Who Do the Candidates Follow?

Lyndsey Franklin
lyndsey@cs.umd.edu

Introduction
Many attempts are made to analyze, understand, and even predict the demographics or actions of voters in presidential elections and much attention is paid to the messages of campaign teams and candidates. An alternative perspective is to analyze the content to which candidates and campaigns subscribe. Rather than focusing on attempts of candidates to persuade others, this exploration attempts to gain insights into the presidential candidates based on the Twitter networks built from those that are followed by major party candidates. To what extent do candidates make use of content from those they follow? Do the candidates look to the same or different sources? Who is influential in their communication towards the candidates?

Dataset:
1.5 degree ego networks of the two presidential candidates and their running mates (Twitter handles: @BarackObama, @JoeBiden, @MittRomney, @PaulRyanVP) pulled from the candidate’s “Following” lists. The “Following” relationship was chosen to focus on who each candidate communicates with, rather than representing who receives updates from the candidate (both candidates and their running mates have impractically large lists of followers, most of whom will not be contacted by the candidates). Networks were limited to 200 nodes from each of the four Twitter handles. Edges were added for each following/followed relation, “replies-to” relation, and “mention” relation. A vertex was added for each person followed by the Twitter handle. An initial exploration of the dataset revealed seemingly glaring edge omissions (such as any relation between @BarackObama and any major news source) and prompted the additional importing of the follows/following, replies to, and mention relationships from several major news sources such as @FoxNews, @WashingtonPost, @USAToday, and others. 385 vertices connected through 3.3k edges appear in the final dataset.
The Great Wall of Media

After grouping by connection, coloring manually by affiliation (blue for Democrats, red for Republican, and purple for news sources), sizing nodes by betweenness centrality, and laying out the network through the Harel-Koren Fast Multiscale algorithm a stark divide can be seen in the Twitter networks of the two major parties. Particularly, a handful of morning news sources provide the connections between Republican and Democratic networks. Figure 1 shows the resulting network visualization.

![Figure 1 A Partisan Twitter Network](http://nodexl.codeplex.com)

This simple grouping also illustrates that the Democratic-connected nodes form a more self-contained follows network while the Republican connected nodes have a smaller follows network and connect more to outside networks such as news media. Both Presidential candidates feature largely as central figures within their respective networks with @BarackObama and @MittRomney having the largest betweenness centralities of the entire dataset (@FoxNews follows as 3rd largest in ranking by betweenness centrality).

A singular node estranged from the Democratic network appears in this visualization and is explored in the next section.
**Almost No One Talks to Beau Biden**

Inspection of the Democratic Twitter network in isolation, an oddity became evident through multiple layout options. Figure 2 represent the Twitter network formed by the follows-relationships of @BarackObama and @JoeBiden. Most nodes connect to @BarackObama as @JoeBiden follows only 15 Twitter users and most overlap with @BarackObama. There is one exception: @BeauBiden is the only Twitter user linked to the Democratic network that does not appear to have any connection to @BarackObama.

![Figure 2 Beau Biden is prominently disconnected from Barack Obama](http://nodexl.codeplex.com)

In fact, there are only four relationships which connect Beau Biden to the Democratic network. They are “Follows” relations between @BeauBiden and @JoeBiden, @Archeleta2012, @JeremyBird, and @jomalleydillon.
The GOP Doesn’t Talk Anymore
Returning to the primary dataset for comparison purposes, the complete graph was filter such that edges with weight less than 2 and nodes with In- and Out-degree less than 5 were removed from the graph. This was done to simplify the network by removing nodes that had few connections and only rare communication. The resulting simplified graph is show in Figure 3.

Fig 3 Though present, Ann Romney and Paul Ryan do not communicate often with Mitt Romney

The three major affiliations become more pronounced without the less-frequent relationships represented. Neither party has any remaining connections to news sources. But, also apparent is the very disconnected Republican grouping of nodes. Notably, neither @AnnDRomney nor @PaulRyanVP communicated with @MittRomney beyond their single “follows” edge in the network. The three remaining frequent communications in the Republican grouping all involve @MittRomney, @LarryKudlow (host of CNBC’s @TheKudlowReport), @RomneyBus (Romney-Ryan campaign account), and @BethMyers2012.

In comparison, more edges remain in the Democratic group than the other two affiliations. Further, not all of the more-frequent communication involves @BarackObama. Several communications between @TruthTeam2012 and @Obama2012 as well as other related campaign staff such as @DavidAxelrod suggest a greater effort to reach larger Twitter audiences.

NodeXL Experience
NodeXL proved to be a powerful tool for social network analysis. Analysis breakdowns fell into two categories: Importing and Sub-Graphing. High-impact features did make the processes such as coordinating windows and autofill made the process easier, however.

Three obstacles impeded importing and dataset formation.

- Twitter does not recognize differences in capitalization when referencing usernames. For instance @barackobama and @BarackObama refer to the same user to Twitter. Within NodeXL however, they are two different vertices and can each have their own edge set which will likely
not overlap. This imposes more clean-up work on NodeXL users who have to choose a capitalization and replace all other instances of a Twitter handle. In this dataset, BarackObama, JoeBiden, and PaulRyanVP all had varying capitalization that needed reconciling.

- NodeXL clears the current workbook on import by default but does not include ego networks when importing by Twitter lists. In order to build this dataset, four separate imports were needed, one for each candidate, so that the 1.5 degree ego network could be obtained. Before discovering the option to disable clearing the current workbook on import, the process of manually copying and pasting from multiple workbooks was proving cumbersome due to multiple confusing dialogs appearing.
- Determining whether or not sufficient “follows” data had been collected was difficult. Specifically, in completing the dataset, three attempts at importing had to be made before edges appeared between @BarackObama and several of the Twitter handles for the major news networks. A manual inspection of @FoxNews confirmed that it was indeed following @BarackObama though this relationship was not represented through two initial datasets.

Creating Sub-graphs was not intuitive
Attempting to directly manipulate and explore a subset of a network in isolation proved to be awkward. It was not possible to create a manipulatable sub-graph within the larger dataset. Instead, to produce Figure 2 and explore a subset of the data, a second workbook had to be created from the larger dataset. More support for subset exploration would have made the process simpler.

Coordinated windows drive analysis again.
When working with the dataset, the coordination between the dataset graph and the underlying worksheets was extremely useful. Particularly, the ability to identify a group on the graph with selection and then scan the vertices sheet for the selected vertices enabled the identification of cluster groups and allowed the assigning of meaningful labels.

Autofill saved time and effort.
The ability to quickly manipulate all nodes within a dataset from a single GUI-based widget alleviated the burden of working with Excel. And, though autofill features would affect the entire dataset, it was still possible to manually adjust individual key nodes which brought context and anchoring to graphs.