All the President’s Tweets:

Studying the Big Data of Twitter Political Communication with a Small Data Approach

Stan Oklobdzija, UC San Diego
and
Thad Kousser, UC San Diego

Abstract. This paper describes and demonstrates an approach to studying political communication on Twitter that combines the flexibility and subtlety of human coding on small datasets with the scalability of big data methods. We apply this method to study the 152,134 tweets sent by candidates and their SuperPACs during the 2016 American presidential election. We begin by outlining the importance of studying this increasingly prominent form of political communication, laying out both the opportunities and obstacles to analyzing the massive volume of social media political texts. We then explain our supervised learning method to classify the ideology and sentiment of tweets, what policy spheres they address, whether they make factual statements or espouse opinions, and other characteristics. To demonstrate what this approach to studying political communication can tell us about a campaign, we then focus on our measure of ideology. We map the candidates on a spatial spectrum based on their social media rhetoric, explore Donald Trump’s unique brand of populism by looking at where he placed himself on different issues, and asked whether he and Hillary Clinton followed the Downsian impulse to moderate their message from the primary to the general election.
The 2016 presidential contest was the “Twitter Election,” marking a qualitative shift in the way that candidates harnessed the power of social media to communicate directly with voters. Speaking to far larger audiences than candidates before them and tweeting more actively and more personally than ever before, @realDonaldTrump, @HillaryClinton and the rest of the field used social media to bypass the traditional media. Twitter blurred the distinction between paid and free media, allowing candidates the complete control over their messages that they’d previously exerted only through costly advertisements. All of this came without spending a cent for airtime and freed candidates from the mercy of the journalists who arbitrate a campaign’s message via earned media. It put voters constantly in touch with candidates, giving them exposure to a campaigner’s rhetoric, positions, and personality that was previously granted mostly just to residents of Iowa and New Hampshire. Twitter in 2016 created an immediate and unfiltered link between political leaders and their followers at a scale rarely seen in American politics.

While Twitter has created new opportunities for scholars to study political communication through social media, it has, at the same time, posed new challenges. As tweeting became a ubiquitous and vital campaign strategy, the 23 candidates with active accounts and 27 SuperPACs supporting them produced a massive number of political messages, sending 152,134 tweets overall, including over 111,000 once officially declared for the race. Twitter provides detailed information

---


about all of these communications, including the exact timing of each tweet and how many times it was retweeted or “liked.” The temporal granularity of Twitter allows us to see how politicians change their messages as they rise or fall in the polls, when they move from a primary to the general election, or after they are inaugurated, all while holding constant their medium of communication. Social media’s feedback mechanisms allow campaigners and scholars alike to gauge the instant reaction of an audience to different types of messages, revealing, for better or for worse, exactly what online followers really want to hear from their leaders.

Yet the ever-expanding scale of Twitter political communication poses a great challenge to studying its rhetorical complexity. One approach is to tackle that complexity through human ingenuity, using research teams to do an in-depth coding of tweets. This approach, used fruitfully in works such as Golbeck et al. (2010), Himelboim et al. (2013), Mejova et al. (2013), and Evans and Clark (2016) necessitates studying a relatively small number of tweets or categorizing only one or two aspects of their content. An alternative is to rely on “unsupervised learning” techniques of automated text analysis to categorize massive number of tweets (Murthy 2015; Obschonka 2017; Bhattacharya et al. 2015) or to infer their ideology using the social networks of politicians and their followers (Barbera 2015a, Barbera et al. 2015, King, Orlando, and Sparks 2015, Alles and Jones 2016).

Our approach seeks to leverage the advantages of in-depth human coding, but to use the machine learning algorithms trained by these coding decisions to categorize tweets on a large scale. We adopt the “supervised learning” approach (see Grimmer and Stewart 2013 or Peterson and Spirling, n.d.), which begins with a team of researchers coding thousands of texts by hand. Human intelligence train artificial intelligence algorithms to map the use of certain words and phrases to
concepts such as the ideology or sentiment of a tweet, replicating the coding choices for tens or hundreds of thousands of tweets.

By combining the scope of big data analysis with the traditions of small data research, the supervised learning approach possesses several clear strengths. It forces researchers to dig into their data, reading document after document after document in order to make coding choices rather than simply applying an unsupervised algorithm to a massive but unexplored dataset. It allows researchers to define which concepts they would like to measure, instead of forcing them to adopt the measures that are available “off-the-shelf” in existing text analysis tools. Perhaps most importantly, it gives researchers a clear check on how well their algorithms perform. With human coding choices to compare machine-generated categorizations against, it is possible to see which measures pass, and which fail, validity tests.

In the first section of this short essay, we describe how we apply supervised learning techniques to create an original dataset of every tweet sent during the 2016 presidential contest by the 23 major candidates and their affiliated SuperPACs. We measure the ideology of these tweets, their sentiment, whether they focused on factual claims or expressed opinions, whether they were personal or political in nature, what policy realm they addressed, and whether the political leaders asked their followers to take certain actions.

In the second section, we focus closely on one of those measures – ideology – in order to demonstrate what can be learned about political communication through this approach. We arrange the major candidates on a common ideological scale based upon the content of their political communication through Twitter, showing just how little difference there was between the messages sent by Hillary Clinton and Bernie Sanders but how widely the leading Republican candidates ranged
in their online ideologies. We focus closely on Donald Trump, first identifying the issues on which took consistently conservative stands and those on which he moderated, presenting a method to learn more about the contours of his particular brand of populism. Finally, we leverage the frequency and precise timing of tweets to chart how the messages of the major party nominees evolved over the campaign, asking whether Donald Trump and Hillary Clinton followed the Downsian impulse to pivot toward the center during the general election or whether one of them bucked conventional wisdom in their political communication.

Section I. Method

Our goal in describing the method that we used to categorize presidential campaign tweets is two-fold. First, we want to present a primer on this application of supervised machine learning, so that other scholars could use this approach to study political communication through social media in other nations, by other American politicians, or in other elections. We have sought to lay out a step-by-step overview of how we proceeded, providing clear metrics of the reliability of all of our variables and passing along lessons we learned that could save other scholars significant labor.

Second, we aim to clearly explain how we collected and coded a dataset that we will share with the scholarly community and update. We will post all 7,525 of our human-coded tweets as well as the machine-coded categorizations of the full set of 152,134 tweets from these candidates and SuperPACs during the election cycle. We have continued to collect, code, and categorize tweets from all of these politicians since the election, and will release this updated dataset at the conclusion of President Trump’s first year in office.
The key idea behind a supervised learning approach to text analysis is to begin with a small sample from the full set of texts being studied, reading and coding them intensively by hand and then training and testing machine learning algorithms on this small dataset before using the algorithms to classify the entire corpus. Starting in October 2015, we used Twitter’s public API to download tweets from each major candidate’s account, and from the account of their affiliated SuperPACs. We also created scripts that automatically downloaded all tweets from these 50 accounts from that point onward, creating a complete dataset of tweets running through January, 2017.\(^3\) Starting in June, 2016, we worked with a team of six graduate and undergraduate research assistants to categorize these tweets. Our coders read only the text of the tweet, receiving no information about who sent it. We created a codebook outlining operational definitions of the concepts that we sought to measure as well as the types of tweets that would fit into each category. We met twice weekly to discuss the concepts and to debate how individual tweets should be categorized, updating our codebook with these difficult cases and the reasoning behind our decisions. We continued these meetings through May, 2017, eventually coding 7,525 tweets.\(^4\)

A great virtue born from the necessity of delving so deeply into the data is that we, along with our coders, began to see patterns in how candidates used social media to communicate throughout the campaign. One of these patterns is illustrated in Figure 1, a “comparison cloud” that focuses only on the tweets of Donald Trump and Hillary Clinton. This type of cloud displays words

---

\(^3\) Our dataset goes back at least as far as the date on which each candidate declared for the 2016 presidential race. For candidates who do not tweet very frequently, it goes back much further, with the earliest tweet coming from Rick Perry in January 2009. Because we have recorded each candidate’s official declaration date, we can exclude earlier tweets from any of our analyses.

\(^4\) Because we began coding tweets before the campaign ended, after drawing a random sample from our full set of tweets generated by June 2016, we drew additional random samples of tweets from the summer and fall of 2016 so that the distribution across time of the tweets in our hand-coded sample reflected the timing of the full corpus of tweets.
in a larger font if one candidate uses them much more frequently than the other. What this figure shows is that Donald Trump had a clear and consistent message in his campaign: “Make American Great Again.” It rang out through his aggregated set of over nine thousand tweets as clearly as it did in his stump speech, demonstrating a level of message discipline in his social media communication that is at odds with his image as a loose cannon on Twitter. By contrast, Hillary Clinton hit on many themes in her campaign – “families,” “women,” “work,” “economy,” and “rights,” – but did not use Twitter to broadcast one clear message.

In Table 1, we provide examples of some of our key measures, as well as tweets that fit into each category of these variables. These three examples include a core concept of political science (ideology), a variable that is often used to capture the tone of online discourse (sentiment), and a variable that we created in order to examine an emerging theme of the 2016 election (whether the tweet primarily makes a factual claim or whether it professes and opinion). The flexibility to create whatever variables a scholar wishes to capture a classic concept, to test a new theory, or to mirror widely accepted political science operationalizations is key strength of the supervised learning approach. Another set of variables that we coded put each tweet into one of the ten subject areas, using the topic list created and applied to democracies around the world by the Policy Agendas Project (The Policy Agendas Project at the University of Texas at Austin, 2017).

In Table 2, we report measures of intercoder reliability for all of our variables, demonstrating strong levels of agreement when different research assistants independently coded the same tweets. This analysis is based on a set of 1,217 tweets, which we assigned to overlapping pairs of coders so

---

5 Because several of the topic areas created by the Policy Agendas Project -- agriculture, labor, energy, transportation, social welfare, housing, domestic commerce, technology, foreign trade, international affairs, public lands, culture – feature zero or very few tweets, we consolidated this handful of tweets into related topic areas.
that each researcher was paired with all others. We report first the rate of agreement between 
coders and then the Cohen’s Kappa, a measure of how much more likely the researchers are to agree 
than we would expect by pure chance. Our rates of agreement range from 75% on our 
three-category sentiment measure to perfect agreement on three of our subject areas, with the 
Cohen’s Kappa measures ranging from “fair” to “almost perfect” agreement levels for all but one of 
our variables. These figures show that humans can reliably record the meaning of political tweets.

Our final step to creating our dataset was to use all of the 7,525 tweets that humans coded 
by hand to train machine learning algorithms. We began by holding out 725 of these as a final 
testing set, providing the check on the performance of these algorithms that we report in Table 3. 
We then “pre-processed” all of our tweets using standard methods for text analysis. We trained 
algorithms on 80% of the tweets, allowing them to learn what words and phrases in a tweet 
corresponded to human categorizations of it as “liberal,” “negative,” “opinion,” etc. We then tested 
the algorithms on the remaining 20% of the tweets, asking how often they could replicate the coding 
decisions made by humans. We selected the best performing algorithms for each variables, trained 
them on all tweets in our training and testing set, and then performed our the final tests on the 725

---

6 This is an important alternate measure of intercoder reliability that takes into account how much more challenging it is 
to reach agreement on three-category variables such as ideology and sentiment than on two-category variables, and also 
how much more likely coders are to agree on variables, such as our subject matter codes, that nearly always take on a 
value of zero. According to Landis and Koeh (1977, p. 165), a Kappa Statistic of below zero indicates a “Poor” strength 
of agreement, a level between 0.00-0.20 is “Slight,” 0.21-0.40 is “Fair,” 0.41-0.60 is “Moderate,” 0.61-0.80 is 
“Substantial,” and 0.81-1.00 is an “Almost Perfect” strength of agreement.

7 We stemmed words so that words like “tax” and “taxes” would be read interchangeably, we made all words into lower 
case, we removed URLs, and we deleted the Twitter handles of candidates from the text of tweets to concentrate our 
analysis on the content that the tweets contained. We turned each tweet into a “term frequency-inverse document 
frequency” vector to highlight words that are used often in one tweet but rarely in others (see Colleoni, Rozza, and 
Arvidsson 2014 for a similar application to Twitter), and allowed words and phrases to enter into our analysis as 
unigrams, bigrams, and trigrams (one-, two-, or three-word phrases).

8 We used a standard set of text analysis algorithms, including multinomial negative binomial (which performed best for 
classifying sentiment), linear support vector machines (which performed best for whether a tweet made a factual claim 
and whether it was political or not), and the bagging classifier (which performed best for the remaining 14 variables)
out-of-sample tweets that we report in Table 3. We then used these algorithms to classify our entire set of 152,134 tweets.

Table 3 shows how reliably, or unreliably for a few cases, the algorithms performed. The first column reports recall rates, the proportion of tweets which our human coders put in a category that the algorithm correctly places in the category. The second column reports the Cohen’s Kappa. For variables that take on two categories, the best performing algorithm typically replicated the human coding more then 90% of the time. With three categories, ideology and sentiment presented more challenging tasks, but the Cohen’s Kappa figures show that the algorithms did much better than we would expect by chance alone. Yet for six of our variables – whether a tweet was personal or political, and whether its subject was Law and Crime, Civil Rights, Education, Environment, or Government Operations – the Cohen’s Kappa was low enough that we do not consider these sufficiently reliable variables to use in our analyses. This points out a final strength of the supervised learning approach: by providing a human check on how well algorithms perform, researchers can make informed decisions about which classifications are meaningful and which are not.

Section II. Ideology

Following the foundational work of Poole and Rosenthal (1985, 1997), political scientists have measured the ideology of politicians on a spatial scale. One challenge to these analyses is to bring together a field of candidates with diverse backgrounds – such as the senators, governors, and business leaders who ran for the presidency in 2016 – into a common ideological space. Two blog postings during that campaign applied peer-reviewed methods to do so, either by measuring the networks of donors who contributed to candidates (Bonica 2014, CROWDPAC 2016) or the
network of Twitter users who followed candidates, news organizations, and interest groups (Barbera 2015a, 2015b).

Our approach measures the ideology that candidates express through their political communication on social media. We present scores for each candidate that represent the mean ideology of their tweets, with a liberal tweet counting as negative one, a neutral tweet as zero, and a conservative tweet as one. Our scores correlate at $r=0.96$ with those reported by CROWDPAC (2016), an important sign of convergent validity. The granularity of tweets allows us to explore the ideological locations of candidates in greater detail. After arranging the major candidates along a general ideological spectrum, we can examine whether they positioned themselves differently in different policy realms and whether Donald Trump and Hillary Clinton shifted over the course of the campaign, following Downsian (1957) logic to drift toward the median voter in the general election.

We display the ideological distribution of candidates based on their Twitter communication in Figure 2. On the Republican side, our ideological score based on tweets has Rand Paul, Ted Cruz and Ben Carson as the furthest right candidates, just as the network-based measures in Barbera (2015a) and CROWDPAC (2016) do. Where we differ from those approaches is in placing Donald Trump toward in the middle of the Republican field, with Jeb Bush and Marco Rubio taking positions that were more moderate than Trump’s on social media. Both of the network approaches placed Trump as the most centrist candidate in either party, but his rhetoric on Twitter positioned him firmly in the midst of the Republican field.

---

9 To compute this correlation and to designate the “major candidates” reported in Figure 2, we focus on the candidates who won delegates to either party’s national convention.
Because we measure both the ideology and the subject matter of each tweet, we are also able to subset a candidate's ideology based on a given topic area. This allows us to examine the nature of the “populism” of Donald Trump. In Figure 3, we examine Trump's ideological positions on the topics of immigration, health care, macroeconomic policy and national security. We expect to see Trump take a harder, more conservative line on immigration and health-care, while adopting a more moderate posture on both trade (where he took protectionist stances that had him often agreeing with labor unions) and defense (where his anti-interventionist views saw him siding with doves more frequently than with hawks). Indeed, subsetting the tweets shows that Trump tweeted closer to the ideological center on national security and macroeconomic policy, (both Trump and Democratic candidate Bernie Sanders opposed the Trans-Pacific Partnership, for instance), while his tweets on immigration, where Trump's border wall proposal perhaps represented the most hard-line view of the campaign, and on healthcare were decidedly to the right of his overall ideological mean. This method of studying political communication can pinpoint the ways in which Trump was able to capture the support of Democratic-leaning voters who opposed free trade and foreign intervention, while holding the Republican base by attacking Obamacare and illegal immigration.

Turning to the Democratic side, our Twitter data depict Hillary Clinton and Bernie Sanders as near-identical ideological twins. Again, this direct measure of their online communication differs from network approaches, which show that Sander's followers and donors were more left-leaning than Clintons. Yet the ideological proximity of their messaging is in keeping with Hillary Clinton’s claim, in her recent books, that she and Sanders both took reliably progressive stands. As Clinton wrote in her post-mortem to the election, “Because we agreed on so much, Bernie couldn’t make an

---

argument against me in this area on policy, so he had to resort to innuendo and impugning my character.” (Clinton 2017). Our ideological scores suggest that she is correct that they stood on common ideological ground – at least in the world of Twitter communications.

Finally, our Twitter data allows us to test the hypothesis that candidates speak to extremists and win over party faithful during a primary, but quickly transition back to the ideological center during the general election (e.g. Polsby and Wildavsky 1976; Schlesinger 1994; Polsby, Wildavsky, and Hopkins 2008). Burden (2004) explores this in Congressional elections, and Acree (2014) uses a supervised learning approach to study ideological drift in the speeches of 2008 and 2012 presidential candidates. Measuring the ideology of tweets lets us track not just primary to general election changes, but to track week-by-week how a candidate's ideological position evolves over the course of a campaign. Did the candidates in the 2016 presidential contest follow their counterparts in past elections or did was this a unique election, with establishment candidates chasing the insurgent outlier Trump throughout the primary and general?

Tracking Hillary Clinton and Donald Trump's average ideological score over the course of 2016, we see two distinct trends emerging in each party's primary process in Figure 5. On the GOP side, Donald Trump began the primary by taking conservative stances, though not as conservative as the Republican field overall. Then he veered sharply toward the center beginning in mid-May of 2016. Although this was two weeks before the last primary was held, it coincided with his becoming the presumptive Republican nominee. According to the Iowa Electronic Markets (2016), the share price for Trump winning the nomination rose from around 40 cents in the middle of April to over 90 cents by the middle of May. At this time, acting surprisingly like a typical Downsian politician, Donald Trump pivoted toward the center when he could afford to look ahead to the general
election. This is clear evidence of ideological drift, and demonstrates that the week-by-week detail provided by Twitter data enables us to spot a trend that might have been hidden in a simple comparison of the primary season to the general election period. In another surprise, Trump began to tweet more extremist messages as Election Day neared. After Labor Day, Trump began a rightward drift that sharpened in the final month of the campaign. Trump’s ideological scores in the last weeks of the campaign were as conservative as they were early in the primary, indicating an attempt to rally the Republican base to mobilize.

On the Democratic side, we observe less fluidity. Hillary Clinton began the election toward the liberal end of the spectrum, and then moved toward the center early in the primary season. From that point until late in the election, Clinton’s ideological position remained stable, showing no clear pivot toward the center once the general election began. Interestingly, in the run-up to Election Day, Clinton moved a bit toward the left, perhaps mirroring Donald’s Trumps strategy to shore up turnout among her base. It is possible that following publication of a letter from former FBI Director James Comey concerning the reopening of a bureau investigation of Clinton’s use of a personal email server and her subsequent polling drop in the final days of the 2016 campaign, Clinton saw a need to boost her support with the most reliable members of her coalition in an election where every vote would count.

Section IV. Conclusion

As this paper illustrates, generating a dataset that extends human-coding to a large universe of Twitter data is feasible for researchers even with modest budgets and few resources. The type of extensive coding that previously took hundreds of hours and armies of research assistants can now be completed in relatively short order by harnessing advances in computing power and
machine-learning. Research questions previously shelved because generating the data proved to arduous a feat can now be addressed. The work of a handful of human coders can be extended to replicate tasks as if there were dozens. As our experience illustrates, maximal returns were gained after hand-coding just 2,500 tweets, an achievable goal for most researchers over the course of a few months.

This paper represents a first look at a project very much in progress. The intriguing results from our measure of ideology will be expanded in further works that will examine the interplay between a candidate’s ideological position and their polling results in a primary election. Further, we will examine when a candidate decides to communicate facts versus when a candidate communicates opinions and what this does to the reach of a candidate’s message. Our tweet collection process is still very much in progress and this data will be made available to other researchers seeking to tackle similar questions. To this end, our data can also be extended towards other political speech from American politicians, allowing its use as training data for the tweets of members of Congress, state politicians, mayors, or any other leaders who have communicate through social media.

As modern political campaign becomes unthinkable without a social media component, understanding how these new technologies influence and shape political discourse becomes vital to understanding modern electoral politics. More than this, however, these new technologies have given us insightful way to analyze and test decades-old theories about how aspiring politicians seek to win voters to their side. More than other social media platforms, the immediacy of Twitter data allows researchers to chart the course of a candidate’s evolving message throughout the ebbs and flows of a campaign cycle—charting the shifts in a candidate’s communication strategy in almost

---

11 We performed a preliminary classification analysis when we had coded only 2500 tweets, and reached similar levels of accuracy to what we report based on 7,525 tweets in Table 3.
real-time. By leveraging new methods and computational techniques, we can examine this new communications medium to see what’s changed about political campaigns in the new millennium and further understand the fundamentals of seeking public office that have remained steady throughout the years.
References


King, Aaron S., Frank J. Orlando, and David B. Sparks. 2015. “Ideological Extremity and Success in Primary Elections: Drawing Inferences From the Twitter Network.” Social Science Computer Review.


Peterson, Andrew, and Arthur Spirling. n.d. “Classification Accuracy as a Substantive Quantity of Interest: Measuring Polarization in Westminster Systems.”


Table 1. Examples of Variables and Tweets in Each Category

<table>
<thead>
<tr>
<th>Ideology</th>
<th>Liberal</th>
<th>Neutral</th>
<th>Conservative</th>
</tr>
</thead>
</table>
| New docs show Bush/Cheney ignored 4th Amend. &amp; Justice Dept on wiretapping. Probably still think they're right. | Wonderful meeting with many prominent leaders of the Asian American and Pacific Islander community in Illinois. | RT @FoxNews: @RealBenCarson on ISIS: "We need to recognize that it is really an existential threat to our nation...the war is against all."
| “Four boys shot my son dead on Christmas Eve.” A moment backstage with a mom who lost her son to gun violence. | ICYMI: Today I appeared on The Real Story with @GretchenCarlson | My conservative alternative to Obamacare focuses on restoring power to patients doctors: |

<table>
<thead>
<tr>
<th>Sentiment</th>
<th>Negative</th>
<th>Neutral</th>
<th>Positive</th>
</tr>
</thead>
<tbody>
<tr>
<td>The sheer size and remoteness of the federal bureaucracy has caused the American people to lose trust in government.</td>
<td>This account will be run by campaign staff from now on but you’ll still see tweets from Hillary. They’ll be signed -H.</td>
<td>Looking forward to joining @jaketapper and @TheLeadCNN today at 4:30pm ET. Tune to hear about our plan to help hardworking families!</td>
<td></td>
</tr>
<tr>
<td>There are some things in Washington we need to burn down. Read more here</td>
<td>For those with a question as to &quot;secret service&quot; protection, neither my team nor I are in the habit of commenting on security.</td>
<td>RT @KevinBingle: BREAKING: The @BostonGlobe @GlobeOpinion endorses Gov. @JohnKasich for President!</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Is the Tweet a Factual Claim or an Opinion?</th>
<th>Factual Claim</th>
<th>Opinion</th>
</tr>
</thead>
<tbody>
<tr>
<td>A top ten state for job creation. Wages growing faster than the nation. Record number of businesses.</td>
<td>&quot;@ToTheTnr @realDonaldTrump You are the closest thing to Teddy Roosevelt. Tough as hell, and speaks his damn mind. And NEVER BS's&quot;</td>
<td></td>
</tr>
<tr>
<td>In fact, single mothers earn $19,000 dollars less per year than married mothers. #empoweredwomen #ceidinner</td>
<td>.@HillaryClinton thinks the law doesn’t apply to her.</td>
<td></td>
</tr>
<tr>
<td>Hundreds turned out this morning in Fairfax to hear Gov. Kasich's message of strength and optimism.</td>
<td>&quot;I am the only one who stood up to 100K protesters, stood up to the big union bosses. I'll stand up to Washington.&quot;</td>
<td></td>
</tr>
</tbody>
</table>
Table 2. Measures of Intercoder Reliability: Humans Agreeing with Humans

<table>
<thead>
<tr>
<th>Measure</th>
<th>Agreement Rate</th>
<th>Cohen's Kappa</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ideology (Liberal, Neutral, or Conservative)</td>
<td>0.78</td>
<td>0.66</td>
</tr>
<tr>
<td>Sentiment (Negative, Neutral, or Positive)</td>
<td>0.75</td>
<td>0.60</td>
</tr>
<tr>
<td>Is the Tweet Political or Personal?</td>
<td>0.91</td>
<td>0.55</td>
</tr>
<tr>
<td>Is the Tweet a Factual Claim or an Opinion?</td>
<td>0.78</td>
<td>0.42</td>
</tr>
<tr>
<td>Topic: Immigration</td>
<td>1.00</td>
<td>0.87</td>
</tr>
<tr>
<td>Topic: Macroeconomics</td>
<td>0.97</td>
<td>0.73</td>
</tr>
<tr>
<td>Topic: Defense</td>
<td>0.96</td>
<td>0.77</td>
</tr>
<tr>
<td>Topic: Law and Crime</td>
<td>0.99</td>
<td>0.86</td>
</tr>
<tr>
<td>Topic: Civil Rights</td>
<td>0.98</td>
<td>0.71</td>
</tr>
<tr>
<td>Topic: Environment</td>
<td>1.00</td>
<td>0.84</td>
</tr>
<tr>
<td>Topic: Education</td>
<td>1.00</td>
<td>0.83</td>
</tr>
<tr>
<td>Topic: Health</td>
<td>0.99</td>
<td>0.82</td>
</tr>
<tr>
<td>Topic: Government Operations</td>
<td>0.98</td>
<td>0.23</td>
</tr>
<tr>
<td>Topic: No Policy Content</td>
<td>0.91</td>
<td>0.78</td>
</tr>
<tr>
<td>Asks for a Donation?</td>
<td>0.99</td>
<td>0.66</td>
</tr>
<tr>
<td>Asks to Watch, Share, Or Follow</td>
<td>0.95</td>
<td>0.65</td>
</tr>
<tr>
<td>Asks for Miscellaneous Action?</td>
<td>0.93</td>
<td>0.57</td>
</tr>
</tbody>
</table>

Note: Based on an analysis of 1,217 tweets coded by rotating pairs of research assistants.
Table 3. Measures of Classification Accuracy: Computers Replicating Humans

<table>
<thead>
<tr>
<th>Measure</th>
<th>Final Testing Accuracy</th>
<th>Cohen's Kappa</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ideology (Liberal, Neutral, or Conservative)</td>
<td>0.66</td>
<td>0.46</td>
</tr>
<tr>
<td>Sentiment (Negative, Neutral, or Positive)</td>
<td>0.65</td>
<td>0.38</td>
</tr>
<tr>
<td>Is the Tweet Political or Personal?</td>
<td>0.87</td>
<td>0.11</td>
</tr>
<tr>
<td>Is the Tweet a Factual Claim or an Opinion?</td>
<td>0.73</td>
<td>0.37</td>
</tr>
<tr>
<td>Topic: Immigration</td>
<td>0.98</td>
<td>0.43</td>
</tr>
<tr>
<td>Topic: Macroeconomics</td>
<td>0.94</td>
<td>0.51</td>
</tr>
<tr>
<td>Topic: Defense</td>
<td>0.92</td>
<td>0.43</td>
</tr>
<tr>
<td>Topic: Law and Crime</td>
<td>0.97</td>
<td>0.00</td>
</tr>
<tr>
<td>Topic: Civil Rights</td>
<td>0.96</td>
<td>0.07</td>
</tr>
<tr>
<td>Topic: Environment</td>
<td>0.99</td>
<td>0.25</td>
</tr>
<tr>
<td>Topic: Education</td>
<td>0.99</td>
<td>0.25</td>
</tr>
<tr>
<td>Topic: Health</td>
<td>0.99</td>
<td>0.57</td>
</tr>
<tr>
<td>Topic: Government Operations</td>
<td>0.97</td>
<td>0.00</td>
</tr>
<tr>
<td>Topic: No Policy Content</td>
<td>0.71</td>
<td>0.40</td>
</tr>
<tr>
<td>Asks for a Donation?</td>
<td>0.99</td>
<td>0.37</td>
</tr>
<tr>
<td>Asks to Watch, Share, Or Follow</td>
<td>0.93</td>
<td>0.62</td>
</tr>
<tr>
<td>Asks for Miscellaneous Action?</td>
<td>0.95</td>
<td>0.41</td>
</tr>
</tbody>
</table>

Note: Based on an analysis of a final testing set of 725 tweets after training on 6800 tweets.
Figure 1. Comparison Cloud of Donald Trump and Hillary Clinton Tweets
Figure 2. Ideological Positions of 2016 Presidential Candidates Based on Twitter Communication

Note: Ideology scores represent mean score for each candidate, with every tweet categorized as either liberal (-1), neutral (0), or conservative (1).
Figure 3. Ideological Positions of Donald Trump, by Issue Area

Note: Ideology scores represent mean score of Donald Trump tweets in each issue area, with every tweet categorized as either liberal (-1), neutral (0), or conservative (1).
Figure 4. Ideological Positions of Donald Trump, Hillary Clinton, and all Candidates in their Parties, over time.

Note: Lines represent LOWESS curves fitting the weekly mean ideology scores of each candidate or group of candidates, with every tweet categorized as either liberal (-1), neutral (0), or conservative (1).