% of tweets) matched our more recent 2016 sample. We classified each of the matched users as being either Conservative or Liberal depending on Barberá's point estimate. We present keyword frequencies by the ideology of the Twitter user in Tables 2 and 3. In Table 2, we see some important differences in the keyword frequencies. We see a higher incidence of tweets about voting machines among conservative Twitter users than among liberal Twitter users in the 2014-2015 data. We also see a higher incidence for the keywords "Ballot" and "Voter id" among conservatives. Among the liberal users in our 2014-2015 sample, we see a higher incidence of the keyword "Early voting."

Next, in Table 3, we provide the same quantities of interest for the data collected during the late 2016 primary season. Here we see that, among conservatives, there was a much higher incidence for the keyword "Ballot." But among the liberal Twitter users, we see a greater keyword incidence for "Early voting" and "Provisional ballot."

One inference we make is the differences between 2014-2015 and the 2016 primary season keyword counts reflect the changing concerns between liberal or conservative Twitter users. Ideologically conservative Twitter users seemed to be more concerned about voting machines and voter id in 2014-2015; ideologically liberal Twitter users seem more concerned about early voting during the same time period. The conservative emphasis in early 2016 shifts to the keyword "Ballot", while the liberal emphasis shifts to "Provisional ballot" and "Early voting." However, at this point inferences like these need to be taken with a grain of salt—we are not comparing any individual-level changes in Twitter behavior, nor are we analyzing the sentiment of the discussion surrounding the keyword in each of the tweets we collect. Such analyses are possible, however, and we are actively involved in each line of research.

5. Election Issues and Election Events

An important facet of our data is that we know the day that each collected tweet was published. We aggregated the tweets we collected by week, and, as in the previous section, divide the data in two distinct periods: the weeks leading up and immediately following the November 4, 2014 U.S. elections and the latter part of the primary season in early 2016. We provide an initial look at all of the tweets for each period in Figure 1.

The data provided in Figure 1 helps to validate the use of data like these for monitoring elections and election administration. The upper panel of the figure gives data that spans the November 2014 general elections. Note that Election Day is in week 13 in the time-series graph—we see a steady buildup of tweets with our selected keywords to the week of the November 13 election, the time-series peaks in that week, and falls off considerably thereafter. This provides validation for this approach—clearly we are collecting information from Twitter that relates to the election cycle in the U.S.

The lower panel provides counts for our Twitter data during the latter part of the 2016 primary cycle. The data presented here runs from April 18, 2016 through June 14, 2016. It covers the latter half of the 2016 presidential primary process, in particular spanning a series of important presidential and statewide primaries. Early in this period, in weeks 2-4, we see a significant amount of tweets mentioning our keywords. During this period, in week two (starting on April 26, 2016) of our series five states held elections: Connecticut, Delaware, Maryland, Pennsylvania, and Rhode Island. In week three, Indiana had its primary; in week four there were republican primaries in West Virginia and Nebraska. We see a second spike later in our series which is consistent with events in this electoral cycle—in week 8 (starting June 6, 2016) six states held primaries or caucuses: California, Montana, New Jersey, New Mexico, North Dakota (Democratic), and South Dakota. Again, we take the correlation between the time-series patterns seen in the lower panel of Figure 1 as validation of this approach to monitoring the discussion about elections in the U.S.

The tweets we collected fall into four general types of election issues, and we have categorized the specific keywords as follows: (1) Election Day Voting (Provisional ballot, Voting machine, Ballot); (2) Polling Places (Polling place line, Precinct line, Pollworker, Poll worker); (3) Remote Voting (Absentee ballot, Mail ballot, Vote by mail, Voting by mail, Early voting); and (4) Voter ID (Voter identification, Voting identification, Voter ID). In Figure 2, we provide a graph that shows the trends of the four issue categories over the period we collected these tweets. This figure shows the number of tweets mentioning keywords that fall in each category during a particular week.

We see that, when observing the raw frequencies in Figure 2, certain issues are discussed more frequently than others, with nearly all tweets in the 2016 sample coming from the Election Day category. As we want to better visualize when specific issues become more or less important temporarily, in Figure 3 we graph the relative frequency of each weekly. We visualize this by using the y-axis to chart the percent of tweets from a category that were sent in that week.

Observing Figure 3, we find that each of the four types of election issues peaked the week before the November 2014 general elections in the U.S. But more importantly, we see that the keywords have specific distributions that track when these particular issues should have been most salient. Note that the polling place tweets appear with relative low frequency through week nine, steadily increase in weeks ten and eleven, and skyrocket in week twelve; thereafter tweets mentioning the keywords associated with polling places drop to essentially zero occurrences per week as a fraction of the total polling place mentions. We also see that Election Day voting keywords have a similar dynamic, though not as sharply increasing in week twelve as polling place tweets.

However, we see that tweets about the keywords in the remote voting category peak after week eight, through week eleven, but then drop considerably in week twelve and thereafter. This makes sense—tweets in that category would concern issues like voting by mail and early voting, and thus we should see most tweets about those topics immediately before the election. Interestingly we see that tweets about voter identification spike in weeks eight and nine of our study period, *before* the election, though we do see a slight upward inflection in the trend line the week of the election. This indicates that the voter identification issue, prevalent during the 2014 November general election in the U.S., was being discussed on Twitter prior to the 2016 election. However, the fact that these discussions peaked before the election itself demonstrates that discussions concerning voter identification might not be related to problems voters faced when they went to cast their ballot, but rather was simply a topic of interest in the weeks preceding Election Day.

6. Tweets by State

Another way we demonstrate the ability to use Twitter data to monitoring election administration problems was to look at patterns of tweets by geography. Many Twitter users provide some geographic location information in their user profiles, and in order to separate users into different states used a series of regular expressions to search for the names of U.S. states or cities, as well as a few common misspellings and abbreviations. Using this approach, we were able to assign a state to approximately a third of the tweets in each of the two databases: 37.0% in the 2014-2015 data, and 37.6% in the 2016 data.[2]

[2] There are of course some potential issues in this approach. Our regular expressions do not involve an exhaustive list of all cities in each state, and some city names could refer to cities in multiple states and/or countries. We experimented with probabilistic classification of more ambiguous geographic location information by trying to use other information in the user profile or the tweets themselves to refine the potential accuracy of the classification, but we do not use those approaches here.

We provide in Figure 4 and Figure 5 maps of the continental U.S., along with various means of visualizing the distribution of tweets across the states. In Figure 4, we provide the data for 2014-2015, while in Figure 5 we provide the maps for the 2016 collection of tweets. The upper four maps provide the raw count of numbers of tweets in each of the four issue areas by state. The bottom four maps provide these counts divided by state population. In each map, the darker the shading the greater the number of tweets (the top four graphs) or the greater the number of tweets per capita (the bottom for graphs).

Starting with Figure 4 and the top panel, not surprisingly we see that much of the tweets are concentrated in some of the larger states. There is a relatively high number of tweets in our 2014-2015 data on Election Day voting in Texas and California, as well as in some other states, especially battleground states like Colorado, Indiana, and Ohio. We see in the other maps in this upper panel many tweets on remote voting and polling places form Texas and California—and a good number of tweets on voter ID coming from Texas, California, and Wisconsin. This latter result is worth noting, in particular as in Texas and Wisconsin voter identification laws have been under considerable debate.

In the lower panel of Figure 4, we see that the per capita presentation is somewhat different. Here we see that a few states stand out in ways that make some sense given the dynamics of the 2014-2015 election cycle. We see a high per capita tweet rate for Election Day voting in Colorado, which is interesting as in 2014 Colorado was experiencing their first major election with same-day voter registration, as well as with voting by mail (note that we do not see Colorado as having a high rate of remote voting tweets, however). We again see that Wisconsin shows a lot of tweets on voter identification, which again makes sense, given the debates in that state over strict voter identification requirements.

Turning to Figure 5, where we present the data for the spring of 2016, some different patterns emerge. In the upper panel, we again see higher levels of tweets about all four of the election issues coming from states with larger populations, in particular Texas and California. But in the lower panels, where we look at the data per capita, we see some different patterns emerge. Recall that during the period that this data was collected, there were important presidential and statewide primaries: in week 2, Connecticut, Delaware, Maryland, Pennsylvania, and Rhode Island; in week 3, Indiana; week 4, West Virginia and Nebraska (Republican); and week 8, California, Montana, New Jersey, New Mexico, North Dakota (Democratic), and South Dakota. Many of these states show high per capita rates of tweeting about election issues, for example Montana (Election Day voting), Indiana (also Election Day voting, also remote voting and polling places), West Virginia (remote voting), and Vermont (Election Day voting, polling places, and remote voting). This is important confirmation that people are using Twitter to express their opinions about the elections occurring in their states.

7. Conclusions

In this paper, we presented the concept of using Twitter for election monitoring. The value of Twitter for this purpose is that it provides real-time information about the experiences and observations of people who are voting, observing the election, or commenting on what they are hearing from friends, family, and acquaintances. We have shown in this paper through a series of different validations that our approach has merit—the keywords we followed produce results that follow temporal, geographic, and ideological lines, which helps to increase our confidence in the use of this approach to collect useful information about elections, while they happen.

Of course, there is much more work that needs to be done to refine the use of social media, like these, for in-depth use to evaluate the integrity of an election. Building a better dictionary of keywords, or a more automated and flexible process for identification of keywords, are needed. Furthermore, research into the effective partitioning of tweets or other social media posts that indicate a problem (or positive experience with the voting process) is necessary. However, we believe that the preliminary evidence provided here indicates that there is utility in the use of social media data for election monitoring efforts, and we encourage the further development of methods to collect and analyze data like these, and to integrate those data with other types of qualitative, quantitative, and forensic tools.

The next steps in the research agenda involve working on a real-time Twitter reporting applications, and deeper analyses of the data we have been collecting since 2014. Of course, we do not see that this type of data serves to replace existing qualitative nor qualitative election monitoring or election analytic activities. Rather, we argue that these data provide another tool that researchers, election officials, and the interested public can use to better understand the integrity of an election, as well as to collect useful information about the experiences of voters so as to help improve election administration in the future.

References

Alvarez, R. M., Ansolabehere, S., & Stewart III, C. (2005). Studying elections: Data quality and pitfalls in measuring the effects of voting technologies. *Policy Studies Journal*, *33*(1), 15-24.

Alvarez, R. M., Atkeson, L. R., & Hall, T. E. (2012). *Evaluating elections: A handbook of methods and standards*. New York: Cambridge University Press.

Alvarez, R. M., Beckett, D., & Stewart III, C. (2012). Voting technology, vote-by-mail, and residual votes in California, 1990-2010. *Political Research Quarterly*, *66*(3), 658-670.

Alvarez, R. M., Levin, I., & Sinclair, J. A. (2012). Making voting easier: Convenience voting in the 2008 presidential election. *Political Research Quarterly*, *65*(2), 248-262.

Ansolabehere, S., & Stewart III, C. (2005). Residual votes attributable to technology. *The Journal of Politics*, *67*(2), 365-389.

Atkeson, L. R., Alvarez, R. M., Hall, T. E., & Sinclair, J. A. (2014). Balancing fraud prevention and electoral participation: Attitudes toward voter identification. *Social Science Quarterly*, *95*(5), 1381-1398.

Atkeson, L. R., Bryant, L. A., Hall, T. E., Saunders, K., & Alvarez, R. M. (2010). A new barrier to participation: Heterogeneous application of voter identification policies. *Electoral Studies*, *29*(1), 66-73.

Barberá, P. (2015). Birds of the same feather tweet together: Bayesian ideal point estimation using twitter data. *Political Analysis*, *23*(1), 76-91.

Bjornlund, E. (2004). *Beyond free and fair: Monitoring elections and building democracy*. Washington, DC: Woodrow Wilson Center Press.

Beauchamp, Nicholas (2017). Predicting and Interpolating State-Level Polls Using Twitter Textual Data. *American Journal of Political Science*. 61(2), 490-503.

Bollen, J., Mao, H., & Pepe, A. (2009). Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena. In Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media.

Bollen, J., Mao, H., & Zeng, X.-J. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, *2*(1), 1-8.

Boulianne, S. (2015). Social media use and participation: A meta-analysis of current research. *Information, Communication & Society*, *18*(5), 524-538.

Caltech/MIT Voting Technology Project. (2001). Voting: What is, what could be. Pasadena, CA: Retrieved from http://vote.caltech.edu/content/voting-what-what-could-be

Davis, C. A., Varol, O., Ferrara, E., Flammini, A., & Menczer, F. (2016). BotOrNot: A system to evaluate social bots. arXiv: 1602.00975 [cs.SI]

Howard, P. N., Bolsover, G., Kollanyi, B., Bradshaw, S., & Neudert, L.-M. (2017). Fake news and bots during the U.S. election: What were Michigan voters sharing over Twitter? COMPROP DATA MEMO 2017.1/26 March 2017. Retrieved from http://politicalbots.org/wp-content/uploads/2017/03/What-Were-Michigan-Voters-Sharing-Over-Twitter-v2.pdf

Hyde, S. D. (2007). The observer effect in international politics: Evidence from a natural experiment. *World Politics 60*(1), 37-63.

Hyde, S. D. (2011). *The pseudo-democrat's dilemma: Why election observation became an international norm*. Ithaca, NY: Cornell University Press.

King, G., Lam, P., & Roberts, M. In Press. Computer-assisted keywords and document set discovery from unstructured text. *American Journal of Political Science*.

Mitchell, A., & Hitlin, P. (2013). Twitter reaction to events often at odds with overall public opinion. Pew Research Center. Retrieved from http://www.pewresearch.org/2013/03/04/twitter-reaction-to-events-often-at-odds-with-overall-public-opinion/

Morstatter, F., Pfeffer, J., Liu, H., & Carley, K. M. (2013). Is the sample good enough? Comparing data from Twitter, streaming API with Twitter's firehose. In Proceedings of ICWSM.

O'Connor, B., Balasubramanyan, R., Routledge, B. R., & Smith, N. A. (2010). From tweets to polls: Linking text sentiment to public opinion time series. In Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media.

Ratkiewicz, J., Conover, M. D., Meiss, M., Goncalves, B., Flammini, A., & Menczer, F. (2011). Detecting and tracking political abuse in social media. In Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media.

Sinclair, B., & Alvarez, R. M. (2004). Who overvotes, who undervotes, using punchcards? Evidence from Los Angeles County. *Political Research Quarterly*, *57*(1), 15-25.

Stewart III, C. (2013). Waiting to vote in 2012. *The Journal of Law & Politics*, *XXVII* (4), 439-463.

Subrahmanian, V. S., Azaria, A., Durst, S., Kagan, V., Galstyan, A., Lerman, K., Zhu, L., Ferrara, E., Flammini, A., Menczer, F., Stevens, A., Dekhtyar, A., Gao, S., Hogg, T., Kooti, F., Liu, Y., Varol, O., Shiralkar, P., Vydiswaran, V., Mei, Q., & Hwang, T. (2016). The DARPA Twitter Bot Challenge. arXiv: 1601.05140 [cs.SI]

Thomas, K., Grier, C., & Paxon, V. (2012). Adapting social spam infrastructure for political censorship. In Proceedings of the 5th USENIX Conference on Large-Scale Exploits and Emergent Threats.

Table 1. Twitter keyword counts.

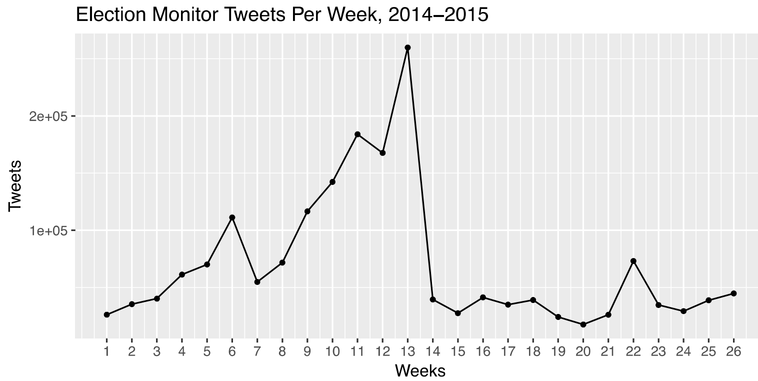
	2014 - 2015		2016	
Keyword	**Count**	**% of Total**	**Count**	**% of Total**
Ballot	12,286,684	66.19	831,830	79.45
Voter id	318,694	17.17	69,657	6.65
Early voting	215,049	11.58	63,987	6.11
Absentee ballot	32,875	1.77	6,424	0.61
Voting machine	27,311	1.47	14,510	1.39
Vote by mail	10,430	0.56	10,215	0.98
Poll worker	8,067	0.43	16,436	1.57
Mail ballot	5,434	0.29	8,621	0.82
Voter identification	3,718	0.20	1,995	0.19
Voting by mail	2,617	0.14	1,744	0.17
Provisional ballot	2,331	0.13	20,866	1.99
Precinct line	530	0.03	106	0.01
Pollworker	489	0.03	561	0.05
Polling place line	60	0.00	4	0.00
Voting identification	12	0.00	1	0.00
Total	1,856,301		1,046,957	

Table 2. Twitter keyword counts, 2014-2015, for Conservatives and Liberals.

	Conservatives		Liberals		Difference
Keyword	**Count**	**% of Total**	**Count**	**% of Total**	
Ballot	166,427	51.42	239,316	49.97	1.45
Voter id	87,460	27.02	117,392	24.51	2.51
Early voting	43,343	13.39	93,131	19.45	-6.05
Absentee ballot	6,401	1.98	11,982	2.50	-0.52
Voting machine	14,020	4.33	2,948	0.62	3.72
Vote by mail	1,403	0.43	5,363	1.12	-0.69
Poll worker	2,439	0.75	2,711	0.57	0.19
Mail ballot	655	0.20	2,562	0.53	-0.33
Voter identification	581	0.18	933	0.19	-0.02
Voting by mail	430	0.13	1,155	0.24	-0.11
Provisional ballot	399	0.12	1,168	0.24	-0.12
Precinct line	12	0.00	20	0.00	0.00
Pollworker	86	0.03	208	0.04	-0.02
Polling place line	5	0.00	38	0.01	-0.01
Voting identification	2	0.00	4	0.00	0.00
Total	323,663		478,931		

Table 3. Twitter keyword counts, 2016, for Conservatives and Liberals

Keyword	Conservatives		Liberals		Difference
	Count	**% of Total**	**Count**	**% of Total**	
Ballot	175,901	80.75	204,016	68.80	11.95
Voter id	17,509	8.04	25,570	8.62	-0.59
Early voting	13,446	6.17	27,917	9.41	-3.24
Absentee ballot	1,066	0.49	2,240	0.76	-0.27
Voting machine	5,191	2.38	4,247	1.43	0.95
Vote by mail	911	0.42	5,258	1.77	-1.35
Poll worker	2,216	1.02	7,664	2.58	-1.57
Mail ballot	572	0.26	5,095	1.72	-1.46
Voter identification	243	0.11	183	0.06	0.05
Voting by mail	88	0.04	985	0.33	-0.29
Provisional ballot	657	0.30	13,044	4.40	-4.10
Precinct line	10	0.00	7	0.00	0.00
Pollworker	31	0.01	308	0.10	-0.09
Polling place line	0	0.00	2	0.00	0.00
Voting identification	0	0.00	0	0.00	0.00
Total	217,841		296,536		



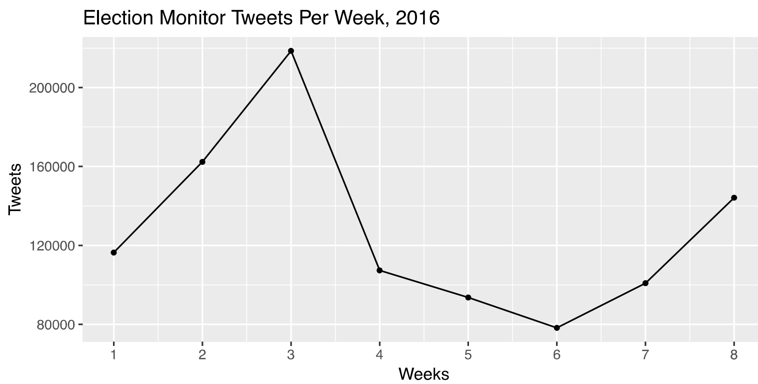


Figure 1. Tweets per week

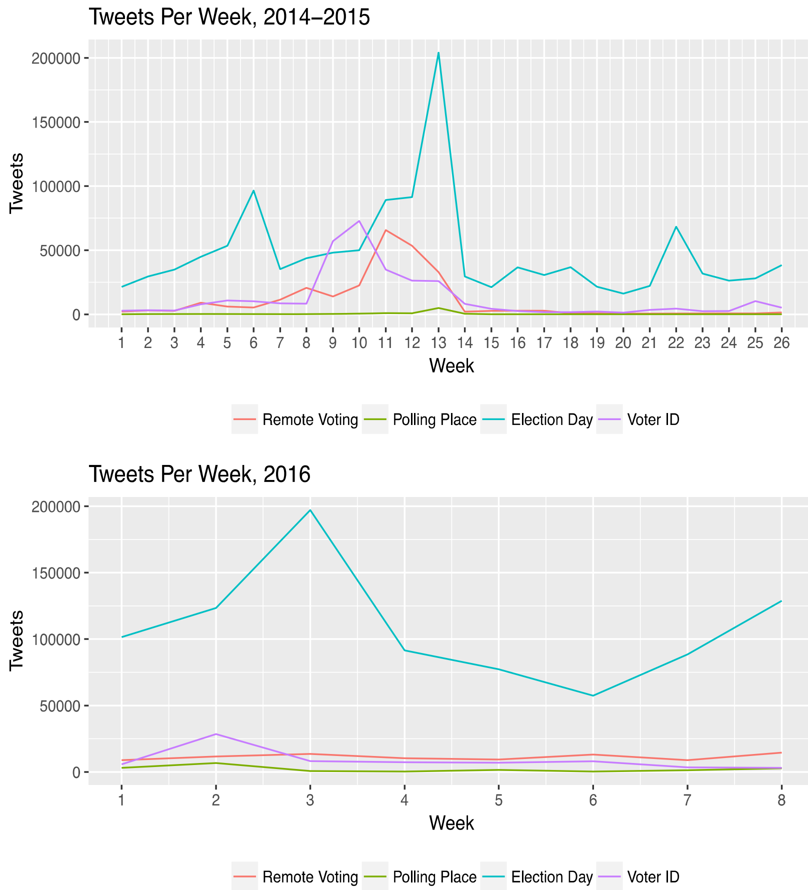


Figure 2. Tweets per week by issue area

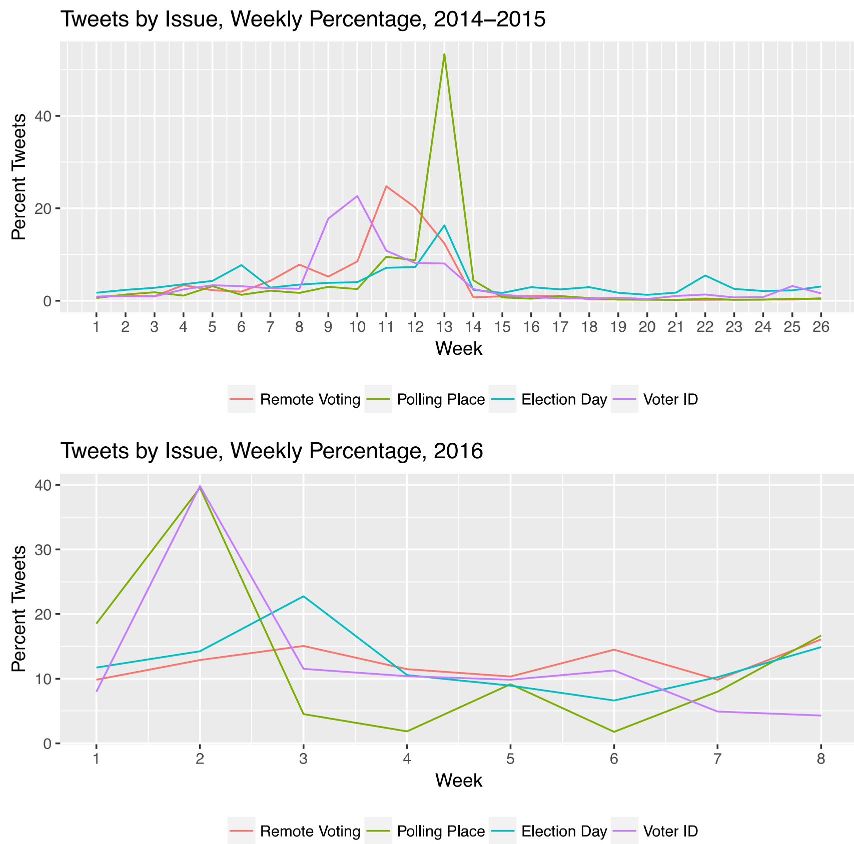


Figure 3. Tweet Frequencies by Week for issue areas

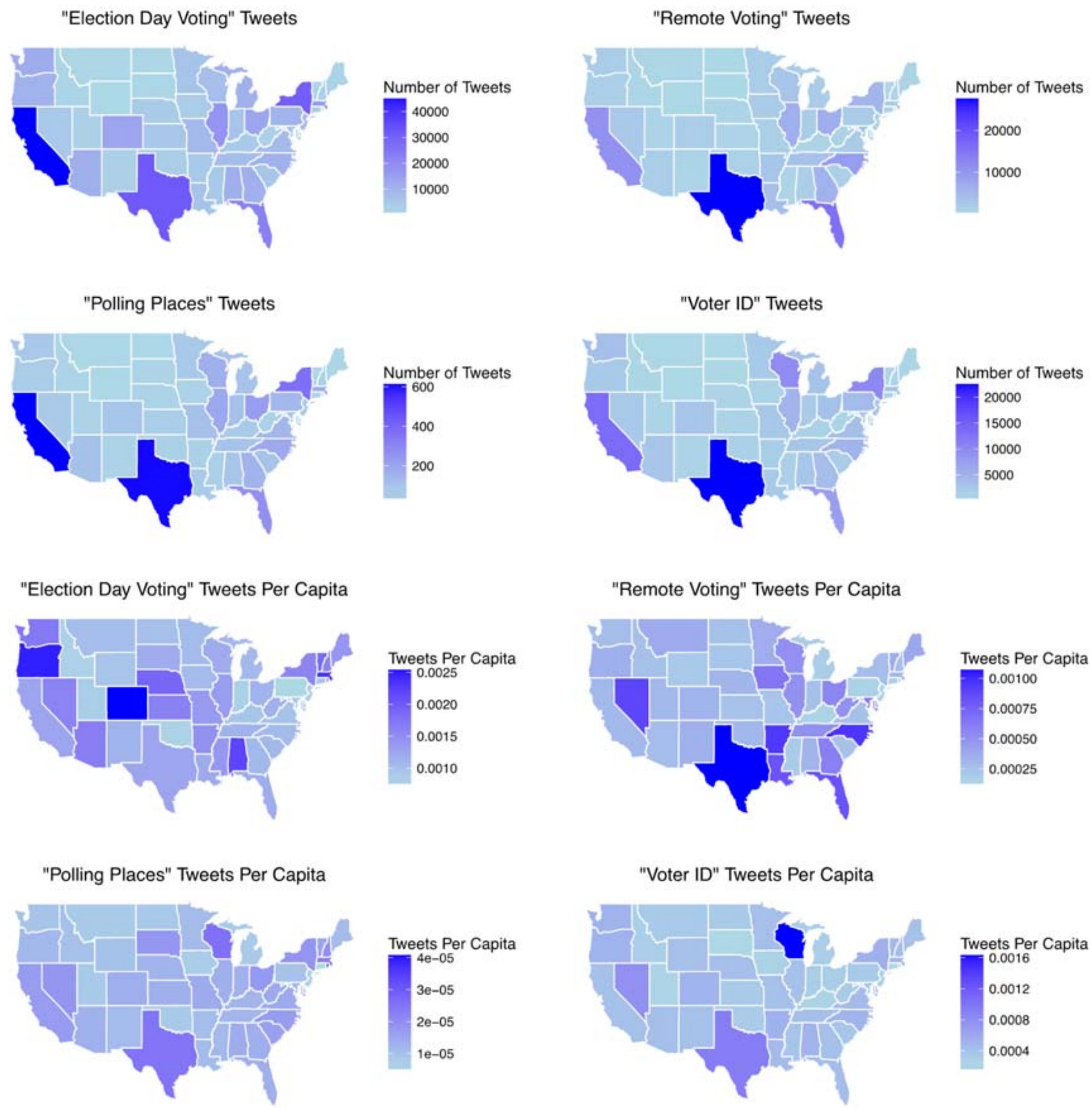


Figure 4. 2014-2015 Data by State

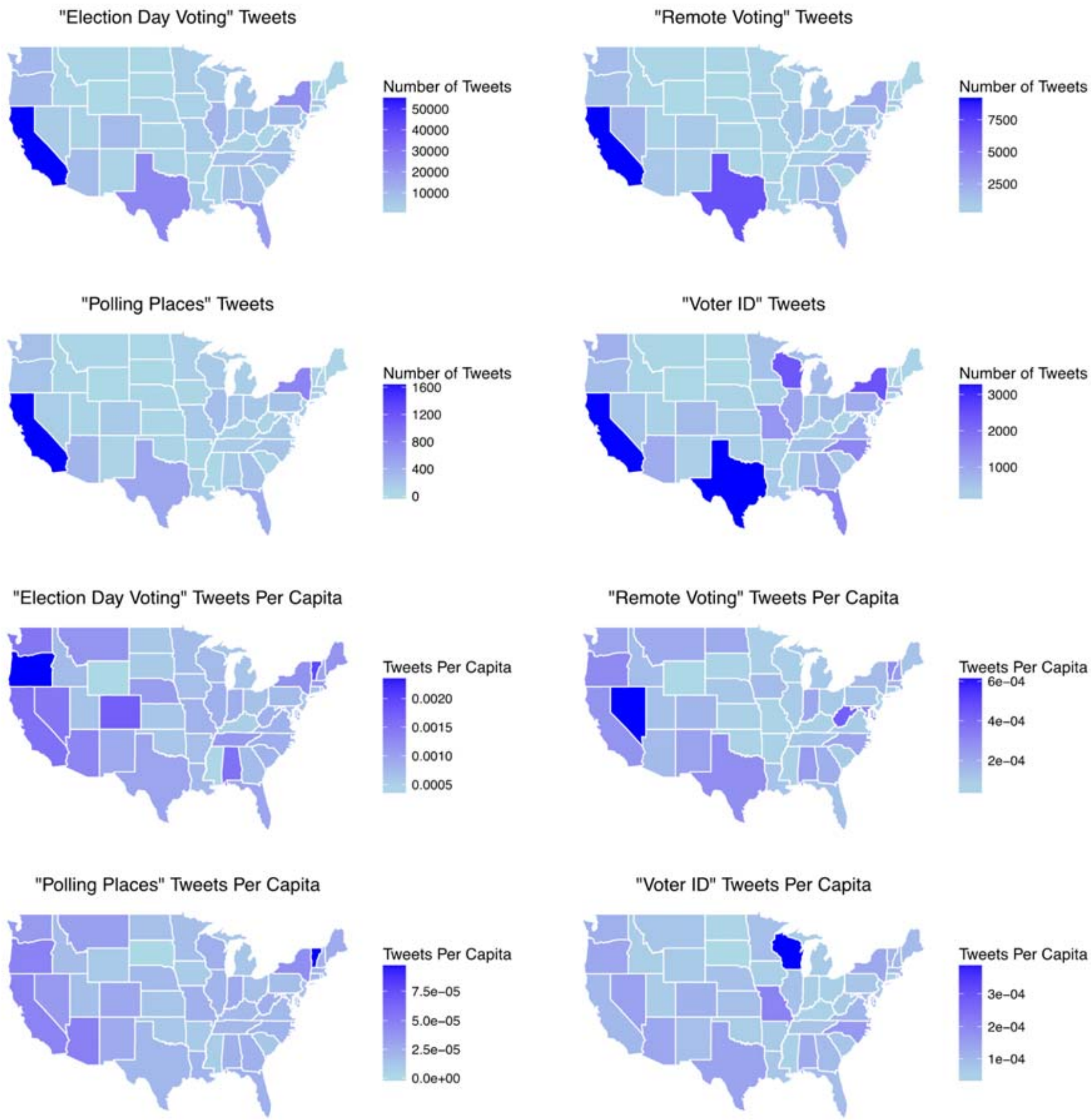


Figure 5. 2016 Data by State

Election Monitoring Using Twitter:

Supplementary Materials

Collecting Tweets About Election Administration Issues

Our approach to collecting Twitter status updates ("tweets") about election administration issues uses a tool written in Python that relies on the Tweepy library, which provides wrappers for Twitter's Streaming and REST APIs.[3] With a custom version of Tweepy's StreamListener class, the tool runs on a server and constantly listens on the stream of newly created tweets, searching for fifteen keywords relating to elections or election administration issues: **voter identification, voting identification, voter id, provisional ballot, absentee ballot, mail ballot, vote by mail, voting by mail, early voting, polling place line, precinct line, voting machine, ballot, pollworker,** and **poll worke**r.

We believe this array of keywords represent topics and ideas that will be present in voters' social media conversations about the problems they encounter when trying to vote. Past research using the U.S. Census Bureau's Current Population Survey, or the periodic Survey of the Performance of American Elections, has consistently found that voters cite an array of problems with voting technologies and election administration, which we summarized with our keywords (Caltech/MIT, 2001; Stewart, 2013). We also focus on issues that were widely discussed in relation to the 2014 general election and 2016 presedntial primaries, such as the use of provisional ballots and issues with voter identification. We consulted with a number of our colleagues who study voting technology and election administration, and they confirmed their expectations that this set of keywords would encompass most of the potential election-related problems that voters might experience in 2014 and 2016. [4]

[3] See http://www.tweepy.org for more information about Tweepy.

[4] While we choose to use a carefully constructed list of keywords based on our knowledge of common issues in previous election cycles, in choosing a static list of keywords, we ran the risk of leaving out important phrases relating to voting in the current election cycle, phrases that might have developed dynamically as the election occurred. Using an automated approach to keyword selection could help solve this issue (see King, Lam, & Roberts, in press), reducing bias. However, as the same types of issues have tended to occur in recent electoral cycles, we believe our approach of using a carefully constructed set of static keywords still has merit in the current application.

Once a set of keywords is created, our tool uses Tweepy to set up a "filter" that automatically pulls from Twitter's Streaming API all information from any tweet sent containing one of our keywords in real time.[5] While we have access to a huge wealth of potential metadata, we limit the data we collect to the most relevant pieces of information for our analyses. Thus, for each tweet that contains one of our keywords, our tool collects the complete character string of the full status update; the user handle and identification number associated with the tweet, and the number of their followers; the time-stamp of the tweet; and whatever geographic information is associated with the tweet or the Twitter user, which includes both user-reported and device-reported GPS location information (if available). If any error is reached in pulling the tweet from the Streaming API, our tool writes a detailed log file, designed such that whenever it encounters a fatal exception it notifies us via an email containing the error's traceback before waiting ten seconds and attempting to restart the StreamListener implementation. Other specialized aspects of its behavior include logging and discarding incomplete responses from Twitter's servers, as well internally storing a set amount of tweets locally in case the process is manually killed.

When Tweepy successfully pulls a tweet containing one of our track words from the Streaming API, our tool automatically uploads this data to a secure MySQL database. More specifically, our tool parses the Twitter data and separates each piece of information into one of two MySQL tables: a table containing the content of each tweet along with the user and tweet id, and a separate table that contains all relevant metadata associated with the tweet. If the tweet was sent by a new user not yet encountered in our stream, a third table storing the user's name and id is updated. The MySQL database was organized thusly to optimize both uploading new data and quickly downloading relevant information for analysis. In particular, MySQL queries allow one to quickly download all tweets sent in various fine-grained time intervals. These can then easily be fed to analysis code in order to compute results such as those presented in this paper.

The tool began collecting tweets August 8, 2014 and has continued to collect tweets thereafter; for the purposes of this study we present data from that start date through February 11, 2015, as well as a more recent

[5] There are limitations to the amount of data one can collect in the Streaming API, and thus one should not consider the data collected by our tool to be the full population of tweets sent containing one of our keywords. However, using the Streaming API represents the closest approximation to the full population of tweets available to most researchers. For more details on the potential limits of using the Streaming API, see Morstatter et al. (2013).

set of tweets that were collected during the 2016 presidential primary period (April 19, 2016 through June 14, 2016).

Bots

With all of the discussion surrounding the 2016 U.S. elections regarding “fake news” and these use of automated systems for disseminating “fake news” on social media outlets in the election, we wanted to determine whether there was evidence that many of the tweets in our databases might have originated with so-called bots.[6] There has been some interesting research in recent years on detecting bots in Twitter data, including the BotOrNot API (Davis et al., 2016).[7] Most approaches to detecting Twitter bots rely upon analyzing the features of the Twitter user (available from their profile), their follower network, and the content of information in their past tweets. The BotOrNot approach uses a large number of these features, and a random forest model, to produce a 0 to 1 classification score for the likelihood that a particular Twitter account is a bot.

To gain some insight into whether or not the Twitter data we have collected using the process described above might be driven in large part by bots, we randomly drew 30,000 Twitter IDs from our 2014-2015 database. Using a Python script we developed, we then passed those 30,000 Twitter IDs through the BotOrNot API, and collected the bot classification score from the API.[8] This process classifies 24,287 of the Twitter IDs that we sampled, the missing data arises because of Twitter accounts that have been deleted, or because the userIDs have changed. While there is no absolute rule as to what classification score would definitively imply that a particular account was a bot, the data we show in Figure 6 to us imply that most of the Twitter users we have in our data are unlikely to be bots.

Figure 6 gives a histogram of the BotOrNot classification scores, and the classification scores have a generally normal distribution, though with a relatively long upper tail. The distribution of bot classification scores

[6] See for example the November 17, 2016 story by John Markoff in the New York Times, “Automated Pro-Trump Bots Overwhelmed Pro-Clinton Messages, Researchers Say”, https://www.nytimes.com/2016/11/18/technology/automated-pro-trump-bots-overwhelmed-pro-clinton-messages-researchers-say.html?_r=0. Academic studies are now focusing on fake news and bots, for a recent example see Howard et al. (2017).
[7] See Subrahmanian et al. (2016) for a review of some of the recent research on Twitter bots.
[8] The Python script passes each Twitter ID to the API, which then classifies the Twitter ID based on current information. This process is subjected to rate limitations, which is why we used a sample of our Twitter IDs, rather than trying to classify all of our Twitter IDs. Also, this approach is assessing the likelihood that a Twitter ID that entered our database in 2014 or 2015 is classified as a bot by the BotOrNot scheme using contemporary data; we believe that it is unlikely that if a Twitter ID was a bot in 2014 or 2015 that it would not continue to be a bot today, furthermore, we think that it is unlikely that a Twitter ID that was not a bot in the past is now a bot. However, we recognize the possibility that this assumption may be incorrect, but again, we remind the reader that we are really here trying to assess the potential proportion of bots in our data, not a definitive classification for every Twitter ID in our data as to whether they are a bot.

has a mean of .3721, with a standard deviation of .1566. In fact, a large fraction of the sampled Twitter IDs have bot classification scores below .50, and nearly all have scores below .75. Based on this analysis, we conclude that there is little evidence that bots produced a large number of the tweets in our election issues keyword databases. This is consistent with other recent research on political discussion on Twitter.[9]

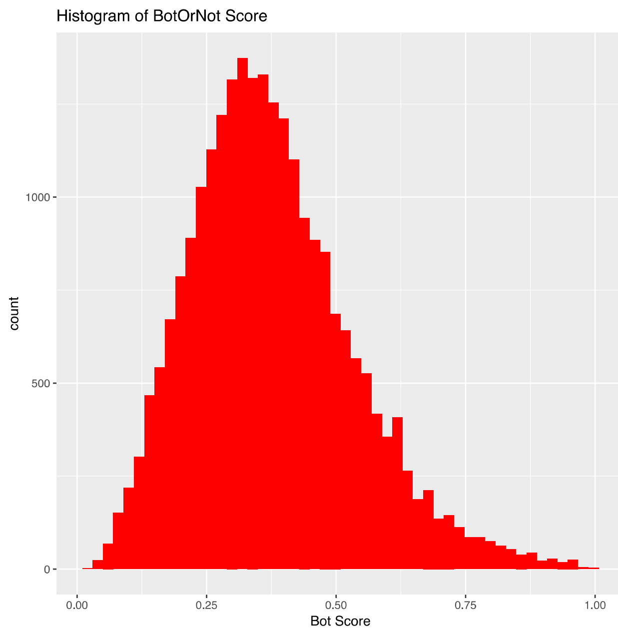


Figure 6. Distribution of BotOrNot Scores

[9] Howard et al.'s (2017) recent study of tweets about the 2016 presidential election in Michigan found in their data that "only 2% of the platforms used to send Twitter traffic were known sources of bots—the rest were platforms that either supported human users or offered very human-like levels of automation" (p. 2).