Assignment 4: U.S. Senators on Twitter

Thomas Brambor

2019-04-10

Network Analysis of U.S. Senate Tweets

Overview

Twitter is a great tool to analyze the public interactions of political actors. For this assignment, I want you to use the information about who follows whom on Twitter as well as past tweets of the current U.S. Senate members to analyze how they interact and what they tweet about.

Data

Twitter Handles of Senators

Twitter does not allow us to search for past tweets (beyond about a week back) based on keywords, location, or topics (hashtags). However, we are able to obtain the past tweets of users if we specify their Twitter handle. The file `senators_twitter.csv` contains the Twitter handles of the current U.S. Senate members (obtained from SocialSeer (https://www.socialseer.com/resources/us-senator-twitter-accounts/) and corrected to include the newest members of the 116th Congress). We will focus on the Senators' *official Twitter accounts* (as opposed to campaign or staff members). I have also added information on the party affiliation of the Senators from here (https://ballotpedia.org/List_of_current_members_of_the_U.S._Congress).

Followers

The file `senators_follow.csv` contains an edge list of connections between each pair of senators who are connected through a follower relationship (this information was obtained using the function `rtweet::lookup_friendships`). The file is encoded such that the `source` is a follower of the `target`. You will need to use the subset of `following = TRUE` to identify the connections for which the `source` follows the `target`.

Tweets by Senators

To make your life a bit easier, I have also already downloaded all available tweets for these Twitter accounts using the following code. You **do not need to repeat this step**. Simply rely on the file `senator_tweets.RDS` in the exercise folder.

```
library(tidyverse)
library(lubridate)
library(rtweet)

# Read in the Senator Data
senate <- read_csv("senators_twitter.csv")

# Get Tweets
senator_tweets <- get_timelines(user = senate$`Official Twitter`,
    n = 3200, ## number of tweets to download (max is 3,200)
    )

saveRDS(senator_tweets, "senator_tweets.RDS")
```

```
# Read in the Tweets
senator_tweets <- readRDS("senator_tweets.RDS")

# How limiting is the API limit?
senator_tweets %>%
  group_by(screen_name) %>%
  summarize(n_tweet = n(),
            oldest_tweet = min(created_at)) %>%
  arrange(desc(oldest_tweet))
```

The data contains about 280k tweets and about 90 variables. Please note, that the API limit of 3,200 tweets per twitter handle actually cuts down the time period we can observe the most prolific Twitter users in the Senate down to only about one year into the past.

Tasks for the Assignment

1. Who follows whom?

a) Network of Followers

Read in the edgelist of follower relationships from the file `senators_follow.csv`. Create a directed network graph. Identify the three senators who are followed by the most of their colleagues (i.e. the highest "in-degree") and the three senators who follow the most of their colleagues (i.e. the highest "out-degree"). [Hint: You can get this information simply from the data frame or use `igraph` to calculate the number of in and out connections: `indegree = igraph::degree(g, mode = "in")`.] Visualize the network of senators. In the visualization, highlight the party ID of the senator nodes with an appropriate color (blue = Democrat, red = Republican) and size the nodes by the centrality of the nodes to the network. Briefly comment.

b) Communities

Now let's see whether party identification is also recovered by an automated mechanism of cluster identification. Use the `cluster_walktrap` command in the `igraph` package to find densely connected subgraphs.

```
# Sample Code for a graph object "g"
wc <- cluster_walktrap(g)  # find "communities"
members <- membership(wc)
```

Based on the results, visualize how well this automated community detection mechanism recovers the party affiliation of senators. This visualization need not be a network graph. Comment briefly.

2. What are they tweeting about?

From now on, rely on the information from the tweets stored in `senator_tweets.RDS`.

a) Most Common Topics over Time

Remove all tweets that are re-tweets (`is_retweet`) and identify which topics the senators tweet about. Rather than a full text analysis, just use the variable `hashtags` and identify the most common hashtags over time. Provide a visual summary.

b) Russia investigation - Dems vs. Reps

One topic that did receive substantial attention in the recent past was Special Counsel Robert Mueller's investigation of the Russian government's efforts to interfere in the 2016 presidential election from 2017 to 2019. Most Democrats were broadly supportive of the effort while several Republican senators supported President Trumps attempts to paint it as a partisan effort to undermine the president.

Try to identify a set of 5-10 hashtags that signal support for Robert Muellers work (e.g. #MuellerInvestigation, #MuellerReport, #MuellerIfYoureListening, #RobertMueller) while other expressed a critical sentiment towards the investigation (e.g. #WitchHunt, #fakenews, #NoCollusion).

The site ritetag.com (https://ritetag.com/best-hashtags-for/mueller) can help with that task. Using the subset of senator tweets that included these hashtags you identified, show whether and how senators from different parties talk differently about the issue of the special counsel investigation.

c) Russia investigation - Barr report

On March 24, 2019, Attorney General William P. Barr published a 4 page summary of the Mueller report, concluding that the Trump campaign did not conspire or coordinate with Russia efforts to influence the 2016 U.S. presidential election. Provide some visualization of how senators responded to the event in their Twitter communication.

3. Are you talking to me?

Often tweets are simply public statements without addressing a specific audience. However, it is possible to interact with a specific person by adding them as a friend, becoming their follower, re-tweeting their messages, and/or mentioning them in a tweet using the `@` symbol.

a) Identifying Re-Tweets

Select the set of re-tweeted messages from other senators and identify the source of the originating message. Calculate by senator the amount of re-tweets they received and from which party these re-tweets came. Essentially, I would like to visualize whether senators largely re-tweet their own party colleagues' messages or

whether there are some senators that get re-tweeted on both sides of the aisle. Visualize the result and comment briefly.

b) Identifying Mentions

Identify the tweets in which one senator mentions another senator directly (the variable is `mentions_screen_name`). For this example, please remove simple re-tweets (`is_retweet == FALSE`). Calculate who re-tweets whom among the senate members. Convert the information to an undirected graph object in which the number of mentions is the strength of the relationship between senators. Visualize the network graph using the party identification of the senators as a group variable (use blue for Democrats and red for Republicans) and some graph centrality measure to size the nodes. Comment on what you can see from the visualization.

c) BONUS ONLY: Who is popular on Twitter?

Using the twitter handles, access the user information of the senators to identify the number of followers they have (obviously, this will require to actually connect to the Twitter server). Re-do the previous graph object but now use the number of followers (or some transformation of that info) to size the nodes. Comment how graph degree centrality (via mentions) and the number of followers are related.

Submission

Please follow the instructions (/Exercises/homework_submission_instructions.md) to submit your homework. The homework is due on Monday, April 22 at 5pm.

Please stay honest!

If you do come across something online that provides part of the analysis / code etc., please no wholesale copying of other ideas. We are trying to evaluate your abilities to visualized data not the ability to do internet searches. Also, this is an individually assigned exercise – please keep your solution to yourself.