Political Hashtag Trends

Ingmar Weber[1], Venkata Rama Kiran Garimella[1], and Asmelash Teka[2]

[1] Qatar Computing Research Institute, Doha, Qatar
{iweber,vgarimella}@qf.org.qa
[2] L3S, Hanover, Germany
teka@l3s.de

Abstract. Political Hashtag Trends (PHT)[1] is an analysis tool for political left-vs.-right polarization of Twitter hashtags. PHT computes a *leaning* for trending, political hashtags in a given week, giving insights into the polarizing U.S. American issues on Twitter. The leaning of a hashtag is derived in two steps. First, users retweeting a set of "seed users" with a known political leaning, such as Barack Obama or Mitt Romney, are identified and the corresponding leaning is assigned to retweeters. Second, a hashtag is assigned a fractional leaning corresponding to which retweeting users used it. Non-political hashtags are removed by requiring certain hashtag co-occurrence patterns. PHT also offers functionality to put the results into context. For example, it shows example tweets from different leanings, it shows historic information and it links to the New York Times archives to explore a topic in depth. In this paper, we describe the underlying methodology and the functionality of the demo.

1 Introduction

Politicians worldwide and in the U.S. in particular have realized the power that social media carries for campaigning. Here, Twitter is on the frontline as it engages users in political debates and, ultimately, mobilizes them for grassroot movements. Within Twitter, *hashtags* play an important role as labels for ongoing debates that other users can "link to". Hashtags are used consciously by key influencers to frame a political debate and to define the vocabulary used in such debates. There are several examples of "hashtag wars" between Democrats and Republicans.[2] Political Hashtag Trends (PHT) is a tool to gain insights into the political polarization of hashtags. It not only assigns a leaning to a hashtag in a given week, but it also shows example tweets for the corresponding leaning, identifies trending hashtags, and links to several external sources such as the New York Times archive or `topsy.com`.

2 Features of the Demo

2.1 Leaning and Trending Information

The core functionality of PHT consists of (i) identifying trending, political hashtags in a given week, and (ii) assigning a leaning to them. The home screen (Figure 1) shows

[1] `politicalhashtagtrends.sandbox.yahoo.com`
[2] See, e.g., `bit.ly/Lkzjwm` or `bit.ly/KTOnUZ`

P. Serdyukov et al. (Eds.): ECIR 2013, LNCS 7814, pp. 857–860, 2013.
© Springer-Verlag Berlin Heidelberg 2013

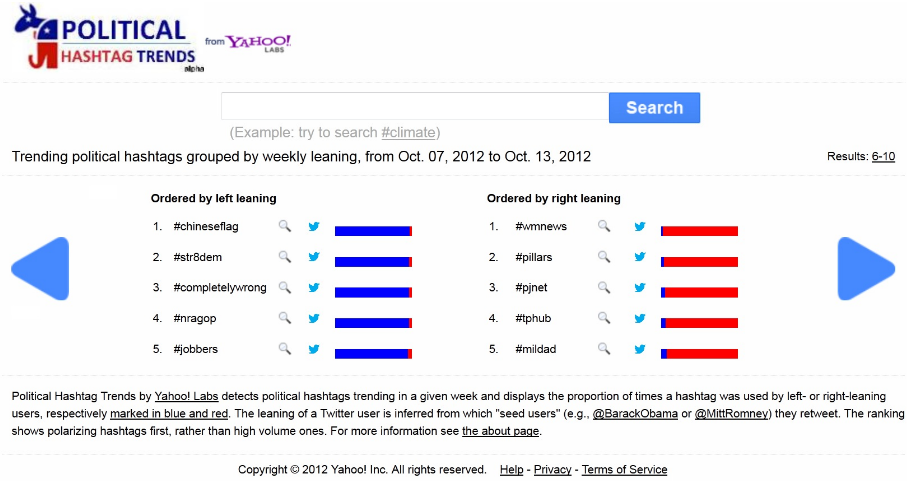

Fig. 1. The starting page of PHT

trending political hashtags for the latest week, arranged by leaning. As an example, the hashtag `#completelywrong` referred to a phrase used in an interview by Romney, leading to his pictures being ranked highest on Google Image Search during that week (`slate.me/Th0yKX`). PHT also gives the user the possibility to go to an older week or to search for hashtags across all weeks. For example, searching for `obamacare` reveals `#obamacares` as a left-leaning tag, with `#obamacaremustgo` showing up as right-leaning.

2.2 Historic Information

Selecting the magnifying glass icon, users can view historic volume and leaning information. This reveals, e.g., that `#bdayreagan` was short-lived and that `#middleclass` has sudden jumps in leaning, indicating hashtag wars and "hijacking" attempts.

2.3 Putting Things into Context

To put the results shown into a broader context, the following features are provided.

Twitter current search: Clicking on a hashtag will present the user with *recent* tweets via URLs such as `twitter.com/#!/search/%23obama`. This provides a quick way to follow the most recent debate on the topic at hand.

Twitter archive search: The user can view tweets for the week of interest for the given hashtag. This is achieved by linking to Topsy's date-specific archive search.

New York Times archive search: Often hashtags are (concatenations of) proper words or names and can be a basis for a search in news archives. To facilitate this, we link each hashtag to a date-specific search on the New York Times archive.

3 How It Works

3.1 Starting Point: Seed Users

Our data acquisition starts with a set of seed users with known political orientation. This set is then expanded using retweet behavior and later cleaned by limiting the geographic scope. The data was obtained using a Ruby wrapper for the Twitter REST API[3], in combination with Apigee[4]. Our seed set of Twitter users contains key official politicians from U.S. politics. To be selected, a Twitter account (i) had to belong to either a political leader in office or it had to be an official party account, (ii) for a person, it had to be the "personal" account rather than an office-related account[5], and (iii) it had to be a verified Twitter account. In total, there were 14 seed accounts for the left and 19 for the right. The ones with the most followers were Barack Obama and Nancy Pelosi (left) vs. Mitt Romney and Newt Gingrich (right).

3.2 Identifying Politicized Users

For each of our seed users we obtained their publicly available tweets. For each tweet we identified up to 100 retweeters. As we observed that, e.g., Canadian Twitter users would retweet U.S. politicians, we limited our analysis further to U.S. users. Concretely, we used Yahoo! Placemaker on user-provided location information and only kept users with a U.S. location. This left us with with 111,813 users. For each week, these users are assigned a *fractional* leaning corresponding to the ratio of their retweets of either left or right seed users. For retweeting users we obtain their public tweets for the given week. Note that our methodology allows for a change in leaning of retweeting users.

3.3 Detecting Political Hashtags

The hashtag `#cutekitten` is non-political and not of interest for our demo. `#russia` might be non-political during the European soccer cup, but will be political during times of protest in Moscow. To tell political from non-political hashtags for a given week w, we look at co-occurrence with a set of hashtags which are deemed to be political. This seed set included hashtags referring to the main political parties and events (`#p2`, `#tcot`, `#teaparty`, `#gop`, `#tlot`, `#sgp`, `#tpp`, and `#ows`) and hashtags containing the strings obama, romney, politic, liberal, conservative, democ, or republic. We then use this seed set as follows. First, we compute the within-week user volumes for each hashtag, i.e., each (user, week, hashtag) triple is only counted once. Then, for each week and for each leaning separately, we keep the top 5% of hashtags in terms of user volume. Note that a user's volume can contribute fractionally to *both* leanings. For each of these hashtags h, we count the number of users who use h at least once in combination with a political seed hashtag. We keep the top 25% in terms of the political-to-all user fractions, again for each leaning separately. In the end, the two lists of left- and right-leaning hashtags are merged and the resulting (h, w) pairs are used in our analysis. We use a high precision approach to get meaningful political hashtags, at the possible expense of recall.

[3] `github.com/jnunemaker/twitter`

[4] `apigee.com`

[5] E.g., `@whitehouse` might change its political leaning but `@BarackObama` would not.

3.4 Assigning a Leaning to Hashtags

We use a voting approach to compute the leaning of hashtags, similar to [1], where both leanings are given an equal voting weight. Concretely, let v_L denote the aggregated user volume of h in w for the left leaning. Let V_L denote the total left user volume of all hashtags in w. Similarly for v_R and V_R. Note that users can contribute *fractionally*, based on their fractional leaning in w. We compute the leaning of h in w as

$$\mathrm{Lean}(h, w) = \frac{\frac{v_L}{V_L} + \frac{2}{V_L+V_R}}{\frac{v_L}{V_L} + \frac{v_R}{V_R} + \frac{4}{V_L+V_R}}, \tag{1}$$

where a leaning of 1.0 is fully left and 0.0 is fully right.

3.5 Assigning Trending Score to Hashtags

To assign a trending score $t(h, w)$ to h in w, we use the burst intensity index from [2] in which a "query" (for us a hashtag) has a high burst index if it has a large relative increase in frequency compared to its past frequencies and the overall frequencies in the given week. In our setting their formula becomes

$$t(h, w) = \frac{f(h,w) / \sum_{h' \in H} f(h',w)}{\sum_{u \leq w} f(h,u) / \sum_{h' \in H} \sum_{u \leq w} f(h',u)}. \tag{2}$$

Here $f(h, w)$ is the user volume for h in w. For a given week we sort hashtags by $t(h, w)$ and, going from the top, assign them to either left (Lean(h,w) ≥ 0.5) or right (Lean(h,w) < 0.5) leaning. For each leaning, we keep the top 20 in terms of trending score and display them, reranked according to Lean(h,w), for the left, or −Lean(h,w), for the right. Note that in all formulas we use user counts, rather than tweet counts, as we observed the former to be much more robust concerning outliers.

4 Related Online Demos

`politicalsearchtrends.sandbox.yahoo.com` [3] classifies Web queries into leaning based on whether they lead to clicks on predominantly left- or right-leaning political blogs. Their web interface is very similar and served as a basis for PHT. `election.twitter.com` computes a sentiment-based "Twitter political index" for the two presidential candidates. `live.votizen.com` shows tweets from registered voters, but does not analyze polarization of hashtags. The spread of memes on Twitter is visualized in `truthy.indiana.edu`, which also incorporates the political leaning of certain Twitter users. `politics.twittersentiment.org/streams` provides insights into political sentiments on Twitter. `google.com/elections/ed/us/trends` analyzes web search volume trends for the two presidential candidates.

References

1. Weber, I., Garimella, V.R.K., Borra, E.: Mining web query logs to analyze political issues. In: WebSci, pp. 330–339 (2012)
2. Subasic, I., Castillo, C.: The effects of query bursts on web search. In: WI, pp. 374–381 (2010)
3. Weber, I., Garimella, V.R.K., Borra, E.: Political search trends. In: SIGIR, p. 1012 (2012)