Psephological Investigations: Tweets, Votes, and Unknown Unknowns in the Republican Nomination Process

Michael J. Jensen
Institute for Government and Public Policy
Autonomous University of Barcelona
109 MRA
08193 Bellaterra, Spain
MichaelJames.Jensen@uab.cat

Nick Anstead
Department of Media and Communications
London School of Economics
Houghton St.
London, UK WC2A WAE
N.M.Anstead@lse.ac.uk

Psephological Investigations: Tweets, Votes, and Unknown Unknowns in the Republican Nomination Process

> *If Twitter were a good predictor of public attitudes, Ron Paul would be the Republican nominee.*[1]
>
> *No, you cannot predict elections with Twitter.* [2]
>
> *Once you have 1 million tweets, what do you do with it?*[3]

Electoral forecasting, particularly in the context of the American presidential primaries, has proven difficult at times. Due to characteristically lower turnout, uneveness in campaign get out the vote efforts, and variance in the ability of campaigns to identify and differentiate their voters among members of the same party, election nights are often filled with surprises. Polling and predictive markets are useful but each entail unique informational biases. Predicting the results of caucuses can be particularly problematic given that the results can be highly dependent on the turnout which requires a higher order of engagement than primaries as caucuses take more time and involve public discussion before votes are cast. Furthermore, trends towards early voting create a highly variable temporal environment under which vote decisions are being made. Consequently, campaigns have a greater need to adjust messaging on a continuous basis over a larger period of time than the immediate lead up to election day. For this reason, much work has gone into reevaluating how polls are conducted as well as supplementing polling with other data. The ease of accessing and processing digital data makes it a particularly intriguing area to explore in an effort to improve voting forecasts. Indeed, some studies have been successful in using Twitter communications and other digital artifacts to predict electoral and market behaviors and political campaigns have invested considerable resources in mining online behavior to gain insights about the electorate (McCoy 2012; Romano 2012). Analyzing changes in real time may be able to alert us to the "unknown unknowns" (Kaushik 2012) within the flow of a campaign: the shifts in preferences and momentum that otherwise we would be unaware of due to the relatively slow pace of polling.

Traditional polling, economic models, and prediction markets are not without limitations First, the history of political polling is filled with divergent vote estimates – a trend that shows little sign of abating. There are differences in sample weights and estimations of likely versus registered voters. Additionally, as response rates dwindle and households are increasingly cellphone only, certain populations become harder to reach, introducing distortions into the samples (Hillygus 2011,

---

1 Tom Rosenstiel of Pew Research qouted in "'Twindex' wants to gauge election 2012's intensity: Will it be successful?" *The Guardian. August 2, 2012. (http://www.guardian.co.uk/world/us-news-blog/2012/aug/02/twitters-twindex-guage-election-2012)*

2 Daniel Gayo-Avello. "I wanted to predict elections with Twitter and all I got was this lousy paper: A balanced survey of election prediction using Twitter." May 1, 2012. (http://arxiv.org/pdf/1204.6441.pdf).

3 Republican pollster Jon McHenry quoted in Stephen Sheppard. "Sorry, wrong number." *National Journal*. July 19, 2012. (http://www.nationaljournal.com/magazine/who-responds-to-telephone-polls-anymore--20120719)

963). These factors make for an increasingly challenging environment in which to conduct political polling. This is particularly true in the US where there are increasing cell-phone only households among key demographic groups and federal laws along with cell phone pricing plans make these individuals more difficult expensive to contact. The deterministic nature of economic models are of little use to campaigns as they provide little actionable information in the face of a largely sealed fate.[4] Predictive markets have shown significant inroads as people are less likely to put money on the line without making an informed guess; however, the dynamics of betting and limited participation in the case of these markets can also produce unreliable results due to information biases (Rothschild 2009). Therefore campaigns and analysts are in need of an economical and efficient means to improve electoral forecasting and make adjustments along the way. Mining social media data may help satisfy those needs as it can provide continuously updated data in real time.

More generally, the field of predictive analytics has taken off recently seizing on the potential of Big Data to create new avenues for commercial and political marketing. And these efforts have not been without success, with inroads in predicting market behavior and election outcomes (Bollen, Mao, and Zeng 2011; Tumasjan et al. 2011). Despite methodological advances, however, this work to date has been largely atheoretical, bereft of connections between online communications and widespread processes of opinion formation and preference articulation. Campaign consultants have little interest in theory building. In the words of one strategist, “Correlations will do, as long as they hold until the election is over” (Nielsen 2012, 141). Hence, these models provide little insight into the conditions under which they hold and the criteria by which to determine if the results are spurious. For this reason, much work is necessary in theoretical development of this field to catch up with the methodological advances that have been made over the last decade. This paper looks into the utility of using information contained within Twitter posts in predicting electoral outcomes. A mention model is developed using the Iowa caucuses and then applied to the ten Super Tuesday nominating contests. Particularly, we are interested in patterns in Twitter communications that can help explain differences between polling and the actual vote. The results find some support for predictive value of Twitter both as an estimate of the overall vote and in terms of otherwise undetected movement in favor of a candidate.

Tweets, Polls, and the Nomination Process

Opinion polls have historically proven unreliable in predicting the outcome of the Iowa caucuses. Several factors distinguish the Iowa caucuses from other the other nomination contests. First, given the state's relatively small population size and its relatively low cost media market, campaign resources are less of a barrier. Due to its large rural and small town population, retail politics all the more feasible as a means of reaching a large percentage of voters. Second, Iowa is the first contest in the nomination process so none of the candidates have any victories which can provide Iowa voters cues. Those who rise from obscurity in Iowa often receive a great deal of media attention and donations allowing a chance to compete later on. Although debate persists over whether the current nominating process or calendar need reform, there is undeniably a significant impact Iowa has on the nomination (Altschuler 2008; Iyengar and Luskin 2004; Stark 1996). Despite frequent polling swings among the many candidates vying for the Republican nomination, Mitt Romeny's support remained steady. Nevertheless, he was narrowly edged out in Iowa by a surging Rick Santorum who outperformed the polls. Furthermore, Santorum's rise in support held to a significant degree, catapulting him to a status as a challenger to Romney's presumed inevitability, a status for Santorum

4 Although many economic models consider factors beyond economic variables, these additional measures such as incumbent popularity, the level of polarization, etc. are factors which campaigns and candidates have very little ability to control. At best, these models reinforce to a campaign the importance of economic stewardship as an issue.

that extended into the Super Tuesday contests in March.

Twitter has been around for a relatively short period of time and its use by members of the public still quite limited. Nevertheless, it can serve an important function in campaign discourse due to its relatively unstructured environment enabling interactions organized around both predefined networks of individuals as well as thematically, bringing together persons who are otherwise relatively unknown. Empirically, this has produced highly decentralized networks of interactions and information cascades (González-Bailón et al. 2011). In relation to electoral forecasting, there are two sense in which we can regard these data. First, we can take these data as a representation of the overall distribution of preference articulation. Second, we can take these data as a particular form of mass observation opening a window into widespread processes of opinion formation indicating how members of the public relate to campaign personalities and themes. Given that our politics increasingly occurs within online spaces, the political information dynamics that emerge here may prove critical to understanding and identifying transformations before they show up elsewhere. This task becomes all the more important as attendant temporal changes in "political information cycles" are perhaps beginning to move faster than opinion polling techniques can provide reliable information to campaigns and publics (Chadwick 2010). For these reasons, mining online communications may not turn out like another *Literary Digest* poll as some critics argue (Gayo-Avello 2011).

Nevertheless, taking tweets as a representation of wider processes in the American public is not without significant theoretical and empirical difficulties. First, those who engage in political communications online tend to be more politically active and hardly representative of the general population even in an election year (Conover et al. 2011a; Conover et al. 2011b; Jensen and Anduiza 2012). Twitter users in particular constitute only 15 percent of the American population and only 8 percent of those use it regularly, of which, only a subset use it for political purposes (Smith and Brenner 2012). Though similar observations may be made about Iowa caucus-goers (Redlawsk, Tolbert, and Donovan 2010), the same can not be generally said about voters in other states or during the general election when turnout is much higher. Furthermore, critics such as Habermas (2006) argue that the highly fragmented digital environments that characterize online communications are disconnected from a unitary public, divorcing them from wider processes of will formation. When combined with the constant strafing on all sides by representatives of the campaigns attempting to shape the discussion, platforms such as Twitter can appear more like an argument war zone rather than an agora (Haberman and Burns 2012; Weigel 2012). Consequently online communications are said to be highly balkanized, distorted, and disconnected, resulting in an errant representation of public opinion.

Although tweeting may be produced under these conditions, at the same time, such an online space may prove to be highly innovative within the context of campaign communication and commentary. Social media platforms like Twitter position users symmetrically with equal agencies as senders and receivers of communication (Castells 2009, 55–57). Hence, the architectures of these platforms provide users with the formal conditions of reciprocal communication capacities. Tweeting as well may become mediatized in the same way polls have come to be as journalists craft accounts of the campaign viewed from Twitter, and the incidence of astroturf can distort the communication flows (Ratkiewicz et al. 2010). Nevertheless, such practices may also be policed and discussed within the Twitter platform. Furthermore, the vulgarization of political discourse is not a necessary consequence of the fluid nature of these communications: the unscripted and nature of tweeting, for example, may lend itself to a modicum of authenticity, inaugurating publics that otherwise would not exist given contemporary levels of political disaffection (Goodnight and Hingstman 1997, 366; Wheeler 2012). Though empirical interactions may not inaugurate a unified public sphere and may be quite polarized, there is evidence that these interactions nevertheless transcend partisan divides rather than reinforce balkanization (Conover, et al. 2011b). In this sense, Twitter communications may function as a testing grounds for campaign arguments and candidates, more broadly reflecting how persons relate to the campaigns at a particular point in time.

Beyond serving as a space of self-referential campaign communications, Twitter communications can provide insights into the overall state of the campaign. Despite the brevity of Twitter messages, limited to 140 characters, evidence points to Twitter communications in particular as a space of preference production and reflexive discourse about campaigns as well as critical reflection on the reporting of political polling (Ampofo, Anstead, and O'Loughlin 2011). Because Twitter-users constitute a relatively small and unrepresentative segment of society one cannot generalize based on the distribution of sentiments articulated in this medium. However, Twitter users may react to the prevailing informational currents in proportion to their overall distribution within the relevant political community. Furthermore, the detection of sudden shifts in the distribution of tweets about candidates and thematics may alert observers to changes in real time that would otherwise remain unknown unknowns. Consequently, both tweets about a candidate as well as retweets may testify to not only to the presumed informational importance of the tweet about a campaign, but also the attendant relative position of the campaign in electoral terms. Finally, in addition to the distribution of campaign commentary, the collective insight of tweets forecasting a winner might reliably function like a predictive market: the candidate most people think will win, based on the variety of information sources they have to draw from, is more likely to win. Hence we offer five hypotheses:

H_1: Candidate mentions will correspond with the overall vote.
H_2: Shifts in candidate mentions will correspond to over or under electoral performance in relation to the polls.
H_3: Relatively high associations between candidates and words relating to gaining momentum will outperform polling predictions.
H_4: Candidates with more often retweeted will have increasing momentum.

Data and Methods

We use two sources of data in this paper. The first set of data are the aggregated polling figures and electoral results posted by Real Clear Politics and the US Election Atlas. This data is supplemented with Gallup national tracking poll data in the lead up to the Super Tuesday primaries and caucuses. Polling data was available for Iowa and all Super Tuesday states apart from Alaska, North Dakota, Idaho, and Massachusetts. Although these polls contain a number of undecided voters, we only consider the confirmed polling support for particular candidates as we are interested in which campaigns are able to capture the undecided vote and perform better than their polling position would otherwise indicate. Conversely, we are also able to see which candidates lose support in the closing days of a campaign and the metrics in Twitter that may alert us to that event.

The second were collected from Twitter's streaming application programming interface over the course of the four days preceding the Iowa Caucuses, held January 3, 2012, and five days prior to the ten selection contests held on "Super Tuesday," March 6, 2012. The Super Tuesday contests included the caucuses in Idaho, North Dakota, and Alaska along with the primaries in Georgia, Massachusetts, Ohio, Oklahoma, Tennessee, Vermont, and Virginia.[5] The Twitter stream was filtered for key terms resulting in a comprehensive sample of all tweets containing up to 1% of the total Twitter stream. That limit was reached at only one point in the lead up to Super Tuesday. Filtering

---

5 Wyoming also has a series of county-level conventions occurring on this day; however, we exclude this state because the conventions are carried out over four days and attended by delegates selected at the precinct level in February. For this reason, Wyoming is not comparable to the other cases either in terms of the composition of the voters or the time frame in which the voting is carried out.

terms included candidate names as well as references to the Iowa caucuses or Super Tuesday, as appropriate. Sources for these tweets are not geographically constrained. This resulted in a dataset of 697,065 and 485,355 tweets in the Iowa and Super Tuesday datasets, respectively. The Iowa dataset was composed of 195,737 separate accounts and the Super Tuesday account composed of 137,945 accounts. The descriptive statistics of each group are presented in Table 1.

[Table 1 here]

Because there are a relatively small number of media and other organizations with very large numbers of Twitter followers and posts, we report the sample medians rather than the means. Iowa drew a smaller but more active and visible population on Twitter as the median number of 10,420 followers (i.e. accounts this user follows), 3,911 friends (i.e. accounts that follow the particular account), and have written a median of 32,960 status posts. The Super Tuesday accounts contained 425,962 individual tweet messages sent from 219,992 unique accounts. These accounts had on average 3,353.80 followers, 955.98 friends, and 13,482.76 status posts.

Candidate selection decisions are neither trivial or without consequences. Candidates outside the political mainstream tend to emphasize their presence online when they are unsuccessful in competing for attention from broadcast media channels (Jensen, Jorba, and Anduiza 2012). Previous research has found evidence of this in the case of Twitter as campaigns such as the Pirate Party in Germany, a smaller party with a manifesto predominantly about regulation of the internet, have garnered a far larger share of attention in this medium than they did either in offline media or the voting results (Jungherr, Jurgens, and Schoen 2011). For this reason, we have selected the candidates that are officially on the ballots in the majority of these states in order to conduct our analysis. For the Iowa contests, these include Michele Bachmann, Newt Gingrich, Jon Huntsman, Ron Paul, Rick Perry, Mitt Romney, and Rick Santorum. By Super Tuesday, this field had been winnowed to Gingrich, Paul, Romney and Santorum, with Santorum and Gingrich failing to qualify for the ballot in Virginia.

The tweet texts were extracted from the metadata and processed to remove extraneous coding. The texts were subsequently normalized in lower case form and tokenized to separate words from punctuation. The tweets were then automatically coded using a natural language processing program scripted in the Python language (Bird, Klein, and Loper 2009). The Iowa and Super Tuesday collection of tweets are coded separately each as an independent corpus as well as segmented on a daily basis. Tweet contents are coded three ways ways. First, they are coded in terms of candidate mentions, a dichotomous variable (0/1) indicating the presence or absence of a candidate mention in each tweet. This measure gauges the value tweets containing information about particular candidates without the distortion introduced by multiple mentions of a candidate that may obscure meanings or overemphasize a candidate. Second, we consider which candidates are more likely to be retweeted, i.e. forwarded onward using a χ^2 statistic as both terms are dichotomous variables. The retweet metric analyzes the extent to which tweets containing information about a particular candidate hold value for recipients that is worth forwarding onward. A third set of measures operationalize predictive terms. This includes the following terms: surge, surging, momentum, gaining, and closing. These correspond to the short-term expectation that a candidate will win a particular contest.

We use two additional measures of changing momentum. The first metric is the change in mentions summed over the observation period:

$$\sum M_t - M_{t-1}$$

The total number of mentions, M, at time t are subtracted from the total number of mentions the previous day (t-1). A second measure divides the summed mentions by the total number of mentions for that candidate. This figure corrects for both front runner status which earns candidates extra attention as well as controls for candidates who have a comparably strong online following. However it also inflates the impact of contenders with the least amount of support as the increasing enthusiasm in the days leading up to the Iowa caucuses constitute a large portion of the overall candidate tweets.

Results

The first sense in which we consider the data from Twitter is as a reflection of the overall distribution of electoral preferences and predictions. To begin, we examine the distribution of tweets and votes in relation to the polling results. The distribution of mentions over the last four days of campaigning leading up to the Iowa caucuses are presented in Table 2 along with the results and average of six polls that were conducted in the two weeks prior. The twitter mentions indicate the percentage of all tweets in the sample in which the candidate was mentioned at least once. The mean absolute error is calculated as the average absolute value of errors for both the Twitter and polling data. Over those two weeks the polling results were relatively consistent indicating a close race between Paul and Romney. The results show the polls were relatively accurate apart from shifts in Santorum's support which ultimately made him the winner.

[Table 2 here]

These results show that, although the Twitter mentions model underestimated Santorum's support as well, it was much closer to the actual result. The mean absolute error between the polling predictions and the vote are slightly smaller than the overall error of the mentions model. The proportion of mentions for each candidate is plotted against the Iowa vote in Figure 1. The linear model has a strong fit with a R^2 of 0.92 on 5 degrees of freedom and a slope of 1.01 indicating an almost perfect linear fit.

[Figure 1 here]

Turning to measures of campaign momentum we look at the increase mentions, claims about changes in candidate position, and retweets of messages containing candidate mentions. We start with the measures of changing mentions. There is an average total increase in mentions of 7545.29 over these last four days. With respect to the total mentions, Table 3 both Paul and Santorum out perform the other candidates by wide margins. The second change in mentions measure, which corrects for the fact Ron Paul had a very active following on Twitter, places the change in momentum towards Santorum in greater relief and Paul's momentum is closer to the mean of this statistic (0.07). However, this metric also inflates the position of Bachmann, Huntsman, and Perry who underperformed their polling predictions Likewise, the χ^2 statistics show tweets mentioning Santorum have the highest likelihood of being retweeted. On this score, both Santorum and Paul beat the average by considerable margins. All three metrics, point to momentum for Santorum in particular. However, only the retweet metric and the measure of momentum are consistent with both Santorum's surge in the closing days before the Iowa caucuses as well as Romney and Paul holding position in relation to the polls. These data suggest that the proportion of Twitter mentions is a reasonable approximation of the overall vote. However, traditional polling provides a better approximation of the vote than Twitter mentions. On that score, the candidate retweet χ^2 statistics provide a superior accounting of shifts in candidate momentum.

[Table 3 here]

Now that we have metrics for predicting the overall vote and changes in the momentum based on the Iowa caucuses, we apply those to the contests on Super Tuesday to determine if the same measures that proved useful in Iowa are also useful when applied to a wider range of electoral contests. Due to the fact that by Super Tuesday the Republican field had narrowed to four candidates, the analysis here is focused on those remaining in the race: Mitt Romney, Rick Santorum, Newt Gingrich, and Ron Paul.[6] Because there is not a representative sample of candidate support within the Super Tuesday states, we use a national tracking poll as a proxy for the total vote in these ten states. Table 4 contains data on the aggregated national vote totals along with along with data on the percentage of tweets mentioning each of the candidates, the Gallup polling data on the preferences of registered Republicans and Republican leaning independents, and differences between both the polling and tweet mention predictions and the actual results aggregated across all ten states. Republican leaners were included along with Republican identifiers as the majority of the Super Tuesday states have open or semi-open primaries that allow independents to vote. The two candidates in the middle, Gingrich and Santorum, over performed based on the polling predictions to a greater extent on average than Paul or Romney.

[Table 4 here]

[Table 5 here]

The national polling margin of absolute error is smaller than that predicted by the percentage of candidate mentions in the Super Tuesday tweet corpus . However, when we calculate the state-by-state margin of error for the seven states in which we have polling data (Table 5), the absolute margin of error at this level is 5.0, just as accurate at the overall predictions based on Twitter mentions. This margin of error treats all contests equal independent of the number of voters in each state as well as the number of persons polled. Nevertheless it is a rough measure indicating the overall performance of polling models. These data show there is a great deal of variance in each of the statewide contests. Of the states where we have polling data, only in Oklahoma did a candidate who was not predicted to win ended up victorious. In this case, Romney came from behind to beat Santorum. Gingrich's relative success relative to his standing in the national polls is not driven by his performance in his home state of Georgia as his vote only improved 3% there.

The candidates' share of the total vote in these ten states is plotted in Figure 2. The linear model has a slope of 0.71 (t=4.426) with an $R^2 = 0.35$. The model fit is not as tight as the Iowa model, however, the national sample of tweets may be too crude a measure to detect and aggregate the particularities of the state-by-state contests in the nomination process. A similar model calculated using the overall national level of the vote and the four candidates produced (not shown), had an R^2 of 0.84 and a shallower slope (0.66).

[Figure 2 here]

In Table 6 we consider the change metrics to determine if they may be useful in accounting for shifts in candidate momentum. The change in mentions fails to hold as, despite the fact that Santorum outperformed the national polls and the polls in 3 out of 5 states for which we have data where he was on the ballot, his mentions overall decline in the closing days before Super Tuesday. The retweeting χ^2 statistics are broadly consistent with the differences between the polls and the overall vote; however, Santorum was more likely than Gingrich to be retweeted and Gingrich outperformed the polls by a wider margin than Santorum. Likewise, Romney was not that far

6 The filtering terms used in the data collection remained the same so there is no change in the data collection between the Iowa caucuses and Super Tuesday.

behind on this metric. The surge measure however is consistent with both Gingrich's performance in relation to the polls as well as the relative lack of movement on the part of Romney and Paul as they both have a small and slightly negative correlation with terms relating to candidate surge.

[Table 6 here]

Discussion and Conclusions

In general these results find a relatively strong correlation between a candidate's presence on Twitter and a candidate's electoral performance, one that comes close to the accuracy of polling. However, the more interesting conclusions stem from points in which the polls and tweets diverge. Although polling needs to continue to innovate to deal with the increasing challenges of reaching many population segments (Goidel 2011; Hillygus 2011), an analysis of candidate mentions on Twitter alone does not improve on the accuracy of traditional polling. The Super Tuesday nomination contests provided a hard case against which to test insights about candidate mentions and changes in Twitter communication flows due to the uniqueness of each state level contest. Such complexities do not normally materialize in a general election where the result more closely tends toward a uniform swing when incumbents are replaced, and stability when they retain office. Nevertheless, the mentions model proved to be generally effective in predicting the vote in the Iowa causes and on Super Tuesday. Furthermore, these data reveal shifts in public discussion that show up in the final voting but not the polling data. We will summarize these findings in relation to the hypotheses posed earlier.

First, in general there is a strong correspondence between the percentage of tweets mentioning a candidate and the percentage of the vote. In Iowa, the tweet model had a 3.1% mean deviation from the electoral result compared with a 2.2% mean average error for the polling data and this model returned similar results based on the aggregated votes across the Super Tuesday states. Though tweet mentions are not quite as reliable as polling, their margins are not far outside that the results predicted by the polls, either. In contrast to the widespread belief that Ron Paul supporters were far more active and vocal online (Sifry 2011; Tau 2011), our results suggest otherwise, and that conversations about the candidates are roughly proportional to their overall level of support within the relevant community of voters. These findings are consistent with the first hypothesis. The crucial question is why might that obtain? If Twitter communications are considered to function like a critically engaged interaction system, then we may have an answer (Ampofo, Anstead, and O'Loughlin 2011). Those participating in this system are not necessarily representative of the preferences across the relevant voting publics, but their communications about specific electoral contests appear to be thematically representative as they are responding to larger currents of conversations and preference distributions.

If we are to take Twitter communications as thematically representative and a participant in the processes of vote choice formation, then we should see evidence of movement corresponding with shifts in support for candidates. Hypotheses two through four deal with shifting support. The second hypothesis asserts that if a candidate outperforms the polls, we should see a trend towards increasing mentions during the final days of the campaign. This model works well to predict the outcome in Iowa where the increase in Santorum's mentions outpaced the rest of the competition by large margins and he finished ahead of Romney and Paul, the projected front runners. That measure is somewhat inconsistent, however, when applied to Super Tuesday. It accurately detects movement in Gingrich's direction, but it also would anticipate large swings in Romney's direction as well, and the only state Romney won that the polls did not predict was Oklahoma. Though he did out perform the polling in 3 additional states where he was expected to win, his overall share of the vote was 38%, the same as his standing in the national polls at that point. Furthermore, the change in mentions overall represents a decline for Santorum and he bested his polling support in three of the six states for which we have polling data and won three of the ten states overall.

The findings are similar in considering the third and fourth hypotheses. Again, there is strong evidence for the surge and retweet metrics that concern the third and fourth hypotheses, respectively in Iowa, but the results are mixed when we turn out attention to Super Tuesday. There is evidence at that point people are discussing a surge in Gingrich's momentum and his surge stands out in contrast to the relatively flat relationship between momentum terms and the other candidates. At the same time, the retweet statistics appear to provide a better account of the overall finish of the candidates with both Santorum and Gingrich out-pacing Romney and Paul's relatively flat statistics on this measure. Hence evaluations of momentum tend to pick up on the most dramatic change, in this case Gingrich winning his home state and gaining in other races outside of Georgia by even larger margins. The fact that this metric is only associated with the candidate with the largest shift in momentum should not be surprising: more than one candidate cannot be said to surge as then they would be keeping pace with each other.

The retweets therefore provide an additional meaningful metric to analyze shifts in support. As communications involve different relationships for the sender and the receiver, commenting on the current state of a political race expresses those themes and candidates the author regards as meaningful. However, retweeting signifies that someone else thought a communication was valuable. These data in both the Iowa case, where there was one candidate who clearly leap frogged the others, and on Super Tuesday where there was both national level movement on the part of two candidates and additional movement within each state competition, this statistic appears to follow the overall arch of change in support consistent for those candidates that moved ahead of the polling as well as those who generally maintained position. Hence, if we are to reconcile this data with the earlier findings for Iowa, significant changes in communications about candidates may only occur under the condition where there is information to report about a change in the position of a candidate.

Given the temporal nature of information, i.e. though a statement may retain meaning, it ceases to be information once it is already known (Luhmann 1995, 67–68), there is an emphasis in tweeting like in all interactive systems to communicate something that is both meaningful and new. In that sense, we may be able to rectify the the diversity of observations regarding changes in candidate mentions. To the extent that a tweet provides information about a candidate, more tweets signal the entrance of more information into the system regarding that candidate. Perhaps at an early stage of a campaign, the entrance of new information indicates the increasing relevance of a candidate. This also holds with respect to estimations of surging by the varied participants in the Twittertariat. However, retweets are an indication not of the introduction of information to an interaction system but the informational value of a message about a candidate that merits its rediffusion. This latter metric perhaps helps explain the robust relevance of retweets in identifying candidate surges: though the communication may not be “new,” the change in momentum of the candidate makes it more relevant ensuring its continued informational value.

Given that we have considered separate electoral events, our findings are not limited to retrospective model fitting of a single electoral event and we are able to draw broader conclusions about the predictive value of Twitter. Nevertheless, there are several caveats on the wider application of Twitter-based predictive analytics. First, there are a wide range of informational distortions that are possible. Automated bots rather than biographical humans may skew the overall distribution of tweets. There is evidence that perhaps as many as 92% of Newt Gingrich's Twitter followers may be bots, though they are relatively inactive (Gawker 2011). Additionally, many political activists sign up to donate their twitter feed to particular groups. Although, these tweets may accurately reflect an opinion, they may also overly amplify the distribution of opinions online. This is an open empirical question. Moreover, to the extent that obtains, researchers, like campaigns that seek to use these data, must develop techniques to identify suspicious accounts in real time and discount them. Nonetheless, mining Twitter may be a valuable tool for identifying change and continuity in the state of an electoral campaign. However, the proper metaphor may be less analogous to traditional polling and more like a canary in a coal mine.

Altschuler, Bruce E. 2008. "Selecting Presidential Nominees by National Primary: An Idea Whose Time Has Come?" *The Forum* 5(4). http://www.degruyter.com/view/j/for.2008.5.4_20120105083452/for.2008.5.4/for.2008.5.4.1206/for.2008.5.4.1206.xml (August 1, 2012).

Ampofo, Lawrence, Nick Anstead, and Ben O'Loughlin. 2011. "Trust, Confidence, And Credibility." *Information, Communication & Society* 14(6): 850–871.

Bird, Steven, Ewan Klein, and Edward Loper. 2009. *Natural Language Processing with Python*. 1st ed. O'Reilly Media.

Bollen, Johan, Huina Mao, and Xiaojun Zeng. 2011. "Twitter mood predicts the stock market." *Journal of Computational Science* 2(1): 1–8.

Castells, Manuel. 2009. *Communication Power*. Oxford: Oxford University Press.

Chadwick, Andrew. 2010. "The Political Information Cycle in a Hybrid News System: The British Prime Minister and the 'Bullygate' Affair." *The International Journal of Press/Politics* 16(1): 3–29.

Conover, M. D, B. Gonçalves, J. Ratkiewicz, A. Flammini, and F. Menczer. 2011a. "Predicting the political alignment of twitter users." In *Privacy, Security, Risk and Trust (PASSAT), 2011 IEEE Third International Conference on and 2011 IEEE Third International Confernece on Social Computing (SocialCom)*, , p. 192–199.

Conover, M. D, J. Ratkiewicz, M. Francisco, B. Goncalves, A. Flammini, and F. Menczer. 2011b. "Political polarization on twitter." In *Proc. 5th Intl. Conference on Weblogs and Social Media*,.

Gawker. 2011. "Update: Only 92% of Newt Gingrich's Twitter Followers Are Fake." *Gawker*. August 2, 2011. http://gawker.com/5826960/update-only-92-of-newt-gingrichs-twitter-followers-are-fake (August 7, 2012).

Gayo-Avello, Daniel. 2011. "Don't turn social media into another 'Literary Digest' poll." *Communications of the ACM* 54(10): 121.

Goidel, Kirby. 2011. *Political Polling in the Digital Age (pdf): The Challenge of Measuring and Understanding Public Opinion*. LSU Press.

González-Bailón, Sandra, Javier Borge-Holthoefer, Alejandro Rivero, and Yamir Moreno. 2011. "The Dynamics of Protest Recruitment through an Online Network." *Scientific Reports* 1. http://www.nature.com/srep/2011/111215/srep00197/full/srep00197.html (August 3, 2012).

Goodnight, G. Thomas, and David B. Hingstman. 1997. "Review essay: Studies in the public sphere." *Quarterly Journal of Speech* 83(3): 351–370.

Haberman, Maggie, and Alexander Burns. 2012. "The 2012 campaign is the smallest ever - Maggie Haberman and Alexander Burns." *POLITICO*. http://www.politico.com/news/stories/0612/77620.html (August 2, 2012).

Habermas, Jürgen. 2006. "Political Communication in Media Society: Does Democracy Still Enjoy an Epistemic Dimension? The Impact of Normative Theory on Empirical Research."

Communication Theory 16(4): 411–426.

Hillygus, D. Sunshine. 2011. "The Evolution of Election Polling in the United States." *Public Opinion Quarterly* 75(5): 962–981.

Iyengar, Shanto, and Robert C. Luskin. 2004. *"Deliberative Public Opinion in Presidential Primaries: Evidence from the Online Deliberative Poll " *.*

Jensen, Michael J., Laia Jorba, and Eva Anduiza. 2012. "Introduction." In *Digital Media and Political Engagement Worldwide: A Comparative Study*, eds. Eva Anduiza, Michael J. Jensen, and Laia Jorba. Cambridge University Press, p. 1–15.

Jungherr, A., P. Jurgens, and H. Schoen. 2011. "Why the Pirate Party Won the German Election of 2009 or The Trouble With Predictions: A Response to Tumasjan, A., Sprenger, T. O., Sander, P. G., & Welpe, I. M. 'Predicting Elections With Twitter: What 140 Characters Reveal About Political Sentiment'." *Social Science Computer Review* 30(2): 229–234.

Kaushik. Avinash. 2012. http://www.youtube.com/watch?v=CrSX97elHDA&feature=youtube_gdata_player (August 5, 2012).

Luhmann, Niklas. 1995. *Social Systems*. Stanford: Stanford University Press.

McCoy, Terrence. 2012. "The Creepiness Factor: How Obama and Romney Are Getting to Know You." *The Atlantic*. http://www.theatlantic.com/politics/archive/2012/04/the-creepiness-factor-how-obama-and-romney-are-getting-to-know-you/255499/ (August 1, 2012).

Nielsen, Rasmus Kleis. 2012. *Ground Wars: Personalized Communication in Political Campaigns*. Princeton University Press.

Ratkiewicz, Jacob, Michael Conover, Mark Meiss, Bruno Gonçalves, Snehal Patil, Alessandro Flammini, and Filippo Menczer. 2010. "Detecting and Tracking the Spread of Astroturf Memes in Microblog Streams." *arXiv:1011.3768*. http://arxiv.org/abs/1011.3768 (July 19, 2012).

Redlawsk, David P., Caroline J. Tolbert, and Todd Donovan. 2010. *Why Iowa?: How Caucuses and Sequential Elections Improve the Presidential Nominating Process*. University of Chicago Press.

Romano, Lois. 2012. "Obama's Data Advantage." *POLITICO*. http://www.politico.com/news/stories/0612/77213.html (August 1, 2012).

Rothschild, David. 2009. "Forecasting Elections Comparing Prediction Markets, Polls, and Their Biases." *Public Opinion Quarterly* 73(5): 895–916.

Sifry, Michael. 2011. "The Ron Paul Paradox" *TechPresident* December 21, 2011. http://techpresident.com/news/21538/ron-paul-paradox (August 7, 2012).

Smith, Aaron, and Joanna Brenner. 2012. "Twitter Use 2012 | Pew Research Center's Internet & American Life Project." http://www.pewinternet.org/Reports/2012/Twitter-Use-2012.aspx (August 2, 2012).

Stark, L. P. 1996. "The Presidential Primary and Caucus Schedule: A Role for Federal Regulation?"

Yale Law & Policy Review 15(1): 331–397.

Tau, Byron. 2011. "Campaigns capitalize on Facebook." *POLITICO*. http://www.politico.com/politico44/2011/12/campaigns-capitalize-on-facebook-108411.html (August 7, 2012).

Tumasjan, Andranik, Timm O. Sprenger, Philipp G. Sandner, and Isabell M. Welpe. 2011. "Election Forecasts With Twitter How 140 Characters Reflect the Political Landscape." *Social Science Computer Review* 29(4): 402–418.

Weigel, David. 2012. "The Tweet Campaign vs. the Real Campaign." *Slate*. http://www.slate.com/articles/news_and_politics/politics/2012/07/barack_obama_and_mitt_romney_are_conducting_a_more_substantive_campaign_than_many_pundits_and_journalists_realize_.html (August 2, 2012).

Wheeler, Mark. 2012. "The Democratic Worth of Celebrity Politics in an Era of Late Modernity." *The British Journal of Politics & International Relations* 14(3): 407–422.

Polling data sources:

http://www.realclearpolitics.com/epolls/2012/president/ia/iowa_republican_presidential_primary-1588.html

http://www.realclearpolitics.com/epolls/2012/president/ma/massachusetts_republican_presidential_primary-2743.html

http://www.realclearpolitics.com/epolls/2012/president/id/idaho_republican_presidential_caucus-3177.html

http://uselectionatlas.org

http://www.gallup.com/poll/153107/Romney-Advances-Lead-Santorum.aspx

Table 1: Twitter Profile Characteristics

	Friends	Followers	Statuses	Total Contributors
Iowa	3911	10420	32960	195737
Super Tuesday	304	258	3771	137945

Table 2: Twitter and Opinion Polling in Relation to Iowa Results

	Vote Percent	Polling Average	Tweet Percent	Difference (Vote-Polls)	Difference (Vote-Tweets)
Bachmann	5.0	6.8	4.9	-1.8	0.1
Gingrich	13.3	13.7	9.7	-0.4	3.6
Huntsman	0.6	2.3	2.7	-1.7	-2.1
Paul	21.4	21.5	17.3	-0.1	4.1
Perry	10.3	11.5	4.1	-1.2	6.2
Romney	24.5	22.8	23.0	1.7	1.6
Santorum	24.5	16.3	20.2	8.2	4.4
Mean Absolute Error				2.2	3.1

Table 3: Twitter Change Metrics for Iowa

Candidate	Change Mentions	Change/Total Mentions	Retweet chiSq	Perceived Momentum
Bachmann	3270	0.09	0.02	-0.02
Gingrich	1151	0.01	343.80***	-0.02
Huntsman	894	0.04	62.49***	-0.00
Paul	12391	0.08	947.68***	-0.02
Perry	2365	0.07	931.10***	-0.01
Romney	4796	0.02	964.44***	-0.02
Santorum	27950	0.18	1764.15***	0.20

Note: ***p<0.000, otherwise insignificant.

Table 4: Super Tuesday Tweets and National Votes

	Percent National Vote	National Polls	Percent Twitter Mentions	Difference (Vote-Polls)	Difference (Vote-Tweets)
Gingrich	23	15	10	8	13
Paul	11	12	8.0	-1	3
Romney	38	38	40	0	-2
Santorum	27	22	29	5	-2
Mean Absolute Error				3.5	5

Note: Poll figures are representative of Republicans and Republican-leaning independents, reported by Gallup's tracking polls conducted February 29-March 4, 2012. All figures have been rounded to the nearest percent as Gallup does not report more exact figures.

Table 5: State-by-State Polling and Results

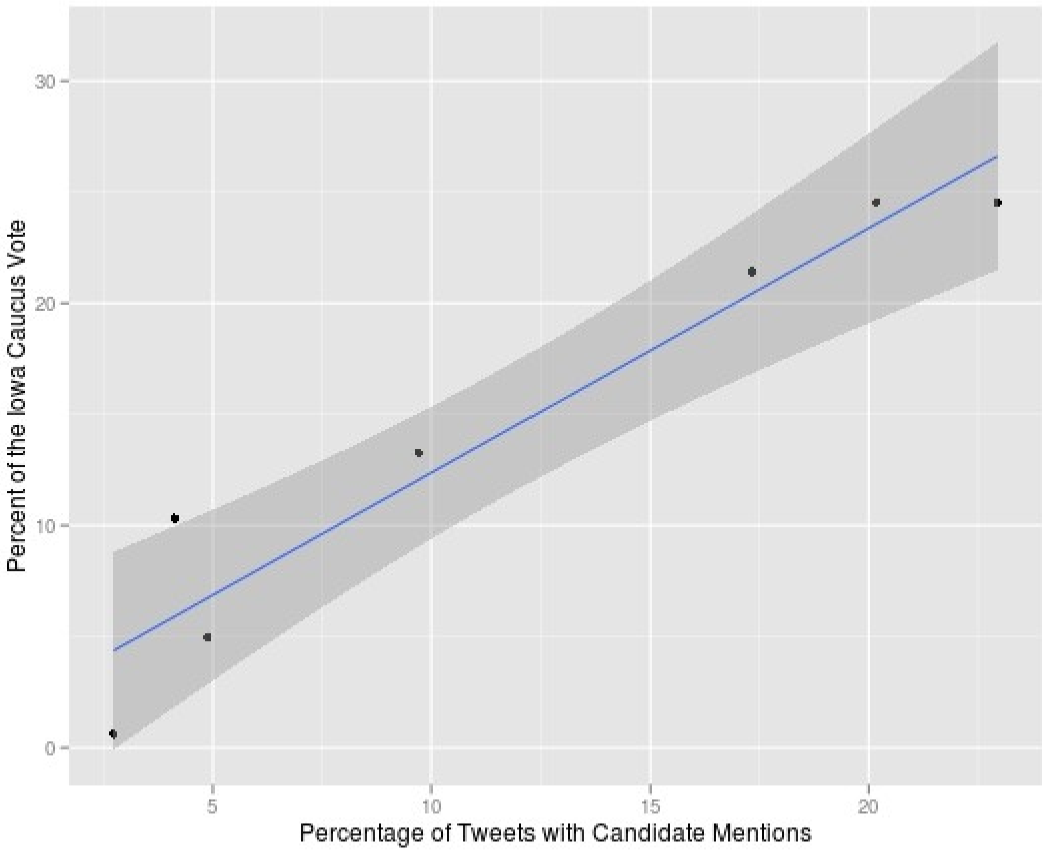
	Polling Average	Vote	Difference (Vote-Polls)
Georgia			
Romney	23.8	25.9	2.1
Santorum	17.4	19.6	2.2
Gingrich	44.4	47.2	2.8
Paul	8	6.6	-1.4
Ohio			
Romney	33.9	38.0	4.1
Santorum	32.7	37.1	4.4
Gingrich	16.4	14.6	-1.8
Paul	11.3	9.2	-2.1
Oklahoma			
Romney	18.0	28.1	10.1
Santorum	43.0	33.8	-9.2
Gingrich	22.0	27.5	5.5
Paul	7.0	9.6	2.6
Tennessee			
Romney	29.7	28.1	-1.6
Santorum	32.3	37.2	4.9
Gingrich	24.7	23.9	-0.8
Paul	9.3	9.0	-0.3
Vermont			
Romney	34	40	6
Santorum	27	24	-3
Gingrich	14	8	-6
Paul	10	26	16
Virginia			
Romney	69	60	-9
Paul	26	41	15
Mean Absolute Error			5

Table 6: Super Tuesday Twitter Change Metrics

Candidate	Change Mentions	Change/Total Mentions	Retweet chiSq	Perceived Momentum
Gingrich	9533	0.10	144.74***	0.06***
Paul	5043	0.08	2.01	-0.01***
Romney	10368	0.03	126.06***	-0.01***
Santorum	-1264	-0.01	162.43***	0.00***

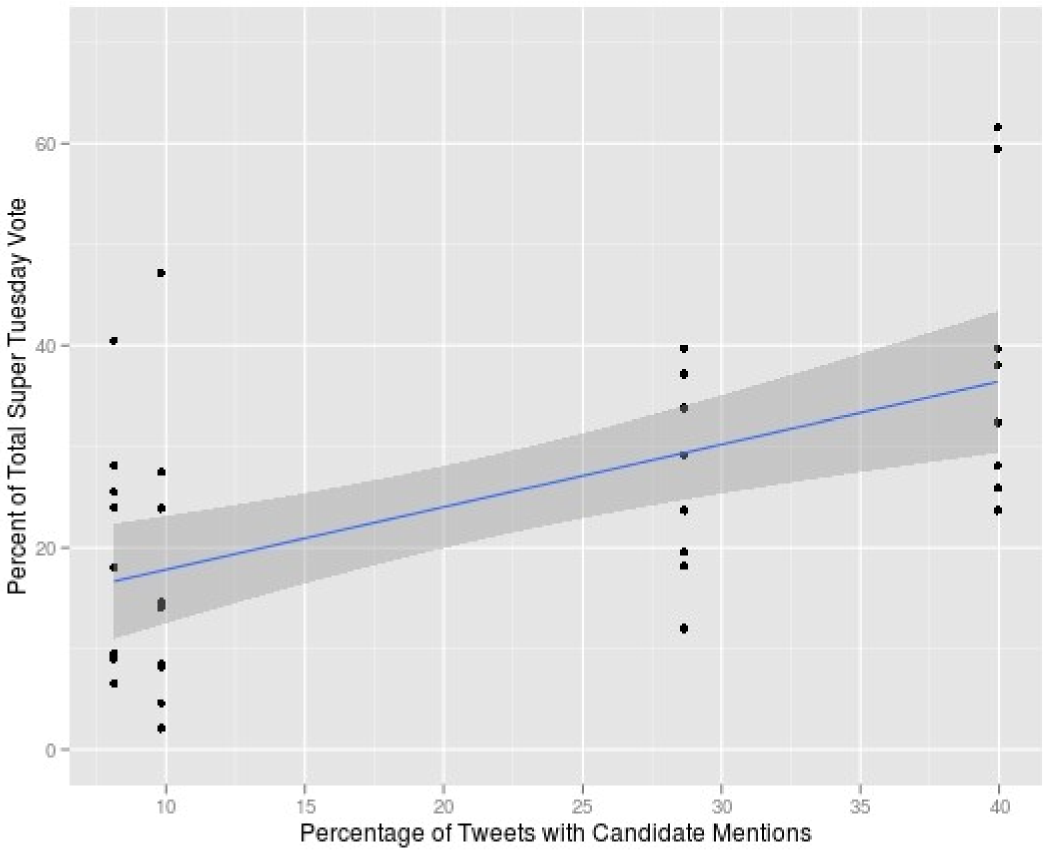
Note: ***p<0.001, otherwise insignificant.

Figure 1: Percentage of Tweet Mentions



Notes: Graph based on a linear ordinary least square model. $R^2 = 0.92$ on 5 degrees of freedom. The line slope is 1.01 and the t value = 7.52 (p=0.000).

Figure 2: Super Tuesday Mentions and Vote (National Model)



Notes: Graph based on a linear ordinary least square model. $R^2 = 0.35$ on 36 degrees of freedom. The line slope is 0.71 and the t value = 4.46.