October 25, 2021

Social Network Analytics, Empirical Exercise #2

Due on Friday, November 5, at 11:59pm

Group structures, clustering, and core-periphery analysis in venture capital co-investment networks

Loading venture capital data

For this exercise, we will analyze groups and clusters in the networks of venture capital firms. Venture capital firms often co-invest with each other on startup venture as part of syndicate investment teams. The ties formed by co-investing together influence the strategy and performance of the venture capital firms and the entrepreneurs they fund, and certain positions in this network are more beneficial than others.

Information about venture capital firms and their investments is contained in several files:

Files generating network data

The two files "Funding_events_7.14.csv" and "Funding_events_7.14_page2.csv" contain information on venture capital investment events, from June 1981 until July 2014.

- In these files, each row represents an investment round made by one or more venture capital firms into a startup company
- The venture capital firms are listed in the column "Investors"
- The startup company is listed in the column "Portfolio Company Name"
- Consider a relationship tie to exist between venture capital firms when the firms invest together in the same round of a portfolio company—the firms show up in the same row together
- Consider firms as tied together if they invest together at least once—ignore multiple instances of a relationship
- Allow relationships to persist over time, so that the network in July 2014 is comprised of all cumulative ties that have occurred up to this point

Files generating venture capital performance data

The file "Venture_capital_firm_outcomes.csv" contains information about the performance of venture capital firms in each year they are active investors and also contains some information about the venture firms themselves.

- The number of successful investments a venture capital firm has in any year is listed in the column "successful_investments"—successful investments represent acquisitions, IPOs, and other events in which the startup generates cash for the venture capital firm
- A venture capital firm goes out of business when it cannot raise a new fund from a limited partner to make new investments—firms that go out of business have a 1 in their most recent observation in the column "out_of_business", and otherwise this column is filled by 0s

Exercises

1. First, perform the Kevin Bacon Hollywood Actor exercise on the venture capital firm network.

   For constructing this network, we want to avoid the possibility that the network simply grows more dense over time, so we want to allow old ties to drop out of the network. We will consider a tie to have "decayed" if it is not renewed within a certain amount of time: a time window greater than 90 percent of all renewal windows for all ties in the network.

These decayed ties should be removed from the current construction of the network. Use this trimmed network for the remainder of the exercises.

Which firm is the center of the venture capital firm network as of July 2014? Consider the most central firm to be the firm with the highest closeness centrality, as in the Hollywood Actor example.

2. Next, we will look at the development of the local group membership of the co-investment network over time. Allow the network to be updated monthly for each month t in the data, adding the new ties that occur through investments in the current month to be added to the existing network of ties that have occurred in previous months.

   In Class Session 4, the figure on Slide 24 plotted over time the industry average of the highest-degree k-core each venture capital firm in the co-investment network belonged to. When a node is a member of a k-core with a high degree, its surrounding ties are very dense. When many nodes are members of k-cores with high degrees, this suggests that there may exist dense subgroups within the network.

   Construct a figure similar to Class Session 4's, plotting the average k-core of each venture capital firm in the network over time. This can be computed using the igraph function `coreness`. On the x-axis should be time. On the y-axis should be the highest-degree k-core each venture capital firm belongs to, averaged over all firms in the network up to that month.

3. Next, we will look at the development of the venture capital firm co-investment network in terms of its global core-periphery structure. Allow the network to be updated monthly, as in Question 2.

   (A) Does the network appear to conform to a core-periphery structure?

   Calculate the co-investment network's concentration scores to determine if it tends towards a core-periphery structure across the data. Illustrate a figure, with one plot for one month from each calendar year in the data, that shows the range of concentration scores for each partition size p–for the range of p in 1 to the number of nodes in the networks–in the network for that month's cross-section. You can exclude the very early period of the data when all of the firms have the same eigenvector centrality.

   (B) Provide one other piece of descriptive evidence outside of the concentration scores to support your conclusion. You can see Slide 29 of Class 4 for some examples of evidence to use.

4. Next, we will compare the core-periphery structure of the network to the potential for a clustered structure. In Slide 25 of Class 4 we saw that the network tended towards a strong core-periphery structure at the end of the animation in the year 2000. Before this point, some clustering structures might be more apparent. Let's try to see if a clustering structure exists in the year 1996.

   If we tried to cluster the network in June of 1996 using partitioning around medoids, what would be the recommended number of clusters? Are any of the clustering solutions suitable under the rule of thumb of achieving an average silhouette width greater than 0.7 (or even the weaker threshold of 0.5)? What does this suggest about the clustering approach versus the core-periphery approach?

   For the clustering, you can use the function `pam` from the `cluster` library. You can use the default distance matrix of the network provided by `proxy::dist` or `dist` as the input for the clustering. Show the silhouette plot for the best-fitting clustering solution for the June 1996 network.

   In addition to the reference guide on Slide 60 of Class 3, you may also find useful this short example using the iris dataset that is built in to R. You can load this data using `data(iris)`.

Extra Credit (2 points)

Last, we can analyze whether different kinds of centrality help venture capital firms and the entrepreneurs they work with to perform better. You may use whichever statistical approach you wish to determine the direction and strength of the relationship between network position and a venture capital firm's performance.

(A) Is being more central in terms of the measures we have encountered so far–degree centrality, closeness centrality, betweenness centrality, eigenvector centrality, and PageRank centrality—related to having more more successful investments in a given year?

The outcome variable of successful investments is a non-negative integer, so the count family models can be useful. Some approaches are described at https://cran.r-project.org/web/packages/pscl/vignettes/countreg.pdf.

(B) Similar to (A), is a venture capital firm being at the center of the network related to being less likely to go out of business?

The outcome variable of going out of business is an event that can happen once, and the likelihood of this event depends on how long a firm has been in business. As a result, the survival family of models can be useful. Some approaches are described at https://www.r-bloggers.com/survival-analysis-with-r/.