15.071: The Analytics Edge

Homework Assignment #9

Due at 5PM, Friday May 3, 2019

You must submit your assignment, including your R code, on Canvas.

Social Network Analysis of the 2017 United States Senate (100 points)

The United States Congress is the legislative (i.e., law-making) branch of the United States federal government. It is composed of two houses: the Senate, which has 100 members (two senators from each of the 50 states), and the House of Representatives, which has 435 members (whose state numbers are in proportion to each state's population).[1] Members of both houses of Congress are elected directly by the people following the American democratic process. Members of the Senate are called "Senators" and are elected by their states to a 6-year term, and members of the House of Representatives are called "Representatives" and are elected by their Congressional districts within each state to a 2-year term.

To pass a new law, or to amend/repeal a previous one, a bill must be introduced which describes the details of the new law or the proposed changes to the old law. All members of Congress, including the non-voting members, may introduce new bills, in either the Senate or the House of Representatives. The Senator or Representative who introduces the bill is known as the bill's "sponsor." Additional members of the sponsor's house may also sign on as "co-sponsors" to indicate their support for the bill. If the bill is passed by both houses of Congress, it then goes to the President to sign, who still has the power to veto the bill.[2]

In this assignment, you will explore the social network of the Senate for the 115th Congress which is the Congress that began on January 3, 2017, and concluded on January 3, 2019.[3] In this assignment,

---

[1] In addition, there are 6 non-voting members in the House: one is from the nation's capital – Washington, D.C. – and there is one from each of the five U.S. territories Puerto Rico, American Samoa, Guam, the Northern Mariana Islands, and the U.S. Virgin Islands.

[2] If the President signs the bill, or both houses of Congress override the President's veto with a 2/3 majority vote, then the bill is passed into law.

[3] There have been a number or membership changes since the Congress began. We include any Senator that served for any length of time during the 115th Congress.

you will examine the social network for the Senate. We have obtained data on all bills introduced in the Senate from `https://www.gpo.gov/fdsys/bulkdata/BILLSTATUS/115` for the entirety of the 115th Congress. Here is your assignment:

1. **Social Network of the United States Senate**
   First consider the following social network of the United States Senate. Each Senator will be a node, and we will assign an edge between two Senators if they have co-sponsored 8 or more bills together.[4] The file `senators.csv` contains a list of all Senators, including their name, state, party affiliation (R - Republican, D - Democrat, I - Independent), and an ID number from $1 - 105$. The file `senateCosponsorship.csv` contains a list of all pairs of ID numbers of Senators who have co-sponsored at least 10 bills together (excluding bills which have 11 or more co-sponsors).

   a) Begin by reading in the two files as two data frames **senators** and **senatorLinks**, respectively. Using the `R` package `igraph`, plot an undirected graph of this social network with `set.seed(144)`, `vertex.label=NA`, and `vertex.size=3`, coloring the nodes Republican = red, Democrat = blue, and Independent = white.

      i. (6 points) Attach a screenshot of this visualization. What are some interesting properties that you observe in this social network?
      ii. (3 points) How many Senators are completely disconnected in this network, and what does this mean?
      iii. (3 points) How separated are the Democrats and Republicans (the two major political parties in the U.S.)?

   b) **Social Senators.**

      i. (6 points) Compute the connected components of the social network. How many Senators are in the largest connected component of this social network?
      ii. (6 points) Plot the largest connected component and attach a screenshot, using the same layout (subsetted to the Social Senators) and `igraph` options as before. Save this layout, which we will use for the rest of the visualizations. This group can be thought of as the "Active" or "Social" Senators who have co-sponsored the most bills together. For the remainder of this problem, we will restrict our analysis to just the Social Senators. Create a data frame **socialSenators** which is only comprised of these Social Senators.

   c) **Degree.** Compute the degree for all of the Social Senators, and add this as a column to **socialSenators**.

      i. (4 points) Who has the highest degree centrality?
      ii. (4 points) How many Republicans/Democrats are in the top 10?
      iii. (6 points) Plot the network of Social Senators, modifying `vertex.size` to be $2 * \sqrt{\text{degree}}$, and attach a screenshot of your plot. Remember for this and the remaining parts to use the same layout as in part (b).

   d) **Closeness.** Compute the closeness and average shortest path for all of the Social Senators, and add these as columns to **socialSenators**.

      i. (6 points) Plot the network of Social Senators, modifying `vertex.size` to be $50 * \sqrt{\text{closeness}}$, and attach a screenshot.

   ---
   [4]We have excluded bills which have at least 11 co-sponsors as not informative of the theme of co-sponsorship.

ii. (4 points) How much variability do you observe in the values of closeness (equivalently, the average shortest path) across the Social Senators?

iii. (4 points) Which metric has more variability among the Social Senators: closeness or degree?

e) **Betweenness.** Compute the betweenness for all of the Social Senators, and add this as a column to **socialSenators**.

    i. (4 points) Who has the highest betweenness score?

    ii. (6 points) Plot the network of Social Senators, modifying `vertex.size` to be $0.2 * \sqrt{\text{betweenness}}$, and attach a screenshot.

f) **"Glue".** (6 points) In social networks, we are often interested in nodes which have high betweenness scores in spite of low degree, because these represent individuals who may have few connections themselves, but act as "glue" holding different parts of the organization together. In the Senate, this may indicate individuals who are more bi-partisan and co-sponsor bills with many different groups of Senators. Which Senators stand out as having high betweenness and relatively low degree? Which political parties are they from?

g) **PageRank.** Earlier in the course, we introduced PageRank as a means of ranking webpages, using the analogy of a taxi driver traversing hyperlinks on the web in some random fashion. In the long-run, the taxicab will visit each webpage some proportion of the time, and the most popular pages with the highest PageRank scores will be visited the most often. In a social network, we can interpret the PageRank of each node as the social influence/prominence of the person in the social network. Compute the PageRank for all of the Social Senators, and add this as a column to **socialSenators**.

    i. (4 points) Which Senator has the highest PageRank?

    ii. (6 points) Compute the correlation between PageRank and degree for the Social Senators. What do you observe?

h) **Community Detection.** Use the `cluster_spinglass()` function with `set.seed(144)` and `spins = 100` to perform community detection on the Social Senators.

    i. (3 points) What is the modularity of the computed community assignment?

    ii. (3 points) How many communities are selected?

    iii. (6 points) Plot the network of Social Senators coloring by community, and attach a screenshot. To do this, you can use the following code to obtain a color palette: `brewer.pal(N, "Spectral")`, where `N` is the number of colors that you would like in your palette. You will need the `RColorBrewer` package in order to run the `brewer.pal()` function.

    iv. (4 points) Add in this community assignment as a new column to **socialSenators**, and explore the different communities created. Which communities are entirely or mostly Republican? Entirely or mostly Democrat? Mainly bipartisan?

i) (6 points) If you could explore one thing in more depth, what would it be and why?