October 20, 2020

Social Network Analytics, Empirical Exercise #1

Due on Wednesday, October 28, 2020 at 8:00am

Analyzing centrality, tie strength, and reachability in a classroom network

Loading classroom data

For this exercise, we will analyze centrality, tie strength, and reachability measures on a different classroom network than the network of MBA students we looked at in class. This network can be read from the file `social_and_task_network.csv`.

In this network, we have two types of ties:

1. `social_tie` indicates whether a student reports having a social relationship with another student
2. `task_tie` indicates whether a student reports working with another student on course-related tasks inside or outside of class

This network has several important characteristics:

- Both of these relationships are directed, that is, not necessarily symmetric or reciprocated. For example, Student A may report having a relationship with Student B, while Student B may not report having a relationship with Student A.
- This network is in the form of an edge list, where each row represents a potential tie.
- "Ego" is the focal node, or vertex, in the network—ego "sends" ties.
- "Alter" is the non-focal node, or vertex, in the network—alters "receive" ties from ego.
- The number given in the `social_tie` and `social_tie` columns represents the strength of each tie—this is a weight. If this number is given as 0, then no tie exists between ego and alter for this type of relationship.

1. First, consider the social and task ties as separate networks.
   (A) Use igraph to generate indegree, outdegree, closeness, betweenness, and PageRank centrality statistics for each individual the social and task networks.
   (B) Compute the correlations of the five centrality measures you generate for the social network with the five measures generated for the task network. Which measures in the task network are most closely related to those in the socializing network? Name at least one insight can you draw from the relationships between these five measures across the two networks.
2. Next, consider the social and task ties together, as two distinct types of ties comprising one network.
   (A) Suppose that a tie is strong if it is above the mean strength for that type, conditional on the tie existing—do not include weights of 0 in the calculation of the mean. Under this definition, does the network satisfy Strong Triadic Closure? Come up with a solution that illustrates this (1) visually, in a plot, as well as (2) programmatically, by giving the number or proportion of ties that are violation of Strong Triadic Closure.
   (B) Now suppose that a tie is strong if it is above the median strength for that type, conditional on the tie existing. Under this definition, does the network satisfy Strong Triadic Closure? What insights does this illustrate about these interactions within the network?
3. Continue to treat the social and task ties as two distinct types ties comprising one network.

(A) It is also possible to compute betweenness on the edges in a network, as well as the vertices. This is a good measure of the flow of information and resources through a network. Calculate the edge-level betweenness for both of the types of tie.

(B) Does it seem like edges with high betweenness tend to be strong or weak ties, according to our two definitions above? Does this result make sense?

4. Continue to treat the social and task ties as two distinct types of ties comprising one network. How many pairs of nodes do not have walks between one another? Find a solution that performs this calculation directly on the matrix—it is possible to verify this solution via igraph afterward.

5. The network-level measure of degree centrality is a good indicator of the dispersion of the degree distribution in a network. Generate and plot a network in R in which the network-level measure of degree centrality,

   $C_{degree} = \sum_{i=1}^{n} \frac{C_{degree}(v^*) - C_{degree}(v_i)}{(n-1)(n-2)}$

   is equal to 1, and another where it is equal to 0. Would this relationship hold true for these networks for other measures of centrality, such as closeness or betweenness?