Nasty, Brutish, and Short:
What Makes Election News Popular on Twitter?

Anonymous

Abstract

Lorem ipsum dolor. Sit amet mauris. Sed ut eu aliquam integer nibh vitae adipiscing diam mauris erat est. Libero amet parturient. Nulla vehicula dui elit cras ante. Lacus imperdiet scelerisque felis bibendum dolorem maecenas amet quis. Magna ultricies faucibus esse convallis quis. Leo pretium habitasse. Nam cursus aptent bibendum nulla per conubia tellus tincidunt. Eget tellus sed ut vel erat. Lorem ipsum dolor. Sit amet mauris. Sed ut eu aliquam integer nibh vitae adipiscing diam mauris erat est. Libero amet parturient. Nulla vehicula dui elit cras ante. Lacus imperdiet scelerisque felis bibendum dolorem maecenas amet quis. Magna ultricies faucibus esse convallis quis. Leo pretium habitasse. Nam cursus aptent bibendum nulla per conubia tellus tincidunt. Eget tellus sed ut vel erat. Lorem ipsum dolor. Sit amet mauris. Sed ut eu aliquam integer nibh vitae adipiscing diam mauris erat est. Libero amet parturient. Nulla vehicula dui elit cras ante. Lacus imperdiet scelerisque felis bibendum dolorem maecenas amet quis. Magna ultricies faucibus esse convallis quis. Leo pretium habitasse. Nam cursus aptent bibendum nulla per conubia tellus tincidunt. Eget tellus sed ut vel erat.

Introduction

In the changing landscape of both journalism and politics, social media is playing an increasingly large role in mobilizing and spreading information to citizens. A Pew Research survey from August 2015 showed that nearly two-thirds of adults in the U.S. who are on Twitter use the platform to get news (Bartherl et al. 2015).

In 2008, President Barack Obama's winning campaign was largely attributed to the success of his social media strategy, the first example of its kind. By establishing an online presence that recruited more than 3 million individual contributors and 5 million volunteers, Obama was able to create a grassroots political movement (Cogburn and Espinoza-Vasquez 2011). Publicity and public sound bites matter– especially when its free and has the potential to go viral.

Although social media messages are less able to be carefully controlled in comparison to paid advertisements, they also have the potential to reach a wider audience. Reactions to an article shared by one potential voter now have the ability to be broadcast and spread to millions of others in a real-time, public sphere.

The popularity of sharing articles on social media also marks an important shift in the role of the news consumer from armchair reader to information propagator. Whereas news used to be broadcast to the reader, now each reader has the potential to broadcast stories to his or her own audience. Sharing a story requires a level of interest and activation on the part of the reader beyond simply reading a story; yet often, this trigger is predictably emotional in nature.

In a 2011 study of the New York Times "most emailed list", Berger and Milkman found that the potential for a news story to go viral is partially driven by physiological arousal, defined as "an excitatory state of sensory alertness, mobilization, or energy" (Berger and Milkman 2012). In short, the desire to share a certain story is often universally impulsive, regardless of context. Yet in the case of political news, this impulse can have a large impact on reach of political messages, an impact that is not always equally distributed.

During the 2016 election year, the New York Times published an article estimating a 2 billion-dollar advantage in free media, including social media, for Donald Trump, all of which has no small impact on the messages broadcast to voters (Confessore and Yourish 2016).

Hypotheses

With the important role of social media in politics and the often emotionally charged nature of content-sharing in mind, we ask the following question in our research:

- Does the emotional vocabulary of political news stories have an impact on its Twitter popularity that persists beyond political affiliation?

To test this question, we focus on three key aspects of stories: length, emotionality, and positivity. We choose these three aspects based on behavioral theories of the Internet and studies of political news, detailed in the section below.

We hypothesize the following behavior in our dataset of stories and tweets:

- **H1:** Story length has a *negative* correlation with Twitter shares, due to the effects of the Internet attention economy and overexposure to political media (Goldhaber 1997).
- **H2:** Emotionality has a *positive* correlation with Twitter shares, consistent for viral content in general (Berger and Milkman 2012).

- **H3:** Positivity has a *negative* correlation with Twitter shares, due to the nature of political news and contrary to generalized findings (Berger and Milkman 2012).

For each of these three independent variables (story length, emotionality, positivity) we repeat analyses across three views of the data: first, the entire dataset; then, by political candidate followed amongst users who follow only one candidate; and finally, by the number of political candidates followed (degree of political engagement), to look for differences amongst different populations of political tweeters.

Literature Review

The (Short) Attention Economy

Although social media has the power to create a flood of free advertising and media for political candidates, the abundance of information on the web has created new challenges and questions about the kind of content being processed by readers. This paradox– between the ease of accessibility to information and the increasingly limited bandwidth of consumers– is described as one of the challenges of being in an *attention economy* (Goldhaber 1997). Moreover, high-impact events like the presidential elections especially intensify this effect– about 60 % of Americans reported feeling exhausted by media coverage of the elections in July of 2016 (Gottfried 2016). To explore the effects of the attention economy on the reading of political news, we examine story length and how it relates to sharing popularity in the analysis to follow.

Negativity in Politics and Online

In addition, the option of anonymity and pseudo-anonymity on a social network like Twitter (along with other traits of Internet communication), is theorized to contribute to increased negative and hostile behavior, potentially increasing tension for the already-fraught subject of politics. This phenomenon is coined as the *online disinhibition effect* (Suler 2004).

In Berger and Milkmans study of story virality, it was found that *positive* content was more likely to be shared than negative content– against conventional belief (Berger and Milkman 2012). Political news, however, is a unique category of news, and this election in particular– where one-in-four Americans report disliking the presidential candidates– appears to have a negative overtone. In a study of responses to the 2012 election campaign on Twitter, it was found that for both candidates, the majority of tweets were far more negative in tone than positive (Mitchell and Hitlin 2013).

To compare the sharing of election news stories versus patterns of general virality in the news, and to examine the extent in which negative sentiment is popular, we calculate the *negativity* of stories, and how that relates to Twitter behavior.

Emotionality

We also examine the effects of the degree of combined emotionality in the content and how that relates to Twitter shares, to see if either more positive or more negative content is more likely to be shared overall than content that ranks low in emotionality. Although positive content was found to be more popular than negative content in the sharing of stories, both highly positive and highly negative content was more likely to become viral, and we expect the same to hold for political news (Berger and Milkman 2012).

Data Collection

Our main dataset is a connected corpus of news articles about the presidential elections and the tweets that share them from January 1, 2016 to May 1, 2016 over 13 news outlets.

News Dataset

For our news dataset, we scraped articles from the RSS feeds of news publications every hour over five months and 13 publications:

- CNN
- Fox News
- The New York Times
- The Wall Street Journal
- The Washington Post
- The Los Angeles Times
- The Associated Press
- Reuters
- McClatchy
- Politico
- Buzzfeed
- The Huffington Post
- NPR

The choices above span a mix of publications. We include sources that:

- Have mostly conservative audiences and mostly liberal audiences, based on a 2014 survey (Mitchell et al. 2014)
- Come from mixed primary media formats (television, paper, online, radio)
- Are viewed as “legacy” (over a hundred years old) and “new” media (founded online within the last 10 years)
- Focus solely on political news (Politico, McClatchy)
- Are newswire services (the Associated Press, Reuters news)

These choices capture a variety of types of election coverage and target audiences.

We look at stories from January 1, 2016 (the start of the election year) to May 1, 2016. This time period captures the bulk of the primary election, when coverage of multiple presidential candidate contenders creates greater variety in news stories for our analysis.

Articles are processed in a 3-step pipeline. After collecting the links to the full content of the news stories from each publication’s RSS feed, we pass each link to a structured content parser that extracts entities and features from the raw HTML. The story text is then passed into a binary classifier for election news, which employs a MaxEnt classifier and performs with a F-score of 0.90 (Vijayaraghavan 2016).

We collect and classify a total of 22,959 articles as election-related with over 80% confidence, an average of 5,700 per month and 191 per day.

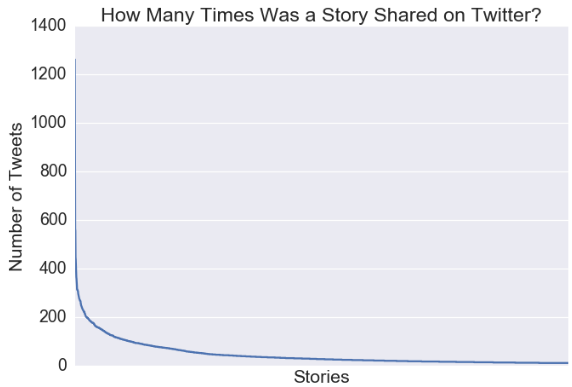

Figure 1: Distribution of Story Shares

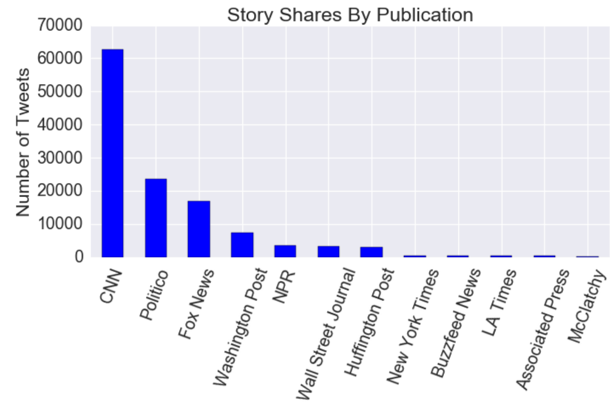

Figure 2: Number of story shares by publication

Tweets Dataset

We start with the firehose of all tweets between January 1, 2016 and May 1, 2016 in our data process.

Tweets pass through a similiar pipeline as news stories. First, we sort all tweets with an election classifier which has been shown to be able to detect election-related tweets with an F-score of 92% (Vijayaraghavan, Vosoughi, and Roy 2016). We then filter by those that share a link (which might potentially be a news story).

We collect and sort a total of 16,667,685 tweets as election-related and containing at least one URL in the text, an average of 4,000,000 per month and 140,000 per day.

Combined Dataset

The final step of our data collection process is to extract, expand and connect the links shared in our election-related tweets with articles in our database.

Twitter automatically formats all links into a shortened "t.co" format, so we first expand all links in tweets (16.6 million), then use regular expressions to see if the final destination of the expanded link matches a query-truncated URL of a story in our database. We checked the validity of 382 billion url-story matches in less than a day by running the processes on the Amazon Web Services cloud computing platform in parallel using the Gnu-parallel command line tool (Tange and others 2011).

Final Corpus

In total, we found that 30% of the election stories we tracked were shared on Twitter during the time period of January 1st through May 1st. There were 137,986 tweets that contained a link to 6,911 unique stories (out of 22,960). Since we chose the story to be the unit of analysis in this thesis, we then eliminated any stories that were shared by less than 10 tweets.

This left a total of 2,650 distinct articles (38%) shared in 123,113 (89%) tweets by 20,956 Twitter users (93%).

Descriptive Findings The vast majority of stories are shared less than 100 times.

Article	# tweets
The One Weird Trait That Predicts Whether You're a Trump Supporter	1260
Donald Trump Is Shocking, Vulgar and Right	901
Biden praises Sanders on income inequality	563
Why I'm voting for Trump	554
Anne Frank's stepsister compares Donald Trump to Adolf Hitler	445
Trump basks in his spotlight	436
Rubio: Law-abiding undocumented immigrants could stay	413
Terrorists use Trump's 'Muslim ban' speech in recruitment video	398
Iowa caucuses: Donald Trump's moment of truth	364
GOP senators: If Cruz wins, we lose	357

Figure 3: Top 10 Most Shared Articles

Story sharing behavior follows an approximate power law distribution. On average, stories are shared 46 times, however, the median (50th percentile) of shares is just 26.

CNN, Politico, and Fox lead in publication popularity with the highest number of stories shared by tweets in our dataset– likely due to the volume and close association to political content of the companies. Because our data pipeline detailed in the previous chapter looks for election-related news, outlets which focus on political news are more likely to show up in our results.

Examining the top 10 most shared stories, it comes as no surprise that outsized personality Donald Trump is by far the most "tweetable" candidate, dominating the list with 7 out of 10 stories featuring his name in the title.

The extent to which he is prominent in articles is clear: by coding each story by the most frequently mentioned candidate mentioned in the body of the text, Trump has nearly three times as much coverage at nearly 60% as the runner-ups, Ted Cruz and Hillary Clinton.

The large number of stories with Cruz as the most-mentioned candidate are likely due to his association to Trump as a Republican runner-up: 96% of stories where Cruz is the most-mentioned candidate feature Trump as the second-most frequently occuring.

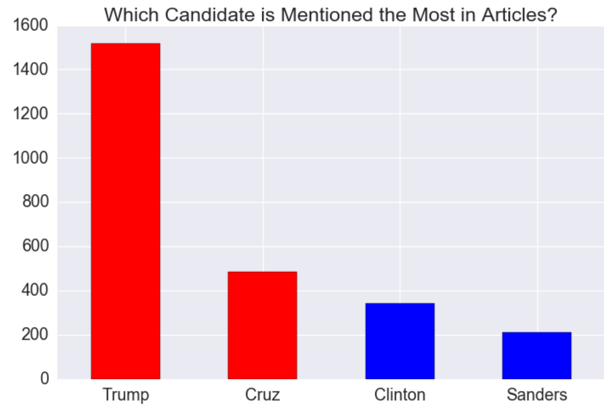

Figure 4: Most frequently mentioned candidate in stories

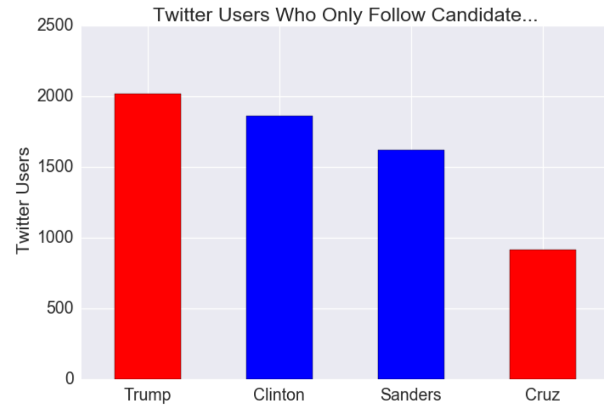

Figure 5: Number of Tweets by Each Segment

Measuring Political and Emotional Engagement

In the following sections, we discuss the tools and methods that we use to analyze our independent variables.

Emotional Coding

For the emotional coding of news articles, we use dictionaries from the Harvard General Inquirer, a lexicon that is popular for computerized content analysis (Stone and Hunt 1963). In Berger and Milkmans study of online virality, automated coding using the LIWC system showed results that were significantly positively correlated with the output of manual coding (Berger and Milkman 2012). The Inquirer is a public-use alternative to the LIWC system.

In particular, we use the *Positiv* and *Negativ* collections, a set of 1,915 well-established words signifying positive outlook (not including words for *yes*) and 2,291 words signifying negative outlook (not including words for *no*), respectively. Repeating the same metrics from *What Makes Online Content Viral?*, we quantify for each document:

$$emotionality = \frac{count(positiv \mid negativ)}{count(words)}$$

$$positivity = \frac{count(positiv)}{count(words)} - \frac{count(negativ)}{count(words)}$$

as independent variables in our analysis.

Followership as a Proxy for Political Engagement

In the following sections, we use Twitter followership as a proxy for measuring degrees of political engagement.

Previous research in network analysis and attempts to predict latent political affiliations of users in the social network has shown that users on Twitter tend to show network homophily within political groups, and that "like follows like". In addition, followership of only Democratic or only Republican official accounts can be used as a reasonable estimator of party loyalty. Those accounts that follow only the officials of one party tend to demonstrate more closeness with other users in their political party than those who do not (Colleoni, Rozza, and Arvidsson 2014).

Due to the highly individual nature of this election, where candidate loyalty does not necessarily imply goodwill towards the party, we look specifically at what candidates users follow.

For general *levels* of political engagement, we look at the number of political candidates a Twitter user follows as a proxy for how likely they are to share political news.

In addition, for single-candidate Tweeters, we divide users by the candidate they follow. At the time of data collection completion (May 1, 2016), the top two candidates by delegate count in each party were Hillary Clinton (D), Bernie Sanders (D) and Donald Trump (R) and Ted Cruz (R), so we split users into these four groups.

Candidate Followership Our dataset contains 6,406 unique single-candidate Twitter users. Trump-only followers lead with about 31%, followed closely by Clinton-only (29%), then Sanders (25%) and Cruz (14%).

Trump's free media advantage becomes clear when looking at the *volume* of tweets each group of users tweet: 37% of tweets sharing articles come from Trump-only followers versus 27% for Clinton-only, 20% for Sanders-only, and 14.6% for Cruz.

We also observe the nature of the content being shared by each group. Again, across all four segments, Republican candidate Trump leads the top number of mentions in stories shared.

Political Engagement

36% of Twitter users in our dataset follow none of the four political candidates, followed by 31% who follow one candidate, 20% who follow two, and 13% who follow three or more candidates.

We see a negative curvlinear relationship between the number of candidates followed (level of observed political engagement) and the ratio of political news tweets per user.

In the following analyses, we segment levels of political engagement into three groups for the sake of comparison:

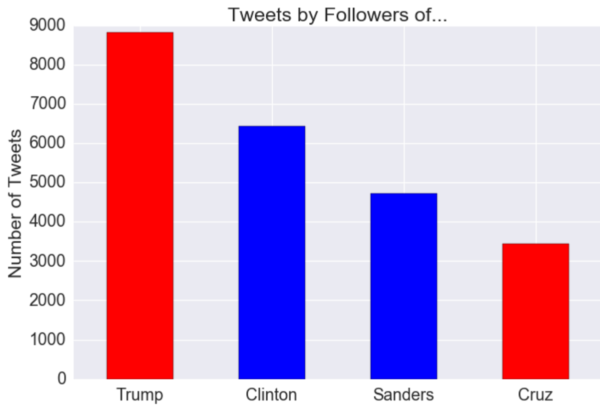

Figure 6: Most frequently mentioned candidate in stories

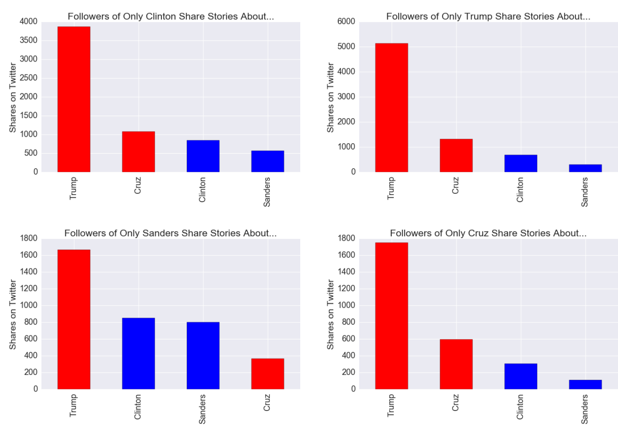

Figure 7: Tweet and User Counts by Followership

Figure 8: Number of candidates followed

Figure 9: Ratio of Tweets to political candidates followed

- *the unaffiliated* (those who follow no presidential candidates, but do tweet about political news)
- *single-candidate* Tweeters (those who follow one and only one presidential candidate, and tweet about political news)
- *political aficionados* (those who follow all 4 (or more) candidates, and tweet about political news).

Analysis

For each of these three independent variables (story length, emotionality, positivity) we repeat analyses across three views of the data: first, the entire dataset; then, by political candidate followed amongst users who follow only one candidate; and finally, by the number of political candidates followed (degree of political engagement), to look for differences amongst different populations of political tweeters.

	Dependent variable:		
	Number of Tweets		
	OLS	*Poisson*	*negative binomial*
	(1)	(2)	(3)
Story length	-0.004^{***}	-0.0001^{***}	-0.0001^{***}
	(0.001)	(0.00000)	(0.00002)
Constant	50.426^{***}	3.936^{***}	3.895^{***}
	(1.819)	(0.005)	(0.027)
Observations	2,650	2,650	2,650
R^2	0.003		
Adjusted R^2	0.003		
Log Likelihood		−65,042.390	−12,778.740
θ			1.346^{***} (0.034)
Akaike Inf. Crit.		130,088.800	25,561.490
Residual Std. Error	60.156 (df = 2648)		
F Statistic	8.452^{***} (df = 1; 2648)		
Note:			*p<0.1; **p<0.05; ***p<0.01

Figure 10: Tweet Volume vs. Story Length, All Tweets

Methodology

Since our dependent variable, tweet volume, is a set of discrete counts that are positively truncated, we use negative binomial regression models for our analysis (Scott Long 1997). The distribution of tweet volume is not a normal distribution, and it is not recommended to perform a log transformation on count data to fit it to an OLS regression unless there is little dispersion in the data (Ohara and Kotze 2010). Poisson models are a subset of negative binomial models without the dispersion parameter. In our analyses, we see that the negative binomial model provides the best fit and that our data is overdispersed, as the dispersion parameter θ is greater than 1.

In each case, we compare our findings to those using linear and Poisson regression models, and are able to achieve the same significant results.

All Data

Story Length Overall, we find a consistently significant negative correlation between story length and Twitter volume ($\beta = -0.0001, p < 0.01$). This aligns with our hypothesis **H1:** that shorter stories are more likely to be shared, due to competing resources in the attention economy.

Emotionality Overall, we find a consistently significant positive correlation between emotionality and Twitter volume ($\beta = 6.019, p < 0.01$).

This confirms **H2:** that emotionality has a *positive* correlation with Twitter shares, consistent for viral content in general (Berger and Milkman 2012).

Positivity Overall, we find a consistently significant negative correlation between positivity and Twitter volume ($\beta = 5.391, p < 0.01$).

This finding supports **H3:** Positivity has a *negative* correlation with Twitter shares, due to the nature of political news

	Dependent variable:		
	Number of Tweets		
	OLS	*Poisson*	*negative binomial*
	(1)	(2)	(3)
Emotionality	305.229**	6.111***	6.019***
	(122.427)	(0.276)	(1.777)
Constant	40.349***	3.714***	3.716***
	(2.685)	(0.006)	(0.039)
Observations	2,650	2,650	2,650
R^2	0.002		
Adjusted R^2	0.002		
Log Likelihood		−65,180.470	−12,779.510
θ			1.345*** (0.034)
Akaike Inf. Crit.		130,364.900	25,563.030
Residual Std. Error	60.181 (df = 2648)		
F Statistic	6.216** (df = 1; 2648)		
Note:			*p<0.1; **p<0.05; ***p<0.01

Figure 11: Tweet Volume vs. Story Emotionality, All Tweets

	Dependent variable:		
	Number of Tweets		
	OLS	*Poisson*	*negative binomial*
	(1)	(2)	(3)
Story length	−0.004***	−0.0001***	−0.0001***
	(0.001)	(0.00000)	(0.00002)
Constant	50.426***	3.936***	3.895***
	(1.819)	(0.005)	(0.027)
Observations	2,650	2,650	2,650
R^2	0.003		
Adjusted R^2	0.003		
Log Likelihood		−65,042.390	−12,778.740
θ			1.346*** (0.034)
Akaike Inf. Crit.		130,088.800	25,561.490
Residual Std. Error	60.156 (df = 2648)		
F Statistic	8.452*** (df = 1; 2648)		
Note:			*p<0.1; **p<0.05; ***p<0.01

Figure 12: Tweet Volume vs. Positivity, All Tweets

and contrary to generalized findings (Berger and Milkman 2012).

By Degree of Political Engagement

Next, we segment levels of political engagement into three groups for the sake of comparison:

- *The unaffiliated* (those who follow no presidential candidates, but do tweet about political news)
- *Single-candidate* followers (those who follow one and only one presidential candidate, and tweet about political news)
- *Political aficionados* (those who follow all 4 (or more) candidates, and tweet about political news)

We repeat the same methods and variables in determining our correlations. Again, we test all three models (OLS, Poisson, and NB) for consistency and report the results of the negative binomial model.

Overall, we find that:

Unaffiliated tweeters show the same patterns as the general dataset with a negative correlation between story length and Twitter shares ($\beta = -0.003$, $p < 0.01$), positive correlation between emotionality and Twitter shares ($\beta = 7.427$, $p < 0.01$), and a negative correlation between positivity and Twitter shares ($\beta = -4.036$, $p < 0.01$).

Single-candidate followers, on the other hand, show a slight positive correlation between story length and number of Twitter shares ($\beta = 0.0002$, $p < 0.01$). We hypothesize that if following a single candidate can serve as a proxy for candidate loyalty, then perhaps the correlation signifies a willingness to read and share more complex content on behalf of the candidate and a deeper degree of political involvement.

We see the same effects for the *political aficionados* group, again, a small but significant positive correlation between story length and number of tweets ($\beta = 0.0003$, $p < 0.01$). Again, this suggests a potential difference in levels of engagement with political news.

By Candidate

We divide Twitter users into four segments: Trump-only, Clinton-only, Sanders-only, and Cruz-only followers and repeat the same regressions within each population. Again, for the sake of brevity, all tables and comparisons of all three models (OLS, Poisson, NB model) can be found in Appendix A, tables A.10 - A.21.

Overall, we were unable to find a significant difference in direction of correlations between any of the three independent variables between candidate groups. However, we did find differences in magnitudes of the coefficients; most notably a strong negative correlation between postivity and number of tweets for Trump-only followers, almost threefold in magnitude ($\beta = -14.692$, $p < 0.01$ compared to the dataset at large ($\beta = -5.391$, $p < 0.01$).

As discussed in the section above, all single-candidate followers that showed a significant correlation between story length and number of Twitter shares showed a slight positive correlation.

Conclusion

We find that, on the whole:

- Shorter stories are more likely to be shared on Twitter (**H1**)
- Stories high in emotional words, both negative and positive, are more likely to be shared on Twitter (**H2**)
- Stories that are less positive are more likely to be shared on Twitter (**H3**)

These results confirm our expectations of reader attention on the Internet, the emotional nature of content virality, and the negative connotation of political media.

We were able to see small but significant differences in the number of political candidates a user followed and the length of the stories that were likely to be shared, suggesting differing degrees of political involvement. However, we were unable to find many significant differences between segments by specific political candidates that users followed.

This suggests that either the characteristics of political news we examine (length, emotionality, positivity) are universal in their effects on motivating readers to share articles, or that our methods of dividing Twitter users do not reveal significant underlying differences between readers.

Limitations & Future Work

Although we take a "big data" approach to our analysis, our dataset is by no means a complete mapping of *all* political news sharing activity on Twitter. Our sources are limited to a small but diverse set of publications, which are by no means representative of all news outlets.

Furthermore, we match tweets with stories optimizing for efficiency and speed by using regular expressions, rather than completeness.

Potential paths of future research include:

- Expanding the set of publications tracked
- Replicating the analysis on a real-time basis
- Using machine learning methods to match tweets with stories in a more intelligent way
- Analyzing more nuanced emotive words in the text, in addition to positive and negative words
- Looking at additional signals in the Twitter data, such as user characteristics of the sharer and network aspects of stories being shared
- Segmenting Twitter users in different ways aside from candidates followed

Still, this analysis provides a first view of article sharing on Twitter in a unique and eventful election year with large responses on social media.

Table 1:

	Dependent variable:
	num_tweets
log(wc)	-0.088^{***}
	(0.017)
emotionality	6.279^{***}
	(1.770)
positivity	-5.978^{***}
	(2.013)
Constant	4.296^{***}
	(0.117)
Observations	2,650
Log Likelihood	−12,761.660
θ	1.361^{***} (0.035)
Akaike Inf. Crit.	25,531.320
Note:	*p<0.1; **p<0.05; ***p<0.01

References

[Bartherl et al. 2015] Bartherl, M.; Shearer, E.; Gottfried, J.; and Mitchell, A. 2015. How do americans use twitter for news? Retrieved from `http://www.pewresearch.org`. (Accessed on 07/12/2016).

[Berger and Milkman 2012] Berger, J., and Milkman, K. L. 2012. What makes online content viral? *Journal of marketing research* 49(2):192–205.

[Cogburn and Espinoza-Vasquez 2011] Cogburn, D. L., and Espinoza-Vasquez, F. K. 2011. From networked nominee to networked nation: Examining the impact of web 2.0 and social media on political participation and civic engagement in the 2008 obama campaign. *Journal of Political Marketing* 10(1-2):189–213.

[Colleoni, Rozza, and Arvidsson 2014] Colleoni, E.; Rozza, A.; and Arvidsson, A. 2014. Echo chamber or public sphere? predicting political orientation and measuring political homophily in twitter using big data. *Journal of Communication* 64(2):317–332.

[Confessore and Yourish 2016] Confessore, N., and Yourish, K. 2016. $2 billion worth of free media for donald trump. Retrieved from `http://www.nytimes.com`. (Accessed on 07/12/2016).

[Goldhaber 1997] Goldhaber, M. H. 1997. The attention economy and the net. *First Monday* 2(4).

[Gottfried 2016] Gottfried, J. 2016. Most americans already feel election coverage fatigue. Retrieved from `http://www.pewresearch.org`. (Accessed on 07/16/2016).

[Mitchell and Hitlin 2013] Mitchell, A., and Hitlin, P. 2013. Twitter reaction to events often at odds with overall public opinion. *Pew Research Center* 4.

[Mitchell et al. 2014] Mitchell, A.; Gottfried, J.; Kiley, J.; and Matsa, K. E. 2014. Political polarization & media habits. *Pew Research Center*. (Accessed on 04/02/2016).

[Ohara and Kotze 2010] Ohara, R. B., and Kotze, D. J. 2010. Do not log-transform count data. *Methods in Ecology and Evolution* 1(2):118–122.

[Scott Long 1997] Scott Long, J. 1997. Regression models for categorical and limited dependent variables. *Advanced quantitative techniques in the social sciences* 7.

[Stone and Hunt 1963] Stone, P. J., and Hunt, E. B. 1963. A computer approach to content analysis: studies using the general inquirer system. In *Proceedings of the May 21-23, 1963, spring joint computer conference*, 241–256. ACM.

[Suler 2004] Suler, J. 2004. The online disinhibition effect. *Cyberpsychology & behavior* 7(3):321–326.

[Tange and others 2011] Tange, O., et al. 2011. Gnu parallel-the command-line power tool. *The USENIX Magazine* 36(1):42–47.

[Vijayaraghavan, Vosoughi, and Roy 2016] Vijayaraghavan, P.; Vosoughi, S.; and Roy, D. 2016. Automatic detection and categorization of election-related tweets. In *10th International AAAI Conference on Web and Social Media*.

[Vijayaraghavan 2016] Vijayaraghavan, P. 2016. Automatic Identification of Representative Content on Twitter. Master's thesis, Massachusetts Institute of Technology.