October 22, 2018

Social Network Analytics, Empirical Exercise #6
Due on Wednesday, December 5, 2018 at 11:59pm

Familiarity versus investor quality in venture capital investing

Setting up the status hierarchy: In the last assignment, we analyzed the influence of investors' status at the firm level on the effectiveness of their diversification strategies. In this exercise, we will analyze the investor network not at the level of the firm, but at the level of the individual. In an ideal world, investors would select entrepreneurs to invest in based on the merits of their quality—for example, their past performance. In reality, however, investors often choose to invest in for other reasons, such as past relationship history, connections to mutual investors or entrepreneurs, or homophily. The ERGM family of models can help us adjudicate between the different reasons an investor might choose to invest in an entrepreneur or not. The ERGM documentation at https://cran.r-project.org/web/packages/ergm/ergm.pdf should be helpful for setting up the data and specifying the terms.

- The file "execs.csv" contains information on startup executives and the startup companies they represent. These are each given by the unique identifiers "PersonId" and "CompanyId."
- The file "investors.csv" contains information on individual investors and the startup companies that they choose to fund.
  - The unique identifiers "PersonId" and "CompanyId" are consistent with the execs file
  - The firm an investor represents is given by the unique identifier "InvestorId"
  - The specific deal that generates the "invests in" relationship between investor and executive is given in "DealId"
- The file "deals.csv" contains information on each individual deal, such as its type and size.
- The file "investor_firm_details.csv" is identical to the "investor_firms.csv" file from the previous assignment, with the new inclusion of each investor firm's geographic location.
- Finally, the file "people.csv" contains background information about each person, such as their education and gender.

1. First, we want to know if network familiarity influences an investor choosing to invest in a particular startup executive. ERGM models allow us to estimate a regression in which the outcome variable is the network itself. We would like to estimate the relationship, "chooses to invest in," which can be represented as a directed network from funding PersonIds to funded PersonIds.

   Consider investments to be all financing deals outside of Buyouts, Mergers/Acquisitions, and IPOs. To make sure that the nodes in the network are likely to be drawn from a more established community, limit the analysis to only deals in which all participating investors are from US cities with at least 1,000 investor firms. For similar reasons, and to limit computational strain, also only consider deals from the year 2000 onward.

   Run an ERGM model, using 20 iterations, that predicts the likelihood of a "chooses to invest in" relationship as a function of the edges in a model, as well as the presence of triangles, which represent the influence of "friends of friends" on the likelihood of funding.

   What do the results suggest about potential network familiarity effects in investment decisions?

2. Next, we want to include information on the investors and executives to determine whether there is homophily in choosing to make investment decisions.

   Run an ERGM model, using 20 iterations, that predicts the likelihood of a "chooses to invest in" relationship as a function of the edges in a model, as well as the presence of triangles. This time, also include the effect of having the same gender and of the investor and the executive both having an MBA.

   What do the results suggest about homophily in investment decisions?

3. Last, we want to know how these results might be affected if we include information about the quality of the startup executives. If investors are choosing based on merit, when we include quality metrics for the entrepreneurs these should dominate over any other predictors.

   Run an ERGM model, using 20 iterations, that predicts the likelihood of a “chooses to invest in” relationship as a function of the edges in a model, the presence of triangles, the effect of having the same gender, and of the investor and the executive both having an MBA. This time, also include the total number of different companies the startup investor has worked for, as well as the total number of successful deals the executive has been a part of. Successful deals can include Buyouts, Mergers/Acquisitions, and IPOs.

   What do these new results suggest about what drives investors’ decisions to invest in an entrepreneur or not? Is this problematic for these industries?

Extra Credit: Dynamic Analysis of familiarity versus investor quality—5 points

The data on individual investors and startup executives grows over time. We can take advantage of this feature to run a more powerful set of models from the Siena family, which uses the longitudinal nature of the data to predict how the network evolves over time. Benefits of the Siena model are that it can predict both ties as well as behaviors, and can take into account nodes joining and leaving the network over time. The model also accounts for ties not reappearing in future periods, rather than assuming that they last forever, so it gives a richer sense of the dynamic decision-making of individuals in the network.

We did not have time to cover Siena in detail in class, but it would be an applicable model for this data. The example script “rsiena teenage drug and alcohol usage.R” walks through an example of using the Siena modeling technique on a dataset that predicts smoking and alcohol consumption at a high school based on the network connections and behavioral similarity of the school’s students. The data for this analysis are included as well.

You can adapt the approach to an analysis of the investor to startup executive network as well. To set up time windows for analysis that will not be too computationally intensive, you can use waves, rather than years, grouping the firms into three waves from 2000-2018. Be creative in your Siena modeling, and see what interesting effects you can find. The algorithm can take some time to run, especially if it needs to be re-run in order to converge, so you may want to run it overnight or during idle time. For Mac users, remember that you can open multiple instances of any application by opening the terminal and typing in “open -n” and the name of the application; for example “open -n “R””.