Bumps and Bruises: Mining Presidential Campaign Announcements on Twitter

Huyen Le
The University of Iowa, USA
huyen-t-le@uiowa.edu

G.R. Boynton
The University of Iowa, USA
bob-boynton@uiowa.edu

Yelena Mejova
Qatar Computing Research Institute, Qatar
ymejova@qf.org.qa

Zubair Shafiq
The University of Iowa, USA
zubair-shafiq@uiowa.edu

Padmini Srinivasan
The University of Iowa, USA
padmini-srinivasan@uiowa.edu

ABSTRACT

Online social media plays an increasingly significant role in shaping the political discourse during elections worldwide. In the 2016 U.S. presidential election, political campaigns strategically designed candidacy announcements on Twitter to produce a significant increase in online social media attention. We use large-scale online social media communications to study the factors of party, personality, and policy in the Twitter discourse following six major presidential campaign announcements for the 2016 U.S. presidential election. We observe that all campaign announcements result in an instant *bump* in attention, with up to several orders of magnitude increase in tweets. However, we find that Twitter discourse as a result of this bump in attention has overwhelmingly negative sentiment. The *bruising* criticism, driven by crosstalk from Twitter users of opposite party affiliations, is organized by hashtags such as #NoMoreBushes and #WhyImNotVotingForHillary. We analyze how people take to Twitter to criticize specific personality traits and policy positions of presidential candidates.

KEYWORDS

Sentiment Analysis; Social Media Analysis; Twitter

ACM Reference format:
Huyen Le, G.R. Boynton, Yelena Mejova, Zubair Shafiq, and Padmini Srinivasan. 2017. Bumps and Bruises:
Mining Presidential Campaign Announcements on Twitter. In *Proceedings of HT '17, Prague, Czech Republic, July 04-07, 2017,* 10 pages.
DOI: http://dx.doi.org/10.1145/3078714.3078736

1 INTRODUCTION

Election campaigns are complex socio-political processes that conclude with the vote on election day. An election campaign's evolution – with its varied attention 'bumps' and 'bruises' – is critical to understand the outcomes. For the U.S. presidential elections, the campaign season now typically extends 18 to 20 months prior to the election day [3]. Election outcomes are sometimes very difficult to foresee during the early and sometimes even at the late stages of campaigns. In 2008, for example, most analysts did not expect Barack Obama to be elected president at the beginning of the election campaign. In 2016, Donald Trump's win was unexpected even just a few days prior to the vote. Irrespective of whether it is possible to predict the outcome of an election, a longitudinal analysis of election data can provide important insights for our understanding of election campaigns.

A presidential election campaign starts with the candidacy announcement. Candidacy announcements are important because these are the time points when the candidates start to seriously introduce themselves to the electorate and invest significant resources into garnering favorable public opinions. Traditionally these announcements were made in mainstream media followed by a 'bump' in public attention. In recent elections, and especially during the 2016 U.S. presidential election, candidates leveraged online social media to launch their campaigns. For example, Ted Cruz was the first mainstream politician to officially announce his candidacy with a tweet. His announcement tweet received a big bump in attention for the candidate with approximately 12.3K retweets and 10.5K likes within one day [2]. Hillary Clinton followed just a few weeks later and also took advantage of Twitter in announcing that she was running. Her announcement tweet received an even higher attention bump with 95.7K retweets and 91.5K likes within one day [1]. Although political insiders knew that both were running before their announcement, Ted Cruz needed attention as did most of the other Democratic and Republican candidates. While announcements and consequent attention bumps were seen in mainstream media in earlier elections, present-day candidacy announcements appear carefully designed to also garner maximum attention on online social media.

In this paper, we focus on these crucial candidacy announcements on online social media and analyze the extent and kinds of public reactions they elicit. Since it is the start of the election process, our aim is not to predict the election outcome but rather to conduct a rich analysis – going beyond simple counts and sentiment analysis – of the tweets about the candidates from around the time of the announcement. We conduct this analysis within the framework of *The American Voter* [10], a seminal work in political science which demonstrated that the most important factors for voters were *party affiliation*, *policy considerations*, and *personality* of the person seeking the office. First, being a Democrat or a Republican, in the context of U.S. politics, was an important factor for individual voters.

HT '17, Prague, Czech Republic

DOI: http://dx.doi.org/10.1145/3078714.3078736

Second, voters' perception of importance of specific policies such as immigration and national security was also a key factor. Third, candidates' personality or character perception was also important for voters. Among these three factors, only party affiliation seemed to consistently impact voters' choices across multiple election cycles. While policies and personality perceptions are important, the specific policies and traits emphasized varied from one election to the next. *The American Voter*, now more than five decades old, set a research tradition that has extended to this point [26]. It is noteworthy that the classic paper and almost all of the follow-up research relied on survey research to analyze these three factors. In contrast, we use the same framework but apply it to online social media data related to these campaigns.

There is little doubt that online social media has played a significant role in elections in the U.S. and elsewhere [21, 38]. Recent elections, and especially the 2016 U.S. presidential election, amply demonstrate the importance of online social media in enabling candidates to directly reach the electorate. First, candidates are routinely active in promoting themselves on online social media. Second, many voters not only discuss political news but also post opinions about candidates and election events on social media [36]. Because of the nature of the medium, there is more opportunity for crosstalk as compared to traditional broadcast media – Democrats can post about Republican candidates and Republicans can post about Democratic candidates. Also in the mix we have journalists, who not only exchange opinions on online social media but also frame communications in traditional broadcast media.

Previously, we adopted *The American Voter* framework to analyze Twitter discourse in the 2012 and 2016 U.S. presidential elections [8, 25]. In [8], we demonstrated the importance of perceptions of personality traits of Obama and Romney during the 2012 U.S. presidential election. In [25], we showed the relevance of all three factors when analyzing Twitter conversations about candidates during the 2016 primaries and several debates. In contrast to our prior work, in this paper we focus on analyzing these three factors at beginning of presidential election campaigns. To the best of our knowledge, an in-depth characterization of carefully design online campaign announcements and the followup public discourse is lacking in prior literature. We summarize our key contributions and findings as follows:

(1) We study Twitter communications around candidacy announcements, an early milestone in the 2016 U.S. election campaign, for six major candidates from both parties. Specifically, we analyze Twitter communications from a week before to a week after these announcements. Of note is that all candidates analyzed announced their run on Twitter that resulted in an instant **bump** in attention, with orders of magnitude increase in tweets.

(2) Our sentiment analysis indicates that an overwhelming majority of tweets about candidates are negative. As soon as the candidates announce, we note a sharp increase in negative tweets – **bruises** – for all candidates. The ratio of negative to positive tweets is very high for some candidates (e.g., 19× for Jeb Bush and 10× for Ted Cruz) and relatively low for others (e.g., 2.8× for Hillary Clinton and 2.6× for Scott Walker).

(3) **Party.** We find that Twitter users of opposite party affiliation dominate the Twitter communication about a candidate. For example, almost 3× more Democrats tweeted about Jeb Bush than Republicans. Such communication is typically organized by hashtags such as #NoMoreBushes and #WhyImNotVotingForHillary.

(4) **Personality.** Personality perceptions also vary across candidates. For example, we find that Jeb Bush, Hillary Clinton, and Rand Paul are considered to be intellectually brilliant whereas Scott Walker and Marco Rubio are perceived as lacking intellectual brilliance. Hillary Clinton and Rand Paul are considered to be machiavellian whereas Marco Rubio and Ted Cruz are perceived as lacking machiavellianism.

(5) **Policy.** We notice that certain policy issues are disproportionately discussed for specific candidates. For example, Marco Rubio leads other candidates on immigration and Hillary Clinton leads others on gay rights. Overall, we find that health care and foreign policy dominates other policy issues.

2 RELATED WORK

We first discuss our previous work on analyzing political communications on Twitter [8, 25] during the 2012 and 2016 U.S. presidential elections. We then discuss prior literature related to analyzing online social media to predict elections. Finally, we discuss prior literature related to analyzing online social media to characterize political affiliation, personality perceptions, and policy issues.

In [8], we proposed a method to systematically track public perception of personality using a core set of 110 traits identified by Simonton [34]. We employed high-precision search templates on an 18-month tweet collection about Obama and Rommney collected during the 2012 U.S. presidential election. Our results showed interesting differences in public perceptions of their personalities. For instance, Romney was perceived as more of an achiever while Obama was perceived as more friendly. We further aggregated the 110 personality traits into 14 broad personality dimensions [34]. Our results showed, for example, that Obama rated far higher than Romney on the moderation dimension and lower on the machiavellianism dimension. We also made observations consistent with party ideology. For example, during time periods when mentions of Obama being conservative increased his Gallup support went down (-0.48 correlation). But when mentions of Romney as conservative increased his Gallup support went up (+0.40 correlation).

In [25], we adopted *The American Voter* and its three factors (party, personality, and policy) as an analysis framework for the 2016 U.S. presidential election. By employing state-of-the-art techniques in political affiliation detection, personality perception measurement, and policy analysis, we analyzed Twitter communications around 10 presidential candidates during multiple 2016 caucuses, primaries, and debates. Including in the data analysis were models built to regress on opinion poll data using a variety of features extracted from Twitter. These models demonstrated, independent of candidate party affiliation, the continuing importance of these three factors. These results further motivate our analysis using *The American Voter* framework in the current paper.

In contrast to our prior work, here we focus on analyzing the beginning of presidential election campaigns. As we discuss next, there is a plethora of prior research on analyzing political discourse as a whole during elections. However, an in-depth characterization of carefully designed campaign announcements on online social media is lacking in prior literature. We aim to fill this gap by analyzing campaign announcements and the public discourse that follows on online social media.

2.1 Election prediction

Using online social media communications to track public opinion and specifically to predict presidential elections is a hot research topic. However, there are conflicting results reported in prior literature. Some researchers [29, 38] found that there is a correlation between public opinion measured from traditional polls and sentiment measured from Twitter. Some researchers [18] claimed that Twitter sentiment analysis cannot accurately predict electoral outcomes, and its performance is only slightly better than a random classifier. Other researchers [17, 23, 32, 35] 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), and the need to control for biases. Some researchers [7, 9, 11, 16] have tried to address these issues, albeit with limited success. In this paper, we measure and analyze tweets collected around candidate announcements but we do not focus on election prediction due to the well-known issues raised in prior literature.

2.2 Party, personality, and policy

Prior work on inferring political affiliation in online social media can be broadly divided into two categories: content-based methods and audience-based methods. Content-based methods (e.g., [14, 30, 43]), as the name implies, address the problem more *directly* by analyzing users' own characteristics such as profile features (e.g., name, location), linguistic features (e.g., tweet text, hashtags), and network features (e.g., followers, retweets, replies). Audience-based methods (e.g., [19, 45]), in contrast, rely on the idea that users have their own ideological biases which are reflected in their sharing and networking behavior [33]. These methods *indirectly* measure political affiliation of users based on whether they post about well-known conservative/liberal issues or follow well-known conservative/liberal users. As we discuss later, our method [24] to infer political affiliation of Twitter users falls in the latter category.

Prior work on analyzing personality perceptions of presidential candidates have relied on broad categorization of sentiment such as positive, negative, and neutral [7, 23, 27, 40]. Tumasjan et al. [38] looked at more detailed sentiment aspects such as anxiety, anger, and sadness for different candidates in the 2009 German national election. More recently, in [8], we proposed a method to systematically measure public perceptions of a candidate's personality using a template-driven approach that measures such perceptions on a continuum for each of 110 personality traits. At one end the continuum represents the perceived absence of a trait and at the other its presence. We use this method in this paper as well.

Public policy issues have been examined at both the macro-level (nation or public as a unit of analysis) [6, 12] and the micro-level (how individuals define issues) [41, 42]. 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 [37, 44]. For instance, Zhang et al. [44] 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, utlizing political scientists' domain expertise to build vocabularies for each topic.

3 METHODS

3.1 Party

We estimate Twitter users' political affiliations based on their connectivity patterns on Twitter with a set of landmark Democrat and Republican accounts. Our method [24] is founded on selective exposure theory [33] which in the context of American politics implies that a user following more Republicans than Democrats is likely to be affiliated with Republicans and vice versa [15, 28]. Thus, by carefully consulting with political scientists we manually curate a set of 30 well recognized Democrats (e.g., Rachel Maddow) and 30 well recognized Republicans (e.g., Sean Hannity) on Twitter as our "landmarks." Generally these landmarks have many followers which help to infer political affiliations for a large number of users. In fact, on average each Democratic landmark has 223,656 followers and each Republican landmark has 277,671 followers. Then, we estimate political affiliation as 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 Other (independent or alternative).

3.2 Personality

We measure the personality perceptions for the 2016 presidential candidates with the high-precision search template method proposed in [8]. This method has two main components. The first is the Adjective Check List (ACL) (consisting of 300 'personality trait' adjectives such as honest and strong) proposed by [20]. Simonton [34] identified a core subset of 110 traits/adjectives and further consolidated these 110 traits into 14 non-orthogonal personality dimensions. Bhattacharya et al. [8] built tweet search templates around these 110 trait adjectives augmented with synonyms and antonyms. This set of forty high-precision search templates is the second component of the method. In general 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. For instance, "[P] is [A]? [T]" is a template where [P] represents a person name (e.g. Hillary Clinton), [A] represents a class of high certainty words (e.g., definitely, very), [T] is a specific trait (e.g., honest or its synonyms), and '?' designates optional. This template retrieves statements such as 'Hillary Clinton is certainly smart' and 'Hillary Clinton is intelligent'. Negation is considered in statements such as 'Hillary Clinton is not decisive', and trait antonyms are also considered such as in 'Hillary Clinton is somewhat unfriendly'. The method calculates a score for each trait using the tweets retrieved by these search templates. Scores are normalized for the number of tweets discussing the trait since individuals may accumulate varying

Candidate	Announce Date	Week Before Announce	Announce Day	Week After Announce	Total
Hillary Clinton	April 12	213,021	342,745	684,721	1,240,487
Ted Cruz	March 23	112,846	269,441	412,460	794,747
Rand Paul	April 7	29,535	227,437	334,134	591,106
Jeb Bush	June 15	95,040	103,710	205,648	404,398
Scott Walker	July 13	57,215	94,640	143,197	295,052
Marco Rubio	April 13	18,659	90,643	177,125	286,427

Table 1: Candidate announcement dates in 2015 and size of tweet collections on the announcement days, the week before, and the week after the announcements. The rows are sorted in the decreasing order of the total number of tweets.

numbers of tweets. These scores allow us to compare candidates in terms of personality perceptions from Twitter users. Further details are in [8].

3.3 Policy

Since public assessments on candidates are also shaped by the candidates' policy preferences, we track Twitter discussions on different policies for each candidate. The list of six policies analyzed include abortion, gay rights, climate change, foreign policy, health care, and immigration. Although not a complete list by any means, these are some of the key issues discussed in our data and in fact were heavily discussed for candidates on Twitter [31]. To identify the list of keywords for each policy, we started with a few well recognized keywords for each policy (e.g., "pro-life" and "pro-choice" for abortion). We then identified other related keywords that co-occurred (e.g., "planned parenthood" was frequently mentioned for abortion). Given the set of keywords for a policy, we extracted all tweets that contain at least one of the keywords.

3.4 Sentiment

In addition to the three factors from *The American Voter*, we also gauge sentiment of the tweets about each candidate, using the more common text mining strategy of sentiment analysis. Given the limited size of tweets (140 characters), we can safely assume that the sentiment detected in political tweets concern some aspect of the entities (in this case, the candidates) mentioned therein [29]. Among several dictionaries for sentiment analysis, we rely on the SENTIWORDNET lexicon [5], 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. In case both scores are equal, we label the tweet as neutral.

4 CANDIDACY ANNOUNCEMENTS

4.1 Data Collection

Based on our anticipation of potential candidates, we decided to monitor Twitter communications for several individuals starting early 2015. Unfortunately, we did not anticipate all individuals who later announced their candidacy. For example, we did not anticipate Bernie Sanders or Donald Trump in early 2015. Therefore, we do not have data for their candidacy announcements. We anticipated other individuals such as Joe Biden and Elizabeth Warren who eventually decided not to run. Table 1 lists six candidates that we study in this work: Hillary Clinton, Ted Cruz, Rand Paul, Jeb Bush, Scott Walker, and Marco Rubio. Note that we excluded some candidates such as Bobby Jindal, Mike Huckabee, and Rick Perry who received very few tweets around their announcements.

We collected tweets concerning each individual posted during the time span of one week before and one week after the announcement. The data were collected using Twitter's streaming API with filter keywords (`statuses/filter`) for each candidate, such as "hillary clinton" for Hillary Clinton. The API tries to provide 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, we are set to capture up to five million tweets per day for each candidate. Note that the highest daily tweet count (for Hillary Clinton) is less than 350K, thus we can safely assume that we are capturing a vast majority of tweets for all candidates. It is also noteworthy that a non-trivial amount of the collected tweets may come from bot accounts on Twitter [39]. In this paper, however, we do not attempt to distinguish between human and bot accounts and leave it as future work. For estimating Twitter users' political affiliations, we also used Twitter's REST API to crawl the follower lists for the Democratic and Republican landmarks.

4.2 Pre and Post Announcement

We note in Table 1 that users tweet more about candidates during the week after their announcement than the week prior. While it is not surprising that the day of the announcement has a big "bump" in the number of tweets, it is noteworthy that this number is larger than the total number received during the entire previous week. For Ted Cruz the number on the day of announcement is more than double for the previous week, while for Rand Paul it goes up by an order of magnitude. Candidate differences in bumps of tweet counts can be explained in part by their popularity. This may also be due to the surprise factor for certain candidates. For example, Ted Cruz (the first major 2016 presidential candidate) was expected to announce his candidacy several days before the announcement day.

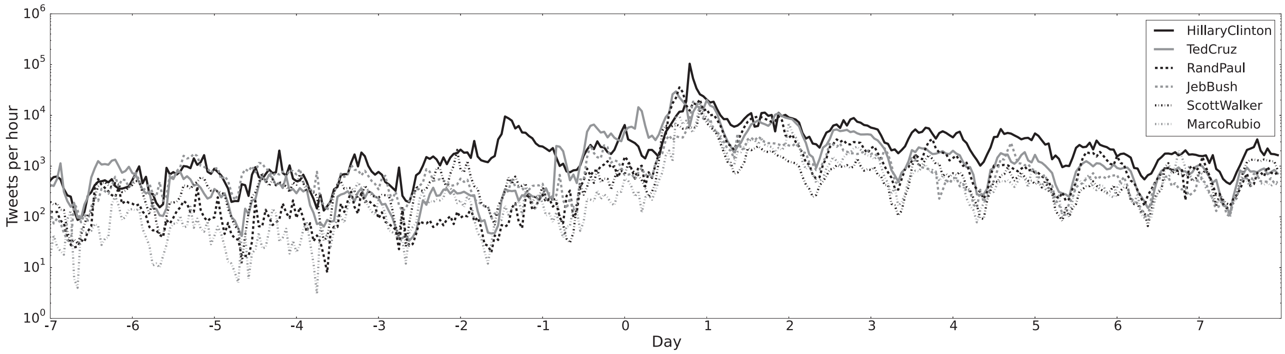
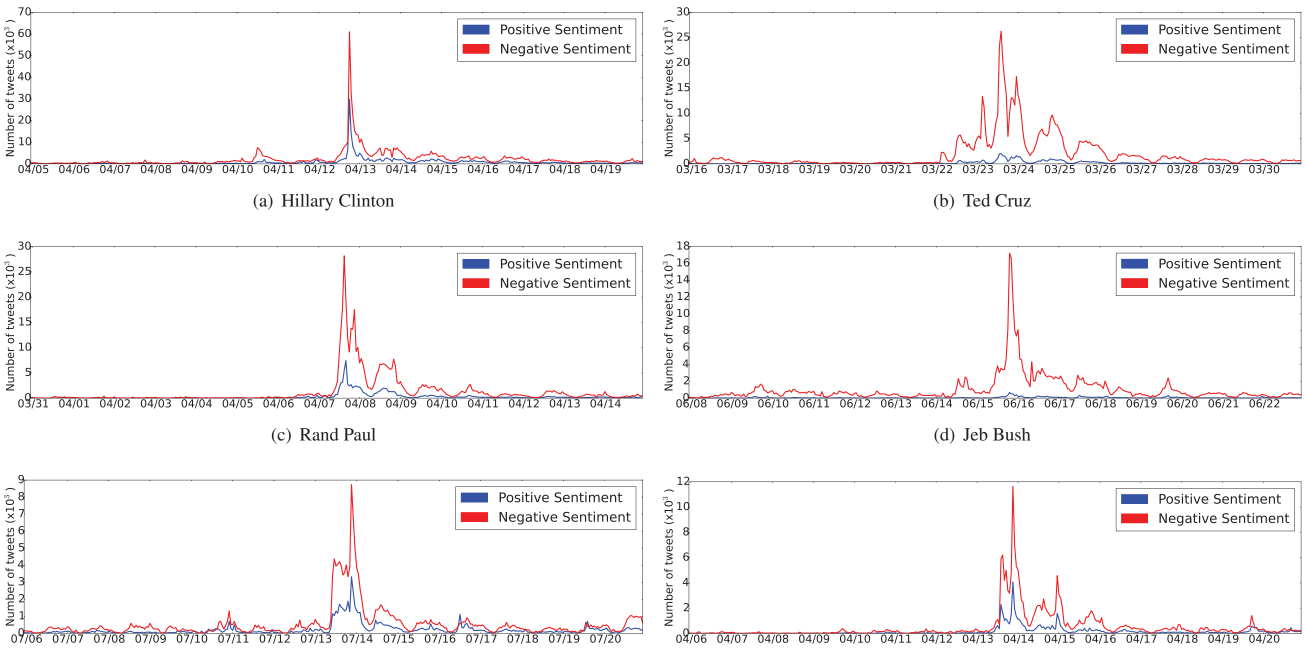


Figure 1: Time series of tweets for candidates a week before and after their candidacy announcements.



(a) Hillary Clinton (b) Ted Cruz
(c) Rand Paul (d) Jeb Bush
(e) Scott Walker (f) Marco Rubio

Figure 2: Time series of positive and negative sentiment for the presidential candidates a week before and after their candidacy announcement. We use the SENTIWORDNET lexicon to classify tweets as positive or negative based on their overall sentiment scores. The y-axis represents the count of tweets labeled as positive or negative.

Figure 1 shows the hourly tweet distribution for the candidates around their announcement dates (note the log scale of the y-axis). The plot confirms that tweet volume peaks at the announcement date (signified as 1 in the Figure). We note sharp bumps most notably for Clinton. A large fraction of bumps comprise of retweets of the candidacy announcement tweets. For example, Hillary Clinton's tweet ("I'm running for president. Everyday Americans need a champion, and I want to be that champion.") was retweeted by more than 100K users. Such retweets for others range from 13K for Ted Cruz, 6K for Jeb Bush, and around 2K for Rand Paul and Scott Walker. The ramp up in tweets to the announcement day (e.g., for Ted Cruz and Hillary Clinton) indicates that announcements were not entirely un-anticipated.

4.3 Sentiment Analysis

We dissect the time series of tweets by applying the sentiment classifier to the captured tweets. Figure 2 shows the resulting volumes of tweets with negative and positive sentiment (excluding neutral, i.e., neither positive nor negative tweets). Interestingly, negative tweets significantly outnumber positive tweets for all candidates, but to a different extent for each. The announcement of Jeb Bush especially prompted a large spike in negative sentiment. Specifically, the ratio of negative to positive tweets was 19.0 for Jeb Bush, 10.9

for Ted Cruz, 4.4 for Rand Paul, 3.4 for Marco Rubio, 2.8 for Hillary Clinton, and 2.6 for Scott Walker. A majority of negative tweets were organized by hashtags. For example, we find 13,743 mentions of #NoMoreBushes for Jeb Bush, 33,458 mentions of #YouCruzYouLose for Ted Cruz, 28,372 mentions of #CantStandRand for Rand Paul, and 9,395 mentions of #WhyImNotVotingForHillary for Hillary Clinton. Overall sentiment analysis indicates bruising negative reactions on Twitter for most candidates.

5 PARTY, PERSONALITY, AND POLICY

5.1 Party

The importance of political affiliation is widely recognized in prior literature [10, 26]. This serves as a key factor in that only a small fraction of people vote against their political affiliations. According to the 2014 Gallup Daily tracking interviews with more than 177K U.S. adults [22], 26% Americans identify as Republicans, 30% Americans identify as Democrats, and 43% American identify as independents. As shown in Table 2, we identify 15.4% of users as Republicans, 20.0% of users as Democrats, and 64.6% of users as Others. Note that while the order of our numbers matches the ordering by Gallup, our method to infer political affiliation of Twitter users based on whether they follow well-recognized Democratic and Republican landmarks offers high precision but low recall. Therefore, the "Other" category may include politically inactive Twitter users whose political affiliation cannot be inferred by our method [13].

Candidate	Democrat (%)	Republican (%)	Other (%)
Hillary Clinton	11.9	15.0	73.1
Ted Cruz	26.1	19.4	54.5
Rand Paul	23.5	14.6	61.9
Jeb Bush	22.0	8.7	69.3
Scott Walker	31.4	16.2	52.4
Marco Rubio	18.9	13.0	68.1
All	20.0	15.4	64.6

Table 2: The party affiliation breakdown of users tweeting for each candidate in percentage.

Table 2 also lists the party breakdown of users who tweet about different candidates. It is interesting to note that the percentage of users from the opposite party significantly outweighs users from the affiliated party for all candidates. For example, 11.9% users who talk about Hillary Clinton are identified as Democrats and 15.0% users are identified as Republicans. As another example, 8.7% users who talk about Jeb Bush are identified as Republicans and 22.0% users are identified as Democrats.

Our findings for political party affiliation help explain the sentiment trends we observed earlier in Figure 2. Recall that the number of negative sentiment tweets clearly outweighed positive sentiment tweets during candidate announcements. Thus, we can conclude that more users negatively talk about the candidates of the opposite party than in support of the candidates of their party. For example, soon after Hillary Clinton's announcement, #WhyImNotVotingForHillary and other negative hashtags started trending on Twitter and it outnumbered #ReadyForHillary and other positive hashtags. Similarly, #NoMoreBushes trended for Jeb Bush and #TedCruzCampaignSlogans trended for Ted Cruz. To mitigate such negative crosstalk at announcements, a possible counter-strategy for campaigns would be to engage their supporters with positive messaging beforehand.

In the following discussion of results for personality and policy, we enrich the analysis by adding the breakdown of results with respect to user party affiliation.

5.2 Personality

Table 3 tabulates the tweet distribution for candidates for the 15 most mentioned traits (of 110 total traits). First, we note that the total number of trait-relevant tweets differs widely across candidates; Cruz has the highest number of tweets and Rubio the lowest. To avoid bias, tweet distribution in Table 3 is shown as percentages instead of counts. The top 15 traits include 9 positive, 3 negative, and 3 neutral traits. The table also breaks down the percentages by traits that are perceived as "present" or "absent" for each candidate. Note that the perceived presence of a positive trait or the absence of a negative trait may be regarded as a *strength*. Similarly, the perceived presence of a negative trait or the absence of a positive one may be regarded as a *weakness*. The top three traits mentioned overall are "easy going" (83.42%), "conservative" (62.68%), and "dull" (38.50%). In fact, easy-going ranks high for three candidates (consistently on the absence side) and conservative for two candidates on the present side. The remaining traits are dominant (10% or greater) for individual candidates. For example, Clinton alone is perceived as deceitful while the tweet percentages of this trait are less than 1% for other candidates.

Figure 3 displays the top ranking 15 traits calculated for each candidate and summary trait scores. The number of tweets matching a trait is listed alongside the trait name. The counts range from a low of 23 (is 'confused' for Walker) to a high of 2,577 (is not 'healthy' for Cruz) and 1,477 (is not 'easy going' for Paul). Scores are from -1 to +1, with -1 (+1) indicating the trait is viewed as absent (present) with high confidence. Puzzled by the "not healthy" perception for Cruz, our further investigation noted some level of noise. For example, we observed tweets such as *Ted Cruz is 'absolutely unfit' for office* and *Ted Cruz is 'sick'*. Clearly unfit and sick are used in a context that is different from physical health.

We focus first on those positive and negative trait bars that touch or cross the 0.5 point in magnitude interpreting these as perceived strengths or weaknesses as described earlier. For Clinton, her top two perceived strengths are that she is poised and persistent. Her top two weaknesses are that she is not seen as natural or handsome. Cruz's top two perceived strengths are that he is not meek and he is equally humorous and aggressive. His leading weaknesses are in not being perceived as moderate or healthy. Discounting healthy because of the noise described earlier, his second weakness is that he is not perceived as being wise. As strengths, Paul is perceived as assertive and not greedy. As weaknesses, he is perceived as rude and not mannerly. Bush's top two strengths are equal; he is not perceived as cold or unscrupulous. His two top weaknesses are that he is perceived as silent and not easy going. Walker's top two strengths are in coming across as dominant and not meek. But he is also perceived as not pleasant and equally not easy going or shrewd. As strengths, Rubio is not perceived as meek or greedy. But then

#	Trait	Clinton		Cruz		Paul		Bush		Walker		Rubio		Total
		pre	abs	pre	abs	pre	abs	pre	abs	pre	abs	pre	abs	
1	**easy going(+)**	0.91	**15.42**	0.05	**13.01**	0.10	**37.71**	0.06	5.47	0.11	7.20	0.14	3.25	**83.42**
2	**conservative (=)**	4.48	0.37	5.87	0.19	**10.13**	0.32	3.76	1.03	5.86	0.33	**30.12**	0.21	**62.68**
3	**dull (-)**	0.05	4.07	0.15	3.18	0.42	0.42	0.80	6.33	**16.52**	3.24	2.56	0.76	**38.50**
4	**healthy (+)**	0.12	2.48	0.86	**23.98**	0.29	0.69	0.06	0.17	0.11	0.28	0.76	0.21	**30.01**
5	**shrewd (+)**	4.03	0.30	3.86	1.21	0.64	0.42	1.43	1.25	0.78	6.98	0.69	3.67	**25.26**
6	**intelligent (+)**	4.09	0.25	4.52	0.74	0.74	0.44	1.60	2.28	0.73	3.91	0.69	3.19	**23.16**
7	**dissatisfied (=)**	0.00	0.14	0.07	0.02	0.00	0.00	**20.52**	0.17	0.28	0.22	0.07	0.00	**21.50**
8	**moderate (+)**	0.05	0.04	0.09	5.91	3.85	1.37	0.46	0.80	0.22	5.69	0.00	0.00	**18.48**
9	**spontaneous (=)**	0.04	9.53	0.00	5.73	0.00	0.22	0.00	0.29	0.00	0.50	0.00	0.48	**16.79**
10	**deceitful (-)**	**12.42**	0.21	0.06	0.25	0.34	0.17	0.23	0.17	0.56	0.00	0.00	0.35	**14.25**
11	**cold (-)**	0.05	0.27	0.07	0.16	0.00	0.44	0.06	**10.21**	0.06	0.28	0.00	1.18	**12.77**
12	**dominant (+)**	0.00	0.02	0.07	0.00	0.07	0.00	0.00	0.06	**11.72**	0.00	0.07	0.00	**12.00**
13	**pleasant (+)**	0.07	1.55	0.02	0.52	1.72	2.23	0.74	1.03	0.00	3.68	0.35	0.07	**11.98**
14	**poised (+)**	**10.31**	0.00	0.02	0.00	0.17	0.00	0.06	0.00	1.00	0.00	0.35	0.00	**11.91**
15	**honest (+)**	0.36	8.28	0.88	0.11	0.22	0.32	0.06	0.34	0.28	0.17	0.48	0.00	**11.49**
#total trait-relevant tweets		**5,604**		**10,376**		**4,079**		**1,754**		**1,792**		**1,444**		

Table 3: Tweet distribution for candidates for 15 most frequently discussed traits (by %). Note that (+) represents a positive trait, (-) represents a negative trait, and (=) represents a neutral trait. The values exceeding 10% are in highlighted in bold.

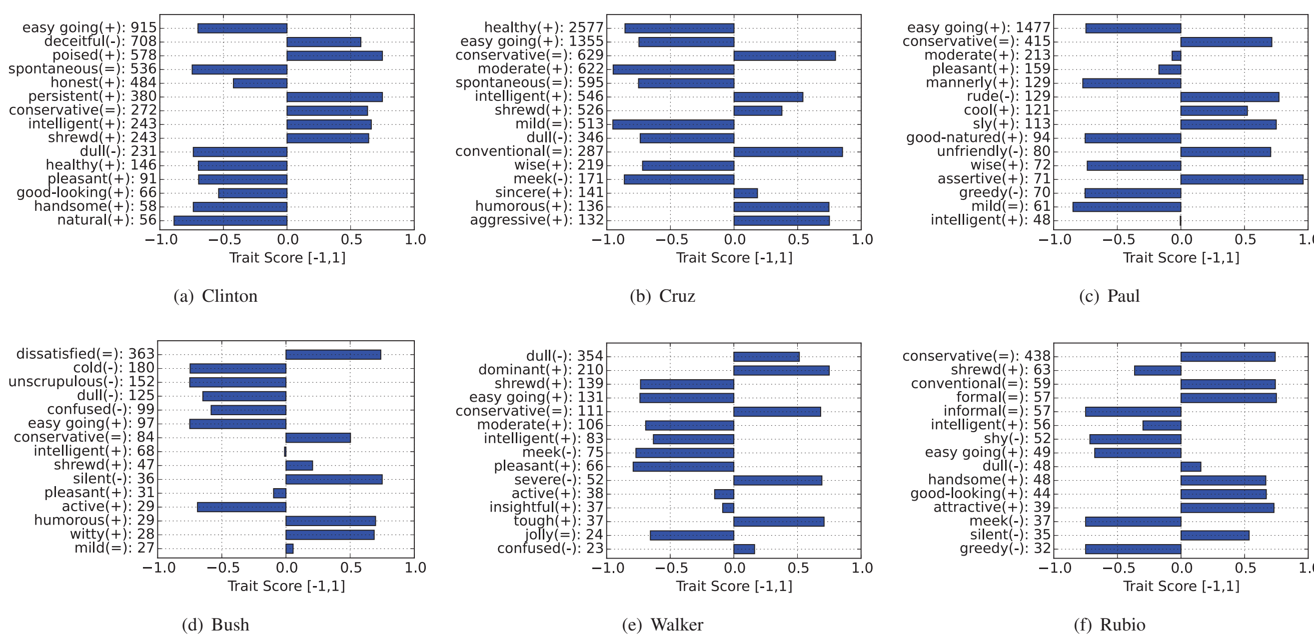
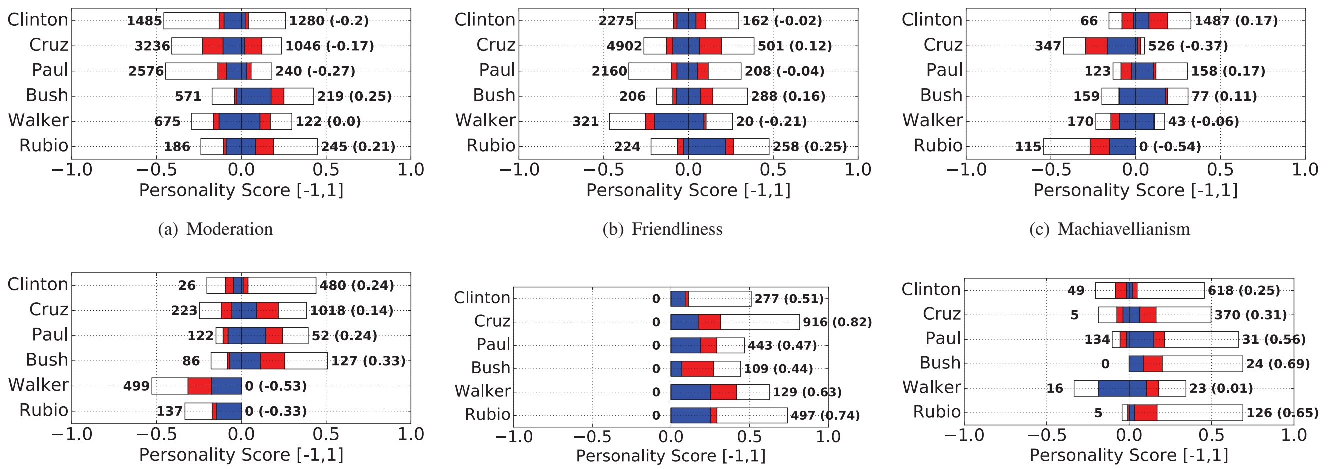


Figure 3: Summary trait scores of top 15 (ranked in the descending order of tweet frequency) personality traits for each candidate. Scores are in the range -1 to +1; -1 (+1) indicates the trait is viewed as absent (present) with high confidence.

he is perceived as silent and not easy going. Overall the perceived strengths and weaknesses are distinct for each candidate.

Figure 4 plots the six most discussed personality dimensions (aggregated from relevant traits) for each candidate. The dimensions are refined with respect to party affiliation and their presence and absence. Scores are in the range [-1,+1]; -1 (+1) indicates the personality is viewed as absent (present) with high confidence. For succinct analysis, we consider personality dimensions whose net scores cross the 0.5 threshold. We note that no candidate stands out in 'Moderation' or 'Friendliness'. Only Walker is seen as lacking in 'Intellectual Brilliance' while only Rubio is seen as lacking



(a) Moderation (b) Friendliness (c) Machiavellianism

(d) Intellectual Brilliance (e) Conservatism (f) Poise and Polish

Figure 4: Personality scores with party breakdown. Scores are in the range -1 to +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 blue, red, white regions indicate tweets from Democrats, Republicans, and Others, respectively.

'Machiavellianism'. Multiple candidates stand out in the remaining dimensions. For example, Paul, Bush, and Rubio are perceived as having 'Poise and Polish'. All candidates are also perceived as conservative but interestingly Bush and Paul do not cross the threshold. Combining party affiliations and personality scores, we note that most candidates received more tweets from Democratic-leaning users than Republican-leaning users. This is most obvious for Walker and Rubio who received about three times more tweets from Democratic-leaning users than from Republic-leaning users. In contrast, Clinton and Paul received about the same numbers of tweets from users affiliated with both parties: around 4% from each group for both candidates.

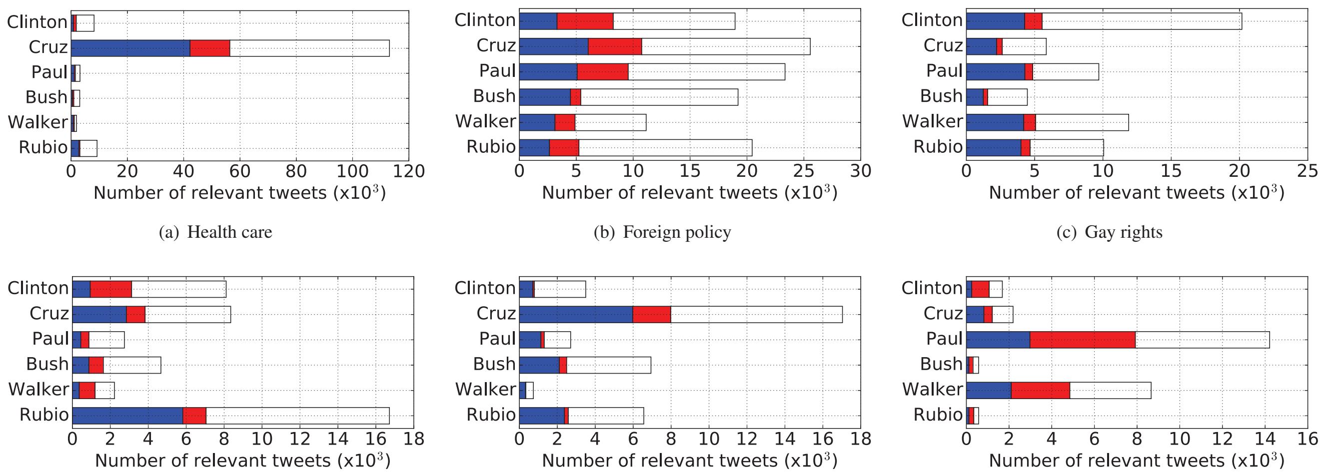
5.3 Policy

The policies advocated by candidates are an important consideration in their public assessment. Out of the policies that we considered, the ones most frequently mentioned in the data are: health care (also known as Obamacare), foreign policy, gay rights (particularly same sex marriage), immigration, climate change, and abortion. We find that about 22.9% of all tweets match at least one of these policies. More specifically, 7.4% of all tweets mention health care, 6.9% tweets mention foreign policy, 2.8% tweets mention gay rights, 2.2% tweets mention climate change, 1.9% tweets mention immigration, and 1.7% of tweets mention abortion. We noticed a mix of partisan perspectives for each candidate on different policies, due to the partisan crosstalk on online social media. Based on our identification of party affiliation, we were able to get a clear picture of how users from both parties are communicating about the policies of the candidates.

Figures 5(a) through (f) compare the breakdown of tweets for different policies. Overall, we note that the blue area is mostly larger than the red area, which means there were more Democratic tweets from those we could identify as Democrats or Republicans. We observed that criticisms are generally more frequent than support for all candidate, with foreign policy being the most popular. That is an unexpected finding as generally foreign policy has not been important in elections [4, 31]. The spike of mentions of health care or Obamacare when tweeting about Cruz is also striking. Clinton leads the way on gay rights and gay marriage. Her position is well-known and it is generally supported. Paul and Walker stand out as most frequently mentioned for their policies on abortion. Rubio is the most frequently mentioned of the candidates about immigration. We next separately discuss our observations for different candidates, achieved via a manual expert analysis of the most frequent tweets.

Hillary Clinton: Foreign policy was one of the two most frequently mentioned, as Clinton's tenure as the Secretary of State as well as other current stances drove the conversation. It was overwhelmingly negative, with the ratio of mentions by Republicans to Democrats at 5 to 3. Her support for the Iran nuclear deal was most frequently mentioned, Republicans being uniformly critical of it. We also observed tweets about other issues such as ISIS, email scandal related to the 2012 Benghazi attack, Iraq, Israel, Cuba. Other international relations, such as those involving Europe and Russia, were not as prevalent. This contrasts with the other most prominent policy area of gay rights and especially her commitment on gay marriage, where Democratic mentions were double the mentions by Republicans, setting an overall positive tone. Most frequently mentioned stance was "same-sex marriage should be a constitutional right," and it was even a part of her announcement: e.g., "Hillary Clinton featured an engaged gay couple in her Presidential campaign announcement."

Ted Cruz: The major focus of tweets about Ted Cruz was about ObamaCare; two-thirds were about health care. This was largely due to the incident in which Ted Cruz, a noted critic of Obamacare, used



(a) Health care (b) Foreign policy (c) Gay rights

(d) Immigration (e) Climate change (f) Abortion

Figure 5: Number of mentioned tweets for different policies with party affiliation breakdown. The blue, red, white regions indicate tweets from Democrats, Republicans, and Others, respectively.

it to buy health insurance for his wife, producing a large spike of derision such as: "Ted Cruz signs up for #Obamacare after vowing to repeal every word of it." Interestingly, he was criticized on this issue by both Democrats and Republicans. Cruz was also a critic of various Obama policies involving relationships with Iran and Cuba, climate change, immigration, women's health care and abortion, often garnering Democratic attention at 2 to 1 compared to his Republican supporters.

Rand Paul: As with the other candidates, foreign policy was mentioned more frequently for Rand Paul than other policy areas. Besides, his position on abortion provoked much criticism from Democrats, making the issue most prominent for Paul compared to other Republican candidates. Health care or Obamacare, climate change, gay rights, and abortion were all mentioned much more frequently by Democrats than by Republicans. The Democrats were critical of his positions for each of these policy areas.

Jeb Bush: Democrats' criticism of Bush's foreign polices concerning Israel, Iran, ISIS, and Cuba (by a 10 to 1 ratio compared to Republican tweeting) made it the most frequent. He was the only candidate for whom Putin was mentioned frequently, and it was negative by more than 2 to 1. Similarly, pungent criticism extended to his positions on climate change, gay rights, and gay marriage.

Scott Walker: Walker's stance on gay rights garnered an overwhelmingly negative discussion, with Democrats tweeting five times more than Republicans about the issue. Further, Democratic critique extended to his positions on the arrangements with Iran, his conception about how to deal with ISIS, and foreign policy in general. Although the mentions of his position on women's health care and abortion were about even between the parties, the opinions were again divided along the partisan lines with Republican support and Democratic opposition.

Marco Rubio: The attention to Rubio's policy positions was overwhelmingly critical. Democrats outnumbered Republicans on his policies on immigration, health care, climate change, and gay rights. Tweets about his positions on foreign policy were almost equally distributed between Democrats, who were critical, and Republicans, who were supportive. Only on policy on abortion was there a substantial number of mentions by Republicans, and that was the least frequently mentioned of the policy areas.

6 CONCLUDING REMARKS

In this paper, we studied Twitter communications of six major presidential campaign announcements, an early milestone in the process of presidential elections in the U.S. We studied how party, personality, and policy impact the discourse on Twitter after candidacy announcements. Our study was designed with two objectives in mind.

Our first objective was to move from traditional survey research to online social media. Specifically, we wanted to characterize online social media communications about the candidates in terms of party, policy, and personality. To this end, we employed text mining techniques to explicate views on candidate personalities and their stance on different policy issues. We were further able to characterize policy issues and personality traits of the candidates by analyzing partisanship. Our methods to measure party, policy, and personality aspects in online social media communications are reusable in the context of different elections.

Our second objective was to establish a starting baseline from which we can track changes in public view of the candidates. The public view of the candidates is expected to change as the campaigns evolve over time. To illustrate this change, we compare our results from the announcement period to a later period that covers several debates and primaries [25]. In terms of personality, for example, our comparative analysis shows that Clinton was perceived as not moderate (moderation personality score of -0.2) when she announced her candidacy but this perception changed to moderate (score of 0.28) afterwards. As another example, Rubio was perceived as not intellectually brilliant (intellectual brilliance personality score of

-0.54) when he announced his candidacy but this perception changed (score of 0.04) later. In terms of policy, gay rights were mostly mentioned for Clinton during the announcement period but they were mostly mentioned for Cruz later. While some of our findings may not generalize for other elections, we envision that our results would serve as a baseline for future election campaigns.

REFERENCES

[1] Hillary Clinton on Twitter: "I'm running for president. Everyday Americans need a champion, and I want to be that champion. H https://t.co/w8Hoe1pbtC". https://twitter.com/hillaryclinton/status/587336319321407488, April 2015.

[2] Ted Cruz on Twitter: "I'm running for President and I hope to earn your support! http://t.co/0UTqaIoytP". https://twitter.com/tedcruz/status/579857596191899648, March 2015.

[3] US presidential candidate announcements: A history of hat throwing. The Economist. http://www.economist.com/blogs/graphicdetail/2015/06/us-presidential-candidate-announcements, March 2015.

[4] J. Aldrich, C. Gelpi, P. D. Feaver, J. Reifler, and K. T. Sharp. Foreign Policy and the Electoral Connection. *Annual Review of Political Science*, 9:477–502, 2006.

[5] S. Baccianella, A. Esuli, and F. Sebastiani. Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In *LREC*, volume 10, pages 2200–2204, 2010.

[6] F. R. Baumgartner, F. R.tner and B. D. Jones. *Agendas and Instability in American Politics*. Chicago : University of Chicago Press, 1993.

[7] A. Bermingham and A. F. Smeaton. On Using Twitter to Monitor Political Sentiment and Predict Election Results. In *Sentiment Analysis where AI meets Psychology*, 2011.

[8] S. Bhattacharya, C. Yang, P. Srinivasan, and B. Boynton. Perceptions of Presidential Candidates' Personality in Twitter. *Journal of the Association for Information Science and Technology*, 2015.

[9] P. Burnap, R. Gibson, L. Sloan, R. Southern, and M. Williams. 140 characters to victory?: Using Twitter to predict the UK 2015 General Election. *Electoral Studies*, 41:230–233, March 2016.

[10] A. Campbell, C. Philip, M. Warren, and S. Donald. *The American Voter*. University of Chicago Press, 1960.

[11] A. Ceron, L. Curini, and S. M. Iacus. 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):340–358, 2014.

[12] R. W. Cobb and C. D. Elder. *Participation in American Politics: The Dynamics of Agenda Building*. Johns Hopkins University Press, 1972.

[13] R. Cohen and D. Ruths. Classifying Political Orientation on Twitter: It's Not Easy! In *Proceedings of the Seventh International AAAI Conference on Weblogs and Social Media (ICWSM)*, 2013.

[14] M. D. Conover, B. Goncalves, J. Ratkiewicz, A. Flammini, and F. Menczer. Predicting the political alignment of twitter users. *Proceedings of the International Conference on Social Computing*, 2011.

[15] M. D. Conover, J. Ratkiewicz, M. Francisco, B. Goncalves, F. Menczer, and A. Flammini. Political Polarization on Twitter. In *AAAI International Conference on Weblogs and Social Media (ICWSM)*, 2011.

[16] J. DiGrazia, K. McKelvey, J. Bollen, and F. Rojas. More tweets, more votes: Social media as a quantitative indicator of political behavior. *PLOS ONE*, 11(8), 2013.

[17] D. Gayo-Avello. "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.

[18] D. Gayo-Avello, P. T. Metaxas, and E. Mustafaraj. Limits of electoral predictions using twitter. In *AAAI International Conference on Weblogs and Social Media (ICWSM)*, 2011.

[19] J. Golbeck and D. Hansen. A method for computing political preference among twitter followers. *Social Networks*, 2014.

[20] H. G. Gough and A. B. Heilbrun. The Adjective Check List Manual. Consulting Psychologists Press, 1983.

[21] S. Hong and D. Nadler. 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):455–461, 2012.

[22] J. M. Jones. In U.S., New Record 43% Are Political Independents. Gallup. http://bit.ly/17uCqDC, January 2015.

[23] A. Jungherr, P. Jrgens, and H. Schoen. Why the Pirate Party Won the German Election of 2009 or The Trouble With Predictions. *Social Science Computer Review*, 2(30):229–234, 2012.

[24] H. Le, Z. Shafiq, and P. Srinivasan. Scalable News Slant Measurement Using Twitter. In *AAAI International Conference on Weblogs and Social Media (ICWSM)*, 2017.

[25] H. T. Le, G. Boynton, Y. Mejova, Z. Shafiq, and P. Srinivasan. Revisiting The American Voter on Twitter. *ACM Conference on Human Factors in Computing Systems (CHI)*, 2017.

[26] M. Lewis-Beck, J. G. William, N. Helmut, and W. F. Herbert. *The American Voter Revisited*. University of Michigan Press, 2008.

[27] Y. Mejova, P. Srinivasan, and B. Boynton. GOP Primary Season on Twitter: "Popular" Political Sentiment in Social Media. In *WSDM*, 2013.

[28] J. S. Morgan, C. Lampe, and M. Z. Shafiq. Is News Sharing on Twitter Ideologically Biased? In *ACM Conference on Computer-Supported Cooperative Work (CSCW)*, 2013.

[29] B. O'Connor, R. Balasubramanyan, B. R. Routledge, and N. A. Smith. From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series. In *AAAI International Conference on Weblogs and Social Media (ICWSM)*, 2010.

[30] M. Pennacchiotti and A.-M. Popescu. A machine learning approach to twitter user classification. *AAAI International Conference on Weblogs and Social Media (ICWSM)*, 2011.

[31] T. Rupar, A. Blake, and S. Granados. The ever-changing issues of the 2016 campaign, as seen on Twitter. http://wapo.st/1OgbJ45, December 2015.

[32] E. T. K. Sang and J. Bos. Predicting the 2011 Dutch senate election results with Twitter. In *Workshop on Semantic Analysis in Social Media (SASN)*, 2012.

[33] D. O. Sears and J. L. Freedman. Selective Exposure to Information: A Critical Review. *Public Opinion Quarterly*, 31(2):194–213, 1967.

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

[35] M. Skoric and N. Poor. Tweets and Votes: A Study of the 2011 Singapore General Election. In *Hawaii International Conference on System Sciences (HICSS)*, 2012.

[36] A. Smith. Cell Phones, Social Media and Campaign 2014. Pew Research Center, November 2014.

[37] S. Stieglitz and L. X. Dang. Social media and political communication: a social media analytics framework. *Social Network Analysis and Mining*, 4(4):1277–1291, December 2013.

[38] A. Tumasjan, T. O. Sprenger, P. G. Sandner, and I. M. Welpe. Predicting elections with twitter: What 140 characters reveal about political sentiment. In *AAAI International Conference on Weblogs and Social Media (ICWSM)*, 2010.

[39] O. Varol, E. Ferrara, C. A. Davis, F. Menczer, and A. Flammini. Online Human-Bot Interactions: Detection, Estimation, and Characterization. In *AAAI International Conference on Weblogs and Social Media (ICWSM)*, 2017.

[40] H. Wang, D. Can, A. Kazemzadeh, F. Bar, and S. Narayanan. 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*, pages 115–120, July 2012.

[41] D. B. Wood and A. Doan. The politics of problem definition: A theory and application to sexual harassment. *American Journal of Political Science*, 47(4):640–653, October 2003.

[42] D. B. Wood and A. Vedlitz. Issue definition, information processing, and the politics of global warming. *American Journal of Political Science*, 51(3):552–568, July 2007.

[43] F. A. Zamal, W. Liu, and D. Ruths. Homophily and latent attribute inference: Inferring latent attributes of twitter users from neighbors. *AAAI International Conference on Weblogs and Social Media (ICWSM)*, 2012.

[44] A. X. Zhang and S. Counts. Modeling ideology and predicting policy change with social media: Case of same-sex marriage. *ACM Conference on Human Factors in Computing Systems (CHI)*, 2015.

[45] D. X. Zhou, P. Resnick, and Q. Mei. Classifying the Political Leaning of News Articles and Users from User Votes. In *AAAI International Conference on Weblogs and Social Media (ICWSM)*, 2011.