Media, Pundits and the U.S. Presidential Election: Quantifying Political Leanings from Tweets

Felix Ming Fai Wong  
Princeton University  
mwthree@princeton.edu

Chee Wei Tan  
City University of Hong Kong  
cheewtan@cityu.edu.hk

Soumya Sen  
Princeton University  
soumyas@princeton.edu

Mung Chiang  
Princeton University  
chiangm@princeton.edu

ABSTRACT

Media outlets and pundits, at least those in the US, have been quick to embrace online social networks to disseminate their own opinions. But pundits' opinions and news coverage are often marked by a clear political bias, as widely evidenced during the fiercely contested 2012 U.S. Presidential elections. Given the wide availability of such data from sites like Twitter, a natural question is whether we can quantify the political leanings of media outlets using OSN data. In this work, by drawing a correspondence between tweeting and retweeting behavior, we formulate political leaning estimation as an ill-posed linear inverse problem with regularization. The result is a simple and scalable approach that does not require explicit knowledge of the network topology. We evaluated our method with a dataset of 119 million election-related tweets collected from April to November, and used it to study the political leanings of prominent tweeters and media sources.

Categories and Subject Descriptors

C.2.0 [General]: Data communications

General Terms

Social Media, Online Social Networks

Keywords

Twitter, Estimation, Political Leaning

1. INTRODUCTION

In an era marked by partisan politics, citizens tend to rely on news media to be informed about political issues and shape their views. But the impartiality of media sources, at least in the US, is often being questioned. However, on the brighter side, today’s online social networks like Twitter are providing a way for the masses to voice their own political opinions by engaging with the media directly. It would therefore seem that one can finally get a fair and balanced view of a controversial issue by sampling tweets from both sides of the argument in the Twittersphere. And in a similar manner, we could also potentially assess the political leanings of various news media sources from analyzing their activities in Twittersphere and the extent of participation and sentiments they tend to generate among tweeters. Unfortunately, there are some key challenges to overcome:

1. Quantification: Is it possible to assign meaningful numerical scores to tweeters about their position in the political spectrum?

2. Scalability: Given Twitter’s large scale and server limitations, how to devise a method that is efficient and scalable?

In this paper, we attempt to answer both questions through a new angle on the correspondence between tweets and retweets. We choose to focus on retweeting behavior primarily because it is a more active form of participation in the dissemination of opinions by tweeters than the passive act of “following” a news source. Framing the quantification problem in the context of elections, we propose a method that is

1. Simple: it does not require explicit knowledge of the network topology, and works within rate limits imposed by the Twitter API;

2. Efficient: computationally efficient because it is formulated as a convex optimization problem, and data efficient because the time required to collect sufficient data to obtain good results is short; and

3. Intuitive: the computed scores have a simple interpretation of “averaging.”

To evaluate our method, we collected a large set of 119 million tweets related to the U.S. presidential election of 2012 over a timespan of seven months. We present a case study of quantifying the political inclination of (a) major media outlets that have a Twitter account, (b) the most prominent tweeters in terms of the number of retweets received, and (c) media outlets studied in the existing works that quantify media bias. The efficacy of our method is demonstrated in our results that both match conventional wisdom and corroborate similar findings in works with other
approaches. We note that the applicability of our methodology goes beyond the context of the political process of elections and can be applied in estimating market segmentation, political census etc.

The organization of the remaining of this paper is as follows. Section 2 reviews related work in studies of Twitter and characterizing political preferences in traditional and online social media. Section 3 motivates and summarizes our proposed approach. Section 4 details our approach in terms of an optimization problem. Section 5 describes our dataset collected during the U.S. presidential election of 2012 together with some basic analysis. We apply our method on the dataset as a case study in Section 6. Then in Section 7 we compare our approach with existing approaches of quantifying media bias. Section 8 concludes the paper with future work.

2. RELATED WORK

There is a large economics and political science literature on quantifying the political bias or positioning of news media outlets. Here we focus only on the empirical approaches that explicitly assign scores to each media outlet.

Lott and Hassett [25] linked the sentiment of newspaper headlines to economic indicators to suggest a Democrat bias in newspapers. Groseclose and Milyo [18] linked media outlets to Congress members by co-citation of think tanks, and then assigned political bias scores to media outlets based on Americans for Democratic Action (ADA) scores of Congress members. Gentzkow and Shapiro [16] did an automated analysis of text content in newspaper articles. They identified phrases that were more commonly used by Republican or Democrat members of the Congress, and quantified media slant based on the frequency of these phrases seen in newspapers.

Explicit information has also been used to quantify political orientations. Ho and Quinn [19] analyzed newspaper editorials on Supreme Court cases to give explicit political positions of major newspapers. Ansolabehere, Lessem and Snyder [3] used 60 years of editorial election endorsements to quantify the political orientation of newspapers, and discovered a shift in political preferences with time. Noel [29] used pundits’ opinions on issues to estimate their positions in an ideological space.

Given its rich structure [10, 23, 22, 20] and publicly accessible data stream, Twitter has proven useful in a number of applications, including prediction [4, 6] and detection and estimation [32, 27, 11]. Our work in quantifying political leaning through Twitter can be considered an application of estimation. Regarding politics, there has been much interest in using Twitter to predict election outcomes [34, 15, 24]. There are also several works on characterizing the political landscape of Twitter [28, 26, 13, 24].

The most relevant are several papers on quantifying political leaning using the Twitter network. An et al. [2] computed pairwise distances between media sources by their mutual follower/subscriber sets, and embedded the media sources on a one-dimensional geometric space, with the fitted positions interpreted as political orientations. King, Orlando and Sparks [21] applied multidimensional scaling to Twitter network data to assign ideology scores to politicians. Barberá [5] proposed an item-response model assuming the Twitter network is probabilistically generated by latent variables representing political orientations. Colbeck and Hansen [17] proposed a graph-based heuristic to propagate ADA scores of Congress members to Twitter to media sources through their followers.

Classification is arguably an easier problem than quantification. Conover et al. [12] considered a range of techniques to predict political alignment of Twitter users. They built a retweet graph based on users’ retweet activity, and showed a simple graph clustering method outperformed machine learning methods based on tweet contents. Other machine learning approaches have been proposed [8, 1, 31], applying a wide range of information as features to aid in classification.

A further comparison between our approach and the above approaches will be made in Section 7.

3. PROPOSED APPROACH

In this section we motivate and explain the intuition of our proposed approach.

3.1 A Motivating Example

To motivate our approach based on retweets, we consider a small example based on some data extracted from our dataset on the presidential election.

Consider a pro-Republican media source A and a pro-Democrat media source B. We observe the number of retweets they received during two consecutive events. During the “Romney 47 percent comment” event (event 6 in Table 1), source A received 791 retweets, while source B received a significantly higher number of 2,311 retweets. It is not difficult to imagine what happened: source B published tweets bashing the Republican candidate, and Democrat supporters enthusiastically retweeted them.

Then consider the first presidential debate. It is generally viewed a pro-Republican event with contrasting performance of the two debaters. This time source A received 3,393 retweets, while source B received only 660 retweets. The situation reversed with Republicans enthusiastically retweeting.

This example provides two hints: (a) The number of retweets received by a tweeter (the two media sources) during an event can be a signal of its political leaning. In particular, one would expect a politically inclined tweeter to receive more retweets during an event favorable to the candidate it supports. (b) The action of retweeting carries implicit sentiment of the retweeter. This is true even if the original tweet does not carry any sentiment itself. The intuition is that tweeters tend to follow and retweet those who share similar political views, e.g., a tweeter is more likely to retweet a newspaper to which it subscribes than any random newspaper, a manifestation of the homophily principle.

3.2 General Description of Method

Our method is built upon the assumption that the two forms of expressing political opinions, tweeting and retweeting, are consistent.

Given a large set of tweets, we group them into sets of relevance: in this paper, we group tweets by events because of simplicity (it can done just by looking at a time series in our case study), but other forms of grouping is also possible, such as by issues (economic, diplomatic, religious). This grouping of tweets allows for a more fine-granular analysis, e.g., tracking change of political leaning over time, and provides more datapoints for our estimation problem.
The next step is to estimate, for every event, a numerical score that quantifies the approval of the candidates by the aggregate Twitter population. This can be done using off-the-shelf sentiment analysis tools. It may seem that the performance of our method will crucially rely on the performance of sentiment analysis, but we will show in our case study that just getting the right trend in sentiment is sufficient. (It has also been shown that Twitter sentiment trends computed with standard techniques correlate with poll results and socio-economic phenomena [30, 7].)

Recall from Section 3.1 that the action of retweeting carries information on the political opinions of the retweeter. We can thus define the political leaning of a retweeted tweeter as the average approval score a person wishes to express when retweeting any of its messages. This political leaning score is on the same scale as the average score (per tweet) from the previous step. Then for every event, we can average over the political leaning scores all retweets in that event.

Now we have one average score found by analyzing tweets, and another found by analyzing retweets. We apply the tweet-retweet consistency assumption to say that they are roughly the same, and this gives an equation per event. Finally, the estimated political inclinations will be the best fit solution to the set of equations.

A formal development of the above ideas will be presented in Section 4.

4. FORMULATION

4.1 Definitions

Consider two political parties or candidates running for an election. During the election campaign there have been $E$ events which attracted considerable attention. We are interested in quantifying the political leaning of $N$ prominent tweeters, e.g., media outlets and celebrities, using Twitter data collected during the $E$ events.

For event $i$, let $U_i$ be the set of users who tweeted about the event, and $T_{iu}$ be the set of tweets sent by user $u \in U_i$ about the event. Also define each tweet $t$ to carry a score $s_t \in [-1, 1]$, such that it is 1 if the tweet shows full support on one candidate, or -1 if full support is shown on the other candidate. Then for user $u$ its approval score is

$$\sum_{t \in T_{iu}} s_t$$

Averaging over all users in $U_i$, the average tweet leaning $y_i$ of event $i$ is

$$y_i = \frac{1}{|U_i|} \sum_{u \in U_i} \sum_{t \in T_{iu}} s_t$$

For source $j$, we quantify its political leaning as $x_j \in \mathbb{R}$, interpreted as the average approval score when someone retweets a tweet originating from $j$.

Now let $V_i$ be the set of users who retweeted any one of the $E$ sources at event $i$, and $R^{(i)}_{uj}$ be the number of retweets sent by user $u$ with the tweet originating from source $j$. Then the retweet approval score of user $u \in V_i$ is the average over all sources it has retweeted:

$$\frac{1}{|V_i|} \sum_{u \in V_i} \sum_{j=1}^{N} \frac{R^{(i)}_{uj}}{|T_{iu}|} x_j$$

and the average retweet leaning is the average over all $u$:

$$\frac{1}{|V_i|} \sum_{u \in V_i} \sum_{j=1}^{N} \frac{R^{(i)}_{uj}}{|T_{iu}|} x_j$$

where $A_{ij}$ is used to denote the inner summation term. Interestingly, the matrix $A$ can be interpreted as a Retweet matrix that captures the tweet-and-retweet response feature in Twitter.

4.2 An Ill-posed Linear Inverse Problem

The main premise of this paper is the behavior of tweeting and retweeting is consistent. Mathematically, we require the average tweet and retweet leanings per event to be similar:

$$y_i \approx \sum_{j=1}^{N} A_{ij} x_j, \quad i = 1, \ldots, E.$$  

We often have many more tweeters (millions) than events ($E = 12$ in our case study), then $N > E$ and the system of linear equations $Ax = y$ is underdetermined, which means there are infinitely many solutions $x$ that can achieve the minimum possible error of 0 in Problem (7). Then the problem becomes an ill-posed linear inverse problem [9]. The challenge of solving ill-posed problems is in selecting a reasonable solution out of the infinite set of feasible solutions. For example, in our case study we tried to solve for the least-norm solution, i.e.,

$$\min_{x} \parallel x \parallel_2$$

subject to $Ax = y$

but the results were unsatisfactory.

4.3 Regularization with a Graph Laplacian

Solving ill-posed problems requires us to incorporate prior knowledge of the problem to rule out undesirable solutions. One such common approach is regularization, and we can change the objective function in Problem (7), $\parallel Ax - y \parallel_2^2$, to $\parallel Ax - y \parallel_2^2 + \lambda f(x)$, where $\lambda > 0$ is a regularization parameter, and $f(x)$ quantifies the “fitness” of a solution such that undesirable solutions have higher $f(x)$ values. For example, the Tikhonov regularization for least-squares uses $f(x) = \parallel x \parallel_2^2$ [9]. In this paper, we propose a regularization term that favors political leaning assignments $x$ with
Figure 1: Number of tweets per day. Numbers on plot indicate events (see Table 1), and dotted lines indicate time periods when significant data were lost due to network outage (five instances).

Let $S_{jk}$ be a similarity measure between tweeters $j$ and $k$ such that $S_{jk} \geq 0$ and $S_{jk} = S_{kj}$. Further, let $S$ be the symmetric matrix whose elements are $S_{jk}$. Then we set

$$f(x) = \sum_{j=1}^{N} \sum_{k=1}^{N} S_{jk}(x_j - x_k)^2,$$

(8)

so that if $S_{jk}$ is large (tweeters $j$ and $k$ are similar), then $x_j$ should be close to $x_k$ to minimize $(x_j - x_k)^2$.

Note that $f(x)$ can be rewritten in terms of a graph Laplacian. Let $D$ be defined as

$$D_{jk} = \begin{cases} \sum_{m=1}^{N} S_{jm} & j = k, \\ 0 & \text{otherwise}, \end{cases}$$

and $L$ be the graph Laplacian defined as $L = D - S$. Then it can be shown that

$$\sum_{j=1}^{N} \sum_{k=1}^{N} S_{jk}(x_j - x_k)^2 = 2x^T Lx.$$

(9)

Before stating our final optimization formulation, we need one more constraint on $x$. If $x = \alpha 1$ for any real $\alpha$, then $x^T Lx$ is minimized to have value 0, and depending on the choice of $\lambda$, solving the regularized optimization problem will mean finding a solution between some feasible solution to $Ax = y$ and $1$, which is not meaningful for our estimation problem. Hence we impose one extra constraint: $x^T 1 = 0$, to prevent the above from happening. Writing $x^T 1 = 0$ as the equivalent form $\sum_{j=1}^{N} x_j = 0$, it also has the interpretation that $x$ has at least one positive element and one negative element, i.e., each candidate or party has at least one tweeter supporting it. Finally, our optimization problem becomes

$$\begin{aligned} & \text{minimize}_x & & \|Ax - y\|_2^2 + \lambda x^T Lx \\ & \text{subject to} & & x^T 1 = 0. \end{aligned}$$

(10)

It is a convex optimization problem and can be solved efficiently. The solution to the optimization problem in (10) can be computed by direct methods, which require $O(N^3)$ flops (since we assume $N > E$), when no structure is exploited. In general, there are more efficient iterative methods to solve (10), e.g., interior-point algorithms [9]. There are standard numerical packages freely available to solve large-scale convex optimization problems, e.g., the CVX software package [14].

4.3.1 Definition of Source Similarity

The choice of $S_{jk}$ is largely independent of the optimization problem itself, and so we defer it to here. In our case study, we compute $S_{jk}$ as follows:

1. Let $a_j$ be the $j$-th column vector of $A$. Then for each $a_j$, compute the detrended version $\tilde{a}_j$ by subtracting the line of least squares fit from $a_j$.
2. Set $S_{jk} = \tilde{a}_j^T \tilde{a}_k / (\|\tilde{a}_j\|_2 \|\tilde{a}_k\|_2)$, i.e., the cosine similarity between the two vectors.
3. Set $S = S - \min(S)$.

Intuitively, if two sources are similar, the retweet response to their tweets should also be similar (recall the simple example in Section 3), and the retweet response is captured by their vectors $a_j$ and $a_k$. Detrending in step 1 is necessary because Twitter activity increases as time to the presidential election decreases, and we need to avoid emphasizing too much on later events when computing similarity. Taking cosine similarity as the similarity measure in step 2 accounts for the variation in popularity of different sources, through normalization with vector magnitudes. Finally, step 3 is needed to make $S$ nonnegative, which is needed for the optimization problem to be convex.

4.3.2 Incorporating Prior Knowledge

Prior knowledge can readily be incorporated into our method through introducing constraints to the optimization problem. Here we consider two examples:

**Anchors.** Suppose we know a certain tweeter $j$ is strongly liberal. We can then set its political leaning $x_j$ to be a fixed value, say +1. In the literature this idea has been used frequently [19, 2, 17].
Minimum pairwise distances. This is our preferred approach. Suppose we know two tweeters $j$ and $k$ have opposite political leanings. We can impose the constraint on the distance of their political leanings as $x_j - x_k \geq c$, where $c$ is a nonnegative constant. We recommend setting $c$ to be moderately small, and let the data decide whether the distance has to be large. In Section 6.3, for each optimization problem we set $c = 0.5$ and impose one such constraint on the most liberal and the most conservative sources.

5. DATASET

In this section we describe the collection and processing of our Twitter dataset of the U.S. presidential election of 2012. Our dataset was collected over a timespan of seven months, covering from the initial phases to the climax of the campaign.

This recent election is highly suitable for our study. The success of Obama's social media campaign in 2008 highlighted the value of social media in elections. As a result, in the 2012 election we see serious effort from both sides of the campaign to leverage Twitter, and thus we have a relatively balanced dataset. Another feature of our dataset is in the consistently high levels of tweeting intensity. Even early events occurring months before the election received millions of tweets, and the later events broke records multiple times in terms of the number of tweets per political event, and the number of retweets of a single tweet.

5.1 Data Collection

From April 8 to November 10 2012, we used the Twitter streaming API to collect 119 million tweets which contain any one of the following keyword phrases: “obama”, “romney”, “barack”, “mitt”, “paul ryan”, “joe biden”, “presidential”, “gop”, “dems”, “republican” and “democrat” (string matching is case-insensitive). A manual inspection of a sample of tweets shows the collected tweets are of sufficiently high quality, i.e., most tweets are indeed about the presidential election. We purposely chose not to do any filtering by language, location etc. to evaluate the robustness of our method against noisy data.

5.2 Event Identification

By observing the time series of tweet counts in Figure 1, we manually identified 12 events as listed in Table 1. We defined the dates of an event as follows: the start date was identified based on our knowledge of the event, e.g., the start time of presidential debates, and the end date was defined as the day when the number of tweets reached a local minimum or dropped below that of the start date. After the events were identified, we extracted all tweets in the specified time interval without additional filtering, assuming all tweets are relevant to the event and those outside are irrelevant.

5.3 Extracting Tweet Sentiment

We applied SentiStrength [33], a lexicon-based sentiment analysis package, to extract the sentiment of tweets. Before using the package, we removed four words from its emotion lexicon: “gay” (as in “gay marriage”), “foreign” (“foreign policy”), “repeal” (“repeal obamacare”) and “battle*” (“battleship”). These four words were considered by the package to carry negative sentiment, but we consider them neutral in the current context of elections. Sentiment analysis was done as a trinary (positive, negative, neutral) classification.

For each tweet in one of the 12 events, we set its score $s_t = 1$ if either (a) it mentions solely the Democrat camp (has “obama”, “biden” etc. in text) and is classified to have positive sentiment, or (b) it mentions solely the Republican camp (“romney”, “ryan” etc.) and has negative sentiment. We set $s_t = -1$ if the opposite criterion is satisfied. If both criteria are not satisfied, then set $s_t = 0$.

Figure 2 shows the values of $y$ due to the above scoring mechanism. The values of all elements, i.e., the average tweet inclination of all events, are all close to 0 even though the possible range is $[-1, 1]$. This indicates the dataset is balanced in terms of praising/bashing both candidates, although it is slightly in favor of Obama. A closer look at the exact values indicates that the sentiment analysis results are reasonable: $y_t$ is smaller for pro-Romney events, e.g., first presidential debate, and larger for pro-Obama events, e.g., Romney’s 47 percent comment.

5.4 Retweet Statistics

Here we describe some statistics of the Twitter users who tweeted, retweeted or were retweeted at least once during the 12 events. Overall, 9.9 million Twitter users are identified in the events. Among these users, 3.7 million only tweeted, 3.5 million only retweeted, and 2.6 million both tweeted and retweeted. Although not in the majority, there is still a significant number of users who both tweeted and retweeted, and we should expect at least this set of users satisfying the consistency assumption.

Figure 3 is a set of complementary distribution functions of users’ tweeting activity and popularity in log-log scale. They all exhibit high skewness, but the skewness is more significant for the number of times retweeted. In particular, its CDF has a line artifact at the tail due to the outlier of Obama. He was retweeted 814,310 times during the 12 events, while the second and third highest retweeted counts are 131k and 117k respectively. In contrast, the maximum number of tweets and retweets made by a user are only 8,703 and 7,619 respectively.

We further analyze the users who both tweeted and re-
Table 1: Summary of events identified in the dataset.

<table>
<thead>
<tr>
<th>ID</th>
<th>Dates</th>
<th>Description</th>
<th># tweets (m)</th>
<th># non-RT tweets (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>May 9 - 12</td>
<td>Obama supports same-sex marriage</td>
<td>2.30</td>
<td>1.35</td>
</tr>
<tr>
<td>2</td>
<td>Jun 28 - 30</td>
<td>Supreme court upholds health care law</td>
<td>1.21</td>
<td>0.78</td>
</tr>
<tr>
<td>3</td>
<td>Aug 11 - 12</td>
<td>Paul Ryan selected as Republican VP candidate</td>
<td>1.62</td>
<td>0.96</td>
</tr>
<tr>
<td>4</td>
<td>Aug 28 - Sep 1</td>
<td>Republican National Convention</td>
<td>4.32</td>
<td>2.80</td>
</tr>
<tr>
<td>5</td>
<td>Sep 4 - 8</td>
<td>Democratic National Convention</td>
<td>5.81</td>
<td>3.61</td>
</tr>
<tr>
<td>6</td>
<td>Sep 18 - 12</td>
<td>Romney's 47 percent comment</td>
<td>4.10</td>
<td>2.55</td>
</tr>
<tr>
<td>7</td>
<td>Oct 4 - 5</td>
<td>First presidential debate</td>
<td>3.49</td>
<td>2.19</td>
</tr>
<tr>
<td>8</td>
<td>Oct 12 - 13</td>
<td>Vice presidential debate</td>
<td>1.92</td>
<td>1.19</td>
</tr>
<tr>
<td>9</td>
<td>Oct 17 - 19</td>
<td>Second presidential debate</td>
<td>4.38</td>
<td>2.67</td>
</tr>
<tr>
<td>10</td>
<td>Oct 23 - 26</td>
<td>Third presidential debate</td>
<td>5.62</td>
<td>3.35</td>
</tr>
<tr>
<td>11</td>
<td>Nov 4 - 6</td>
<td>Elections (before Obama projected to win)</td>
<td>7.50</td>
<td>4.40</td>
</tr>
<tr>
<td>12</td>
<td>Nov 7 - 9</td>
<td>Elections (after Obama projected to win)</td>
<td>6.86</td>
<td>4.43</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td></td>
<td>48.90</td>
<td>30.28</td>
</tr>
</tbody>
</table>

Figure 3: Complementary distribution functions of user tweet and retweet activity.

Figure 4: Histogram of tweet-retweet ratio.

tweeted. Figure 4 is a histogram of the ratio of 

\[
\frac{\# \text{ tweets}}{\# \text{ tweets} + \# \text{ retweets}}
\]

among these users. Somewhat surprisingly, a significant portion of these users have a balanced number of tweets and retweets, i.e., a ratio of close to 0.5. This seems to suggest that the users tend to use an equal number of tweets and retweets to express their opinions, supporting our consistency assumption.

6. EXPERIMENTAL RESULTS

6.1 Quantifying Major Media Sources

From the American National News Media list of Mondo Times,\(^6\) we extract a list of major media sources to which we apply our quantification method. We consider only media sources that are marked as popular, excluding radio shows (low popularity compared to other media), news aggregators (not consistent in reporting style) and news agencies.

Table 2 shows our results in quantifying the political leaning of these media sources. We caution that the results should not be considered as definitive proof of media bias, but as we can see, the numbers quantify conventional wisdom on which are liberal or conservative media sources.

\(^6\)http://www.mondotimes.com, a news media directory used in [19].

\(^7\)From Mondo Times. C: conservative/leans right, N: no bias, L: liberal/leans left.

Table 2: Twitter political leaning scores of major media sources.

<table>
<thead>
<tr>
<th>Media Source</th>
<th>Score</th>
<th>Classification(^7)</th>
</tr>
</thead>
<tbody>
<tr>
<td>US News &amp; World Report</td>
<td>-0.164</td>
<td>C</td>
</tr>
<tr>
<td>CNBC</td>
<td>-0.159</td>
<td>C</td>
</tr>
<tr>
<td>Fox News</td>
<td>-0.128</td>
<td>C</td>
</tr>
<tr>
<td>Washington Times</td>
<td>-0.102</td>
<td>C</td>
</tr>
<tr>
<td>CBS News</td>
<td>-0.076</td>
<td>L</td>
</tr>
<tr>
<td>HLN</td>
<td>-0.069</td>
<td>C</td>
</tr>
<tr>
<td>Newsweek</td>
<td>-0.053</td>
<td>N</td>
</tr>
<tr>
<td>The Week</td>
<td>-0.049</td>
<td>C</td>
</tr>
<tr>
<td>Chicago Tribune</td>
<td>-0.016</td>
<td>C</td>
</tr>
<tr>
<td>Christian Science Monitor</td>
<td>-0.015</td>
<td>N</td>
</tr>
<tr>
<td>LA Times</td>
<td>-0.011</td>
<td>L</td>
</tr>
<tr>
<td>ABC News</td>
<td>-0.002</td>
<td>L</td>
</tr>
<tr>
<td>MSNBC</td>
<td>0.006</td>
<td>L</td>
</tr>
<tr>
<td>USA Today</td>
<td>0.009</td>
<td>N</td>
</tr>
<tr>
<td>Wall Street Journal</td>
<td>0.039</td>
<td>L</td>
</tr>
<tr>
<td>Washington Post</td>
<td>0.053</td>
<td>L</td>
</tr>
<tr>
<td>Time</td>
<td>0.062</td>
<td>N</td>
</tr>
<tr>
<td>CNN</td>
<td>0.104</td>
<td>N</td>
</tr>
<tr>
<td>NY Times</td>
<td>0.205</td>
<td>L</td>
</tr>
<tr>
<td>Huffington Post</td>
<td>0.364</td>
<td>L</td>
</tr>
</tbody>
</table>
We single out two outliers and explain the unexpected results using their tweet contents:

**CBS News.** Compared to other events, which normally results in hundreds to low thousands of retweets, we observe a spike of 12,000 in its number of times being retweeted during event 10 (third presidential debate), and the most retweeted tweet was an instant poll result (“BREAKING: who won the debate? ...”). The debate had a rather mixed review, as seen from the correspondingly low average tweet leaning \( y_{10} \), and as a result the estimated political leaning of CBS News is skewed towards the negative side.

**Wall Street Journal.** Somewhat surprisingly, most of its retweeted tweets are actually quite neutral. This can be explained by the claimed separation between the Journal’s news section and editorial section [18], and from the tweet contents, we do find most of the tweets coming from news reports, rather than editorials. Our result agrees with the results in multiple works [18, 25], which ranked Wall Street Journal as the most and the second-most liberal media outlet respectively.

### 6.2 Quantifying Prominent Tweeters

We rank Twitter users by their total number of times being retweeted during the 12 events, and identify the top 1,000 of them. Figure 5 shows the histogram of the computed scores for \( \lambda = 10^{-5} \) together with a number of notable tweeters by where they lie in the score spectrum. The results are qualitatively the same for \( 10^{-4} \leq \lambda \leq 10^{-7} \). These results are discussed as follows:

**Parody accounts.** Among the top 1,000 tweeters we identify three parody Twitter accounts (FiredBigBird, BigBirdRomney, BIGBIRD) created in response to the Big Bird comment in event 7, and one account (InvisibleObama) created in response to the “invisible chair” skit during event 4. These accounts, being sarcastic in nature, are against the Republican camp but were found to have pro-Republican scores. The reason is that they received most of the attention during the event of interest, and these events are pro-Republican. Accounting for this type of behavior will require further content or network analysis.

**Candidates.** The results for candidates’ accounts appear correct, including their election campaign accounts and the accounts of their political parties (TheDemocrats, RepublicanGOP). Biden’s account appears to be the only exception. Again, part of the reason is the skew in attention to the vice presidential debate, which is computed to have a low average tweet score \( y_{10} \) (see Figure 2).

**News media.** Compared to politicians and celebrities, most media sources are concentrated at the center of the score spectrum, with moderate variation according to their political leanings. This suggests media sources tend to be objective relative to other prominent Twitter users.

### 6.3 Comparison with Existing Results

We apply our method to the three sets of media sources used in the empirical studies of [2, 18, 19]. We exclude sources that either (a) no longer exist due to a merge or a change in TV program host, or (b) have less than 25 retweets (as a comparison, the least popular included sources have more than 150 tweets). For all three papers we report the resultant Kendall’s \( \tau \) statistic, and for [2, 18] we also report their Spearman’s \( \rho \) and Pearson correlation coefficients because we have access to their actual ADA scores.

<table>
<thead>
<tr>
<th>With paper</th>
<th>Kendall’s ( \tau )</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>An et al. [2] ((N = 19))*</td>
<td>0.46</td>
<td>0.0052</td>
</tr>
<tr>
<td>Groseclose and Milyo [18] ((N = 16))</td>
<td>0.50</td>
<td>0.0064</td>
</tr>
<tr>
<td>Ho and Quinn [19] ((N = 18))</td>
<td>0.39</td>
<td>0.026</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Spearman’s ( \rho )</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>An et al.</td>
<td>0.60</td>
</tr>
<tr>
<td>Groseclose and Milyo</td>
<td>0.62</td>
</tr>
</tbody>
</table>

Table 3: Correlation test results.

![Figure 6: Twitter political leaning scores vs ADA scores estimated in [2] with least squares fit.](image)

Table 3 shows the results of statistical tests. We can see that the correlations between the rankings computed by our method and those of existing results are statistically significant. Comparing p-values, our results are in better agreement with those of An et al. [2], which is not surprising because their results are also derived from Twitter data. On the other hand, the media sources in [19] mostly consist of traditional and regional newspapers, which received fewer retweets. Hence our method has the least agreement with their results, although the correlation is still high. Figures 6 and 7 show scatter plots of our political leaning scores against the ADA scores in [2] and [18] at optimal \( \lambda \).

**Sensitivity to regularization parameter.** We vary \( \lambda \) and observe the change in the test statistics. The results in Figures 8, 9 and 10 show that they do not vary much over a wide range of \( \lambda \) (for [2]: 0.01 \( \leq \lambda \leq 0.1 \); for [18]: \( \lambda > 0.01 \); for [19]: \( \lambda > 0.03 \)), and fine-tuning \( \lambda \) may not be necessary.

*We added datapoints from the paper’s figures and demo webpage.*
6.4 Time Dynamics

One advantage of using tweet-retweet response to infer political leaning is in the ability to do fine-granular temporal analysis. Here we illustrate with a simple qualitative analysis.

We quantify the political leanings of the set of 1,000 tweeters in Section 6.2, but this time we vary the amount of information used for estimation. We start by running our method using data from events 1 to 2, then we use events 1 to 3, and so on. Then each Twitter user has a sequence of 11 political leaning scores, and we plot the trajectories of a select set of users in Figure 11. Note the plots are on the same scale.

Candidates and media sources. For candidates, we take the Twitter accounts of the four (vice-)presidential candidates and their respective political parties, which results in a total of 6 accounts. For media sources, we use the sources in Section 6.1 that also appear in the top 1,000 tweeter list. The result is a set of 12 accounts. From Figure 11, we can see the qualitative behavior is similar for the two types of accounts. The scores initially drift but eventually stabilize at around 6 to 8 events, suggesting that a sufficient number of events is needed to ramp up the accuracy of estimation. Afterwards any change to the scores is gradual. This observation is reasonable because all candidates and most media sources have predefined political stances.

Parody accounts. Besides the four accounts mentioned in Section 6.2, we manually inspected the top 100 tweeters...
of the top 1,000 list and added six more accounts. Figure 11 shows the political leanings of these accounts appear to be more chaotic.

7. COMPARISONS

Comparison with content-based analysis. Here we compare with media bias studies in economics and political science [25, 18, 16, 19, 3, 29] which analyze news media content directly. Except for [16], all studies require some form of manual coding and analysis, which is expensive and time-consuming. A more fundamental problem is the scarcity of data. Because the amount of data available for analysis is limited by how fast the media sources publish, researchers may need to aggregate data created over long periods of time, often years, to perform reliable analysis.

Analyzing media sources through their outlets in online social networks, e.g., Twitter, offers many unprecedented opportunities. Communication in social media involves many more participants and happens at much shorter timescales as compared to print or broadcast media. Hence data are generated at much higher rates, and we can quickly collect sufficient data for analysis (seven months in our case study). Social media sources also provide a range of data not previously available, such as timestamps and citations, to support richer analysis.

Comparison with graph-based analysis. Although incorporating graph information is often useful, the huge sizes of most online social networks means it is difficult for an average researcher to obtain an up-to-date snapshot of a network. In the context of Twitter, the rate limiting mechanisms set by the Twitter API prevents crawling the network to any reasonable size. This problem is highlighted in [5], which had to analyze a random sample of 12k out of 377k users because of rate limiting. In contrast, our method requires only one connection to the real-time Twitter stream.

We also argue that using retweets is more robust than using the Twitter graph to infer political leanings. Retweeting is an explicit act of approval, but following (a tweeter) is not. A Twitter user may follow two media sources with opposite political stances because of rate limiting. In contrast, our method requires only one connection to the real-time Twitter stream.

As of the time of writing, each authenticated client can make 350 requests/hour, i.e., crawl the neighbors of at most 350 Twitter users.
To initiate further discussion, we pose two open questions:

1. There is no golden standard in quantifying political leaning, bias, orientation, slant etc. because they are fuzzy concepts. How can we compare results across different methods or algorithms?

2. For any algorithm that can track political leanings with time, how can we know that a stabilization in scores is due to convergence properties of the algorithm (an artifact) or a real shift in political leaning?

9. ACKNOWLEDGMENT

This work was in part supported by an ARO MURI grant.

10. REFERENCES


