Homework/Problem Set 2

Statistical Analysis of Networks (CRJ) 605

This assignment has two parts. This assigment is **DUE** 3/14.

Part A:

For the first part of the assignment, you will work with networks from the PINS study. The networks are:

- The *Get Along With Most* network, in edgelist format (available at: https://justicecenter.la.psu.edu/research/projects/files/pins-get-along-with-edgelist)
- The *Power/Influence* network, in edgelist format (available at: https://justicecenter.la.psu.edu/research/projects/files/pins-data)

For *each* network (i.e. *Get Along With Most* and *Power/Influence*) do the following:

1.a. Import the edgelist into *R* and create a directed graph that is an object of class `network`.

2.a. Calculate the indegree centrality, outdegree centrality, and betweenness centrality for each actor.

3.a. Calculate the standardized centrality scores (i.e. indegree, outdegree, and betweenness) for each actor.

4.a. What is the correlation between indegree centrality and betweenness centrality? What does the correlation tell us, conceptually?

5.a. Find two nodes that have the same indegree centrality, but differ with respect to their betweenness centrality. Why is there a difference? (In other words, why is there variation in betweenness centrality among these two nodes but not variation in indegree centrality?)

6.a. Calculate the mean centrality scores (i.e. indegree, outdegree, and betweenness).

7.a. Interpret the value of the mean for each centrality measure *and* compare the means for each measure.

8.a. Calcuate the graph centralization for indegree, outdegree, and betweenness centrality.

9.a. Compare the three graph centralization scores.

Now, compare the networks:

10.a. How do the indegree, outdegree, and betweenness centrality scores differ for the *Get Along With Most* and *Power/Influence* networks?

11.a. What do the differences in the graph centralization scores for these networks tell us?

12.a. Overall, how are the *Get Along With Most* and *Power/Influence* networks different with respect to centrality?

BONUS Question:

13.a. Symmetrize each network using the `rule="weak"` argument in the `symmetrize()` function. Examine the coreness of each symmetrized network using the `kcore()` function. How do these networks differ with respect to their *cohesiveness*?

Part B:

For this part of the assignment, you will work with one network created from the course survey data and one network created from the Boston Special Youth Project. The two networks are:

- The *3 Shows Watched on TV* network, in adjacency matrix format (available at: https://www.jacobtnyoung.com/uploads/2/3/4/5/23459640/crj_605_networks_spring_2019_tv_w1_adjancency.csv). These data represent students' responses to watching television.
- The *Boston Special Youth Project Affilation* network, in edgelist format (available at: https://www.jacobtnyoung.com/uploads/2/3/4/5/23459640/edgelist.bipartite.gangs.csv). The data come from social worker field notes collected during the Special Youth Project, a gang intervention conducted in Boston during the 1950s. As a feature of the study, social workers recorded events and who was present at the events over a 3 year interval. Each edge represents an individual's (i.e. gang member's) presence at an event as recorded by a social worker. There are 166 unique gang members in the first mode and 33,487 unique events in the second mode.

For *each* network (i.e. *3 Shows Watched on TV* and *Boston Special Youth Project Affilation*) do the following:

1.b. Import the edgelist into *R* and create a `bipartite` graph that is an object of class `network`.

2.b. Generate a plot of the network using the `gplot()` function where nodes in the first set are a different color, a different size, and a different shape than nodes in the second set.

3.b. What is the density of the graph? What does it mean?

4.b. Plot the degree distribution for each mode. Describe the properties. How are the distributions different?

5.b. Calculate the standardized degree centrality scores for each node set.

6.b. Calculate the mean degree centrality for each node set. Interpret each mean.

7.b. Calculate the dyadic clustering coefficient using the `reinforcement_tm()` function in the `tnet()` package. Interpret the coefficient.

Now, compare the networks:

8.b. How do the degree centrality scores differ for each node set of the *3 Shows Watched on TV* and *Boston Special Youth Project Affilation* networks?

9.b. How do the dyadic clustering coefficients differ across the networks? What does the difference mean, conceptually?

BONUS Question:

10.b. Calculate the graph centralization score for degree centrality of each node set for each network. Interpret the scores.