

Visualization of the Public's Opinion on Politically Influential Tweets

Tanyoung Kim and Carl DiSalvo

Georgia Institute of Technology
tanykim@gatech.edu and carl.disalvo@lcc.gatech.edu

Abstract

We present a Twitter-based website *The Political Grid Project* that focuses on harnessing the potential of Twitter as a political medium. On this site, the general public can create new datasets additionally to existing Twitter data, and investigate the data through diverse visualizations. This site lists tweets posted by political entities, and a user votes on each tweet based on the two standards—how much she agrees with the idea expressed in the tweet, and how important it is to her. More importantly, our site provides users with various interactive visualizations that represent both individual and collective voting data accompanied by data accessed from Twitter APIs. While exploring the visualizations, users can closely examine various aspects of politics, including their political preferences, the ties among the voted entities based on the analysis of the collective data, and temporal or geographical patterns. Ultimately, we expect our users to reflect on the political issues in tweets that are otherwise ephemeral and to better understand candidates in upcoming elections. Furthermore, our site allows users to investigate the public's aggregated opinion and their political stance among the public.

Background and Problems

Twitter, the most popular micro-blogging service, has become a crucial medium for the contemporary politics. Politicians and political parties take advantage of Twitter to spread their voices. So do journalists and their news channels to response to these voices and drive the public's opinion.

In the vast Twitter-sphere, many third party applications exist that focus on these voices from politics and journalism. (e.g., washingtonpost.com/mention-machine, wefollow.com/twitter/politics, 2012twit.com). These sites show influential Twitter users based on the number of followers and tweets that are freely acquired through Twitter APIs. Additionally, the APIs enable users to retrieve data about the spread of a single tweet including

the number of re-tweets and replies. However, these general descriptions do not adequately describe the complexity of the publics' responses to the political voices and their tweets. For example, a follower of a political Twitter user may re-tweet or reply when she likes it as well as when she does not like it. We believe that what people really think about the tweet is more important. To our knowledge, there are no Twitter applications that enable end users to create new types of data, which can explicitly express the users' opinion, besides consuming and adding to the Twitter datasets.

Twitter data is a unique and interesting resource for research in social sciences (social media studies, journalism, and public policy) and engineering (data mining and information visualization). Researchers in these fields exploit the data to discover insights about human behavior and social networks that are politically meaningful; investigating the effects of Twitter on an election (Tumasja 2010), discovering the network segregation according to political orientations (Conover 2011), and computing the political preference of the citizens (Golbeck 2011) to list a few. Through complex qualitative and statistical methods, these projects demonstrate that Twitter can be a barometer to predict diverse phenomena in the creation of public opinions and the effects on related political events. However, these research projects were conducted with data previously obtained for academic purposes, and did not leverage the real-time participation of the public.

The Political Grid Project

The approaches of third-party applications and research projects are summarized as either re-presentation of the existing Twitter data or only pertaining to academic concerns. Three factors differentiate our project from the previously mentioned applications and projects. First, our site presents *politically influential tweets*, not the real time stream of Twitter user's followings. Second, we do not

simply calculate the data accessed through Twitter APIs, but enable users to *create additional data through voting*. This action gives users a chance to rethink about the messages and to express their opinion on them. Third, the aggregation of the voting data is the useful resource for the further data analysis. The analysis results do not stay in the researchers' laboratory but open to the users as a form of *interactive visualizations allowing their self-investigation*.

Technology

Because we do not only simply retrieve data from Twitter APIs but also collect users' voting data, we have built our own database using MySQL. We develop the web pages with PHP and utilize CSS and jQuery for the dynamic user interface and cross-browser/platform support (Figure 1). In the database, we set up tables that contain the following information:

- **Voices:** Politically influential Twitter users (name, screen name, and the number of followers, following, and tweets)
- **Citizens:** Twitter users who signed in our site and vote (name, screen name, voting history, and favorite voices)
- **Tweets:** Voted tweets (Voice ID, Tweet ID, post date, and the text of Tweet)
- **Votes:** All the votes done by the citizens (Tweet ID, voting time, and voting points)

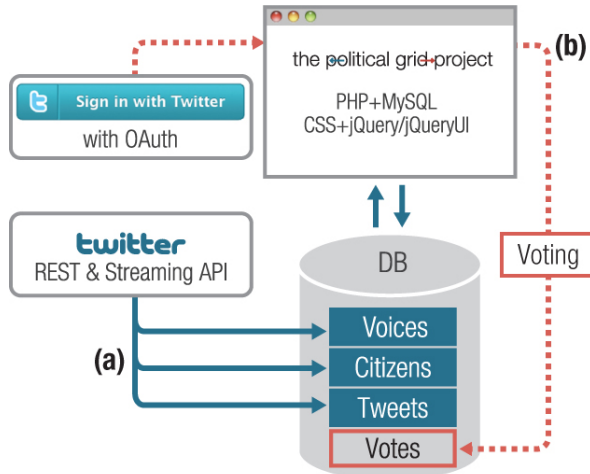


Figure 1. System of *The Political Grid Project*: We collect data from two sources, (a) Twitter APIs and (b) Sign-in users' voting.

Reflecting and Voting on Tweets

A user of our site signs in with his Twitter ID and reviews the tweets by political entities that were selected based on their popularity. We collect more data by allowing a user to *vote* on the individual tweets; we ask two questions—how much she *agrees* with the tweet, and how *important* it is to her—and she responds to those on a numerical grid

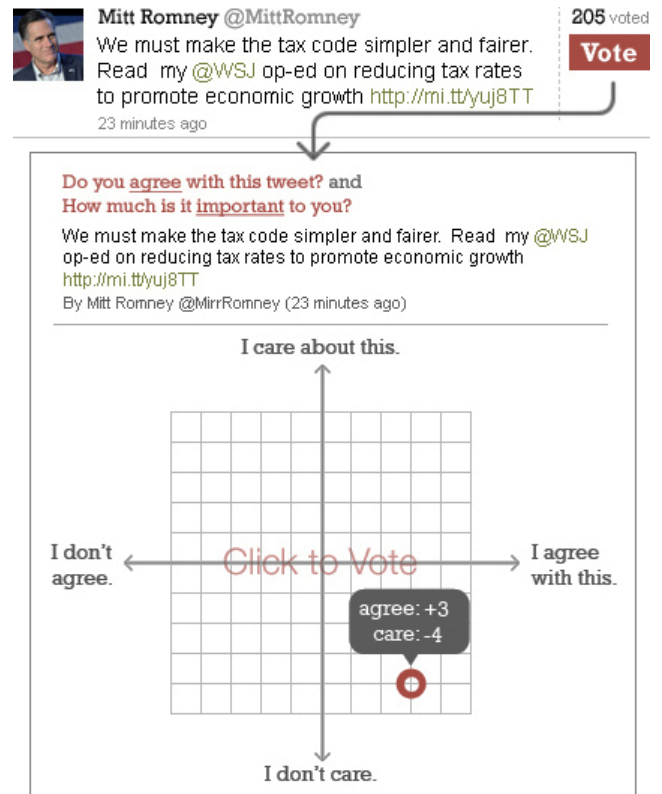


Figure 2. Voting on a single tweet: A user clicks on an interaction of the “grids.” Along X axis, he rates from -5 to 5 according to the degree of agreeing with the tweet. Similarly, Y axis has the ratings for the importance of the issue.

(Figure 2). Inferring this participatory action, we named our website *The Political Grid Project* (political-grid.org, a tentative domain name).

We begin by retrieving the list of Twitter users that are influential on politics, specifically the U.S. presidential elections in 2012 from Twitter's own suggestions (https://twitter.com/#!/who_to_follow/interests/us-election-2012). In this paper as well as our system, we call these politically influential Twitter accounts “Voices.” The users of our site can also suggest other Twitter users as Voices. Once a newly suggested voice is reinforced repeatedly by other users, this voice will be included in the public pool of voices, so that all users can see their tweets and vote on them.

The list of tweets on Twitter.com is updated in real-time, thus the tweets are meant to be extremely ephemeral. By contrast, we do not let tweets flow fast, but keep the important tweets visible. In doing so, we expect users to actually think about the messages in the tweets, not simply scan them. Users see the tweets that are presented in three different lists. First, on the very first page after sign-in, users see the list of *the top twenty tweets that have received the most votes in the past twenty-four hours*; we assume these tweets have more important issues. Second, while we

encourage users to vote on tweets posted by a variety of voices, we also support users' possible needs of focusing on their interested voices; they can "favorite" voices so users can examine only *the list of tweets only from their favorite voices*. To enable the examination of the real-time issues, the last list shows *the current stream of all public voices* just like Twitter.com.

Engaging through Visualization

A series of visualizations represents the diverse aspects of the data and allows users to investigate the data. At a micro level, a user can examine her political preference and the voting results of each tweet or voice in a greater detail. She can also explore the networks of the voices and the citizens from a macro perspective.

Tweet-centric Visualization: After voting, the results of all the users' voting are immediately displayed. We visualize the results on the same voting panel, in which each circle is filled with gray-scale. The more votes an intersection (a pair of two points) gets, the darker the gray is (Figure 3). To advertise our site, we encourage users to reply to the tweet including the information of our site, which will be displayed to the user's followers.

Single Citizen-centric Visualization: Users examine their political orientation and preference based on the voting data on the voices. Since we collect the voting data from two questions—how much they agree, and how much they care, users filter the standards (Figure 4).

Voice-centric Visualization: We also design a visualization that focuses on an individual voice. While adopting a line graph on a timeline, we present the temporal change of the voices' voting results (Figure 5). Users also examine the voted tweets posted in the given time period.

Social Network Diagrams: Analyzing the collective opinion on a myriad of tweets, we create network diagrams that represent *social ties and segregation among the votes*. We adopt the visualization techniques from other social network and visualization research such as NodeXL (Hansen 2011) and Vizster (Heer 2005). A user switches the view by filtering favorite voices. She also selects multiple voices and compares citizens' opinions on them.

In a similar way, we present the *relationship between citizens*, which is calculated based on the similarity in the choice of favorites and voting patterns. The network diagram enables users to investigate their relative political orientation as well as the polarization of the citizens accordingly. In this diagram a user compares multiple citizens; for example, if a user selects nodes representing citizens, then the voices that have high preferences from all the selected citizens appear. This interactive feature *connects both voices and citizens*. We expect that visualizations that centralize the citizens—not the voices—



Figure 3. Visualization idea of a single Tweet: The collective voting result of all the users are presented on the grids.

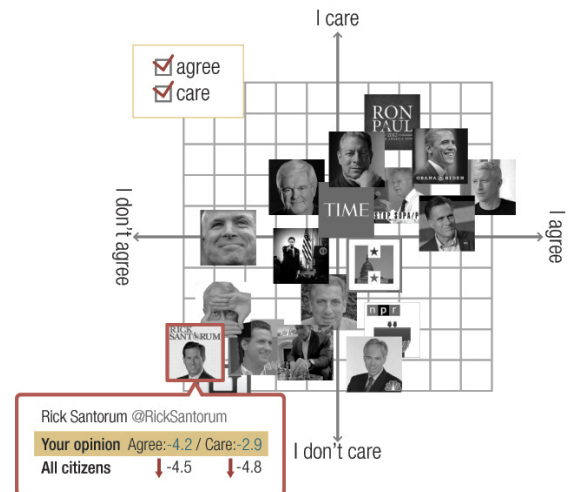


Figure 4. Visualization idea of the all voting result of a single user: A user examines her voting results and compares her data with all citizens'.

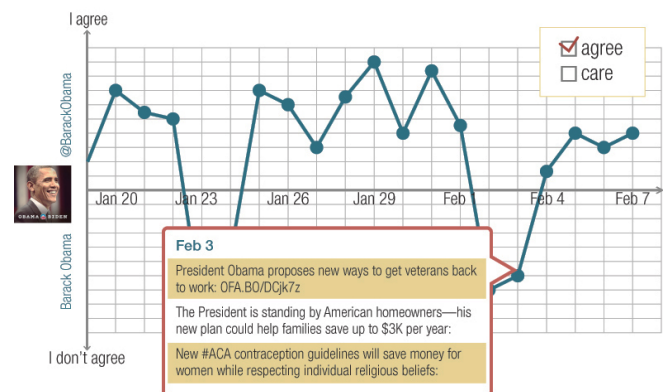


Figure 5. Visualization idea of a single voice: temporal trend of the voting results

will encourage users' active participation (as a compensation of being "famous" in the sphere) and the creation of community (as a tool to discover people with similar political stance).

Location-based Visualization: Twitter API provides location data through Twitter users' location setting in their profile or tagging location information on each tweet. Although not all users open their location information due to privacy concerns, we can exploit the accessible location information to analyze the voting patterns according to geographical context (e.g., Who has more attention and where? Who is more popular and where?).

Evaluation and Further Research

The Political Grid Project allows the general public to express their opinion on influential tweets by allowing users to *vote on the individual tweets* and by investigating the public's opinion through *visualizations of the data*. These novelties will prompt further research that evaluates the effects of our site and utilizes the collected data.

On-site Engagement: Researchers in Information Visualization often follow traditional HCI methods when they measure the social aspects of open visualization systems. For example, both quantitative (log file analysis) and qualitative (interviews) methods can be combined as in the evaluation of *Many Bills*, a website that visualizes congressional legislation (Assogba 2011). These methods typically divide user groups according to their activities and identify each group's behavioral traits.

We combine these two methods to analyze user behavior in greater detail and to reveal factors that drive different use patterns. For example, we can detect the users' general interest in politics and their political stance through additional surveys and analyze possible biases in the collected data.

Listening to Political Voices: We hope to learn how on-site experiences affect users' understanding of political issues. Through surveys we can ask following questions. Do they read the tweets more carefully on our site than the normal Twitter timeline? Do they learn more about the politicians or political parties and their arguments? Specifically, do they become more knowledgeable about the candidates for upcoming elections? Does our site help them choose the right candidate? We can also ask whether the users' self-measurement of political orientation is similar to our analysis. If not, what are the causes of the deviance? If this analysis reveals some limitations of our system, how can we improve it?

Comparing Different Contexts: Our first focus for *The Political Grid Project* is on the presidential election in the U.S. targeted at the largest demographic group of Twitter.com. Later we plan to support other languages and

countries. For example, in South Korea, Twitter has been a popular medium for politicians, and the presidential election will be held in December 2012. We can conduct comparative research to see the possible differences in voting patterns and other behaviors due to the differences in political situations, the political polarization of Twitter users, and even cultural differences in general.

Focusing on a Specific Political Event: To effectively compare voting patterns and results on politicians, we will choose a specific political event to track user behaviors with our site. This approach is similar to existing event-oriented Twitter research (Starbird 2010, Tumasjan 2011). One example of such an event is a major party's convention during the presidential nomination process. Furthermore, it will be possible to compare the voting patterns and results on politicians on our site with the actual results of such events.

Current Status and Plan: The site has been under development as of March 2012. Since the most crucial part of the site is the exploration through visualizations, the primary concern of the development phase is the optimal presentation of visualizations that treat potentially sizable datasets. We plan a closed beta test in May 2012, thus we will be able to demonstrate the site fully at the workshop. After the small-scale beta test, we aim to open it to the general public in summer 2012, targeting the presidential election in both U.S. and South Korea.

Acknowledgements

Special thanks to Tom Bellitire, Michael Dandy, Thomas Lester, and Justin Roberts for PHP/MySQL programming.

References

- Assogba, Y., Ros, I., DiMicco, J., and McKeon, M. Many Bills: Engaging Citizens through Visualizations of Congressional Legislation. In *Proc. of CHI 2011*.
- Conover, M. D., Ratkiewicz, J., Francisco, M., Goncalves, B., Flammini, A., and Menczer, F. Political Polarization on Twitter. In *Proc. of ICWSM 2011*.
- Golbeck, J. and Hansen, D. computing Political preference among Twitter Followers. In *Proc. of CHI 2011*.
- Hansen, D., Smith, M., and Shneiderman, B. EventGraphs: Charting Collections of Conference Connections. In *Proc. of Hawaii Int'l Conference on System System*. 2011.
- Heer, J. and boyd, d. Vizster: Visualizing online social networks. *IEEE Transactions on Visualization and Computer Graphics*. 2005.
- Starbird, K., Palen, L., Huges, A., and Vieweg, S. Chatter on the Red: What Hazards Threat Reveals about the Social Life of Microblogged Information. In *Proc. of CSCW 2010*.
- Tumasjan, A., Sprenger, T.O., Sandner, P.G., and Welp, I.M. Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment. In *Proc. of ICWSM 2011*.