Social Network Analytics, Empirical Exercise #3
Due on Tuesday, November 7, 2017 at 8:00am

Clustering and core-periphery structures in venture capital co-investment networks

1. There are two files, "Funding_events_7.14.csv", and "Funding_events_7.14_page2.csv", containing information on venture capital investment events, from June 1981 until July 2014. Perform the small-world exercise for the venture capital co-investment network, and determine which firm is the center of the network as of July 2014. Compute the average shortest path length between all firms in July 2014. Why is this number so high?

2. Reproduce the figure from class that plotted the coreness of the venture capital network over time, using monthly snapshots to update the network. On the x-axis should be the age of the network in months. On the y-axis should be the highest-degree k-core each venture capital firm belongs to, averaged over all firms for that monthly snapshot.

3. Recreate the figure from Question 2, but now allow ties to "decay." Remove ties from the network if they are not renewed within 10 years. Does the figure appear much different than before? What does this suggest about the nature of relationships in the network?

4. Let's also take a look at a clustering approach. For the June 1991 network, use the iterative correlation clustering method, which takes the correlation of the original adjacency matrix and keeps taking the correlation of this result until it converges to a matrix of 1s and -1s. Use this method to split the network into clusters of cliques, and plot the result. Perform this calculation directly on the matrix.

5. Do you think the venture capital tends to exhibit more of a core-periphery or clustered/component structure? Describe a couple pieces of evidence for whichever opinion you choose.

Extra challenge problem: Another method for clustering networks relies on the betweenness scores of the network edges. The method calculates the betweeness centrality score for all edges, and then removes the edge with the highest score. It then recalculates the betweenness for the remaining edges, and again removes the one with the highest score. For the June 2001 graph, would this method produce the same number of clusters as the correlation method?