Revisiting The American Voter on Twitter

Huyen T. Le* **G.R. Boynton*** **Yelena Mejova‡** **Zubair Shafiq*** **Padmini Srinivasan***
*The University of Iowa, Iowa City, IA, USA ‡Qatar Computing Research Institute, Qatar
{huyen-t-le,bob-boynton,zubair-shafiq,padmini-srinivasan}@uiowa.edu, ymejova@qf.org.qa

ABSTRACT
The American Voter – a seminal work in political science – uncovered the multifaceted nature of voting behavior which has been corroborated in electoral research for decades since. In this paper, we leverage *The American Voter* as an analysis framework in the realm of computational political science, employing the factors of *party*, *personality*, and *policy* to structure the analysis of public discourse on online social media during the 2016 U.S. presidential primaries. Our analysis of 50 million tweets reveals the continuing importance of these three factors; our understanding is also enriched by the application of sentiment analysis techniques. The overwhelmingly negative sentiment of conversations surrounding 10 major presidential candidates reveals more "crosstalk" from Democratic leaning users towards Republican candidates, and less vice-versa. We uncover the lack of *moderation* as the most discussed personality dimension during this campaign season, as the political field becomes more extreme – Clinton and Rubio are perceived as moderate, while Trump, Sanders, and Cruz are not. While the most discussed issues are *foreign policy* and *immigration*, Republicans tweet more about *abortion* than Democrats who tweet more about *gay rights* than Republicans. Finally, we illustrate the importance of multifaceted political discourse analysis by applying regression to quantify the impact of party, personality, and policy on national polls.

Author Keywords
Election; Political Affiliation; Sentiment Analysis; Twitter

ACM Classification Keywords
K.4.3. Organizational Impacts: Computer-supported collaborative work

INTRODUCTION
The American Voter [16], published in 1960, demonstrated that the most important factors for voters when choosing their President were: *party or partisanship*, *policy considerations*, and *personality* of the person seeking office. Partisanship – being a Democrat or a Republican – was extremely important and it was the one factor that extended from one election to the next. Though policies shifted across elections, people's assessment of the importance of specific policies influenced their votes. The third factor was how the candidates were perceived – their personality or character. *The American Voter* was based on the largest collection of survey research of its day, covering presidential elections from 1948 through 1956. The study became an instant classic and set a research tradition that has extended to this point. *The American Voter Revisited* [40] replicated the original research in the 2000s and found among other things that once again party, policy, and personality remain the three most important factors in understanding who people vote for in presidential elections.

Our aim is to use the classic *The American Voter* study as the basis for a principled analysis of Twitter conversations around the 2016 U.S. presidential election. Specifically, we assess partisanship, discussions on policies, and perceptions of candidates' personality using Twitter communications. While our approach derives in a principled way from *The American Voter*, our work is novel in that it bridges from traditional survey research to mining large-scale publicly available online social media communications.

Social media has already played a significant role in elections in the U.S. and elsewhere [53, 38]. It was pivotal in the 2016 U.S. presidential election as well [6]. Growing numbers of the general public, especially the younger demographic, follow elections on social media [51, 34]. Given this trend, most candidates and their campaigns are actively trying to attract and engage social media users. Taking candidacy announcement as an example, Ted Cruz was the first mainstream politician to officially announce his candidacy with a tweet. A few weeks later, Hillary Clinton also took advantage of Twitter to announce she was running. Candidates differ in their social media strategies and successes [37]. Donald Trump posted inflammatory tweets to dominate the news cycle and drive up attendance at his rallies. Bernie Sanders's campaign used #FeelTheBern to gather and rally a large grassroots movement.

The analysis of social media activity in prior literature has primarily focused on counts (tweets, retweets, comments, likes, etc.) and sentiment as indicators of public engagement, reach, and opinion. Content tends to be considered only in outlier conditions such as when a post goes viral. In contrast, our goal is to use computational methods to explore *content* about policy discussions and personality perceptions in addition to party-specific engagement on Twitter. Our key contributions are the following.

- We study Twitter communications around the 2016 U.S. presidential candidates using 50 million tweets collected from November 15, 2015 to February 29, 2016. This period covers Iowa, New Hampshire, Nevada, and South Carolina caucuses and primaries as well as several debates.

CHI 2017, May 06-11, 2017, Denver, CO, USA
 ISBN 978-1-4503-4655-9/17/05$15.00
DOI: http://dx.doi.org/10.1145/3025453.3025543

- We implement computational methods for tracking political discourse about party, personality, and policy on Twitter.
- We conduct statistical analysis of electoral polls to show the importance of party, personality, and policy in the modeling of political deliberation.

The unique contribution of our work is in applying computational methods to an established and long line of election research in political science. Specifically, we contribute to the stream of research initiated by *The American Voter* by moving it forward from survey research to the realm of online social media.

RELATED WORK

The American Voter

Although published several decades ago, *The American Voter* [16] continues to serve as the baseline for researchers to understand voting behavior. The original study used survey data collected during three U.S. presidential elections (1948, 1952, and 1956). The central argument of the study is the funnel model, which claims that party affiliation shapes voters' attitude towards policy considerations and personality perception of presidential candidates. The study analyzed how voters form their own party identification and how the psychological attachment between voters and their party determines political attitudes.

Based on the original study, *The American Voter Revisited* [40] in 2008 attempted to understand voting behaviors during the U.S. presidential elections in 2000 and 2004. The authors found that, as compared to the 1950s, more voters identified with their party affiliations in the 2000s. More importantly, the authors also found that voting outcomes are still explained by party identification, short-term policy issues, and perceptions of candidates. The authors concluded that even though the influence of these factors on voting outcomes may have changed over time, the three-pronged paradigm of party, policy, and personality remains intact.

We ask whether these dimensions are present in current social media data, whether it is possible to track them, and whether they still relate to the politicians' success at the polls. Below, we discuss prior research on measurement and analysis of party, policy, and personality, especially in online social media.

Party

Identification of political affiliations is a well-researched area. Prior political science research primarily relies on interviews to explicitly ask users about their political affiliations. On online social media, however, only a small fraction of users report their political affiliation. For example, less than 10% of U.S.-based adult Facebook users self-report their political affiliation [8]. It is challenging to infer users' political party affiliations at a large scale using online social media data. Cohen and Ruths [20] showed that it is difficult to infer political orientation for "normal" Twitter users who rarely discuss politics. Some researchers have used machine learning techniques such as LDA (Latent Dirichlet Allocation) [21], SVM (Support Vector Machine) [21, 57, 45], and BDT (Boosted Decision Tree) [45] to infer users' political affiliations based on profile features (e.g., name, location), linguistic features (e.g., tweet text, hashtags), and network features (e.g., followers, retweets, replies). Some researchers have used label propagation techniques where a user's political affiliation is inferred based on whether they post about well-known conservative/liberal issues or follow well-known conservative/liberal users. For example, Zhou et al. [59] inferred political leaning of users on Digg based on how they voted (equivalent to like or share) on news articles with labeled political ideology using Amazon Mechanical Turk (AMT) workers. More recently, Golbeck et al. [33] inferred political leaning of Twitter users based on whether they follow a seed group of well-known political personalities (e.g., Congress members). We use a similar method to infer political affiliation of Twitter users in our work.

Prior political science research has shown that the political party affiliation's impact on voting outcomes may not be straightforward. Using survey data on political figures and events, Bartels [10] applied a Bayesian model to study opinion change with respect to partisan bias. He found that partisanship is not just a simple "running tally", but rather shapes voters' attitudes and reactions to politics, resulting in sharp differences in opinions between Democrats and Republicans towards various political events. Other studies have also looked at the marginal and joint impact of political affiliation and other factors on voting preferences (e.g., [12, 25, 36]), which we also examine here.

Personality

Perception of politicians' personality reflects both the campaign's success in framing their candidate and the values of the electorate projected onto these public personas. Pew Research Center and the Washington Post conducted surveys to study how personalities of political leaders are perceived by the public over time [27, 28]. Through phone interviews, U.S. adult voters were asked for the one word that comes to mind when a politician is named. Comparing Obama and Romney, voters' most common perception of Obama were *Good*, *Trying*, *President*, and also *Failure* and *Incompetent* and for Romney were *Honest*, *Businessman*, *Rich*, *Good*, and *Conservative*. However, survey-based personality research cannot rival the scale of online social media. Prior studies on personality perceptions of presidential candidates from online social media communications have relied on broad categorization of sentiment such as positive, negative, and neutral [13, 39, 54, 41]. Tumasjan et al. [53] looked at more detailed sentiment aspects such as anxiety, anger, and sadness for different candidates in the 2009 German national election. More recently, Bhattacharya et al. [14] proposed a method to systematically measure personality traits suggested by *The American Voter* [16] using a template-driven approach that measures the personality trait on a continuum, measuring either its presence or absence. As state-of-the-art, we use their method in our work.

Further, party affiliation has been repeatedly shown to relate to the rhetoric of the politicians in question, as been shown by [23, 32, 29], who have studied the influence of politicians' personality traits on their leadership and decision-making styles. For example, Gallagher et al. [29] found that the personality traits of political leaders shape their choices and their level of consistency in policy making. Further, Benoit [12] showed that Republicans discuss character more, and policy less, than Democrats. While examining debates, television spots, and acceptance addresses from 1948 to 2000, he found that Democrats emphasized more on traits like 'empathy' and 'drive' while Republicans emphasized more on traits like 'sincerity' and 'morality'. The automated analysis framework we present in this paper provides parallel insights of the politicians' character as perceived by social media users.

Policy

Policy issues have been examined at both the macro-level (nation or public as a unit of analysis) [19, 11] and the micro-level (how individuals define issues) [55, 56]. In order to track policy-related discussions on online social media, researchers typically create a lexicon of relevant terms of each policy and track their occurrences within the content [58, 52]. For example, Zhang et al. [58] manually identified relevant keywords, phrases, and hashtags related to same-sex marriage on Twitter, community wikis, and news articles to predict policy changes on the issue. We follow a similar high-precision approach, engaging political scientists' domain expertise to build vocabularies for each topic.

Prior political science research has explored the impact of policies on voters' evaluation of presidential candidates. Benoit [12] investigated presidential elections between 1948 and 2000 by quantitatively analyzing the texts of primary and general debates, television spots, and acceptance addresses. The key conclusions were that Democrats discuss policy more than Republicans, and that Democrats focus more on issues such as education while Republicans focus more on issues such as national security. Dolan [25] examined the American National Election Study (ANES) election data between 1992-2006 and concluded that while voters may have a different view on abortion than their party, people tend to go with their party affiliation when voting in elections. Finding that more Democrats consider abortion an important issue than Republicans, Dolan showed that the relationship between policy and party is not homogeneous. A similar recent study by Highton [36] of ANES data spanning three presidencies (H.W. Bush: 90-92, Clinton: 92-96, and W. Bush: 00-04) further showed a strong partisanship effect, but one which over time changes with economic and cultural attitudes. In this work we aim to capture the interplay between partisanship and views on policy and on politicians' characters, however incorporating economic and cultural variables (which is beyond the scope of this work) is an exciting future direction.

Election Prediction

Due to the popularity of social media, using Twitter data to track public opinion and specifically to predict presidential elections has been an active research area. There are conflicting results reported in prior literature for election prediction using sentiment analysis of social media communications. In one of the seminal attempts, Tumasjan et al. [53] analyzed the content of more than 100K tweets published prior to the German national election. They found that the share of tweets for six different parties closely matched the election results with an error of less than 2%. O'Connor et al. [43] studied the correlation between public opinion measured from traditional polls and sentiment measured from Twitter. They found that while Twitter sentiment correlates with consumer confidence and presidential job approval polls, there is not a strong correlation with the 2008 presidential election poll. Gayo-Avello et al. [31] also evaluated the power of Twitter data in predicting the 2010 U.S. election outcomes. They also found that Twitter sentiment analysis is not accurate in prediction and its performance is only slightly better than a random classifier. Other researchers [39, 30, 50, 47] have also highlighted issues with using Twitter to predict elections, such as the need of methodological justification in terms of accuracy, the need to produce a true forecast (i.e. issued prior to the election), the need to control for biases, etc. Later, some researches [13, 24, 18, 15] tried to address these problems albeit with limited success. Due to these various issues, in this paper we do not attempt to predict the outcome of the election. Quite the opposite, we provide a complex view of the electorate's decision landscape that cannot be reduced to a single metric.

METHODOLOGY

Our aim is to use state-of-the-art computational approaches to track party, policy, and personality in online social media communications. Each approach is driven by a lexicon or a seed list, supplied by subject matter experts, and is then exploited to produce quantitative measurements for further analysis.

Party

We infer political affiliations of Twitter users using a method that is both simple and efficient. This method is supported by the theory of selective exposure [48] which implies that online social media users tend to follow other users with similar beliefs or ideology. Specifically, in the context of American politics, Democrats are more likely to follow other Democrats and Republicans are more likely to follow other Republicans [22]. Thus, a user following more Republicans than Democrats is likely to be affiliated with Republicans. Similarly, a user following more Democrats than Republicans is likely to be affiliated with Democrats. We manually curated sets of 30 well recognized Democrat (e.g., Rachel Maddow) and 30 Republican (e.g., Sean Hannity) accounts on Twitter as the "landmarks." The curation was done in consultation with political scientists. We intentionally include several journalists with recognized affiliations as landmarks because of their large Twitter followings. On average, each Democratic landmark has 223,656 followers and each Republican landmark has 277,671 followers. Political affiliation is then a function of the number of landmark Democrats and Republicans that each user follows on Twitter: $\frac{\#Republicans - \#Democrats}{\#Republicans + \#Democrats}$. The output is in the range of $[-1, 1]$, where -1 indicates Democratic affiliation, $+1$ indicates Republican affiliation, and 0 indicates Others (independent or alternative). Our method to infer political affiliation of Twitter users based on whether they follow well-recognized Democratic and Republican landmarks is high precision and low recall. Therefore, the "Others" category may include politically inactive Twitter users whose political affiliation cannot be inferred by our method [20, 42].

Personality

To characterize personality perceptions, Bhattacharya et al. [14] used the Adjective Check List (ACL) [35], which has 300 adjectives or traits commonly used to characterize a person's personality. The ACL covers a wide variety of traits such as *intelligent*, *creative*, *determined*, *cheerful*. Traits are viewed as either positive (e.g., honest), negative (e.g., anxious), or neutral (e.g., jolly). Simonton reduced them to a core set of 110 traits with factor analysis and used these traits to characterize the personality of 39 American presidents [49]. Simonton further consolidated these 110 traits into 14 non-orthogonal personality dimensions, which include *moderation*, *friendliness*, *intellectual brilliance*, *machiavellianism*, *poise and polish*, *achievement drive*, *forcefulness*, *wit*, *physical attractiveness*, *pettiness*, *tidiness*, *conservatism*, *inflexibility*, and *pacifism*. Subsets of the 110 traits are given loadings on a continuous scale of [-1,1] that show how traits positively or negatively contribute to a particular personality dimension. For example, the *intellectual brilliance* dimension is composed of *artistic* (.84), *inventive* (.76), *curious* (.74), *intelligent* (.64), *sophisticated* (.62), *insightful* (.54), *wise* (.46), *dull* (-.71), and *commonplace* (-.41).

In our study, we characterize perceptions on the personalities of the 2016 presidential candidates using Simonton's 110 traits and 14 personality dimensions. We use a set of forty high-precision search templates [14] to identify tweets expressing these personality perceptions. There are two types of templates, one to retrieve tweets stating that a trait is present and the other to retrieve tweets saying that a trait is absent. An example template is: [P] is [A]? [T]. Here [P] is a variable representing a person name such as Hillary Clinton, [A] represents a class of high certainty words (e.g., definitely, very) and [T] is a specific trait such as honest (or its synonyms). The '?' designates item being optional. This template retrieves statements such as 'Hillary Clinton is certainly smart' and 'Hillary Clinton is intelligent'. Another example is [P] is [S] [T], where [S] represents words that are only somewhat certain (e.g., sort of, somewhat, kinda). It retrieves statements such as 'Hillary Clinton is sort of decisive' and 'Hillary Clinton is somewhat friendly'. We consider negation in statements ('Hillary Clinton is not decisive') and trait antonyms ('Hillary Clinton is not unfriendly'). We further manually examine tweets retrieved by the search templates to eliminate false positives due to sarcasm. Using the tweets retrieved by these search templates, we calculate a score for each trait. Since different presidential candidates may accumulate varying numbers of tweets, we normalize this score for the number of tweets discussing the trait. Further details are in [14].

Policy

To understand how policy preferences of candidates impact their perception, we track tweets related to 11 different policy categories for each candidate. The list of policies includes abortion, gay rights, climate change, foreign policy, health care, immigration, gun control, education, economy, veterans, and miscellaneous. While not a complete list by any means, these are some of the key issues discussed in our data, and were identified as most commonly discussed around candidates on Twitter by Rupar et al. [46]. In consultation with political scientists, we compiled the list of keywords for each policy by starting with a few well recognized keywords for each policy, e.g., "pro-life" and "pro-choice" for abortion. We then identified other related keywords with which they co-occurred (e.g., "planned parenthood" was frequently mentioned for abortion). Further, the miscellaneous policy category includes other keywords such as "blacklivesmatter" and "sex discrimination" which could not be neatly included in other policy categories. Given the set of these highly precise keywords for a policy, we extracted all tweets that contain at least one of the keywords. Being the most topic-dependent dimension, policy tracking is most in need of expert supervision and keyword curation, and we leave an automated keyword-based extraction of such topics to future work.

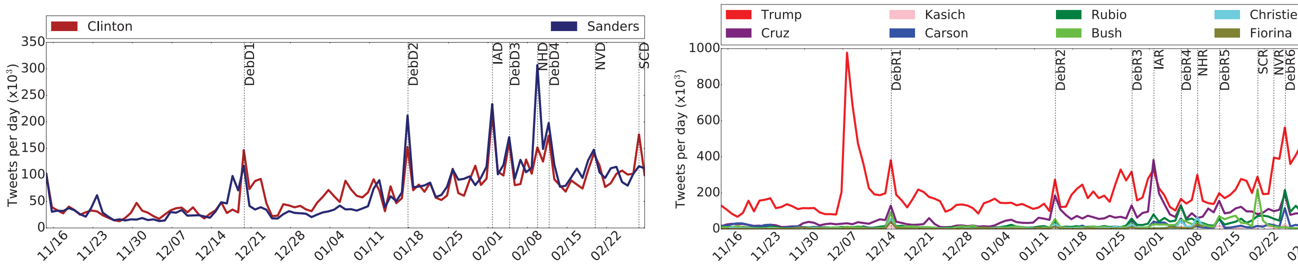
Sentiment

While our main points of focus are the three factors from *The American Voter*, we also analyze our data using the more common text mining strategy of sentiment analysis. The goal here is to gauge the general negativity or positivity in the discussion surrounding each presidential candidate. Given the limited size of tweets (140 characters), we can safely assume that the sentiment detected in political tweets concerns the entities (in this case, the candidates) mentioned therein [43]. For sentiment analysis, we rely on the SentiWordNet lexicon [7], which assigns positive and negative scores to each *synset* (set of synonyms) of WordNet (containing around 117K synsets). To this end, we split a tweet's text as separate sentences, remove symbols such as "< [^>]∗ >", tokenize, and stem before matching to the SentiWordNet lexicon. To quantify the overall sentiment of each tweet, we use the common approach which is to sum up positive and negative sentiment scores of the matched tokens. If the positive sentiment score is larger than the negative sentiment score, we label the tweet as positive, and similarly for negative. If both scores are equal, we label the tweet as neutral.

DATA

Data Collection

This work uses Twitter collections around 10 major presidential candidates listed in Table 1. Two of these candidates are running as Democrats (Clinton and Sanders) and the remaining eight are running as Republicans. For each candidate, we collected tweets posted from November 15, 2015 to February 29, 2016. The data were collected using Twitter's streaming API with filter keywords (`statuses/filter`) for each candidate. We used full names of candidates such as "hillary clinton" for Clinton. This API provides all tweets related to the filter keywords, but caps the tweets at 1% of all public tweets. Since more than 500 million tweets per day are posted on Twitter [1], we are set to capture up to five million tweets per day for each candidate. Note that the highest daily tweet count (for Trump) is less than one million, thus we can safely assume that we are capturing a vast majority of tweets for all candidates. Moreover, for computing Twitter users' political affiliation, we also used Twitter's REST API to crawl the follower lists for the sixty landmark accounts.



(a) Democratic candidates (b) Republican candidates

Figure 1. Time series of tweets for candidates. The vertical black lines indicate the following major election events during the data collection period. (1) DebR1: CNN Republican debate, (2) DebR2: Fox Business Republican debate, (3) DebD2: NBC News Democratic debate, (4) DebR3: Fox News Republican debate, (5) IAD: Democratic Iowa caucus, (6) IAR: Republican Iowa caucus, (7) DebD3: MSNBC Democratic debate, (8) DebR4: ABC News Republican debate, (9) NHD: Democratic New Hampshire primary, (10) NHR: Republican New Hampshire primary, (11) DebD4: PBS Democratic debate, (12) DebR5: CBS Republican debate, (13) NVD: Democratic Nevada caucus, (14) SCR: Republican South Carolina primary, (15) NVR: Republican Nevada caucus, (16) DebR6: CNN Republican debate, (17) SCD: Democratic South Carolina primary.

We decided not to discount retweets versus original tweets when counting tweets. We find retweets to usually be quoted as-is, without modification from the user – our data is 58.15% retweets, and of those 99% are quoted verbatim. It may be the case that retweeting is less powerfully associated with user opinion. It would be an interesting future work to find a suitable discount factor for retweets.

We did not explicitly attempt to geographically filter tweets because only a small fraction are GPS located. When examining the GPS located tweets in our data, we find 85.3% to be from the USA and 3.3% from the UK in a distant second place. Our data may include some tweets from international Twitter users, although the lexicons used in our analysis should at least limit them to English language tweets.

Data Statistics

As shown in Table 1, Trump has a clear lead with about 21 million tweets about him from more than 5 million users. Moreover, about half of the users and tweets were posting about Trump. This is roughly 3X the Twitter attention given to Clinton or Sanders who have similar numbers of tweets and users. Cruz follows with about 6.5 million tweets from close to a million users. The remaining candidates trail behind to varying degrees.

Figure 1 shows the time series of tweets for all presidential candidates. We observe sharp spikes for both Democratic and Republican candidates. These spikes typically correspond to major election events. Our data collection covers the Iowa, New Hampshire, Nevada and South Carolina caucuses and primaries as well as several debates. Some events happened on the same day for both parties. For example, both Democrats and Republicans held the Iowa caucus on February 1 and New Hampshire primary on February 9. Others happened on different days for each party. For example, the Nevada caucus for Democrats was held on February 20 while that for Republicans was held on February 23. Different events happened on the same day for both parties as well. For example, the Nevada caucus for Democrats and the South Carolina primary for Republicans were held on February 20.

We observe that all candidates received spikes on major election event days. The magnitudes of spikes differ depending on how well candidates performed at that event. For example, Clinton received 51% more tweets than Sanders on the day she defeated him at the South Carolina primary. As another example, Sanders received 102% more tweets than Clinton when he defeated her in the New Hampshire primary. However, we also note that losing candidates sometimes received more tweets than winning candidates. For example, even though Clinton won the Iowa caucus, Sanders received 8% more tweets. This difference is small as compared to the previous examples possibly because Sanders lost to Clinton by a very small margin. While Trump received more spikes than other Republican candidates, Cruz received the largest spike when he unexpectedly won the Iowa caucus instead of Trump. Cruz received 16% more tweets than Trump on the day of event. As another example, Bush received the second largest spike for the South Carolina primary, even though Rubio and Cruz are ranked second and third respectively after Trump. The spike was because Bush announced that he would quit the race after his disappointing performance in the South Carolina primary. Bush received 166% more tweets than Rubio while only 32% fewer tweets than Trump on the event day.

Besides the spikes for all candidates on days of major election events, candidates sometimes had spikes on specific days based on their own campaign activities. For example, Trump received many more tweets than other candidates starting December 7 (which then gradually faded after 5-6 days) when he called for a ban on Muslims entering the U.S. [2]. While not obvious in Figure 1 due to Trump's large tweet counts, Kasich had a spike on November 19 when he announced plans to create a new federal agency that would promote Judeo-Christian values [26]. Fiorina also received many more tweets on November 29 when she told the Fox News that Obama is "delusional" for saying that climate change is a major national security threat [4].

Table 1 also dissects the tweet data by detected sentiment. Negative tweets significantly outnumber positive tweets for all candidates, but to a different extent for each candidate. For the whole time period, Bush received the highest ratio of negative tweets to positive ones (13.4×), followed by Cruz (7.6×) and Trump (4.3×) while Carson (2.5×), Christie (2.6×), and Sanders (2.7×) received the lowest ratio.

Candidate	Number of Users	Number of Tweets	Tweets from Republicans				Tweets from Democrats				Tweets from Others			
			Positive	Negative	Neutral	Total	Positive	Negative	Neutral	Total	Positive	Negative	Neutral	Total
Hillary Clinton	1,391,829	7,224,615	217,549	764,536	86,469	1,068,554	306,149	802,342	122,813	1,231,304	1,097,466	3,249,497	577,794	4,924,757
Bernie Sanders	1,500,362	7,215,240	88,652	291,246	34,781	414,679	402,260	1,079,684	158,993	1,640,937	1,239,066	3,348,727	571,831	5,159,624
Donald Trump	5,096,698	21,351,629	423,049	1,508,704	172,275	2,104,028	333,269	1,447,187	133,512	1,913,968	2,956,020	12,968,070	1,409,543	17,333,633
Ted Cruz	995,067	6,478,005	223,283	1,398,221	111,358	1,732,862	86,020	693,479	41,525	821,024	399,699	3,298,447	225,973	3,924,119
John Kasich	159,999	425,123	17,887	59,017	7,632	84,536	15,533	49,837	6,060	71,430	64,406	182,385	22,366	269,157
Ben Carson	510,513	1,518,287	52,007	157,262	18,537	227,806	62,615	162,681	19,452	244,748	279,862	677,408	88,463	1,045,733
Marco Rubio	495,516	2,727,233	158,753	491,231	63,297	713,281	63,786	230,133	24,245	318,164	348,875	1,176,074	170,839	1,695,788
Jeb Bush	560,798	1,807,891	22,258	238,806	11,882	272,946	20,534	225,332	12,427	258,293	77,658	1,145,254	53,740	1,276,652
Chris Christie	199,998	522,164	19,833	51,696	6,665	78,194	28,108	78,277	8,234	114,619	86,414	218,409	24,528	329,351
Carly Fiorina	132,672	294,669	9,701	30,878	3,951	44,530	14,015	52,480	5,962	72,457	37,794	123,691	16,197	177,682
Total	11,043,452	49,564,856	1,232,972	4,991,597	516,847	6,741,416	1,332,289	4,821,432	533,223	6,686,944	6,587,260	26,387,962	3,161,274	36,136,496

Table 1. Statistics of candidate tweet collections from November, 15 2015 to February 29, 2016. Close to half the users and tweets in the dataset were posting on Trump. Trump's numbers reflect a 3X lead in tweets and users compared to Clinton and Sanders. Negative sentiment tweets significantly outnumber positive and neutral ones. The sentiment charged nature of the dialog is also indicated by the low prevalence of sentiment neutral tweets (about 8%). Same party tweets outnumber competing party tweets for seven out of ten candidates. Democrats tweeted more about Trump than about Clinton or Sanders. Republicans were more interested in Clinton than in Sanders.

PARTY, PERSONALITY, AND POLICY

Party

The party affiliation serves as a key factor in determining the electorate's perception of candidates and voting decisions. In the U.S., it is usually dichotomized into Democrats vs. Republicans, although alternative or independent affiliations have been on the rise [17]. For those affiliated, prior literature has shown that only a small fraction of people vote against their party affiliations [16, 40]. Here, we analyze the breakdown of tweet volume and its sentiment with respect to party affiliations of the users as Republicans, Democrats, and Others. In total, our data contains almost 50 million tweets from more than 11 million Twitter users. It is noteworthy that the total number of tweets from Republicans (6,741,416) and Democrats (6,686,944) are almost equal.

Interestingly, the breakdown of tweets by party in Table 1 shows that tweets from users affiliated to the same party outnumber tweets from the competing party for all candidates, except for Fiorina, Christie, and Carson. For example, 23% (1,640,937) tweets mentioning Sanders are from Democrats while 6% (414,679) tweets are from Republicans. As another example, 27% (1,732,862) tweets mentioning Cruz are from Republicans while 13% (821,024) tweets are from Democrats. Overall, Trump received most tweets across all political affiliations as compared to other candidates.

Table 1 also shows the presence of some "crosstalk". While Republican leaning user tweets were mostly about Republican candidates (78%), Democrat leaning users were tweeting slightly more (57%) about Republican candidates than about candidates from their own party. Thus, we find far more "crosstalk" from Democratic leaning users than from Republican leaning users [9], suggesting an aggravated echo-chamber effect on the Republican side. Republican leaning users tweeting about Democrats were far more intereseted in Clinton (2.6× more) than in Sanders. Whereas when tweeting about Republicans, Democrats mostly talked about Trump, followed by Cruz in the distant second.

Table 1 also provides the sentiment breakdown of tweets for all party affiliations. As indicated earlier, negative tweets – almost three-fourths of all tweets – dominate. This observation holds for each candidate as well. The highly charged nature of the Twitter election dialog is also seen in the low numbers of sentiment neutral tweets. Around 8% of the tweets posted by each user group have neutral sentiment.

Personality

Table 2 provides an overview of more than 316 thousand tweets in our data that convey perceptions of candidates' personality along fourteen personality dimensions. The table is again broken down by party affiliation, Republican leaning or Democrat leaning. The most discussed personality dimension is moderation, which accounts for almost one-third of all personality-related tweets. This is followed by friendliness, machiavellianism, intellectual brilliance, pacifism, and wit. The remaining eight personality dimensions are discussed infrequently and are summarized under "Other Dimensions".

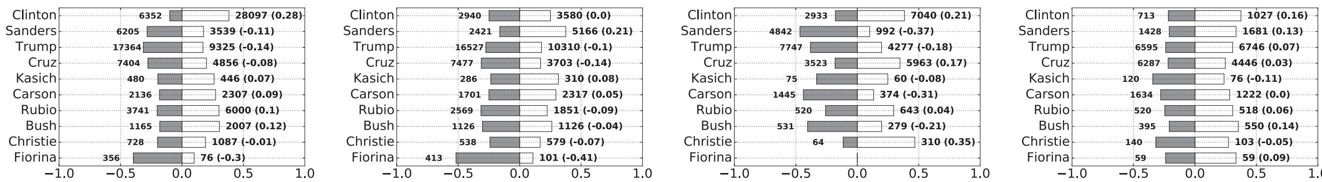
Comparing the types of posts made by users leaning towards each party we find a few differences. For example, Democrats discuss friendliness of candidates more than Republicans. In contrast, Republicans discuss machiavellianism and pacifism of candidates more than Democrats. Overlaying sentiment, we note that while negative tweets account for about four-fifths of all personality-related tweets overall, Republicans tend to be more negative than Democrats (83% vs. 77%). Finally, as compared to Republicans, Democrats are more positive about pacifism and more negative about machiavellianism and wit.

Figure 2 shows the candidate-level breakdown of the four most frequently discussed personality dimensions. We excluded candidates with very few tweets for some personality dimensions from our analysis. For systematic analysis, we divide candidates into high-frequency and low-frequency groups based on their tweet counts for each personality dimension and limit candidate comparisons to each group.

Moderation. Clinton and Rubio are perceived as moderate while Trump, Sanders, and Cruz are perceived as not moderate in the high-frequency group. Clinton received the highest score of 0.28 for moderation. For example, Clinton received more than 11 thousand tweets containing phrase

Personality Dimension	#Tweets	%Tweets	Republicans				Democrats			
			# Tweets	% Tweets	# Positive Tweets	# Negative Tweets	# Tweets	% Tweets	# Positive Tweets	# Negative Tweets
Moderation	103,671	32.8%	15,788	31.8%	1,738 (11%)	13,348 (85%)	16,521	35.3%	1,996 (12%)	11,971 (72%)
Friendliness	65,041	20.6%	8,159	16.4%	1,394 (17%)	6,239 (76%)	10,334	22.1%	1,499 (15%)	8,203 (79%)
Machiavellianism	41,692	13.2%	8,930	18.0%	1,873 (21%)	6,788 (76%)	6,091	13.0%	926 (15%)	4,985 (82%)
Intellectual Brilliance	34,319	10.8%	5,694	11.5%	662 (12%)	4,787 (84%)	5,885	12.6%	755 (13%)	4,910 (83%)
Pacifism	23,202	7.3%	3,879	7.8%	115 (3%)	3,737 (96%)	1,211	2.6%	67 (6%)	852 (70%)
Wit	12,751	4.0%	1,112	2.2%	150 (13%)	926 (83%)	1,769	3.8%	174 (10%)	1,540 (87%)
Other Dimensions	35,717	11.3%	6,080	12.3%	668 (11%)	5,150 (85%)	4,966	10.6%	1,125 (23%)	3,666 (74%)
Total	316,393	100.0%	49,642	100.0%	6,600 (13%)	40,975 (83%)	46,777	100.0%	6,542 (14%)	36,127 (77%)

Table 2. Statistics of tweets conveying personality perceptions and their breakdown across party affiliation and sentiment.



(a) Moderation (b) Friendliness (c) Machiavellianism (d) Intellectual Brilliance

Figure 2. Scores for four most discussed personality dimensions. Scores are in the range [-1,+1]. -1 (+1) indicates the personality is viewed as absent (present) with high confidence. The number of tweets for absence (presence) of each personality dimension are provided on the left (right) of the bar plot. The net score is included in the parenthesis. The bar plots are shown for candidates receiving a minimum of 100 tweets on a dimension.

"Hillary Clinton Is *Calm*, *Cool* and Effective". Keywords calm and cool are positively aligned with moderation. On the other hand, Trump received -0.14 score for moderation – he received more than a thousand tweets including "Donald Trump is *terrible* for disrespecting Moslems" (terrible being negatively aligned with moderation). In the low-frequency group, Bush, Carson, and Kasich are perceived as moderate while Fiorina is perceived as not moderate.

Friendliness. Sanders is perceived as friendly, while Cruz and Trump not, whereas Clinton is seen as neutral in the high-frequency group. Kasich and Carson are perceived as friendly while others, led by Fiorina, are perceived as not friendly in the low-frequency group. In particular, Sanders was considered by many as cute ("Bernie Sanders is so *cute*"), whereas, for instance, Cruz received more than a thousand tweets saying "Cruz is *nasty*" and "Ted Cruz *isn't likeable*". Although having no direct relation to the political strengths of the candidates, surprisingly, thousands of tweets referred to their personal amiability.

Machiavellianism. Machiavellianism is associated with being deceitful and unscrupulous. In the high frequency group, Clinton and Cruz are perceived as machiavellian, while Sanders and then Trump are not. For example, Cruz received hundreds of tweets such as "Ted Cruz is *untrustworthy*" and "Ted Cruz is so *dishonest* that TV stations won't run his ads because they're worried about legal culpability". On the other hand, Sanders received more than a thousand tweets such as "100% of 18–29 year olds think Bernie Sanders is *honest* and *trustworthy*" which result in -0.37 machiavellian score for him.

Intellectual Brilliance. In the high-frequency group, Trump and Cruz and both are perceived as intellectually brilliant. For example, Trump received more than six thousand tweets such as "Donald Trump Is *Smart* To Remind Voters Of Clinton Drama". All other candidates are in the low frequency group and most of them (Clinton followed by Bush, Sanders, and then Fiorina) are perceived as intellectual brilliant. Only Kasich and Christie are perceived as not intellectually brilliant. Kasich, for example, even received tweets like "John Kasich is a very *stupid* man" contributing to his low score.

Policy

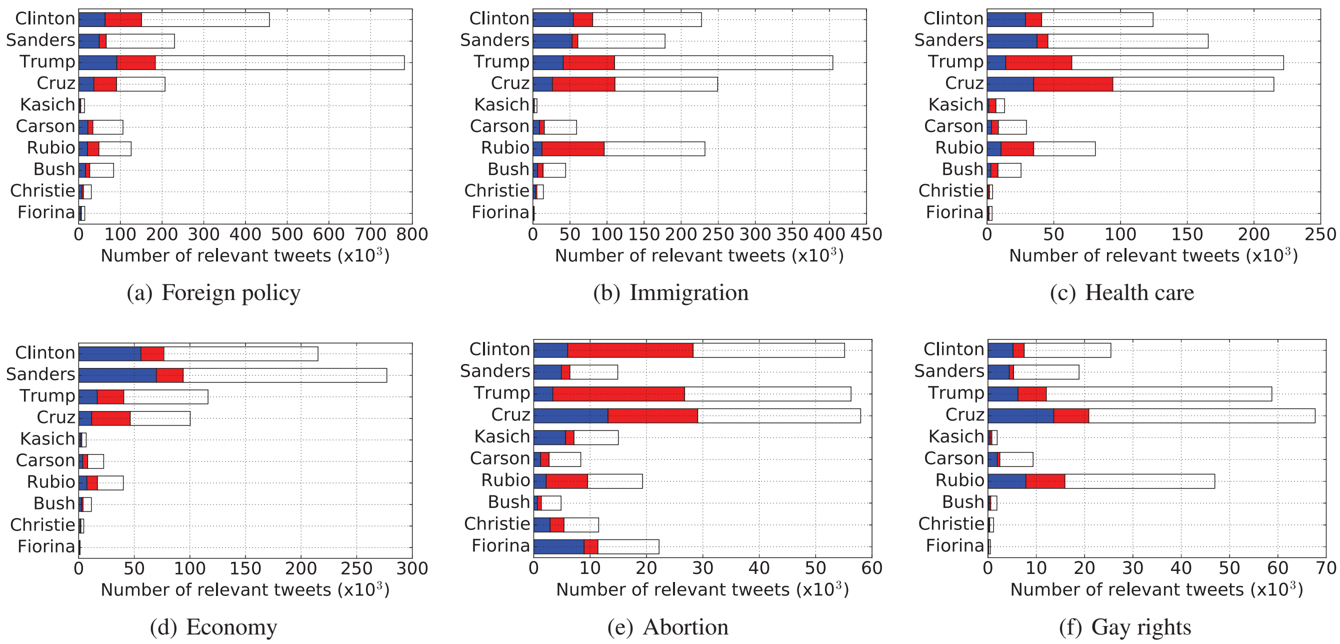
Table 3 presents an overview of the more than six million tweets about policy in our data. The table is limited to the six most discussed policies and the remaining five policies are summarized under "Other Policies". Foreign policy (33%), immigration (23%), healthcare (14%), and the economy (13%) are the most discussed policies. Democrat leaning users express most interest in foreign policy compared to immigration which ranks second in drawing their attention (11% difference). In contrast, the two policies draw equal attention from Republican leaning users. Republicans tweet more about abortion, health care and immigration (3-6% difference) than Democrats while Democrats tweet more about gay rights, foreign policy, and economy (2-6% difference) than Republicans. Overlaying sentiment, we again note that negative tweets account for about four-fifths of all policy-related tweets. While Republicans are more positive than Democrats for gun control, veterans, and gay rights (6-8% difference) the order is reversed for abortion (6% difference).

Figure 3 shows the candidate-level breakdown of the six most frequently discussed policies. We note that Clinton, Sanders, Trump, Cruz, and Rubio received much more policy-related tweets than the remaining candidates. Thus, we limit our analysis to these top candidates.

Foreign policy. Trump is by far the most discussed candidate for foreign policy, with almost twice relevant tweets than Clinton in the second place. The most mentioned foreign policy keyword is *ISIS* for both candidates, with Trump

Policy	#Tweets	%Tweets	Republicans				Democrats			
			# Tweets	% Tweets	# Positive Tweets	# Negative Tweets	# Tweets	% Tweets	# Positive Tweets	# Negative Tweets
Foreign Policy	2,050,648	33.2%	307,254	28.4%	51,039 (17%)	239,913 (78%)	316,859	31.2%	51,856 (16%)	249,622 (79%)
Immigration	1,416,459	22.9%	289,251	26.8%	30,712 (11%)	244,987 (85%)	207,076	20.4%	21,280 (10%)	178,188 (86%)
Healthcare	884,371	14.3%	171,476	15.9%	31,931 (19%)	125,392 (73%)	134,343	13.2%	22,674 (17%)	102,493 (76%)
Economy	795,784	12.9%	120,462	11.1%	12,188 (10%)	104,678 (87%)	171,748	16.9%	16,780 (10%)	148,904 (87%)
Abortion	265,593	4.3%	79,007	7.3%	11,060 (14%)	65,250 (83%)	49,127	4.8%	9,823 (20%)	36,389 (74%)
Gay Rights	232,515	3.8%	25,832	2.4%	10,641 (41%)	12,422 (48%)	40,310	4.0%	13,969 (35%)	23,760 (59%)
Other Policies	537,551	8.6%	87,467	8.1%	30,101 (34%)	50,454 (58%)	97,239	9.5%	27,747 (29%)	60,542 (62%)
Total	6,182,921	100.0%	1,080,749	100.0%	177,672 (16%)	843,096 (78%)	1,016,702	100.0%	164,129 (16%)	799,898 (79%)

Table 3. Statistics of tweets discussing different policies, and their breakdown across party affiliation and sentiment.



(a) Foreign policy (b) Immigration (c) Health care

(d) Economy (e) Abortion (f) Gay rights

Figure 3. Number of tweets for the six most discussed policies with party affiliation breakdown. The blue and red regions indicate tweets from Democrats and Republicans, respectively. The white region represents tweets from Others.

having about 200 thousand tweets and Clinton having about 150 thousand matching tweets. However, after *ISIS*, the most mentioned keyword for Trump is *Paris* (78 thousand tweets) and that for Clinton is *Benghazi* (75 thousand tweets). This observation can be explained by Clinton's email scandal as part of hearings conducted by the House Select Committee on Benghazi.

Immigration. Trump is the most discussed candidate for immigration, followed by Cruz, Rubio, and Clinton. In general, for these candidates three most mentioned keywords are *refugee* and *immigrants*. However, immigration-related tweets for Rubio mostly had the keyword *amnesty*. This observation can be explained by Rubio's change his stance on amnesty. While the other Republican candidates were consistently against amnesty, Rubio changed his stance on amnesty when he ran for the presidency.

Healthcare. Trump and Cruz received most healthcare related tweets, followed by Sanders and Clinton. The most mentioned keyword in healthcare is *ACA* (Affordable Care Act) which outnumbers *ObamaCare*.

Economy. Democratic candidates are most discussed with respect to the economy, with Sanders leading Clinton. The most mentioned keywords in economy related tweets are *Wall Street*, *tax*, *social security*, and *poverty*.

Abortion. Three most discussed candidates about abortion are Cruz, Trump, and Clinton – they received roughly equal number of tweets. There most mentioned keyword about abortion is *Planned Parenthood*. Each of the top three candidates received about 30 thousand tweets containing *Planned Parenthood* with much more tweets from Republicans than Democrats.

Gay rights. Republican candidates are the most discussed about gay rights, with Cruz in the lead followed by Trump and Rubio. The most mentioned keyword in gay rights is *gays*, with Cruz and Trump each receiving more than 50 thousand tweets. Other popular keywords include *gay marriage*, *same-sex marriage*, and *marriage equality*. All candidates received more tweets about gay rights from Democrats than Republicans.

Feature	β	Std. Error
Tweets from Republicans on Abortion	1.48	0.20
Tweets from Republicans on Veterans	1.33	0.3
Tweets on presence of Pacifism	1.30	0.17
Tweets from Democrats on Gay rights	1.18	0.24
Tweets from Republicans on Healthcare	1.05	0.28
Tweets on presence of Machiavellianism	0.85	0.19
Tweets from Democrats on Gun control	0.58	0.16
Tweets from Democrats on Abortion	-0.96	0.18
Tweets from Democrats on Miscellaneous	-1.14	0.23
Tweets from Republicans on Climate change	-1.55	0.21
Tweets from Others on Veterans	-1.65	0.30
Score of presence of Achievement Drive	-2.78	0.79
Tweets from Republicans on absence of Intellectual Brilliance	-3.37	0.96

Table 4. List of statistically significant variables (p-value < 0.001) in our regression model for Democratic candidates. The features are sorted in descending order with respect to estimated coefficient values. Adjusted $R^2 = 0.908$

Feature	β	Std. Error
Tweets from Others on Miscellaneous	1.35	0.21
Tweets from Others on Climate Change	1.27	0.21
Tweets on absence of Conservatism	1.17	0.22
Tweets from Republicans on Foreign policy	1.12	0.14
Tweets from Others on Gun control	1.10	0.17
Tweets from Republicans on Abortion	0.95	0.16
Tweets from Democrats on Gay rights	0.54	0.13
Tweets from Others on Foreign policy	-0.58	0.16
Tweets on presence of Pacifism	-1.07	0.15
Tweets on absence of Forcefulness	-1.07	0.31
Tweets from Democrats on Gun control	-1.09	0.21
Tweets from Democrats on Climate change	-1.15	0.20
Tweets from Democrats on absence of Moderation	-2.65	0.80

Table 5. List of statistically significant variables (p-value < 0.001) in our regression model for Republican candidates. The features are sorted in descending order with respect to estimated coefficient values. Adjusted $R^2 = 0.883$

IMPACT OF PARTY, PERSONALITY, AND POLICY ON VOTING OUTCOMES

Heretofore, we showed that the factors of party, personality, and policy outlined in *The American Voter* can be tracked in social media using computational methods, and that the output of these methods is directly interpretable in the context of the ongoing presidential campaigns. However, to what extent do these signals correspond to the outcomes of electoral polls? To answer this question, by which we may confirm or disconfirm the validity of *The American Voter*'s framework in the social media-based electoral opinion modeling, we quantify the effect each factor has on electoral outcomes.

Using the methods described above, we build features for the factors of party, personality, and policy:

- 3 features for party (Democrat, Republican, Other),
- 14 features for different personality dimensions, which are separately computed for presence and absence as well as for party affiliations,
- 11 features for different policy categories, which are separately computed for party affiliations,
- raw tweet counts and sentiment-labeled tweet counts,

in total resulting in 184 variables available to regress on poll numbers. Since a large number of polling organizations conduct polls at different time intervals, we rely on the RealClearPolitics (RCP) [3] aggregation of reputed national polls. Specifically, we use RCP's daily average of national polls for all presidential candidates.

First, we preprocess this set of 184 variables for use in regression analysis. We log-transform all volumetric variables (tweet counts) to alleviate the skew. Also, many of the features are interdependent, making any model built with such variables suffer from multi-collinearity. Thus, we use a stepwise selection process to eliminate features with high VIF (Variance Inflation Factor) and AIC (Akaike Information Criterion) scores. The process results in 72 variables for the Democratic candidates and 61 for the Republican candidates.

Due to data scarcity, we are not able to separately run regressions for individual candidates (however such models would be very interesting for individualized analysis). Instead, we train linear regression models on RCP's average national poll numbers separately for Democratic and Republican candidates. We then rerun the regression on the subset of features with p-value < 0.05. The final model for Democratic candidates has 24 variables and that for Republican candidates 28 variables.

The model for Democratic candidates (Clinton and Sanders), which uses the 24 variables, explains 0.908 proportion of variance (adjusted $R^2 = 0.908$).[1] Table 4 lists the subset of 13 variables with p-value < 0.001. The model includes 4 personality-based variables and 9 policy-based variables, most of which have integrated the party aspect. For personality, we note that tweets on presence of pacifism and machiavellianism have a positive impact on poll numbers. Interestingly, tweets from Republicans on absence of intellectual brilliance have a negative impact on poll numbers. For policy, the impact of party aspect is more obvious on policy-related variables with four from each party. It is noteworthy that tweets from Republicans on abortion have a positive impact, whereas tweets from Democrats on abortion have a negative impact on poll numbers. Similarly, tweets from Republicans on veterans have a positive impact whereas those from Others have a negative impact on poll numbers of Clinton and Sanders.

The model for Republican candidates, which uses the 28 variables, explains 0.883 proportion of variance (adjusted $R^2 = 0.883$). Table 5 lists the subset of 13 variables with p-value < 0.001. The model also includes 4 personality-based variables and 9 policy-based variables, most of which have integrated the party aspect. For personality, we note that tweets on absence of conservatism has the most positive impact on poll numbers, reflecting the preference for political outsiders. In contrast, tweets on presence of pacifism as well as on absence of forcefulness have a negative impact. It is interesting to note that tweets from Democrats on absence of moderation has the most negative impact on poll numbers. For policy, we note that tweets from Others on climate change and gun control have a positive impact, whereas those from Democrats have a

[1] Note that the goal of our regression model here is descriptive. Therefore, we tolerate some degree of over-fitting to the data.

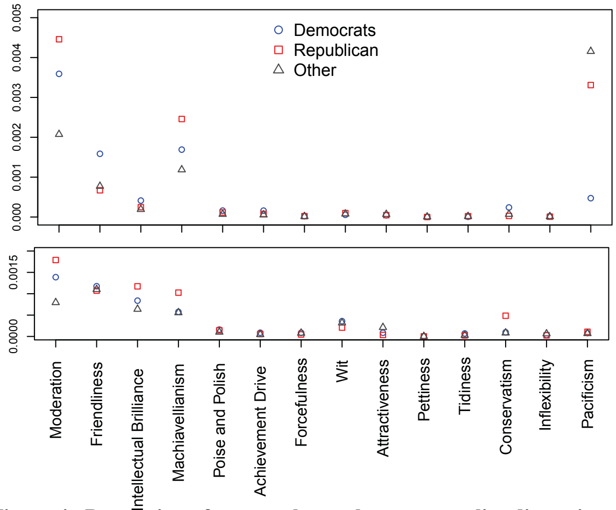


Figure 4. Proportion of tweet volume about personality dimensions of Clinton (above) and Trump (below) from users identified as Democrats (blue circles), Republican (red squares), and Other (grey triangles).

negative impact on poll numbers. Moreover, tweets from Republicans on foreign policy have a positive impact, whereas those from Others have a negative impact on poll numbers.

The comparison of these models for both parties reveals several key similarities and differences. Among the most significant variables with p-value < 0.001, both models include 4 personality-based variables and 9 policy-based variables. Furthermore, both models include policy-based variables about abortion, gay rights, gun control, and climate change. The interesting differences between these models are related to the sign of variable coefficients. For example, tweets on presence of pacifism have a positive impact ($\beta = 1.30$) on poll numbers of Democratic candidates while they have a negative impact ($\beta = -1.07$) on poll numbers of Republican candidates. Similarly, tweets from Democrats on gun control have a positive impact ($\beta = 0.58$) on poll number of Democratic candidates while they have a negative impact ($\beta = -1.09$) on poll numbers of Republican candidates.

DISCUSSION

In our exploration of *The American Voter* factors on Twitter, we find that all three factors still play an important role in the electoral discourse. The interplay between these factors is especially fruitful for further analysis and validation of established political theories. For instance, in Figure 4 we plot the proportion of tweets from users of different political affiliations about personality dimensions of the two eventual nominees – Clinton (above) and Trump (below). We can see that the personality dimension of moderation is much more emphasized for Clinton, however it is only the Other (perhaps more independent) and Republican users who emphasize her pacifism (which includes keywords such as *weak*). Further, only users marked to be Republican emphasize Trump's conservatism, whereas the Other group – potentially the independent vote – not at all.

Further examining partisanship, we were able to identify 572,316 and 893,048 of the 11 million users as Republican leaning and Democrat leaning respectively. Although the seed list used in our approach limits the reach of our classification, the Twitter conversation is dominated by a vast majority that is not following these established opinion leaders. Even among those we detected as interested in partisan opinion leaders, we observed considerable crosstalk about other party's candidates. However, when combined with the overall negative tone, this possibly indicates the presence of partisanship for those already aligned with a party, as found, for example by [5].

Moreover, our study corroborates findings reported in prior literature and also highlights some new trends. For example, Benoit [12] reported that Republicans emphasized more on personality traits like 'sincerity' while Democrats emphasized more on 'drive'. In corroboration, we also find that Republicans discuss the personality dimension of machiavellianism (which includes *sincere* trait) more than Democrats, who discuss the personality dimension of achievement drive more than Republicans. Benoit [12] also reported that Democrats focus more on policy issues such as education while Republicans focus more on national security. Our analysis shows that Democrats not only discuss education more than Republicans, but also issues such as gay rights and climate change. On the other hand, Republicans discuss policy issues such abortion and veterans more than Democrats.

Methods presented in this paper were able to uncover more than 6 million policy-related tweets and 316 thousand personality-related tweets, but we caution the reader not to focus on raw volumes, as they depend on the sensitivity of the lexicons used wherein. Instead, cross-policy and cross-personality trait analyses would present more insight, as presented in this paper. The crafting of more accurate and up-to-date lexicons is a task involving subject-matter experts, and thus can be only partially automated, however attempts have recently been made to automatically augment lexicons, such as in crisis situations [44].

CONCLUSION

The multi-factor framework of *The American Voter* brings a structure to the rich political discourse on social media. Our use of computational methods allows the evolution from survey-based electoral research to encompass the automated analysis of new rich sources of user-generated media. Our findings reveal the partisan divides in the perceptions of policy positions and personality traits of the presidential candidates. Further, using statistical analysis of *party*, *personality*, and *policy* factors we show the importance of each one in the modeling of electoral deliberation.

No social media study would be complete without the standard cautionary statement that social media analysis is not meant to replace the traditional survey techniques, having neither comparable participant sampling, nor the affordances of a (sometimes long-form) survey. However, the fact that political figures are increasingly using Twitter to reach their supporters indicates its increasing importance, at least in the eyes of the politicians themselves. This study establishes a baseline from which the views of the candidates can be studied, views which we fully expect to change not only in this election cycle but also in future elections.

REFERENCES

1. 2013. New Tweets per second record, and how! https://blog.twitter.com/2013/new-tweets-per-second-record-and-how. (August 2013).
2. 2016. Donald Trump: Ban all Muslim travel to U.S. http://www.cnn.com/2015/12/07/politics/donald-trump-muslim-ban-immigration/. (December 2016).
3. 2016. Election 2016 Presidential Polls. http://www.realclearpolitics.com/epolls/latest_polls/president/. (2016).
4. 2016. Fiorina: Obama 'delusional' about magnitude of climate change as security threat. http://www.foxnews.com/politics/2015/11/29/fiorina-obama-delusional-about-magnitude-climate-change-as-security-threat.html. (November 2016).
5. Alan Abramowitz and Steven Webster. 2015. All politics is national: The rise of negative partisanship and the nationalization of US house and senate elections in the 21st century. In *Annual Meeting of the Midwest Political Science Association Conference*.
6. Hunt Allcott and Matthew Gentzkow. 2017. Social Media and Fake News in the 2016 Election. (January 2017). NBER Working Paper No. 23089.
7. Stefano Baccianella, Andrea Esuli, and Fabrizio Sebastiani. 2010. SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining.. In *LREC*, Vol. 10. 2200–2204.
8. Eytan Bakshy, Solomon Messing, and Lada Adamic. 2015. Exposure to Ideologically Diverse News and Opinion on Facebook. *Science* 348, 6239 (May 2015), 1130–1132.
9. Pablo Barber, John T. Jost, Jonathan Nagler, Joshua A. Tucker, and Richard Bonneau. 2015. Tweeting From Left to Right: Is Online Political Communication More Than an Echo Chamber? *Psychological Science* (August 2015).
10. Larry M. Bartels. 2002. Beyond the Running Tally: Partisan Bias in Political Perceptions. *Political Behavior* 24, 2 (Jun 2002), 117–150.
11. Frank R. Baumgartner, F. R.tner and Bryan D. Jones. 1993. *Agendas and Instability in American Politics*. Chicago : University of Chicago Press.
12. William M. Benoit. 2004. Political Party Affiliation and Presidential Campaign Discourse. *Communication Quarterly* 52, 2 (2004), 81–97.
13. Adam Bermingham and Alan F. Smeaton. 2011. On Using Twitter to Monitor Political Sentiment and Predict Election Results. In *Sentiment Analysis where AI meets Psychology*.
14. Sanmitra Bhattacharya, Chao Yang, Padmini Srinivasan, and Bob Boynton. 2015. Perceptions of Presidential Candidates' Personality in Twitter. *Journal of the Association for Information Science and Technology* (2015).
15. Pete Burnap, Rachel Gibson, Luke Sloan, Rosalynd Southern, and Matthew Williams. 2015. 140 Characters to Victory?: Using Twitter to Predict the UK 2015 General Election. *Cardiff University (working paper)* (2015).
16. Angus Campbell, Converse Philip, Miller Warren, and Stokes Donald. 1960. *The American Voter*. University of Chicago Press.
17. Pew Research Center. 2015. A Deep Dive Into Party Affiliation. http://www.people-press.org/2015/04/07/a-deep-dive-into-party-affiliation/. (April 2015).
18. Andrea Ceron, Luigi Curini, and Stefano M. Iacus. 2014. Every tweet counts? How sentiment analysis of social media can improve our knowledge of citizens political preferences with an application to Italy and France. *New Media & Society* 16, 2 (2014), 340–358.
19. Roger W. Cobb and Charles D. Elder. 1972. *Participation in American Politics: The Dynamics of Agenda Building*. Johns Hopkins University Press.
20. Raviv Cohen and Derek Ruths. 2013. Classifying Political Orientation on Twitter: Its Not Easy!. In *Proceedings of the Seventh International AAAI Conference on Weblogs and Social Media (ICWSM)*.
21. Michael D. Conover, Bruno Goncalves, Jacob Ratkiewicz, Alessandro Flammini, and Filippo Menczer. 2011a. Predicting the political alignment of Twitter users. *Proceedings of the International Conference on Social Computing* (2011).
22. Michael D. Conover, Jacob Ratkiewicz, Matthew Francisco, Bruno Goncalves, Filippo Menczer, and Alessandro Flammini. 2011b. Political Polarization on Twitter. In *AAAI International Conference on Weblogs and Social Media (ICWSM)*.
23. Edmond Costantini and Kenneth H. Craik. 1980. Personality and politicians: California party leaders, 19601976. *Personality and Social Psychology* 38, 4 (1980), 641–661.
24. Joseph DiGrazia, Karissa McKelvey, Johan Bollen, and Fabio Rojas. 2013. More tweets, more votes: Social media as a quantitative indicator of political behavior. *PLOS ONE* 11, 8 (2013).
25. Patrick Dolan. 2011. Abortion and American Politics: Can a single policy issue lead to party switching? Breithaupt American Government Paper Competition, Roanoke College. (April 2011).
26. Adam Edelman. 2016. 2016 candidate John Kasich: I would create a NEW government agency that would

'promote Judeo-Christian values'. http://www.nydailynews.com/news/politics/kasich-create-agency-push-judeo-christian-values-article-1.2439022. (November 2016).

27. Pew Research Center for the People & the Press. 2012a. Obama in a Word: From 'Inexperienced' and 'Change' to 'Good' and 'Trying'. http://www.people-press.org/2012/09/05/obama-in-a-word-from-inexperienced-and-change-to-good-and-trying/. (September 2012).

28. Pew Research Center for the People & the Press. 2012b. Romney in a Word: 'Honest,' 'Businessman,' 'Rich'. http://www.people-press.org/2012/08/29/romney-in-a-word-honest-businessman-rich/. (August 2012).

29. Maryann E. Gallagher and Susan H. Allen. 2014. Presidential Personality: Not Just a Nuisance. *Foreign Policy Analysis* 10, 1 (January 2014), 1–21.

30. Daniel Gayo-Avello. 2012. "I Wanted to Predict Elections with Twitter and all I got was this Lousy Paper" – A Balanced Survey on Election Prediction using Twitter Data. *arXiv: 12046441* (2012).

31. Daniel Gayo-Avello, Panagiotis T. Metaxas, and Eni Mustafaraj. 2011. Limits of Electoral Predictions Using Twitter. In *AAAI International Conference on Weblogs and Social Media (ICWSM)*.

32. Alexander L. George and Juliette L. George. 1998. *Presidential personality and performance*. Boulder, CO: Westview Press.

33. Jennifer Golbeck and Derek Hansen. 2014. A method for computing political preference among Twitter followers. *Social Networks* (2014).

34. Jeffrey Gottfried, Michael Barthel, Elisa Shearer, and Amy Mitchell. 2016. The 2016 Presidential Campaign - a News Event That's Hard to Miss. http://www.journalism.org/2016/02/04/the-2016-presidential-campaign-a-news-event-thats-hard-to-miss/. (February 2016).

35. Harrison G. Gough and Alfred B. Heilbrun. 1983. The Adjective Check List Manual. Consulting Psychologists Press. (1983).

36. Benjamin Highton. 2012. Updating Political Evaluations: Policy Attitudes, Partisanship, and Presidential Assessments. *Political Behavior* 34, 1 (March 2012), 57–78.

37. Ryan Holmes. 2016. How Bernie Sanders Beats Donald Trump at Social Media. http://fortune.com/2016/04/18/bernie-sanders-donald-trump-social-media/. (April 2016).

38. Sounman Hong and Daniel Nadler. 2012. Which candidates do the public discuss online in an election campaign? The use of social media by 2012 presidential candidates and its impact on candidate salience. *Government Information Quarterly* 29, 4 (2012), 455–461.

39. Andreas Jungherr, Pascal Jrgens, and Harald Schoen. 2012. Why the Pirate Party Won the German Election of 2009 or The Trouble With Predictions. *Social Science Computer Review* 2, 30 (2012), 229–234.

40. Michael Lewis-Beck, Jacoby G. William, Norpath Helmut, and Weisberg F. Herbert. 2008. *The American Voter Revisited*. University of Michigan Press.

41. Yelena Mejova, Padmini Srinivasan, and Bob Boynton. 2013. GOP Primary Season on Twitter: "Popular" Political Sentiment in Social Media. In *WSDM*.

42. Jonathan Scott Morgan, Cliff Lampe, and M.Z. Shafiq. 2013. Is news sharing on Twitter ideologically biased?. In *ACM Conference on Computer Supported Cooperative Work (CSCW)*.

43. Brendan O'Connor, Ramnath Balasubramanyan, Bryan R. Routledge, and Noah A. Smith. 2010. From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series. In *AAAI International Conference on Weblogs and Social Media (ICWSM)*.

44. Alexandra Olteanu, Carlos Castillo, Fernando Diaz, and Sarah Vieweg. 2014. CrisisLex: A Lexicon for Collecting and Filtering Microblogged Communications in Crises. In *AAAI International Conference on Weblogs and Social Media (ICWSM)*.

45. Marco Pennacchiotti and Ana-Maria Popescu. 2011. A Machine Learning Approach to Twitter User Classification. *AAAI International Conference on Weblogs and Social Media (ICWSM)* (2011).

46. Terri Rupar, Aaron Blake, and Samuel Granados. 2015. The ever-changing issues of the 2016 campaign, as seen on Twitter. http://wapo.st/1OgbJ45. (December 2015).

47. Erik T. K. Sang and Johan Bos. 2012. Predicting the 2011 Dutch senate election results with Twitter. In *Workshop on Semantic Analysis in Social Media (SASN)*.

48. David O. Sears and Jonathan L. Freedman. 1967. Selective Exposure to Information: A Critical Review. *Public Opinion Quarterly* 31, 2 (1967), 194–213.

49. Dean K. Simonton. 1986. Presidential Personality: Biographical Use of the Gough Adjective Check List. *Journal of Personality and Social Psychology* 51, 1 (1986), 149–160.

50. Marko Skoric and Nathaniel Poor. 2012. Tweets and votes: A study of the 2011 Singapore general election. In *Hawaii International Conference on System Sciences (HICSS)*.

51. Aaron Smith. 2014. Cell Phones, Social Media and Campaign 2014. Pew Research Center. (November 2014).

52. Stefan Stieglitz and Linh X. Dang. 2013. Social media and political communication: a social media analytics framework. *Social Network Analysis and Mining* 4, 4 (December 2013), 1277–1291.

53. Andranik Tumasjan, Timm O. Sprenger, Philipp G. Sandner, and Isabell M. Welpe. 2010. Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment. In *AAAI International Conference on Weblogs and Social Media (ICWSM)*.

54. Hao Wang, Dogan Can, Abe Kazemzadeh, Franois Bar, and Shrikanth Narayanan. 2012. A System for Real-time Twitter Sentiment Analysis of 2012 U.S. Presidential Election Cycle. In *50th Annual Meeting of the Association for Computational Linguistics*. 115–120.

55. Dan B. Wood and Alesha Doan. 2003. The Politics of Problem Definition: A Theory and Application to Sexual Harassment. *American Journal of Political Science* 47, 4 (October 2003), 640–653.

56. Dan B. Wood and Arnold Vedlitz. 2007. Issue Definition, Information Processing, and the Politics of Global Warming. *American Journal of Political Science* 51, 3 (July 2007), 552–568.

57. Faiyaz A. Zamal, Wendy Liu, and Derek Ruths. 2012. Homophily and latent attribute inference: Inferring latent attributes of Twitter users from neighbors. *AAAI International Conference on Weblogs and Social Media (ICWSM)* (2012).

58. Amy X. Zhang and Scott Counts. 2015. Modeling Ideology and Predicting Policy Change with Social Media: Case of Same-Sex Marriage. *ACM Conference on Human Factors in Computing Systems (CHI)* (2015).

59. Daniel X. Zhou, Paul Resnick, and Qiaozhu Mei. 2011. Classifying the Political Leaning of News Articles and Users from User Votes. In *AAAI International Conference on Weblogs and Social Media (ICWSM)*.