Understanding US Primary Candidates' Communication Strategies and Topics of Interest Through Twitter

Abstract

This paper examines the Twitter communication strategies of six major candidates seeking their party's nomination for the 2016 U.S. presidential election. Firstly, we assess the top down communication strategy of the candidates. We look at the volume and type of activity while analyzing the text of the tweets to better understand the approaches of these candidates, such as their mentions of spouses and rivals. Using sentiment analysis, we uncover the level of positivity in each of the candidates' tweets and the level of anger displayed by each of the candidates. Finally, in assessing the topics discussed, we note who, and what, policy areas the candidates choose to talk about. We conclude by discussing the advantages and challenges in assessing communication strategies using Twitter data.

The Impact of Twitter on Elections

Politics has radically changed in just a generation. From the Internet being a receptacle upon which to dump policy positions that only political junkies consulted, online communication is now central to political strategy. The Internet represents a way to go direct to voters, to drive the campaign agenda with well-crafted messages without the expenditure of massive amounts of campaign funding. Twitter exemplifies this; a medium that encourages pithy (140 characters or less) communications direct to millions of followers at, essentially, no variable cost per tweet.

The importance of Twitter in politics and everyday life has led to the microblogging site being widely studied in political marketing (Bode and Dalrymple 2015; Jackson and Lilleker 2011; Aharony 2012; Conway, Kenski, and Wang 2013). Political candidates value Twitter as it allows access to voters unfiltered by the traditional media. That said, Twitter can also interact with traditional media to amplify a message's impact. Done especially well, or especially badly, tweets earn traditional media mentions. For example, Donald Trump's controversial tweets about the other candidates, debate moderators, and even the Pope, garnered widespread television and newspaper coverage in the U.S and beyond.[1]

Strategies in U.S. primaries are particularly interesting as the candidates have personal styles which can be quite distinct from the traditional approaches of their party. This leads to questions about the role of party label as useful information, and the clear control over the political brand that often seems vital to political marketing strategy (Knuckey and Lees-

---

[1] http://www.nytimes.com/interactive/2016/01/28/upshot/donald-trump-twitter-insults.html?_r=0, July 29th, 2016, *The New York Times,* The 250 People, Places and Things Donald Trump Has Insulted on Twitter: A Complete List, by Jasmine C. Lee and Kevin Quealy, Accessed August 10th, 2016.

Marshment 2005; Marland 2016; Scammell 1999). Indeed, some candidates may even be thought to damage the party brand. Donald Trump, in particular, was seen as a candidate who used the direct access Twitter allowed him to bypass his party's establishment and engage directly with voters (Heffernan 2016). He was able to convey the image of an outsider that saw him easily defeat establishment candidates to gain the Republican nomination.

Our data is all the tweets, and retweets, of the major party, i.e. Democratic and Republican, candidates. Traditionally, political marketing research uses case studies (Miller 2013), interviews (Parsons and Rowling 2015), and manual coding (Graham et al. 2013). We employ text mining and sentiment mining, specifically using the Syuzhet package for the R program, and its associated lexicons (Jockers 2016a). Our theoretical approach seeks to understand the communications strategies of candidates in primaries, and how these strategies differ by candidate. Do the candidate's focus upon themselves, their spouses, their rivals? We thus look at a traditional paradigm of political communications, messages sent by politicians, even though the online space we use can be used more easily for co-productive activities than traditional media (Lilleker and Jackson 2010).

We examine topics, such as healthcare and immigration, in the tweets from each candidate to better understand the policies that the candidates discuss. We will assess which terms are relatively popular in each candidate's tweets by comparing the popularity of terms within a given candidate's tweets against the popularity of terms in the entire corpus of all the candidates' tweets. This will show how topics highlighted differ within, and between, parties (Enli and Skogerbø 2013).

Our Data and Method

The Tweets and The Campaign Length

We captured the Twitter activity of six candidates for the nomination of the two major U.S. political parties, the Republicans and Democrats, from the start of January 25th, 2016 to the end of May 4th, 2016. January 25th is the first full day of trading on the Iowa Electronic Market, a prediction market which allows users to place real money on the outcome of the election -- the contract price on this market we will use as our measure of candidate success. January 25th is a week before the first actual voting of the primary season, the Iowa Caucus on February 1st. The final date, May 4th, is when John Kasich, Donald Trump's last rival in the Republican Primary suspended his campaign, in effect conceding that Trump had won the nomination. At this point, while there were still significant states to vote, it also seemed nearly certain that Hillary Clinton had secured the Democratic nomination. May 4th thus marks a natural break when the attention of the candidates' and media effectively turned to the general election in November.

Our data collection process involved REST APIs, which are publicly available to Twitter developers. This provides our R codes authorized access to read Twitter data that are open to public. We capture data at discrete intervals throughout the specified dates because the REST APIs impose a maximum search limit of 3,200 most recent tweets per profile. Using discrete intervals means that tweets can be deleted by the candidates between collections and our data may omit a small number deleted by the campaigns. While this is a possible source of bias, we find this is less of a concern than it might be given these candidate's Twitter feeds are monitored constantly by the press and public. Campaigns know that anything they post becomes public record. In essence, deletion gains a candidate little; a tweet's effects have already been caused as soon as they are posted.

The Candidates

We focus on the tweets of six major candidates for the U.S. presidential election. Our definition of major candidates is those who have their own contract price on the Iowa Electronic Market. We will explain this market later, but having a contract price suggests that the candidate is seen by those hosting the market as being a candidate likely to secure the nomination. This includes two Democratic candidates -- Hillary Clinton, the eventual winner, and Bernie Sanders, her main rival. The choice of these two candidates would have been relatively easy to make. After the vice-president, Joe Biden, declared that he was not running in autumn 2015[2], few suspected a candidate other than Clinton or Sanders would win the Democratic nomination. The Republican primary always seemed more uncertain but by late January, a number of initially plausible candidates appeared very unlikely to win. The most obvious omission from the list of likely winners is Jeb Bush, the early favourite[3], who by January appeared to be struggling and was, as it turned out correctly, thought by late January not to be a strong contender.[4] John Kasich, Trump's last rival also did not have a contract price as, despite his persistence, his campaign never seemed to have a plausible path to victory through electoral means (as opposed

---

[2] http://www.cnn.com/2015/10/21/politics/joe-biden-not-running-2016-election/, October 21st, 2015, *CNN*, Joe Biden won't run for president, by Stephen Collinson, Accessed August 11th, 2016.

[3] http://www.nytimes.com/2015/06/16/upshot/jeb-bush-is-still-the-favorite-the-markets-say.html, June 15th, 2015, *The New York Times,* Jeb Bush Is Still the Favorite, the Markets Say, by David Leonhardt, Accessed August 11th, 2016.

[4] https://www.washingtonpost.com/news/the-fix/wp/2016/01/31/will-jeb-bush-win-the-republican-nomination/, February 8th, 2016, *The Washington Post,* Will Jeb Bush win the Republican nomination?, by Aaron Blake, Accessed August 11th, 2016.

to back stage machinations).[5] (Kasich won only his home state of Ohio). Therefore, the four Republican candidates with contract prices, i.e. plausible winners, are Donald Trump (the eventual winner), Ted Cruz, Ben Carson, and Marco Rubio.

First, we consider the basic descriptive statistics of our data (Vergeer, Hermans, and Sams 2011). Table 1 shows the Twitter activity of each candidate. Ted Cruz was the most active with an average of 34.8 tweets a day, followed by Bernie Sanders who tweeted nearly once an hour. Ted Cruz specialized in pithy emails, leaving Bernie Sanders the clear winner for most words tweeted. Ben Carson was the least active on Twitter. Indeed, Carson was the first of these candidates to suspend his campaign – we will later see the impact of this on his Twitter activity. Essentially, Carson's activity dried up as his campaign petered out.

[5] http://www.latimes.com/politics/la-na-trailguide-04202016-john-kasich-continues-to-believe-despite-math-not-1461180421-htmlstory.html, April 20th, 2016, *Los Angeles Times,* John Kasich continues to believe, despite immense odds, by Kurtis Lee, Accessed August 11th, 2016.

Table 1. Descriptive Statistics: Candidate Twitter Use

	Democrats		**Republicans**			
	Hillary Clinton	Bernie Sanders	Ben Carson	Ted Cruz	Marco Rubio	Donald Trump
Number of Followers~	6,112,796	2,092,378	1,311,590	1,089,765	1,350,812	7,908,579
Total Tweets~	5,456	8,285	3,031	16,571	5,426	31,830
Date Joined Twitter?	April 2013	November 2010	February 2013	March 2009	August 2008	March 2009
Suspended Campaign	NA	NA	March 4th	May 3rd	March 15th	NA
Birth State	Illinois (3)	New York (43)	Michigan (0)	Alberta, Canada (0)	Florida (47)	New York (34)
Home State	New York (27)	Vermont (6)	Florida (1)	Texas (47)	Florida (47)	New York (34)
Name of Spouse	Bill (0)	Jane (4)	Candy (7)	Heidi (35)	Jeanette (5)	Melania (10)
Tweets in Our Data	1,865	2396	539	3512	884	1406
Tweets per Day^ (101 days in data)	18	24	5	34.8	9	14
Words in Our Data*	16,327	21,510	4,486	22,408	6,558	11,885
Words Per Tweet	8.75	8.98	8.32	6.38	7.42	8.45

Notes: ^ includes time after suspending campaign for Rubio, Carson and Cruz

* After removing URLs, non-words, and stopwords (uninformative words, e.g., "the")

~ as at May 4th, 2016, data accessed using http://www.trackalytics.com/

Text Cleaning and Sentiment Analysis

We cleaned the collected tweets in R, a statistical program with a number of packages written for this purpose. A key element of cleaning Twitter data is removing uninformative characters. These include URLs, characters that do not form words, and stopwords. Stopwords are common and uninformative words present in nearly all writing but that do not distinguish between texts. These include words such as “the” and “and”. Stopwords are provided by commonly available text mining packages for R. We used the ‘tm’ package, which accesses the dictionary provided by the Journal of Machine Learning Research (Lewis et al. 2004).

Upon this dataset of cleaned words, we ran sentiment analysis using the publicly available Syuzhet package for R (Jockers 2016b). Of the many types of sentiment analysis offered by the package, we have specifically chosen the NRC Word-Emotion Association Lexicon (Mohammad n.d.). We use its associated dictionary comprised of 14,182 words, each of which is evaluated for the eight emotions and two sentiments. By sentiment mining each candidates’ tweets, for example, we can compare Hillary Clinton’s and Bernie Sander’s positivity. As our data is composed of the candidates’ tweets, our analysis does not examine, as might be more common in research, the mood of the general Twitter users in reaction to the political campaigns. Instead, we uncover the emotions displayed by each candidate through their messages as this is important to understanding each of the candidate’s campaign strategy. We are interested in candidate Twitter strategy, and not the response of readers as such.

We also analyzed the text for topics discussed by candidates and detail this method later. The topics addressed in the tweets allow us to get a better idea of the substantive policies focused on. (We will note substantial challenges with candidate’s outlining policy positions on Twitter. Additionally, our technique is not effective at uncovering the precise policies advocated rather we can assess simply the topic raised.)

Measure of Performance

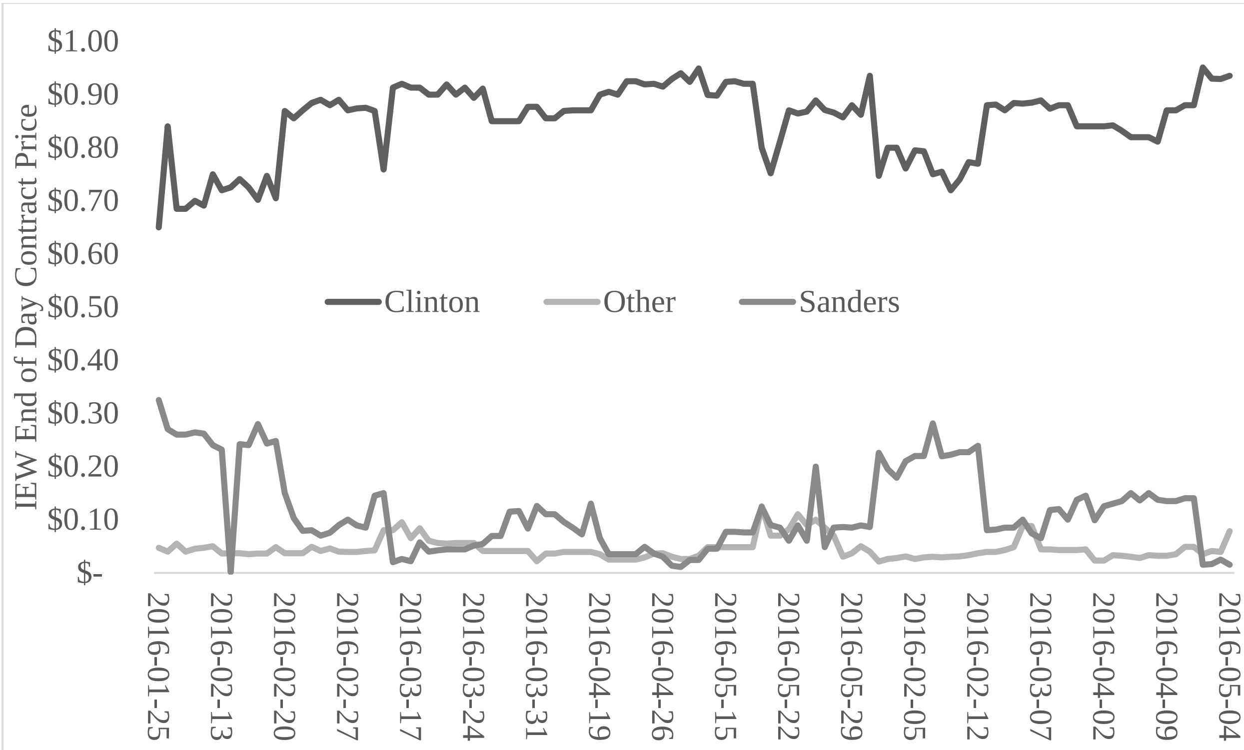
We are interested in how performance impacts campaign strategy. We operationalize performance as the perceived likelihood of a candidate winning the primary, with higher likelihood being assessed as positive performance of a candidate's campaign. We compiled a complete database of the Iowa Electronic Nomination Market prediction prices from the market's first full day, January 25th 2016. This provides daily "wisdom of the crowd" assessments of the likelihood of candidates winning the eventual nomination and gives us the ability to see changes in perceived performance. Performance changes can register even when no voting occurs. For example, strong or weak performance in a debate might move the markets, changing the perceived performance of the candidate even though no votes had been cast or delegates gained. Analyzing this data, we shall see how perceived progress, or setbacks, impact messaging.

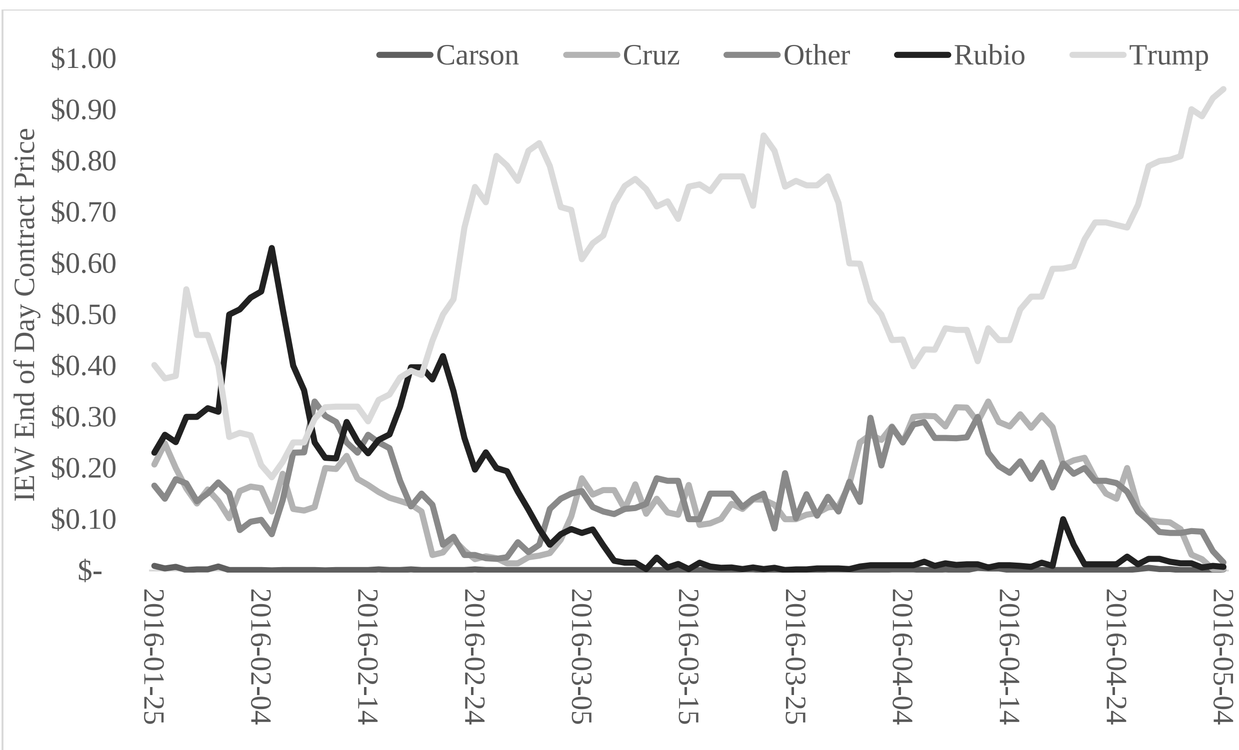
Our measure of the likelihood of winning is the Iowa Electronic Markets (IEM) contract price which arises from a free trade in candidate "futures". Those who register on the IEM site can purchase, using real money, a contract that pays out one (U.S.) dollar if the candidate wins the nomination of their party. The contract price can therefore be conceptualized as an approximate probability of any candidate winning the primary, as assessed by the wisdom of the crowd. (Given this market trades continuously there are multiple prices each day, we focus on the end of the day price). Using this measure, we can see how the fortunes of candidates change. For example, a contract that paid $1 if Donald Trump won the nomination could be purchased for 40.1 cents on January 25th but the price that had risen to 94 cents on May 4th when his final competitor dropped out reflects the increased likelihood of Trump securing the nomination.

(There remained talk of a convention plot to oust Trump so the perceived chance of Trump not gaining the nomination remained 6%, $1 minus 94 cents, despite the fact he no longer faced formal rivals). The market price shows the volatility in the Republican primary, Trump's contract price dropped to 18.2 cents on February 5th a few days after Ted Cruz won the Iowa Caucuses. (The standard deviation of end of day prices on a Trump contract is relatively high at .196). Figure 1 gives the contract prices for the candidates.

The Democratic primary was much more stable. Hillary Clinton started at 65 cents, her lowest contract price, but this price gradually rose to 93.5 cents by the end of the Republican race. Her contracts were almost the same price as a Trump contract despite Clinton still facing active resistance from Bernie Sanders. The relative stability of the Democratic primary can be seen from the fact that the standard deviation of her contract price is much lower than Trump's at only .067.

Figure 1. Prediction Market Prices for Each Candidate





Communication Strategy

Communication strategies are (often) meticulously crafted by political campaigns. Indeed, modern political marketing (Lees-Marshment 2001a; b) which has impacted a number of countries, e.g., (Marland 2012; O'Cass 1996) has been criticised as potentially losing the substance of political discourse amongst the strategic repositioning in response to voter preferences (O'Shaughnessy 2001; Savigny 2008). In this research we first look at communications strategy, essentially the messages sent in the text, such as mentions of rivals. In many ways here we study the selling of political candidates, (McGinniss 1988), rather than the more modern broader idea of political marketing as engaging and responding to voters (Lees-Marshment 2004).

Mention their wife/husband?

One method candidates use to humanize themselves is by showcasing their family. The picture of the candidate embracing family members is a staple of political life. We sought to understand how spouses feature in Twitter strategies. Spouses were a significant feature of the 2016 primary campaign, most notably when Donald Trump seemed to use Twitter denigrate the appearance of Ted Cruz's wife (East 2016). It is unclear why spouse appearance is important to how a candidate will do the job but it does illustrate the schoolyard nature of some communications.

Referring back to Table 1 we can see that Ted Cruz made most mentions of his spouse. It is likely that Cruz saw his relationship with Heidi as a strength, compared to his main rival, Donald Trump, who was on his third marriage which might not appeal to some primary Republican voters, especially conservative Christians.[6]

One of the notable features of this particular campaign is that Hillary Clinton's spouse, Bill, had already served as president. This created a specific challenge for the Clinton campaign. Hillary wanted to highlight her experience, she had unique in-depth personal knowledge of what it takes to be president, while avoiding having her spotlight diverted to her spouse by a press still eager to report on the ex-president. Interestingly, Hillary Clinton does not mention Bill in her tweets. The campaign strategy seems to seek to avoid any possibility of Hillary being overshadowed by the former president who, while popular with many primary voters, could still be controversial.[7] Hillary, as the first female candidate with a strong chance of winning the presidency, must have been especially keen to emphasize her independence. The Clinton's strategy had changed notably from 1992, Bill's first presidential campaign, when they explicitly ran as a team, so voters could get "two [Clinton's] for the price of one" (Dowd 1992).

Self Mentions

The candidates for U.S. presidency in 2016 showed evidence of having strong belief in themselves that critics might regard as excessive. (That said, a requirement for the job might be excessive self-belief). A focus on oneself may be more than a feature of a candidates' personalities; it may be a campaign strategy. Donald Trump, especially, had throughout his

[6] http://www.dailymail.co.uk/news/article-3378905/Trump-says-personal-indiscretions-including-cheating-wife-fair-game-politics.html, December 30th, 2015, *Daily Mail*, Trump says his own personal 'indiscretions' – including cheating on his first wife – are fair game in politics, by David Martosko, Accessed August 11th, 2016.
[7] http://www.cnn.com/2016/01/04/politics/bill-clinton-hillary-clinton-2016-campaign, January 4th, 2016, *CNN*, Bill Clinton on the trail: The good, the bad, and the ugly, by Gregory Krieg, Accessed August 11th, 2016.

career developed a strong personal brand that he leveraged in multiple arenas. To asses which candidates are the most self-focused, we measured the number of times that a candidate's tweets mention their own last name, relative to the total number of words tweeted by the candidate.

Figure 2. Self-Mentions as % of Total Words

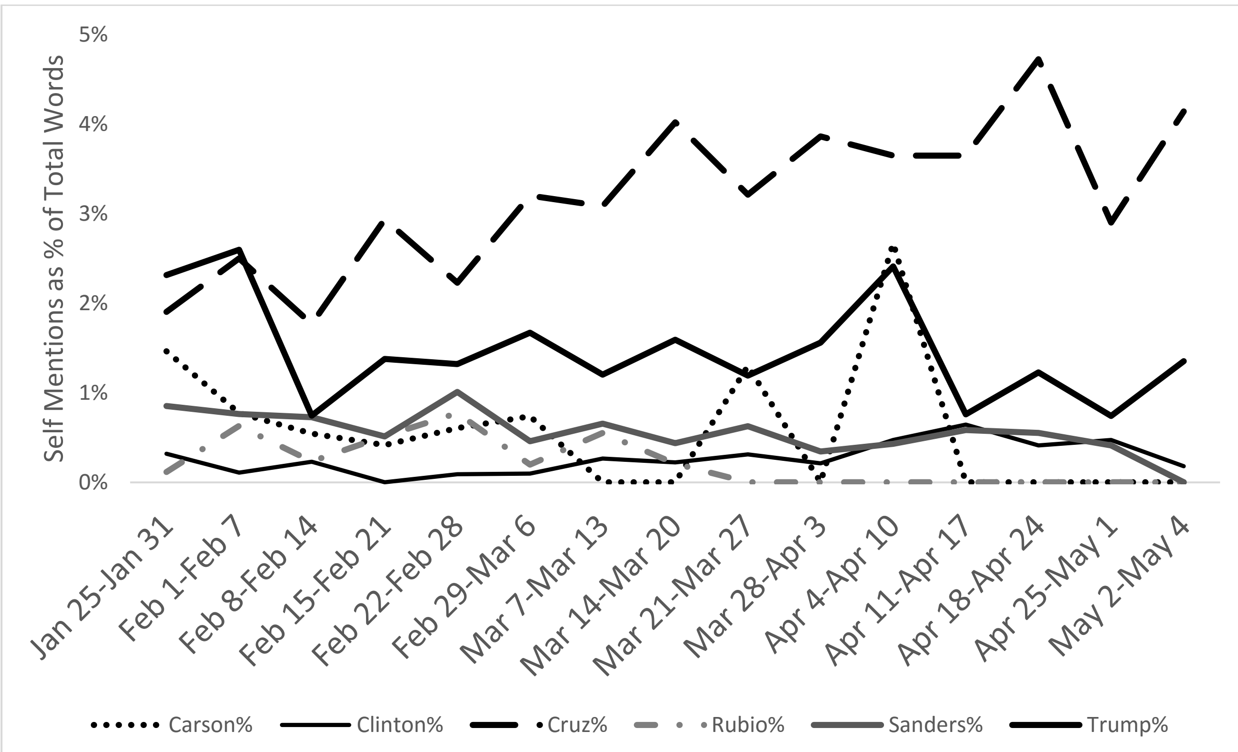
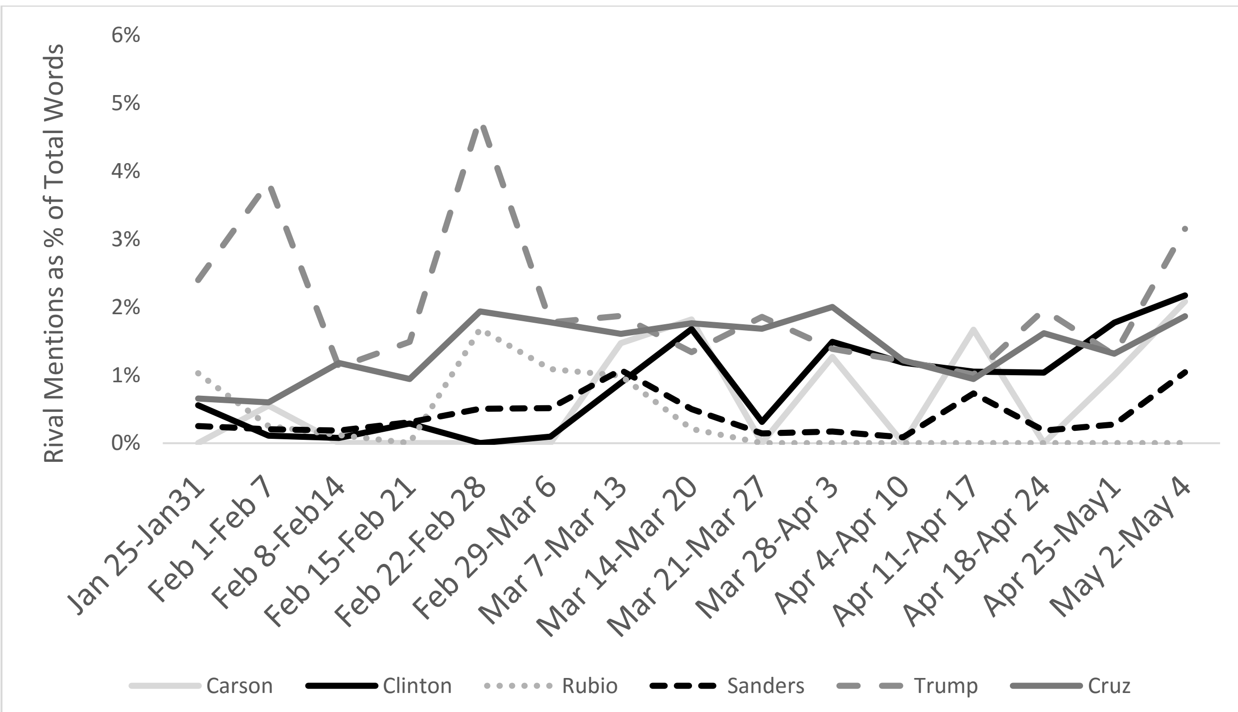


Figure 2 shows the candidate's level of self-mentions, the percentage of total words that were the candidate's last name, by week. Interestingly, Hillary Clinton shows a consistently low level of self-mentions. Her campaign appears to have decided that the Clinton name was already well-known enough and may invoke unwanted associations with Bill. The campaign instead focused more on using the more unique identifier of 'Hillary'. While Donald Trump showed high levels of self-mentions, it was his main rival, Ted Cruz, who seemed to have focused most relentlessly on getting his name across.

Mentions of Rivals

Figure 2 shows the mentions of candidates' rivals as the percentage of total words that were all the other candidates' last names by week. It is noticeable here that Sanders seems to mention his rivals less compared to the other candidates. Donald Trump was the most active in mentioning his rivals, suggesting that his campaign may be more aggressive in attacking the other candidates.

Figure 3. Mentions of Rivals as % of Total Words



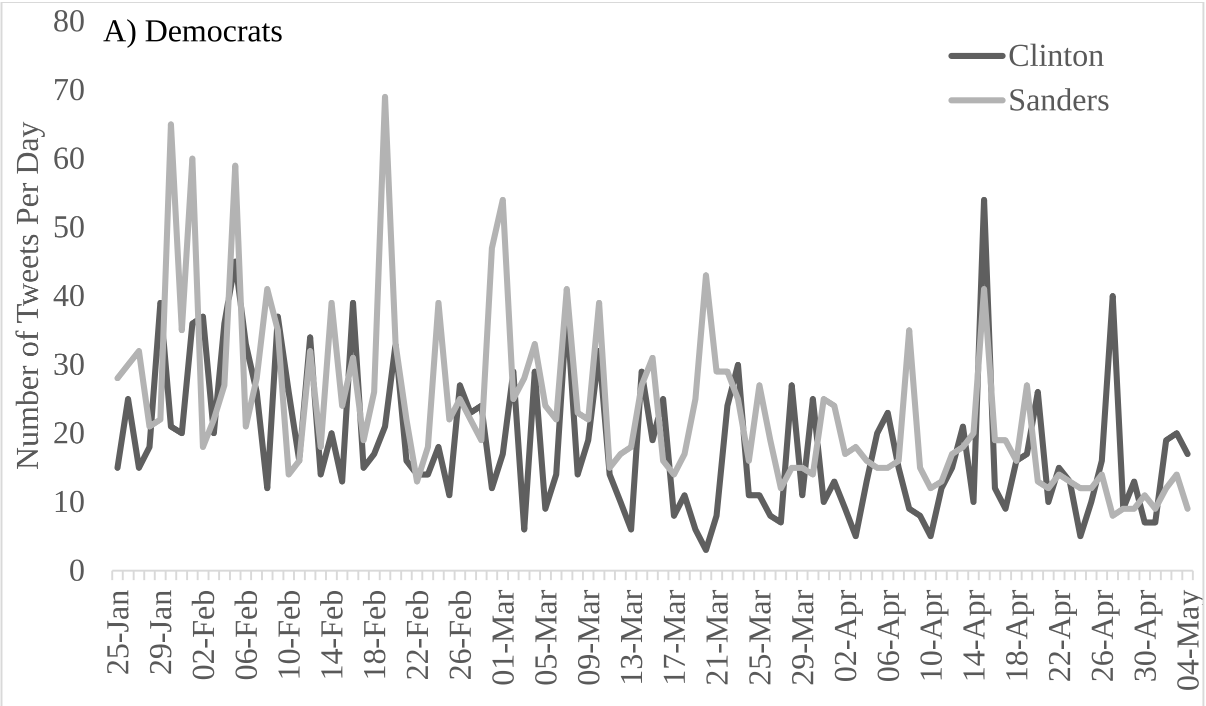
The Level of, and Changes in, Activity

Next, we turned to assessing the level of activity. We suggest that this provides a proxy for the amount of effort each candidate invests in his or her Twitter strategy. One might argue that the more comprehensive the Twitter strategy, all else being equal, the better managed the online campaign. Twitter is a medium that relies on constant engagement. From this, we can theorize about the attention that candidates are paying to Twitter, and how much the platform has impacted each candidate's marketing strategy.

Figure 4 shows the Twitter activity of the candidates. Panel A) features the Democrats: Clinton and Sanders. For these candidates, we see a relatively high level of activity but lessening over time. Moving average is one of many smoothing techniques used to remove random variation in the data to show a trend. Shown in Figure 5, the moving average, calculated in intervals of 7 days, confirms the lessening of activity over time. This may be associated with the campaign having settled into a routine. While Sanders' supporters were keen to emphasize that his chance of success was in line with theoretical advice to campaigns (Abramowitz 1989; Abramson et al. 1992), the longer the voting went on, the harder it became to credibly support Sanders winning mathematically.

Panel B of Figure 4 shows the Republicans. Here we clearly see the constant high levels of Ted Cruz's activity into May. The most obvious feature of the figure is the precipitous fall in activity of candidates leaving the race. (For campaign funding reasons ending a campaign is formally described as "suspending" the campaign). When a candidate drops out they typically take drastic actions to reduce the spending of the campaign and the staff helping to craft the candidate's tweets are likely to find themselves out of a job.

Figure 4. Level of Activity Over Time



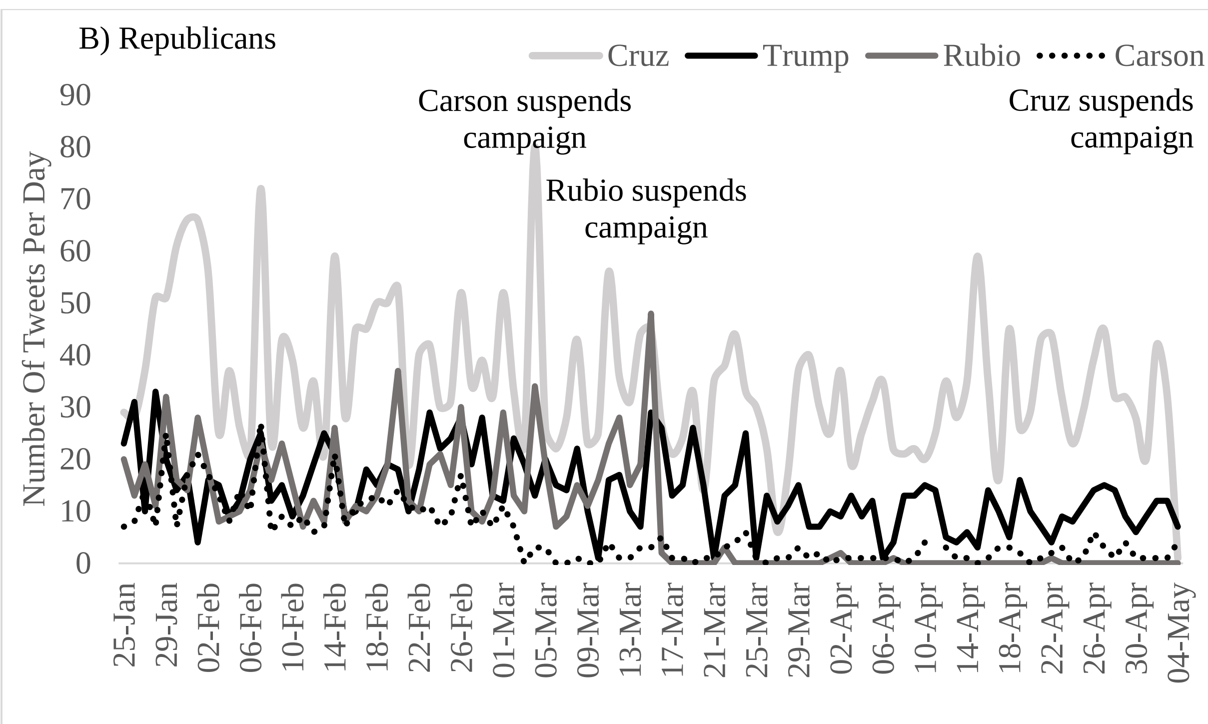
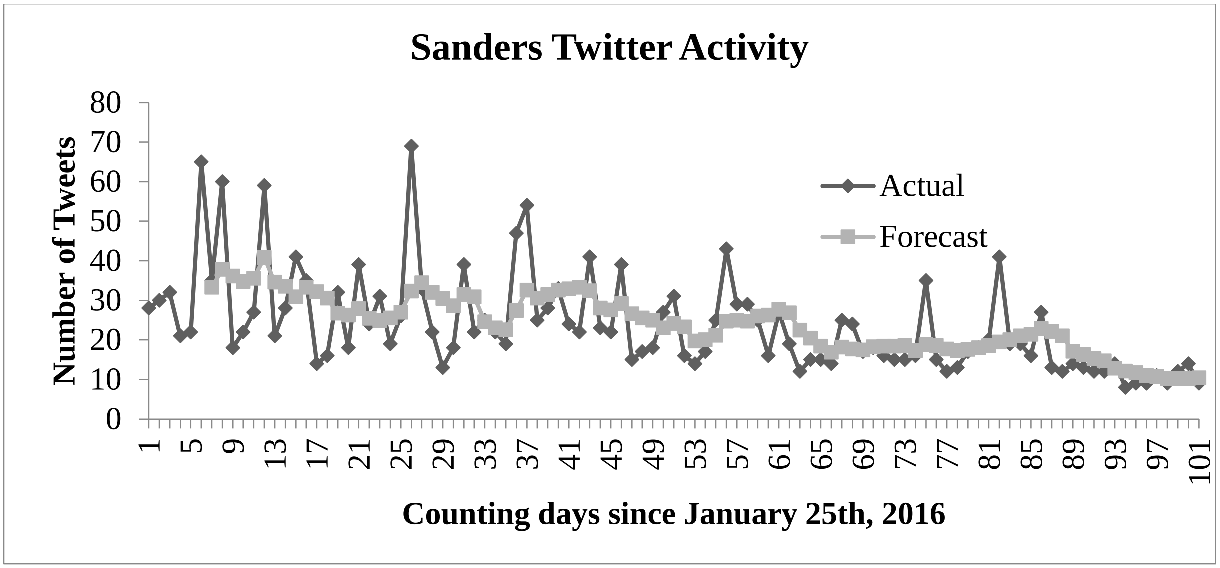
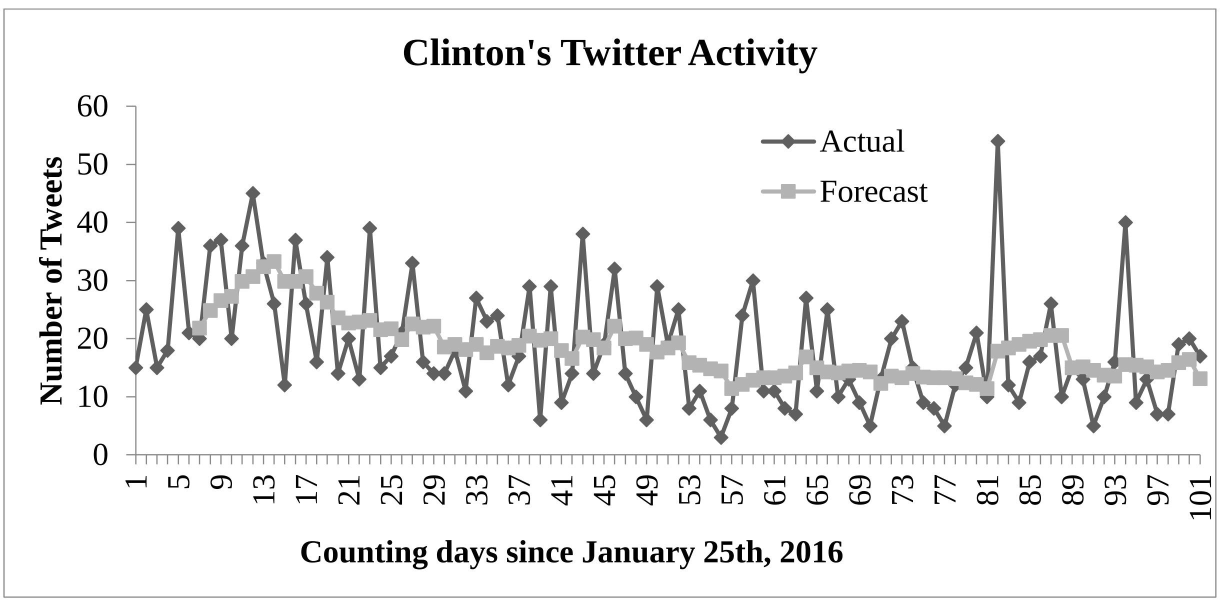


Figure 5. Moving Average of Twitter Activity

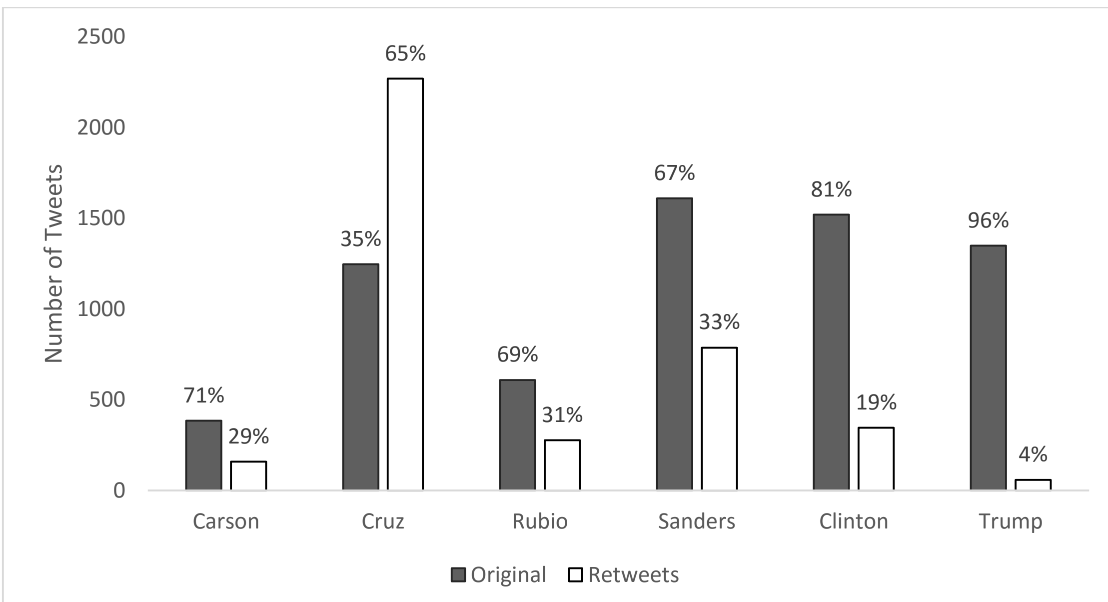




The Use of Retweets

Figure 6 shows the composition of the tweets gathered for each candidate. It is noticeable that Trump tweets proportionally more original content than retweets compared to any other candidate. While it is perhaps unsurprising that a candidate who prided himself on having a huge amount of newsworthy things to say should generate a lot of original content, this may also reflect Trump's challenges in retweeting. (He re-tweeted material from White Supremacists -- an example of a quickly deleted tweet)[8] This sharply contrasts with his closest rival, Ted Cruz, who retweets other peoples' content much more than any other candidate. Donald Trump seems to have a desire for more control over the content of his tweets or, perhaps, less confidence in the campaign's ability to spot material to re-tweet.

Figure 6. Original Tweets vs. Retweets



[8] https://www.washingtonpost.com/blogs/plum-line/wp/2016/07/04/trumps-white-supremacist-tweets-arent-the-problem-theyre-a-symptom-of-the-problem/?utm_term=.c75e95eacc9a, July 4th, 2016, *The Washington Post*, Trump's white supremacist tweets aren't the problem. They're a symptom of the problem, by Paul Waldman, Accessed August 11th, 2016.

Sentiment Analysis

Analysing their tweets, we can also hope to gain an insight into the image that the candidates are hoping to project to the world. To do this, we measured the emotions and sentiments in the candidates' tweets using the Syuzhet R statistical package. The process was as follows:

We used the NRC Word-Emotion Association Lexicon method, which was made publicly available for research purposes. It involves a dictionary that consists of 14,182 words, which have been evaluated by Mohammad and Turney (2013) for eight emotions: Anger, Anticipation, Disgust, Joy, Fear, Sadness, Surprise, and Trust. They also look at two sentiments: Positive and Negative. A word can belong to multiple components, but the ten components stand on their own. Although Mohammad and Turney found that emotions and sentiments are often correlated, they do not completely overlap. For example, while words associated with disgust are often also associated with negative sentiment, the emotions do not all fit neatly into sentiments, e.g., anticipation is not necessarily positive. After cleaning the data, the words in a tweet are matched to the words listed in this dictionary. (This process loses the order of the words, thus the analysis is often referred to as using a "bag of words" rather than analyzing text with a beginning, middle and end). Each word is given a value of 1 for the components that it matches, and a value 0 for those components it does not match. Note that a single word can represent one component, e.g., an emotion, several components, e.g., an emotion and a sentiment, or no components at all.

We then aggregated the number of sentiment related words used per day by each candidate and divided this by the total number of words tweeted that day to give a measure of each candidate's sentiment on a given day. This gave us a total of 549 emotion and sentiment observations. The Democrats supplied 202 days and each candidate posted tweets in all 101 days we analyse. The Republicans gave us only 347 observations, not 404 (4 candidates each on 101 days) as might be expected. This is because the four Republican did not always post every day. This was especially true of Ben Carson and Marco Rubio who dropped out relatively early in the race.

Table 2. Correlation of Sentiment

%	**Anger**	**Anticipation**	**Disgust**	**Joy**	**Fear**	**Sadness**	**Surprise**	**Trust**	**Positive**	**Negative**
Anger	1									
Anticipation	-0.01	1.00								
Disgust	0.47	0.00	1.00							
Joy	-0.05	0.66	-0.09	1.00						
Fear	0.49	0.03	0.50	-0.08	1.00					
Sadness	0.55	0.14	0.52	0.06	0.46	1.00				
Surprise	0.11	0.24	0.15	0.37	-0.08	0.22	1.00			
Trust	0.02	0.36	0.03	0.40	0.07	0.11	0.14	1.00		
Positive	-0.11	0.50	-0.13	0.57	-0.03	0.00	0.08	0.47	1.00	
Negative	0.81	-0.10	0.53	-0.15	0.52	0.61	0.01	-0.07	-0.18	1.00

Table 2 displays the correlations of the entire data set, and indeed shows that there are different emotion and sentiment pairs that are strongly correlated, such as positive and joy, and negative and anger.

Who Uses the Most Positive Language?

We next considered who conveyed the more positive image on Twitter. We created a scaled variable called net valence, which is simply the positive sentiment score minus the negative sentiment score, divided by the sum of the two scores for that day. We then created linear regression models to check the association between the candidate, and party and the level of net valance.

One factor that might cause the candidates to exude a more positive tone is their level of performance. We therefore controlled for the last price of the candidate at the day of the tweet. Higher-end of day prices is assumed to reflect greater success, which in turn might improve the positive sentiment of the candidate and campaign staff and be reflected in the increased net valence of the tweets. Negative sentiment may also intensify when last prices decreases.

Table 3. Predicting Net Valence by Political Party

	Estimate	**Std. Error**	**t value**	**Pr (> |t|)**
Intercept	0.34437	0.03238	10.634	< 2e-16 ***
LastPrice	-0.13848	0.04634	-2.988	0.002933 **
Republican	0.10648	0.03137	3.394	0.000739 ***

Significance codes: *** p<0.001 ** p<0.01

Table 3 shows that Republicans were significantly more positive against the baseline, which is the Democratic score, in the sense of having a higher net valence in their tweets. Unexpectedly, higher last prices actually predicted lower levels of positivity. If anything, success is associated with negativity. We suspect that this is likely to be a result of the characteristics of those candidates who we found were most successful, and so had higher last prices rather than a direct causal relationship between being negative on Twitter and being successful. We therefore turned to individual candidates.

Table 4. Predicting Net Valence by Candidate

	Estimate	Std. Error	t value	Pr (> |t|)
Intercept	0.32024	0.10359	3.091	0.00209 **
LastPrice	-0.01576	0.11669	-0.135	0.89265
Sanders	-0.07027	0.09568	-0.734	0.46305
Rubio	0.22359	0.09052	2.470	0.01381 *
Trump	-0.10602	0.05422	-1.956	0.05103 ^
Carson	0.12964	0.10880	1.192	0.23395
Cruz	0.21062	0.09243	2.279	0.02307 *

Significance codes: ** p<0.01 * p<0.05 ^ p<0.1

In Table 4 we see that Rubio and Cruz were considerably more positive than the baseline, which we chose to be Hillary Clinton. On the other hand, Donald Trump was significantly less positive than Hillary Clinton.

Who is the angriest candidate?

2016 was widely described as an election that revealed the anger of the electorate on both sides of the aisle[9]. The Democrats gave surprisingly strong support to an anti-establishment candidate, Bernie Sanders, who, despite his lengthy service in Congress, only joined the party ("registered as a Democrat") in September 2015, which was well into his campaign to be the nominee of the party[10]. Prior to this, although he had caucused with the Democratic party, he had preferred to remain independent and was happy to be described as a "socialist", a term not widely used by the U.S. political class. If anything, the Republican disdain for the establishment

[9] http://www.bbc.com/news/magazine-35406324, February 4th, 2016, *BBC*, Why are Americans so angry?, by Vanessa Barford, Accessed August 12th, 2016.
[10] http://www.dailykos.com/story/2015/9/17/1422309/-BREAKING-Bernie-to-officially-Register-as-a-Democratic, September 17th, 2015, *The Daily Kos*, BREAKING: Bernie to officially Register as a Democrat, by cdub24, Accessed August 5th, 2016.

was even more obvious. The candidate who was eventually chosen, Donald Trump, had no prior political experience. His closest rival, Ted Cruz, was a senator but one widely disliked by his own party who did not see him as a team player. (His colleague, Republican Senator Lindsey Graham, joked that one could murder Ted Cruz on the floor of the Senate, and as long as the trial was held in the Senate "nobody would convict you".[11])

We sought to examine if the anger allegedly felt by the electorate was mirrored in the Twitter activity of the candidates. Do the candidates match the reported anger of the electorate? Table 5 shows the results of a linear regression model predicting anger, given the political party and controlling for last price (our proxy for performance). The dependent variable is simply the % of all words tweeted by the candidate that are classified as angry. The output shows that the Republican dummy variable is highly significant, perhaps surprisingly given the sitting President was a Democrat the Republicans were less angry, but the last price variable is not.

Table 5. Anger Expressed by Political Party

	Estimate	**Std. Error**	**t value**	**Pr (> |t|)**
Intercept	0.043595	0.002579	16.904	< 2e-16 ***
LastPrice	-0.001109	0.003691	-0.300	0.764
Republican	-0.013796	0.002499	-5.521	5.21e-08 ***

Significance codes: *** p< 0.001

We then looked at individual candidate's expressions of anger. Table 6 shows that last price is significant, and that increased performance is associated with less angry language. It is notable that Hillary Clinton, who is the baseline used in the analysis, used angrier language than

[11] http://www.cnn.com/2016/02/26/politics/lindsey-graham-ted-cruz-dinner, February 26th, 2016, *CNN*, Lindsey Graham jokes about how to get away with murdering Ted Cruz, by Catherine Treyz, Accessed August 12th, 2016.

Carson, Cruz and Rubio. (They used significantly less angry language than the baseline). This may be somewhat surprising, given the president was from her party, but this may reflect a campaign strategy to use her anger to focus her supporters on the disagreements that she has with the Republicans. Her strategy seems consistent with the theoretical strategic advice given to more centrist candidates (Clinton was widely perceived to be more centrist than her rival Bernie Sanders) to focus their parties externally -- on the problems in the other party's policies -- not on the benefits of one's own policy platform (Bendle and Nastasoiu 2014; Bendle 2014).

Table 6. Anger Expressed by Candidates

	Estimate	**Std. Error**	**t value**	**Pr (> |t|)**
Intercept	0.060593	0.008447	7.173	2.41e-12 ***
LastPrice	-0.024505	0.009515	-2.575	0.010278 *
Sanders	-0.011401	0.007802	-1.461	0.144534
Rubio	-0.028453	0.007381	-3.855	0.000130 ***
Trump	-0.008105	0.004421	-1.833	0.067305 ^
Carson	-0.030154	0.008871	-3.399	0.000726 ***
Cruz	-0.035343	0.007537	-4.689	3.47e-06 ***

Significance codes: *** p< 0.001 * p<0.05 ^ P<0.1

Major Topics Addressed in Candidate Communication

Our final in-depth investigation considers what topics the candidates were choosing to discuss on Twitter. In addition to reflecting what candidates are thinking about, candidates might actively manage what they talk about in an attempt to nudge the political agenda through online communication (Sayre et al. 2010). The idea might be to influence what the public focus upon, what the public think matters to other people (Mutz 1989), or what the media covers (Shapiro and Hemphill 2016).

To do this we first needed to draw up a short list of the major policy topics in the election. We retrieved this list from a New York Times article that outlined what topics the candidates would be discussing.[12] (When using this list, we decided to modify "Syrian refugees" to "Syria" to broaden the scope.) Having a list of eight topics -- Immigration, Gun Control, Climate Change, Syria, Healthcare, Abortion, Death Penalty, and Taxes -- we could then consider which candidates sought to focus more on which specific topics.

We next needed to develop a list of words associated with each topic. Each of the topic words was searched through Twitter, paired with the word "America". The code we used in R searches 2,000 of the latest public tweets in live time that contains both search words. Filtering through the gathered tweets as a bag of words, i.e. just looking at word usage not the context of usage, lists of associated words were created according to their frequency. From this we formed the list of words that would be bundled together for each of the topics. Six to nine words were manually selected for the topics to be searched through the dataset as shown below.

[12] http://www.nytimes.com/interactive/2016/us/elections/candidates-on-the-issues.html, December 15th, 2015, *The New York Times*, Where the Candidates Stand on 2016's Biggest Issues, by Wilson Andrews and Thomas Kaplan, Accessed August 12th, 2016.

Table 7. Topics and Words Associated with each Topic

Topics	**Words Associated with Topic**
Immigration	Immigration, Islamic, Muslim, illegal, immigrant, immigrants, Mexican, border
Gun Control	Gun control, guns, rifle, NRA, shooting, violence, amendment
Climate Change	Climate change, fracking, energy, ice, fuel, gas
Syria	Syria, Syrian refugee, terrorist, terrorism, terrorists, ISIS, crisis, Islam
Health Care	Health care, insurance, Obamacare, health, social, policy
Abortion	Abortion, parenthood, prolife, prochoice, reproductive, right
Death Penalty	Death penalty, capital punishment, innocent, trial, justice, murder
Taxes	Tax, taxes, offshore, evading, business, money, rich, economy, jobs

We recorded mentions of these “associated” words in each candidate’s tweets. The mentions of these words were then grouped into months, then divided by the total words tweeted by each candidate that month. This procedure yielded rough estimates for the focus on a candidate on a specific topic. Figure 7 shows the themes that each candidate focused on.

Figure 7. Candidate Focus on Topics

There are several notable features of this analysis. Both Democrats focused heavily on words associated with the topic of healthcare, this being an important topic to them, and perhaps less so to the Republicans. From both a policy and campaigning standpoint this is interesting. Healthcare is clearly a major policy issue which is reflected in its importance to the candidates. From a political campaigning aspect, it is interesting because the topic is one which the Republicans had campaigned heavily since the introduction of Obamacare. We do not have previous data to compare to but it is possible to conjecture that the Republican focus may be moving away from healthcare this cycle.

We can see a clear difference in focus between the candidates within party. Hillary Clinton seems to have focused more on gun control compared to Bernie Sanders, who spoke more about climate change. Both of these topics are important to followers of the Democratic party but the candidates emphasised them differently. Ben Carson, arguably the most atypical candidate, seemed to be especially willing to focus most on hard to handle, sensitive and contentious topics that were relatively important to the Republican base, i.e. Immigration and Syria.

The discussion of abortion shows the strengths and limitations of this type of analysis when trying to understand the policy views of candidates. Abortion seems to be a relatively important topic across the U.S. political spectrum. We can say that all these candidates used words relevant to the topic of abortion. However, our analysis cannot reveal exactly what the candidate's were saying about abortion., Sanders and Cruz show similar levels of focus on this topic but, given the divisive nature of the topic, they likely have said quite different things which is not picked up in our analysis.

Most notable is Donald Trump's relative avoidance of discussions of abortion. This may have been because he was perceived as somewhat out of step with his party having previously taken positions more accepting of Abortion than the typical modern Republican candidate's policy position. It is also likely that Trump wanted to focus on what he perceived as his strength -- his business background. It appears that although Trump's quantity of tweets related to many of the more divisive topics are small compared to the other candidates, the ones that do get tweeted are controversial and propagated by the media.[13]

Challenges and Further Research

Twitter provides a wonderful repository of data for the researcher. It allows insights into the character of the candidate that only personal contact might have provided in the past. Twitter also records, in easy to access form, the messages sent by campaigns that previously might only have been available after careful study of the text of numerous speeches given at campaign stops all around the country. That said, Twitter is far from a perfect source of data. Traditionally, campaigns may have undervalued their online communications. Any remnant of this that remains will leave the Twitter strategy relatively neglected, suggesting that the communications in this medium may be unrepresentative of the overall strategy – perhaps an intern was doing the posting. This is unlikely to be the case in the major political campaigns that we study, but it may be the case for more minor campaigns such as campaigns for state senator.

---

[13] Trump seemed to perform poorly with Hispanic voters because of, amongst other issues, his views on immigration policy. His attempts to reach out to Hispanic voters using campaign messages were not always successful. http://www.huffingtonpost.com/entry/donald-trump-taco-bowl_us_572bf20be4b096e9f090e1c5, May 6th, 2016, *The Huffington Post*, Donald Trump's Taco Bowl Hilariously Backfires, by Ed Mazza, Accessed August 12th, 2016.

The regression models presented in this paper predicted net valence and anger as the dependent variables, while controlling for the candidate's performance with the Iowa Electronic Markets future prices. While this paper focused on emotions and sentiments, future research could build on the presented topic analysis, and measure the depth of disseminated messages. For an example, campaign performance could impact the information richness of online messages, and future research could shed light on the effectiveness of discussing policy on Twitter.

Twitter is restricted to 140 characters, meaning the medium is poorly suited for detailed policy pronouncements. This is a drawback when studying candidate policy but it also forces campaigns to condense ideas into easily digestible forms. When examining Twitter usage, we see what the campaigns think it is important to communicate with the voters. We can see what campaigns think voters are interested in (Bendle and Cotte 2016) even if we cannot be sure that this is what the voters are actually interested in.

We specifically address two of the more common challenges with analysing Twitter strategy.

Do candidate's write their own tweets?

One question that may be asked is whether candidates write their own tweets. U.S. primary election campaigns are major operations and so one would expect that staff would be employed to tweet on the candidate's behalf. Indeed, the sheer volume of activity of the candidates, for example the regular tweeting of Ted Cruz, suggests that each candidate must receive assistance. It is hard to imagine how Ted Cruz could find the time to do all of his own tweets, even if he wished to. This means tweets from a candidate's account must be approved --

either by formal sign off on each tweet or simply by security limitations placed on who can access the candidate's twitter account -- rather than being actually written by the candidate. While who composes the tweets is an interesting organization question, we would suggest that it doesn't matter too much for our analysis. We are interested in campaign strategy as revealed by Twitter activity. Any staffer composing tweets on behalf of a candidate in effect speaks for the candidate. One may see the writer of tweets as conceptually similar to a speech writer – the candidate still gives a speech despite having staff paid to write the speech. Writers for a candidate on Twitter should seek to adopt the tone and communication strategy of the candidate. The writer should talk about the topics that the candidate wants to discuss and highlight policy positions the candidate wants highlighted. This means that even if a candidate does not write every word, and probably no serious candidate ever does this, the candidate still is responsible for the output. If a writer diverged from the candidate's desired approach, then they would soon be brought back into line. Thus, throughout this research paper we have described, for example, the sentiment attached to Ben Carson's tweets. A lengthier, but more precise form of wording might be: the sentiment in tweets by the campaign under the control of Ben Carson. We maintain the fiction that candidates write their own tweets mainly for ease of exposition.

Is the Online strategy the same as offline?

Candidate online strategies may diverge from their offline strategies, and indeed online the Facebook strategy might diverge from the Twitter strategy. Given our exclusive focus on Twitter, we cannot authoritatively comment on other communications strategies used by the candidates. That said, we suggest that modern U.S. presidential primary campaigns are run by professional staffers who will work on ensuring a clear and coherent message, even though this might be tailored to different platforms. Furthermore, many of the topics we see being embraced in the tweets resonate with the general communication strategy of the candidate -- healthcare was an important policy priority for the Sander’s campaign which was reflected in the Twitter strategy. While research to compare different types of strategy would be useful, our current research is helpful as it allows us to comment on what is now a critical part of candidate communication.

Conclusion

We have sought to provide a detailed overview of the Twitter strategies of the major candidates in the 2016 primary elections using text analysis of the candidates’ tweets. This allowed us a better understanding of the communications undertaken by the candidates. This involved gaining a greater understanding of the basic details of the campaigns, such as number of tweets, and the strategic approaches used, such as self-mentions. We analysed the text for the net valence of the sentiment used in the tweets. We also checked specifically on the anger expressed by the candidates. Finally, we looked at the policy topics addressed in the candidates’ tweets.

Overall, we suggest that not only does Twitter provide candidates a way to communicate with potential voters, it is also an interesting source of data for better understanding campaign communication strategy and the policy topics that the candidates wish to see highlighted.

References

Abramowitz, Alan I (1989), "Viability, Electability, and Candidate Choice in a Presidential Primary Election: A Test of Competing Models," *The Journal of Politics*, 51 (4), 977–92.

Abramson, Paul R, John H Aldrich, Phil Paolino, and David W Rohde (1992), "'Sophisticated' Voting in the 1988 Presidential Primaries," *The American Political Science Review*, 86 (1), 55–69.

Aharony, Noa (2012), "Twitter use by three political leaders: an exploratory analysis," *Online Information Review*, 36 (4), 587–603.

Bendle, Neil and June Cotte (2016), "Assumptions of Rationality in Political Marketing: The Case of the Republican Autopsy," *Journal of Nonprofit & Public Sector Marketing*, 28 (1), 66–83.

Bendle, Neil T and Mihaela-Alina Nastasoiu (2014), "Primary Elections And US Political Marketing," in *Political Marketing In The United States*, J. Lees-Marshment, B. Conley, and K. Cosgrove, eds., New York, NY: Routledge USA, 85–111.

Bendle, Neil T. (2014), "Reference Dependence in Political Primaries," *Journal of Political Marketing*, 13 (4), 307–33.

Bode, Leticia and Kajsa E. Dalrymple (2015), "Politics in 140 Characters or Less: Campaign Communication, Network Interaction, and Political Participation on Twitter," *Journal of Political Marketing*.

Cameron, Michael P., Patrick Barrett, and Bob Stewardson (2015), "Can Social Media Predict Election Results? Evidence From New Zealand," *Journal of Political Marketing*.

Conway, Bethany Anne, Kate Kenski, and Di Wang (2013), "Twitter Use by Presidential Primary Candidates During the 2012 Campaign," *American Behavioral Scientist,* 57 (1596-1610).

Dowd, Maureen (1992), "THE 1992 CAMPAIGN: Candidate's Wife; Hillary Clinton as Aspiring First Lady: Role Model, or a 'Hall Monitor' Type?," *New York Times,* Accessed August 5th 2016.

East, Kristen (2016), "Trump: Tweeting picture of Heidi Cruz was a 'mistake,'" *Politico*, (accessed July 18, 2016), [available at http://www.politico.com/story/2016/04/donald-trump-heidi-cruz-mistake-221499#ixzz4EkqVsa9R].

Enli, Gunn Sara and Eli Skogerbø (2013), "Personalized Campaigns In Party-Centred Politics Twitter And Facebook As Arenas For Political Communication," *Information, Communication & Society*, 16 (5), 757–74.

Graham, Todd, Marcel Broersma, Karin Hazelhoff, and Guido van 't Haar (2013), "Between Broadcasting Political Messages And Interacting With Voters," *Information, Communication & Society*, 16 (5), 692–716.

Heffernan, Virginia (2016), "How the Twitter Candidate Trumped the Teleprompter President," *Politico Magazzine,* (accessed July 18, 2016), [available at http://www.politico.com/magazine/story/2016/04/2016-heffernan-twitter-media-donald-trump-barack-obama-telemprompter-president-213825#ixzz4EkUsIY7P].

Jackson, Nigel and Darren Lilleker (2011), "Microblogging, Constituency Service and Impression Management: UK MPs and the Use of Twitter," *The Journal of Legislative Studies*, 17 (1), 86–105.

Jockers, Matthew (2016a), "syuzhet: Extracts Sentiment and Sentiment-Derived Plot Arcs from Text," *The R-Project*, (accessed August 13, 2016), [available at https://cran.r-project.org/web/packages/syuzhet/index.html].

——— (2016b), "Introduction to the Syuzhet Package," (accessed August 13, 2016), [available at https://cran.r-project.org/web/packages/syuzhet/vignettes/syuzhet-vignette.html].

Knuckey, Jonathan and Jennifer Lees-Marshment (2005), "American Political Marketing: George W. Bush and the Republican Party," in *Political Marketing: A Comparative Perspective*, D. G. Lilleker and J. Lees-Marshment, eds., Manchester, UK: Manchester University Press.

Lees-Marshment, Jennifer (2001a), "The Marriage of Politics and Marketing," *Political Studies*, 49 (4), 692–713.

——— (2001b), *Political marketing and British political parties: the party's just begun*, Manchester ; New York; New York: Manchester University Press; Distributed exclusively in the USA by Palgrave.

——— (2004), *The political marketing revolution: transforming the government of the UK*, Manchester: Manchester University Press.

Lewis, D. D., Y. Yang, T. Rose, and F. Li (2004), "RCV1: A New Benchmark Collection for Text Categorization Research," *Journal of Machine Learning Research*, 5, 361–97.

Lilleker, Darren G and Nigel A. Jackson (2010), "Towards a More Participatory Style of Election Campaigning: The Impact of Web 2.0 on the UK 2010 General Election," *Policy & Internet*, 2 (3), 67–96.

Marland, Alex (2012), "Amateurs versus Professionals: The 1993 and 2006 Canadian Federal Elections," in *Political Marketing in Canada*, A. Marland, T. Giasson, and J. Lees-Marshment, eds., UBC Press, 59–75.

——— (2016), *Brand Command: Canadian Politics and Democracy in the Age of Message Control*, Vancouver, BC: UBC Press.

McGinniss, Joe (1988), *The Selling of the President*, New York, N.Y.: Penguin Books.

Miller, William (2013), "We Can't All Be Obama: The Use of New Media in Modern Political Campaigns," *Journal of Political Marketing*, 12 (4), 326–47.

Mohammad, Saif (n.d.), "NRC Word-Emotion Association Lexicon," (accessed August 13, 2016), [available at http://saifmohammad.com/WebPages/NRC-Emotion-Lexicon.htm].

Mohammad, Saif M. and Peter D. Turney (2013), "Crowdsourcing a word–emotion association lexicon.," *Computational Intelligence*, 29 (3), 436–65.

Mutz, D.C. (1989), "The Influence of Perceptions of Media Influence: Third Person Effects and the Public Expression of Opinions," *International Journal of Public Opinion Research*, 1

(1), 3–23.

O'Cass, Aron (1996), "Political marketing and the marketing concept," *European Journal of Marketing*, 30 (10/11), 37–53.

O'Shaughnessy, Nicholas (2001), "The marketing of political marketing," *European Journal of Marketing*, 35 (9/10), 1047–57.

Parsons, Michael and Martyn Rowling (2015), "Social Media and the Paradox of Fear: An Exploratory Study of Political Relationship Marketing Within South Wales," *Journal of Political Marketing*.

Savigny, Heather (2008), *The problem of political marketing*, New York: Continuum International Publishing.

Sayre, Ben, Leticia Bode, Dhavan Shah, Dave Wilcox, and Chirag Shah (2010), "Agenda Setting in a Digital Age: Tracking Attention to California Proposition 8 in Social Media, Online News and Conventional News," *Policy & Internet*, 2 (2), 7–32.

Scammell, M (1999), "Political marketing: lessons for political science," *Political studies*.

Shapiro, Matthew A. and Libby Hemphill (2016), "Politicians and the Policy Agenda: Does Use of Twitter by the U.S. Congress Direct New York Times Content?," *Policy & Internet*, Online 4 A.

Vergeer, Maurice, Liesbeth Hermans, and Steven Sams (2011), "Online social networks and micro-blogging in political campaigning: The exploration of a new campaign tool and a new campaign style," *Party Politics*, 19 (3), 477–501.