A Critical Evaluation of Tracking Surveys with Social Media: A Case Study in Presidential Approval

Robyn A. Ferg[1]

Frederick G. Conrad[2]

Johann A. Gagnon-Bartsch[1]

[1]University of Michigan Department of Statistics

[2]University of Michigan Program in Survey Methodology

Abstract

Relationships found between public opinion polls and data extracted from social media have led to optimism about supplementing traditional surveys with these new sources of data. However, many initial findings have not been met with usual levels of scrutiny and skepticism. Our goal is to introduce a higher level of scrutiny to these types of analyses. We focus on presidential approval, because we believe signals relating to politics might be some of the strongest in social media, and provide an illuminating test case. Our first contribution is to develop a framework to interpret the strength of relationships found between public opinion poll surveys and tweets containing a given keyword. Following methods that exist in the literature, we measure the association between survey based measures of presidential approval and tweets containing the word "Trump." We then implement placebo analyses, in which we perform the same analysis as with the "Trump" tweets but with tweets unrelated to presidential approval, and we conclude that the relationship between "Trump" tweets and public opinion polls is not strong. As our second contribution, we suggest following social media users longitudinally. For a set of politically active Twitter users, we classify users as a Democrat or Republican and find evidence of a political signal in terms of frequency and sentiment of their tweets around the 2016 presidential election. However, even in this best-case scenario of focusing exclusively on politics and following users who are politically engaged, the signal found is relatively weak. For the goal of supplementing traditional surveys with data extracted from social media, these results are encouraging, but cautionary.

Keywords: social media, Twitter, surveys, big data and surveys, Twitter sentiment, presidential approval

1 Introduction

Surveys are critical for understanding public opinion and setting public policy. While asking survey questions to samples designed to represent the entire population has been very successful for many years, surveys are becoming increasingly costly to perform and response rates are declining (e.g. de Leeuw and de Heer (2002)). One proposed alternative to traditional surveys, as laid out by the AAPOR task force on big data (Murphy, et al., 2014), is to use data gathered from social media to supplement or in some cases replace traditional surveys (Hsieh & Murphy, 2017).

Early analyses were promising, finding high correlations when tracking public opinion surveys with tweets containing a given keyword. For example, O'Connor, Balasubramanyan, Routledge, & Smith (2010) found high correlations between sentiment of tweets from 2008-2009 containing the word "jobs" and survey-based measures of consumer confidence, as well as a high correlation between the sentiment of tweets from 2009 containing the word "Obama" and survey-based measures of presidential approval. Cody, Reagan, Dodds, & Danforth (2016) found similar correlations using more recent tweets through 2015. Daas & Puts (2014) found high correlations between sentiment of various subsets of Dutch social media messages and consumer confidence in the Netherlands. These findings suggest there may be an underlying relationship between data extracted from social media and public opinion surveys.

However, inconsistencies in these initial analyses warrant skepticism in underlying relationships between social media data and survey responses. In O'Connor et al. (2010), a high correlation is observed between Obama's standing in 2008 presidential election polls and the frequency---but not sentiment---of "Obama" tweets. Surprisingly, however, O'Connor et al. (2010) also found a positive correlation between Obama's standing in election polls and the frequency of tweets that contain the word "McCain". O'Connor et al. (2010) did not find a relationship between "job" (as opposed to "jobs") tweets or "economy" tweets and consumer confidence, raising concerns about the robustness of the findings. Further confusing this issue, Cody et al. (2016) did find a relationship between "job" tweets and consumer confidence, resulting in a set of subtly contradictory findings. Daas & Puts (2014) found correlations between Dutch consumer sentiment and various subsets of Dutch social media messages (such as messages containing pronouns, messages containing the most frequent spoken and written words in Dutch, and messages containing the Dutch equivalents of "the" and "a/an") that were just as strong as messages containing words about the economy, raising red flags for whether the economic tweets were truly capturing consumer confidence.

Upon further analysis, the initial relationships that appear strong between Twitter data and public opinion surveys can easily fall apart. Conrad, Gagnon-Bartsch, Ferg, Schober, Pasek, & Hou (2019, online) further investigated the relationship between sentiment of "jobs" tweets and consumer confidence, finding that seemingly small changes in sentiment calculation can drastically change the strength of the resulting relationship. Neither sorting "jobs" tweets into various categories (e.g. news/politics, job advertisements) (Conrad, et al., 2019) nor weighting survey responses to reflect the population of Twitter users (Pasek, Yan, Conrad, Newport, & Marken, 2018) restored the relationship. Furthermore, correlations between sentiment of "jobs" tweets and consumer confidence were found to be unstable over time (Conrad, et al., 2019; Pasek, et al., 2018). Conrad et. al. concluded that correlations between consumer confidence and sentiment of "jobs" tweets as reported in O'Connor et al. were likely spurious.

With the benefit of hindsight, it is perhaps not surprising that public opinion for select topics, such as the economy, can be difficult to obtain from social media. For example, even if a user's "jobs" tweet is about the economy (as opposed to, for example, Steve Jobs), the user's opinion about the economy is not always clear from the tweet. Tweets about politics, on the other hand, are often quite clear with regard to who or what a user supports or opposes. Therefore, if there is a strong, reliable signal present in Twitter that might be used to supplement traditional surveys, we might reasonably expect to find it in the political realm. In addition, there is some evidence that non-probability online survey panels produce plausible estimates of Americans' political affiliation and ideology, despite very different sampling practices. Kennedy, Mercer, Keeter, Hatley, McGeeney, & Gimenez (2016) compared the estimates of political affiliation and ideology derived from responses to a questionnaire administered to samples from nine non-probability panels. All told essentially the same story about political affiliation (all somewhat overestimated the proportion of Democrats and somewhat underestimated the proportion of Independents) and ideology (Democrats were likely favor a government that does more, within seven points of a gold standard based on telephone surveys of representative samples, and Republicans were likely to believe the government does too many things, within eight points of the gold standard). For these reasons, we focus our attention in this paper on tracking presidential approval, which we regard as "best-case scenario" for the goal of using social media data to supplement traditional surveys.

There are two main contributions in this paper. Our first contribution is methodological. If social media are to be reliably used to track public opinion, there needs to be a method of evaluating the strength of associations between social media data and public opinion surveys. While the results of Conrad et al. (2019) and Pasek et al. (2018) cast doubt on the credibility of previously observed relationships between Twitter sentiment and public opinion surveys, there remains a need for a systematic framework to interpret the strength of such relationships. To address this we propose the use of *placebo analyses*. The idea behind a placebo analysis is to replicate the primary analysis but using variables that are known to have no true relationship with the response. As an example of a placebo analysis, DiNardo & Pischke (1996) revisited a previous study that claimed wage differentials were due to computer use in the workplace. When replacing the variable for computer use in the analysis with pen/pencil use, the estimated effect of pencil use on wage differentials was similar to the estimated effect of computer use. This casts doubt on the original claim that computers in the workplace were causing the wage differential since the true effect for the placebo variable (pencil use) should be zero. The implication of an estimated non-zero effect is that the original analysis was not credible, see Athey & Imbens (2017) for further details. We develop a framework to evaluate and interpret the strength of observed correlations between social media sentiment and public opinion surveys by essentially performing multiple placebo tests. In the context of presidential approval, we first calculate the correlation between survey-based measures of presidential approval and the sentiment of tweets that contain the word "Trump". In doing so, however, we adjust smoothing and lag parameters to obtain the best possible correlation, as is typically done in similar analyses (O'Connor et al. 2010, Conrad et al. 2019). Because we optimize over these parameters, it is difficult to interpret the strength of the resulting correlation. We therefore compare our observed correlation to other correlations that are calculated in a similar way, but which are assumed to be spurious. Using this framework, we conclude that while there may be a signal when tracking sentiment of tweets containing the word "Trump", it is small and not obviously useful. These results cast doubt on whether Twitter data can reliably be used as a replacement for traditional surveys.

Our second contribution deals with the method in which social media data are obtained. As an alternative to the commonly used method of simply collecting tweets that contain a given keyword (e.g., "Trump") irrespective of who is posting them, we propose following a set of politically active Twitter users over time. This method of collecting tweets is similar to Golder & Macy (2011), who tracked mood using up to 400 tweets for each of millions of users. By collecting tweets in this manner we can track changes in sentiment among a fixed set of users. We classify politically active Twitter users as a Democrat or Republican and find evidence of a political signal when tracking both the frequency and sentiment of these users' tweets around the 2016 U.S. presidential election.

2. Relationship between "Trump" Tweets and Presidential Approval

We obtain survey based measures of presidential approval from the website FiveThirtyEight.com, which aggregates multiple presidential approval surveys and weights each survey by sample size and pollster quality rating (based on historical accuracy in predicting election results and methodological standards) to obtain an overall measure of daily presidential approval (Silver, 2017).

We scrape 1000 tweets per day containing the word "Trump" during the time period from January 20, 2017 through August 25, 2019. This particular interval started with the first day of the Trump administration and covered the following 31 months. Sentiment of individual tweets is calculated using Vader, a rule-based sentiment method trained on tweets and shown to perform well at assessing sentiment of tweets (Hutto & Gilbert, 2014). Vader assigns a continuous sentiment score between -1 and 1 to each individual tweet.

There is much variation in mean Twitter sentiment day-to-day. This variation is intrinsic to Twitter (that is, it cannot be simply attributed to our limited sampling of 1000 tweets per day; see Appendix A for details). To address this daily variation, we introduce a smoothing parameter k: the smoothed Twitter sentiment for a given day is calculated by taking the average sentiment of that day and previous k-1 days. We also introduce a lag term L, shifting survey responses ahead or behind by L days. This tells us whether Twitter sentiment leads or lags presidential approval. We allow k to be in $\{1, 2,\ldots,45\}$ and L to be in $\{-30, -29,\ldots,29, 30\}$. We choose k and L such that we obtain the highest correlation between sentiment of "Trump" tweets and presidential approval. We choose k and L in this manner for three reasons: (1) it is not clear a priori whether social media lags survey responses or vice versa and it is not clear what the optimal smoothing might be, (2) we want to give the political signal the best chance of emerging, and (3) similar methods were performed in previous analyses (e.g. O'Connor et al. (2010) and Cody et al. (2016)). An optimal smoothing of 45 days and lag of 30 days (meaning that Twitter sentiment lags presidential approval by 30 days) gives the maximum correlation of 0.516 between sentiment of "Trump" tweets and presidential approval. While this is not as high as previously observed correlations between "Obama" tweets and presidential approval (0.73 in O'Connor et al. (2010) and 0.76 in Cody et al. (2016)), the correlation of 0.516 might still seem to suggest there is a relationship between sentiment of "Trump" tweets and presidential approval from 2017 through mid-2019.

The observed correlation of 0.516 appears to be moderately strong. However, we optimized over the smoothing and lag parameters, and trends in time-series data can artificially inflate correlations, so it is unclear how to interpret the strength of the 0.516 correlation. To accurately interpret the strength of this observed correlation, we want to know how large the correlation would be if there were no underlying relationship between "Trump" tweets and presidential approval. To do this, we use a random sample of 5000 tweets per day from the same time frame. We first extract all words and symbols (such as emojis and numbers) that appear in at least one tweet per day. After removing stop words (e.g. "the", "an"), we are left with 495 words and symbols. We call these placebo words, as the only relationships between sentiment of tweets containing a given placebo word and presidential approval are presumably spurious. There are some "Trump" tweets in our random sample of all tweets, but they constitute a small percentage of our random sample. For each of these placebo words we repeat the same analysis as we did with the "Trump" tweets. That is, using tweets that contain a given placebo word, we adjust smoothing and lag such that we obtain the maximum absolute correlation between sentiment of tweets containing the placebo word and presidential approval. Further discussion of optimal smoothing and lag parameters is given in Online Appendix B. This results in 495 placebo correlations. We call the set of these correlations the reference distribution. Figure 1 gives the reference distribution. The reference distribution is bimodal. This is because we manipulate the smoothing and lag parameters to find the optimal correlation (in absolute value) between sentiment of tweets containing each of the placebo words and presidential approval. To assess the strength of the relationship between "Trump" tweets and presidential approval, we compare the observed correlation in relation to the reference distribution. If there truly is a relationship between sentiment of "Trump" tweets and presidential approval, the observed correlation should be much larger than nearly all of the placebo correlations. Our observed correlation of 0.516 is represented by the dashed vertical line in Figure 1 and is larger than many of the placebo correlations, but not considerably so. About 5.3% of the placebo correlations are larger in absolute value than the correlation between presidential approval and "Trump" tweets (see Online Appendix B for further details). However, none of the placebo words with maximum absolute correlations greater than 0.516 are meaningfully related to presidential approval, e.g., "giveaway", "17", "enough", "city", and "name" are the five words with the highest maximum absolute correlation with presidential approval. While there appears to potentially be a signal, if anything it is a very weak signal, and a signal that is not significantly stronger than ones found with a random sample of tweets unrelated to politics.

[Figure 1 about here]

Note that this placebo analysis framework can be used to evaluate the strength of any measure of association and any pre-processing of sentiment between messages containing some keyword and survey responses, not just correlation when adjusting for smoothing and lag in the context of presidential approval.

3. Longitudinal Analysis of Twitter Users

The results of the previous section raise concern on the utility of tracking public opinion with tweets that contain a given word over time. We propose an alternative: instead of tracking tweets containing a given word (e.g. "Trump"), we follow a group of users longitudinally. A longitudinal study of Twitter users performed in this manner may have several advantages. For example, when following the word "Trump" over time, we cannot be sure as to what extent the demographics of users tweeting about Trump are changing over time. Our goal in this section is to determine whether we can convincingly detect any signal when tracking the tweets of a specific set of users over time. To see whether there is a believable political signal in the tweets of these users, we examine their tweets around what we assume to be one of largest signals on Twitter for this set of users: the outcome of the 2016 presidential election.

First we gather a set of politically active users. We define a user as politically active if they produced at least 20 original (non-retweet) tweets in 2016, at least 10 of which were political (determined by whether a tweet contained at least one word from a hand-created list of political words). We had a total of 4189 politically active users. See Online Appendix C for further details on gathering our set of politically active users.

We would ideally like to know each user's political party affiliation. We create a training set of users with known political affiliation, Democrat or Republican, by hand-classifying users whose self-provided profile description contained a political word. Our training set consisted of 170 Democrats and 393 Republicans. Using this set of users we build a classifier to predict political affiliation of the remaining users. As covariates for the classifier we used the list of 3040 accounts that at least 30 of the users with known political affiliation follow. A random forest is used as the classifier. The random forest appears to perform well, with only 2.66% of users with known political party being incorrectly classified and the most important accounts for classification being either politicians, political commentators, or family members of politicians. A confusion matrix and variable importance plot can be found in Appendix E. We use the trained random forest to predict political party for the remaining politically active users with unknown political party and apply an 80% cutoff rate (meaning a user is classified as a member of a given political party if at least 80% of the trees predict the user to be a member of that party), which gives 489 total Democrats and 996 total Republicans that we use going forward.

We consider two metrics for tracking the tweets of our set of Democratic and Republican users: frequency and sentiment. Frequency tells whether or not our set of users are tweeting about political events, and sentiment tells us their reaction to those events. We first consider the frequency of all original (i.e., non-retweet) tweets sent by our set of Democratic and Republican Twitter users. Figure 2 shows the frequency of original tweets for Democrats and Republicans from 2016 through mid-2017. The solid vertical lines on these plots represent election day (November 8, 2016) and inauguration day (January 20, 2017) and the dashed vertical lines represent the top four days with the highest frequency of tweets. The top four days with the highest frequency of tweets for Democrats, in order of frequency, are October 10, 2016; November 9, 2016; October 20, 2016; and September 27, 2016. These days correspond to the day after the election and the days after the three presidential debates between Hillary Clinton and Donald Trump. The top four days for Republicans are November 9, 2016; October 20, 2016; October 10, 2016; and November 8, 2016. These days correspond to the day after the election, days after the third and second debates, and election day. The frequency of tweets is clearly politically driven for both Democrats and Republicans.

[Figure 2 about here]

After observing a fairly convincing signal that our set of users are tweeting about political events, we next consider sentiment of original tweets, measuring how the users reacted to those events. We find that while frequency of tweets among our politically active users is mainly driven by political events, sentiment for both Democrats and Republicans is driven by both political and nonpolitical events. Large daily spikes in average sentiment for all tweets from Democrats and Republicans correspond to holidays, such as Christmas and Thanksgiving, and a large daily drop is likely in response to a mass shooting, as can be seen in Figure 3.

[Figure 3 about here]

Many of the events that affect the sentiment of tweets of both our Democrats and Republicans occur outside of the political realm. Therefore, with the idea that Democrats and Republicans react to holidays and tragedies with similar sentiment, we are instead interested in the difference in sentiment between Democrats and Republicans. Figure 4 shows the daily difference in the mean sentiment of Democratic and Republican tweets from two months before the election through two months after the election. There is a clear drop the day after the election, and there appears to be an overall change when comparing difference in sentiment from before the election to after the election: Democrats are generally happier before and Republicans happier after. Presumably because the election results were a surprise for many, the notable change in difference in sentiment between Democrats and Republicans was immediate as opposed to gradual.

[Figure 4 about here]

While Figure 4 suggests a genuine difference in sentiment between our set of Democrats and Republicans from before the election compared to after the election, this change in sentiment is arguably relatively small. We look specifically at users who are vocal about politics and have fairly clear political party affiliation. We thought that the 2016 presidential election would be one of the largest signals on Twitter for these users, and the signal observed in Figure 4 is less pronounced than we might have imagined for such a set of users.

4. Discussion

If social media data is to be used to supplement or replace surveys tracking public opinion, there must be sufficient evidence that the social media data is indeed a valid way of measuring public opinion. This includes evidence that we are indeed tracking the signal of interest, a high signal to noise ratio, and stability of the relationship over time. We address these issues in accomplishing our two main goals: developing a framework to interpret an observed relationship between surveys of public opinion and tweets containing some keyword, and finding evidence of a political signal when following Twitter users longitudinally.

We found the correlation between sentiment of "Trump" tweets and presidential approval, 0.516, by optimizing smoothing of sentiment and lag between survey responses and tweets. We developed a framework to interpret the strength of this observed correlation by comparing it to 495 placebo correlation obtained by performing the same analysis, but with tweets containing everyday words. The correlation of 0.516 was not especially strong in comparison with the reference distribution. This shows that there is a high level of noise in Twitter data; many of the placebo correlations, which should consist of nearly pure noise, were as high as the correlation between “Trump” tweets and presidential approval. As an alternative method to tracking tweets that contain the word “Trump”, we proposed following politically active users longitudinally over time. We found evidence of a political signal when classifying users as Democrat or Republican based on the accounts they follow. When tracking the frequency of their tweets over time, we found a clear political signal, with frequency of tweets spiking at political events. The difference in sentiment between Democrats’ and Republicans’ tweets also changed immediately following the 2016 election. Noticeable changes in the tweeting patterns of our set of users around political events confirms that we are indeed capturing our political signal of interest. This is consistent with previous results that found events in Twitter data, for example frequency of “Obama” and “Romney” tweets leading up the 2012 presidential election (Barberá & Rivero, 2015) and sentiment of “Obama” tweets spiking on Obama's birthday (Pasek, McClain, Newport, & Marken, 2019). However, given that the election was what we assumed to be one of the clearest signals on Twitter for this particular set of users, the change in sentiment is relatively small.

While we only considered social media data extracted from Twitter, similar methods can be applied to data extracted from other social media platforms. For example, we can interpret the relationship between Reddit posts containing the word “Trump” and presidential approval using our placebo analysis framework. Following social media users from other platforms over time may also a valid and fruitful method of extracting posts to analyze.

Creating a post on social media is in many ways different from responding to a survey question (Schober, Pasek, Guggenheim, Lampe, & Conrad, 2016), involving different psychological processes, reasons for posting, and considerations of the audience. As one example, the demographics of social media platforms do not reflect the demographics of the general population (Wojcik and Hughes, 2019). All of these differences have the potential to introduce bias, and completely removing this bias from social media data is perhaps a nearly impossible task.

While we have found no evidence that tweets containing a given keyword reliably track public opinion, we still believe there is potential for social media data to be utilized for this purpose. The results of our longitudinal analysis suggest that there is a real, if weak, signal in Twitter data, and a future line of work could make use of that signal. This seems unlikely to replace traditional public opinion surveys, but could potentially supplement surveys. Smith and Gustafson provide an example of supplementing election polls with Wikipedia page views of candidates to more accurately predict election results (Smith & Gustafson, 2017). Many challenges lie ahead, but with the right methods, there is potential for social media data to improve upon traditional methods of capturing public opinion.

Author Information

Robyn A Ferg: fergr@umich.edu

Johann A Gagnon-Bartsch: johanngb@umich.edu

Frederick G Conrad: fconrad@umich.edu

Data Availability

Presidential approval was downloaded from the website FiveThirtyEight, available at https://projects.fivethirtyeight.com/trump-approval-ratings/?ex_cid=rrpromo . Data and scripts for replicating all analyses in this paper can be found at https://github.com/robynferg/Tracking_Presidential_Approval_with_Twitter. The Twitter data available online used in the placebo analysis gives the daily average sentiment for tweets containing each of the placebo words. To protect the privacy of the politically active users, we have blinded the user name and tweet content in the data set available online.

Software Information

Sentiment calculations using Vader were performed in Python version 3.65. All other analyses were performed in R version 3.5.1.

Funding

This work was supported by the National Science Foundation [grant number 1646108] to [RAF]

References

Athey, S., & Imbens, G. W. (2017). The State of Applied Econometrics: Causality and Policy Evaluation. *Journal of Economic Perspectives, 31*(2), 3-32.

Barberá, P., & Rivero, G. (2015, December 1). Understanding the Political Representativeness of Twitter Users. *Social Science Computer Review, 33*(6), 712-729.

Cody, E. M., Reagan, A. J., Sheridan Dodds, P., & Danforth, C. M. (2016, August 5). *Physics.* Retrieved from arXiv.org: https://arxiv.org/pdf/1608.02024.pdf

Conrad, F. G., Gagnon-Bartsch, J. A., Ferg, R. A., Schober, M. F., Pasek, J., & Hou, E. (2019). Social Media as an Alternative to Surveys About the Economy. *Social Science Computer Review*. doi:https://doi.org/10.1177/0894439319875692

Daas, P. J., & Puts, M. J. (2014, September). Social Media Sentiment and Consumer Confidence. *European Central Bank Statistics Paper Series*(5).

De Heer, W., & De Leeuw, E. (2002). Trends in household survey nonresponse: A longitudinal and international comparison. Survey nonresponse, 41, 41-54.

DiNardo, J. E., & Pischke, J.-S. (1996). The Returns to Computer Use Revisited: Have Pencils Changed the Wage Structure Too? *NBER Working Paper Series*.

Golder, S. A., & Macy, M. W. (2011, September 30). Diurnal and Seasonal Mood Vary with Work, Sleep, and Daylength Across Diverse Cultures. *Science, 333*(6051), 1878-1881. doi:10.1126/science.1202775

Hsieh, Y. P., & Murphy, J. (2017). Total Twitter Error. In P. Biemer, E. de Leeuw, S. Eckman, B. Edwards, F. Kreuter, L. Lyberg, . . . B. West, *Total Survey Error in Practice* (pp. 23-46). Hoboken, New Jersey: Wiley.

Hutto, C., & Gilbert, E. (2014). VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. *Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media.* Ann Arbor, MI.

Kennedy, C., Mercer, A., Keeter, S., Hatley, N., McGeeney, K., & Gimenez, A. (2016). Evaluating online nonprobability surveys. Pew Research Center. Available at: http://www. pewresearch. org/2016/05/02/evaluating-online-nonprobability-surveys/(accessed September 2016).

Murphy, J., Link, M. W., Hunter Childs, J., Langer Tesfaye, C., Dean, E., Stern, M., . . . Harwood, P. (2014). Social Media in Public Opinion Reserach: Executive Summary of the AAPOR Task Force on Emerging Technologies in Public Opinion Research. *Public Opinion Quarterly, 78*(4), 788-794.

O'Connor, B., Balasubramanyan, R., Routledge, B. R., & Smith, N. A. (2010). From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series. *Proceedings of the Fourth International AAAI Conference on WEblogs and Social Media*, 122-129.

Pasek, J., McClain, C. A., Newport, F., & Marken, S. (2019). Who's Tweeting About the President? What Big Survey Data Can Tell Us About Digital Traces? *Social Science Computer Review*.

Pasek, J., Yan, H. Y., Conrad, F. G., Newport, F., & Marken, S. (2018). The Stability of Economic Correlations Over Time: Identifying Conditions Under Which Survey

Tracking Polls and Twitter Sentiment Yield Similar Conclusions. *Public Opinion Quarterly, 82*(3), 470-492.

Schober, M. F., Pasek, J., Guggenheim, L., Lampe, C., & Conrad, F. G. (2016, January 13). Social Media Analyses for Social Measurement. *Public Opinion Quarterly, 80*(1), 180-211. doi:https://doi.org/10.1093/poq/nfv048

Silver, N. (2017, March 2). *How We're Tracking Donald Trump's Approval Ratings*. Retrieved from FiveThirtyEight: https://fivethirtyeight.com/features/how-were-tracking-donald-trumps-approval-ratings/

Smith, B. K., & Gustafson, A. (2017, May 06). Using Wikipedia to Predict Election Outcomes: Online Behavior as a Predictor of Voting. *Public Opinion Quarterly, 81*(3), 714-735. doi:https://doi.org/10.1093/poq/nfx007

Wojcik, S., & Hughes, A. (2019). Sizing up Twitter users. Pew Research Center. Available at: https://www.pewresearch.org/internet/2019/04/24/sizing-up-twitter-users/(accessed June 2020).

Figures

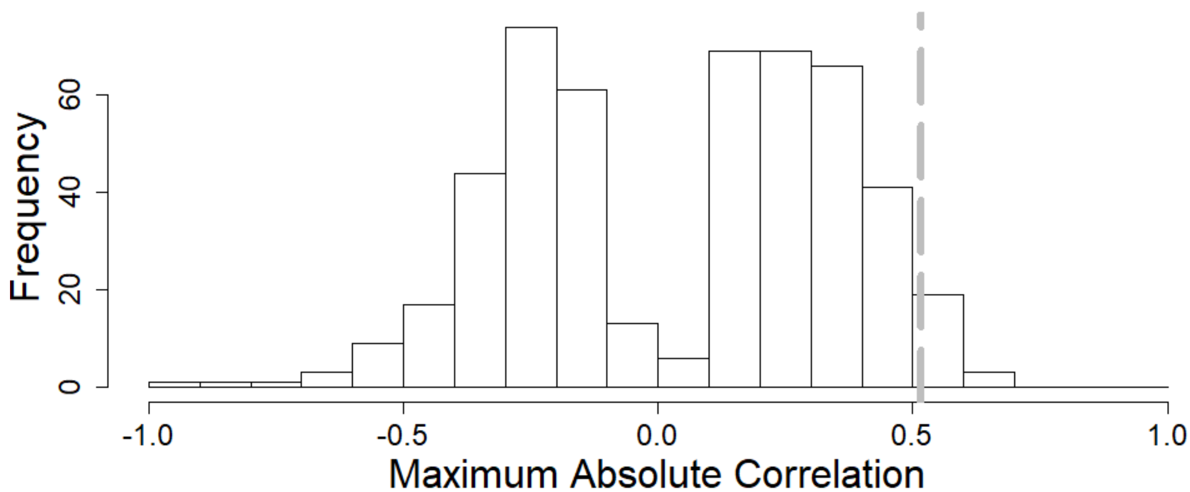
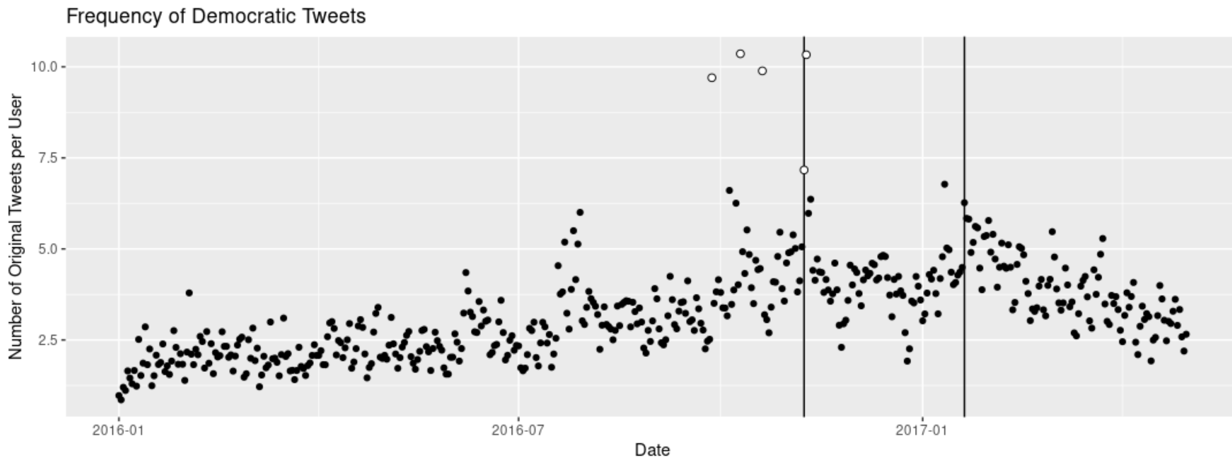


Figure 1: Reference distribution of maximum absolute correlations between presidential approval and sentiment of 495 placebo words with *k* in {1,…,45} and *L* in {-30, -29,…, 29, 30}, with bin widths of 0.1. Maximum correlation between sentiment of "Trump" tweets and presidential approval, 0.516, is denoted by the vertical dashed line.



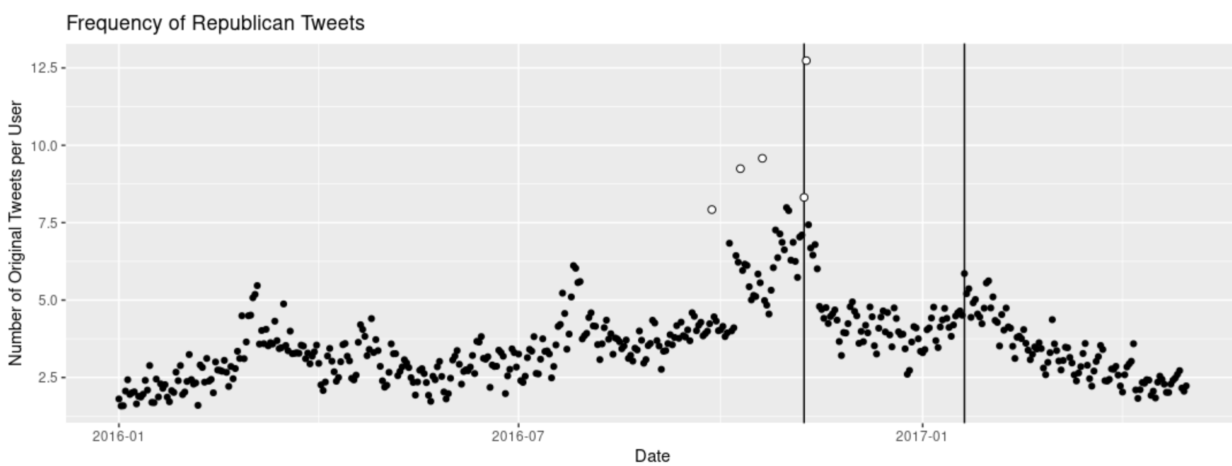


Figure 2: Average number of original tweets per day per Democrat (top) and Republican (bottom) users from 2016 through mid-2017. Vertical lines represent election day (November 8, 2016) and inauguration day (January 20, 2017). White points are the days with the highest frequency of tweets for Democrats and Republicans.

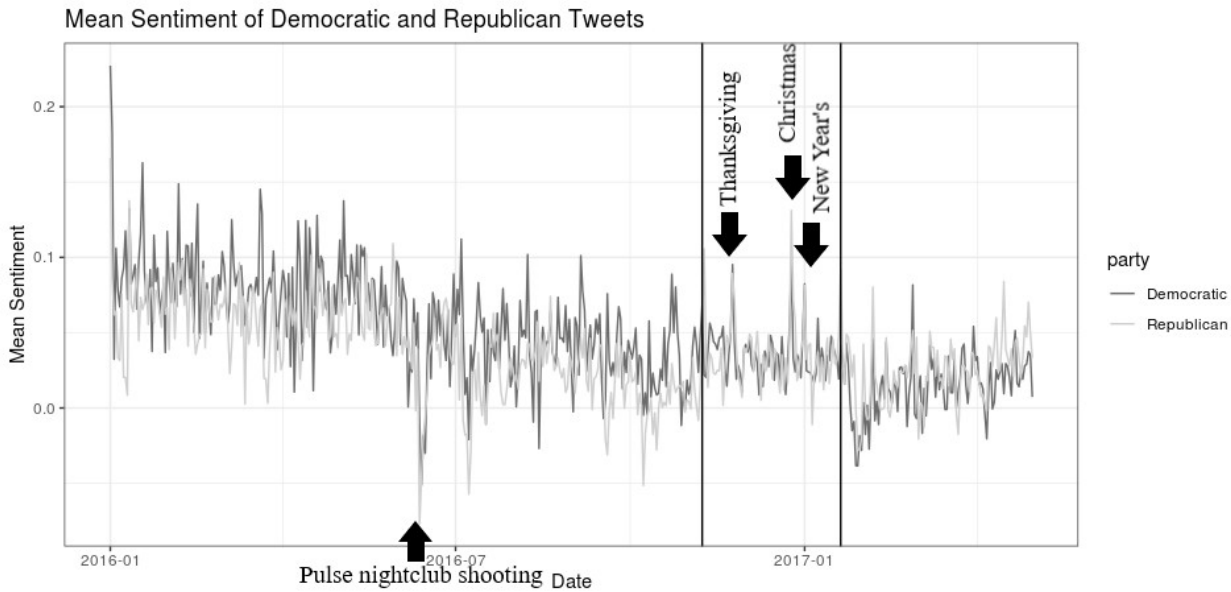


Figure 3: Average daily sentiment for Democrats (dark grey line) and Republicans (light grey line) from May 2016 through May 2017. Vertical lines represent election day (November 8, 2016) and inauguration day (January 20, 2017).

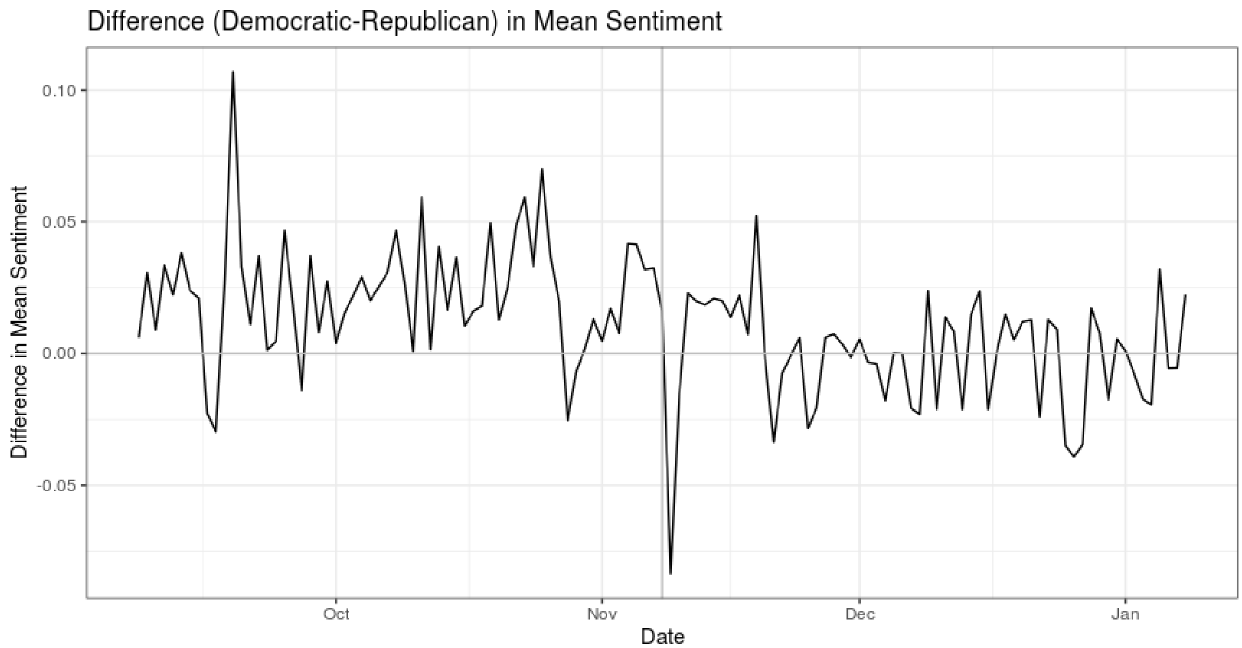


Figure 4: Difference in average sentiment between Democrats and Republicans (Democrats minus Republicans) from two months before the election (September 8, 2016) through two months after the election (January 8, 2017). The vertical line is election day (November 8, 2016).