Town Hall Tweets

Analyzing Influential Tweeters During a Presidential Debate

Introduction

Presidential debates are an instrumental part of presidential elections in the United States. Before every presidential election, a series of presidential debates take place, one regarding foreign policy, one regarding any subject (usually in town hall format), and one regarding domestic issues. This year, during the second debate, which was held in town hall format, audience members posed questions to the candidates. This town-hall debate took place on October 16, 2012 at Hofstra University. To collect relevant tweets that occurred during this debate, I used Topsy.com’s API to search for mentions of “Obama” and “Romney” in tweets. [3] I used NodeXL to analyze and visualize the tweet's authors and their networks. [4] I chose the second debate specifically for analysis because a range of issues was discussed and it was not just limited to just domestic or just foreign policy issues. Because it was a more general debate, I hypothesize that it gaged more general interest and thus would be more interesting for analysis than it's two counterparts.

Data

Topsy is a service that allows users and researchers to search past tweets for keywords. The current twitter API only allows searching the Twitter timeline for keywords in real-time. If a researcher is interested in a past event, the twitter API is almost useless. Topsy was an important and useful resource in this project. Using the Topsy API, I queried twitter for tweets that happened on October 16 during the second presidential debate. Topsy does not provide all the tweets in a given time period, but it yields the most “influential” tweets and provides an influence score for each of those tweets. My query resulted in 1000 total tweets. Each tweet returned from Topsy consists of an author, the tweet content, and an "influence score" which is calculated by Topsy. From these tweets, there were 437 unique tweeters. To create a graphml file that was compatible with NodeXL, I collected all of the unique nodes. If there was more than one tweet by an author, I took the highest "influence" tweet. I then used the twitter API to check for follow relationships between those 437 nodes. This search resulted in 1202 edges. My
resulting graphml file consisted of the following attributes: author, tweet, influence score, and whether Obama, Romney or both candidates were mentioned in the tweet.

**Analysis**

**Overview**

Upon loading the graphml file, I grouped the vertices by whether their tweet had mentioned Obama, Romney or both candidates. I then applied the Group in a Box layout and discovered that the biggest group consisted of Romney mentions.

The Fruchterman-Reingold algorithm was used to layout the network. The size of each vertex is related to the influence score. Each vertex is labeled with the author name. The largest group is Romney mentions, while the light blue group is mentions of Both Barack Obama and Mitt Romney. The right bottom group represents those who mentioned Obama.
Headlines

Headline 1: Huffington Post is the Bridge between Some of the Most Followed Political Tweeters

The graph below was filtered for nodes that have in-degrees above 15. This graph is significant because it shows the role of a key player among these influential tweeters. The Huffington Post serves as a bridge between many of the news agencies and important personalities. Of the 437 vertices, 10 have in-degrees greater than 15. These 11 are represented in the graph below. Five of eleven vertices are news agencies or news personalities. This shows the most followed people in this particular network are news agencies or news personalities.

The Huffington Post follows both personalities (Bill Maher, Russell Simmons, and Anderson Cooper) as well as other sources of news (Breakings News, and Time). While the Huffington Post follows Breaking News and Time, it does not follow the New York Times or CNN breaking news. This reveals an interesting phenomenon in news agency relationships. Often news agencies like to pride themselves as being the first break a story. Potential rivalries between these agencies could explain the lack of edges between them.
Headline 2: During Most Political Events, The most influential tweets come from Entertainers (the Music Industry, the Sports Industry, and Comedians)

Each of the tweets by the individuals had an “influence score” ranging from 8-14. Topsy assigns influence scores to tweets based on the amount of influence a certain tweet exerts. Their algorithm for calculating this metric is not completely transparent, however it is an interesting way of filtering the graph. I filtered the graph, so that it only included vertices with the highest “influence score” (between 13-14). What immediately became apparent was that celebrities had the highest “influence” on twitter during the second presidential debate. Only one politician appeared in this list and that was Barack Obama. Of the 9 vertices, two are comedians, two are sports players, two are singers, one is a website/service (YouTube), one is a social activist/director (Michael Moore) and one is a politician (President Barack Obama).

*"Influential" is based on the score from Topsy.com
image also shows that comedians play an influential role during this debate as two of the nodes that appear above are comedians (Liduval and billmaher). While Michael Moore is not a comedian, it is important to note that many of his tweets have comedic tendencies. Thus, an alternative headline for the graph above could be, "Comedians exert most influence during debates."

Headline 3: Barack Obama is more closely connected to Celebrities on Twitter than to News Agencies

In the graph below, the Clauset-Newman-Moore algorithm was applied to create groups, followed by the application of the "Group-In-A-Box" layout. Intergroup edges were hidden and the nodes were filtered for those that had more than 2 in-degrees, an influence score above 10 and a page rank above 1. In the image below there are three groups: the light blue group consists mainly of celebrities, the dark blue group consists mainly of new agencies, news personalities, or editors of news agencies, and the green group consists mainly of comedians, with the exception of Elizabeth Banks and Think Progress. Barack Obama is more closely connected to the group of celebrities than the News Agencies. Similarly, Donald Trump is also
connected to the celebrities as well. The reason for this clustering is that more celebrities follow Barack Obama than do the news agencies. In an attempt not to look biased, I hypothesize that news agencies avoid following the president of the United States. Celebrities on the other hand can freely broadcast their political affiliations without having their validity and accuracy questioned, while news agencies do not have the same luxury.

**User Experience**

To analyze and visualize this data, I utilized both Gephi (another network analysis tool) and NodeXL. The biggest usability issue I faced with using NodeXL was related to the fact that it does not run across all platforms. I own an Apple computer and have partitioned my computer into two partitions. I mainly work on my Mac side of my partitions so it was tedious to have to switch partitions just to use this software.

Additionally, when I set the shape of the vertex to image and attempted to do Group in a Box Layout, not all the images would be imported. This was a serious issue, especially when the
spreadsheet indicated that the selected vertices should be presented as images and not disks. I would then have to go in manually and select the nodes I wanted to be presented as images and manual indicate that by right clicking. This proved to be both tedious and time-consuming. Below is an image of what happened when not all of the nodes were represented the same.

Another complaint I would express regarding NodeXL is the inability to undo one's actions. After spending a good amount of time arranging vertices so that they are visible, if a button is accidentally pressed that causes the graph to be refreshed; all of that work is lost. This was exceedingly frustrating and happened to me a few times during the course of this project.

Because NodeXL works as a template in excel, it was easy to learn to use. If a user is familiar with Excel, it is very simple to experiment in the NodeXL ribbon in order to discover the tool for which you are looking. The drop down menus in the NodeXL window were clear and easy to understand. Additionally, I prefer NodeXL over Gephi because of its easy to understand interface and available options. Initially, my graphml files had some errors. When I loaded it into Gephi, I received an error message regarding a null exception. However, NodeXL gave me
an error message, which immediately pointed me in the right direction and allowed me to quickly solve my problem.

References


