MY461: Social Network Analysis

LT 2019

Take Home Exam

Instructions to candidates

This paper contains five questions. Answer all questions. All questions will be given equal weight (20%). Responses for each question should be a maximum of 500 words excluding tables and figures. Please include a bibliography with any cited sources (this does not count towards the word count).

The exam questions will be released on March 27, 2019. The exam is due on <u>May 1, 2019 at 9am</u>.

Submission will be done through Moodle. Please submit your answers in a PDF file. You will be evaluated based on your responses to the 5 prompts. To help us determine where any errors were made, however, you must additionally submit an annotated R file that presents the code used to arrive at your responses.

Background information:

For this exam, we'll be focusing on a network representing bill co-sponsorship among the senators of the 110th Congress (we are not looking at the House of Representatives; just the Senate), which was from January 2007 to January 2009 – the last two years of George H. W. Bush's presidency. At this time, the Democrats had a majority in both the House and the Senate.

You can find some background on the senate and its members at that time <u>here</u> and <u>here</u>.

We are providing you with two files:
The first file (110_sen.csv) has basic information on each of the 102 senators. The variables are as follows:
- Name
- Party – the political party of the senator. (I have recoded some Democratic-Farmer-Labor, Independent Democratic, and Independent parties as Democratic, since they are sufficiently similar for our purposes)
- Gender
- Religion
- Class – this specifies when their seat will be up for election (see more <u>here</u>)
- State – each of the 50 US states has two senators (our dataset has two additional members, who were replacing someone else: John Barrasso replaced Craig Thomas after his death, and Roger Wicker replaced Trent Lott after his resignation; all four are Republicans).
- CensusRegion – this divides the states into 9 divisions, following the UC Census Bureau designations (see <u>here</u>).
  - 1 = New England
  - 2 = Mid-Atlantic
  - 3 = East North Central
  - 4 = West North Central
  - 5 = South Atlantic
  - 6 = East South Central
  - 7 = West South Central
  - 8 = Mountain
  - 9 = Pacific
- PriorExperience – previous job held
- Education – highest degree(s) and the degree-granting institutions

- FirstTookOffice – the year they first took office
- Born – the year they were born

The second file (110_billspon.csv) records which bills each senator sponsored. Each column represents a senator, and each row represents a bill. When there is a value in a cell, that represents a senator sponsoring a bill. You may notice that there are different values in the cells; for our purposes, we simply leave the file as is and disregard the values. (In other words, we want you to leave the values as they are and not try to change them by, e.g., recoding them all as 1, *but* we also do not expect you to try to interpret any meaning from the particular values).

Using these files, create a network of the senators, with edges representing the number of bills they cosponsored. Additionally create a network that only retains edges between senators when the number of bills they co-sponsored is in the top quartile (so, at or above the 75th percentile; with this, we are retaining only the edges between senators that are especially strong, as they co-sponsor many bills together). Below, we refer to this new network as the "simplified network," and the main network with all edges as the "full network.")

With these networks in hand, answer the following questions.

1   Consider the overall metrics (density, average path length, transitivity) of both the full and the simplified networks. How do the two networks compare to each other? Compare the simplified network to random networks created with the Erdős–Rényi and the configuration models. What do these comparisons tell you about the nature and structure of the relationships among the senators? In your answer, make sure to define each of the metrics and give an intuitive interpretation for them.

2   Who do you see as the most influential senators in the 110th Congress? Identify two potential meanings of "influence," as proxied by different centrality measures. Justify your choice of each centrality measure. In that justification, present clear interpretations of what each centrality measure is capturing about the position of senators. For each of your chosen centrality measures, identify the senator who has the highest value in the simplified network. Discuss why you think that senator is the most central in the network. Make explicit reference in your response to the concepts of social capital and brokerage/structural holes covered in the course material. [Note that if you use edge weight for a centrality measure, you need to identify how the calculation interprets those weights (i.e., do higher values mean greater closeness or greater distance?); in those cases where the measure assumes that higher values means greater distance, you should use 1/weight in the calculation.]

2   How does political party membership influence bill co-sponsorship? Calculate the assortativity by political party and the probability of a co-sponsorship tie within and between each party (using a blockmodel approach) for the simplified network. Run the Louvain community detection algorithm on the full network and see how the resulting communities align (or not) with political party. Plot the full network twice (with the same layout) with nodes coloured by (1) the results of the community detection algorithm and (2) the vertex attribute which you identify as most closely aligning with the resulting communities (include a figure label). Discuss what the results of these analyses imply about the political parties at this time.

4   Do different senators seem to fulfil different roles in the senate? Evaluate the structural equivalency of the senators in the full network and use this to divide the senators into four equivalency classes. Interpret the four groups identified, drawing on information regarding party and leadership. Discuss what this implies about the senators, drawing on the concepts of position and role.

5   What helps predict whether senators cosponsor bills? Consider the exponential random graph model for the simplified network shown below. Interpret each term in the ERGM (except edges) – how does each term influence whether or not two senators are connected? In your discussion, make explicit

reference to the concepts of homophily and transitivity. Calculate the fitted probability of a tie between: two women from the same region, who joined in the same year, and who have both cosponsored a bill with one other senator; once, when they are both Democrats and once when they are both Republicans.

```
========================================
                            Model 1
----------------------------------------
edges                       -6.742 ***
                            (0.715)
nodefactor.party.Republican -1.204 ***
                            (0.035)
nodematch.party              1.841 ***
                            (0.069)
nodefactor.gender.Male      -1.110 ***
                            (0.128)
nodematch.gender             0.647 ***
                            (0.150)
nodematch.region             0.663 ***
                            (0.112)
absdiff.joined              -0.007
                            (0.004)
gwesp.fixed.0.6              3.724 ***
                            (0.375)
----------------------------------------
AIC                         4050.097
BIC                         4102.473
Log Likelihood              -2017.049
========================================
*** p < 0.001, ** p < 0.01, * p < 0.05
```

Estimates (with standard error in parentheses) of the coefficients for each term in the exponential random graph model, which predicts the log-odds of a tie in the simplified network. The GWESP term has an alpha value of 0.6.