Nobemver 12, 2018

Social Network Analytics, Empirical Exercise #5
Due on Tuesday, November 20, 2018 at 12:00pm

Diversification and social status in venture capital

Bringing in the status hierarchy: Earlier in the course, we looked at how the co-investment networks of venture capital firms influenced different aspects of their performance. In this exercise, we will analyze how the co-investment networks of venture capital firms influence their strategies, in terms of the types of startup companies that they invest in. Investors are constantly trying to find the next, best project to work on, but how to diversify into new industries is often unclear. Diversification is challenging because it requires new and distinct skills and expertise. Firms with high social status may be more effective at managing firms in disparate industries because they are able to leverage their position in order to get more recognition for their ventures, even if they require distinct sets of skills.

First, there are two files with information about the investor firms and the startup companies that they invest in.

- The file "investor_firms.csv" contains information on investment firms and the number of successful investments, in terms of acquisitions and IPOs, managed by each firm. Each investor firm is also given a unique key that can be used to relate it to the other data files.
- The file "startup_companies.csv" contains information about the industry area the firms that are invested in by venture capital firms are in, and the status of whether the startup is earning revenue or not.

Next, there are two files that help relate the investor firms and startup companies to one another.

- The file "investors_and_deals.csv" provides the relation between investor firms and the deals that they are a part of.
  - Investors are listed by their unique identifier
  - Deals that investors participate in are listed by another unique identifier
  - The column "lead_investor" indicates whether a firm was a lead investor on a deal
- The file "startups_and_deals.csv" provides the relation between startup companies and the deals that fund them.

1. First, we want to know if diversification is an effective strategy for venture capital firms.
   (A) Create a plot, such as a scatterplot or loess curve, that indicates whether firms with more diversified portfolios have more successful investments or not. Define diversification as the number of industry areas a firm is invested in, divided by its total number of investments.
   (B) Run a appropriate regression predicting the number of successful investments as a function of diversification. Is diversification related to having more successful investments?
2. Next, we want to know if higher-status firms are more successful as well. A firm that serves as a lead on investments takes on a more visible role and is responsible for mobilizing financial, as well as social, support for the investment. As the lead, it can exert influence over the direction of the venture as well as the other investors that are involved with the project.

   We can define a status relationship for a pair of firms A→B as the proportion of times that Firm A has served as a lead investor in deals it has participated in with Firm B:

   $$status_{AB} = \frac{count(deals\ with\ B\ in\ which\ A\ is\ lead)}{total\ deals\ with\ B}$$

   Let this proportion be the entries of a matrix representing relationship between each of the investors—that is, a network. Then, each investor's status can be represented as its eigenvector centrality in this network. In this way, an investor's social status is represented by its ability to be a leader on deals, as well as to be connected to other firms that lead their own deals.

(A) Create a plot, such as a scatterplot or loess curve, that indicates whether firms with higher status have more successful investments or not.

(B) Run a appropriate regression predicting the number of successful investments as a function of status. Is status related to having more successful investments?

3. Which investors are more effective at diversifying their portfolios? In a regression, it is possible to determine how two variables jointly predict an outcome by specifying an "interaction" term. This term can be included in the model by including a term that multiplies the variables together—`variable1*variable2`—in the formula of the model. If this term is positive, it indicates that there is a synergistic effect of the two variables—high levels of both together have a positive effect on the outcome variable, above and beyond the effect of each variable individually on its own.

   (A) Run a appropriate regression predicting the number of successful investments as a function of the interaction of status and diversification. Is this interaction related to having more successful investments?

   (B) Create a plot that illustrates the interaction effect predicted by the regression. We can accomplish this using a 3D scatterplot generated by the fitted values of the model. Assign the model that you estimated in Part A to an object, below called "model." The plot can be then executed with the following code:

```
# set up scaled grid of (x,y) values
diversification = seq(0,1000, by=20)
status = seq(0,1000, by=20)
values = expand.grid(diversification=diversification, status=status)

# prediction from the model
values$successful_investments = predict(model,newdata=values)

# regular 3d plot
scatterplot3d(values$diversification, values$status, values$successful_investments)

# interactive 3d plot you can move around
plot3d(values$diversification, values$status, values$successful_investments)
```

   The command `scatterplot3d` is from the package "scatterplot3d", and the command `plot3d` is from the package "rgl."

4. Last, it is worth investigating whether the benefits to diversification are symmetric to both investment firms and startup companies. The incentives for investors and their firms are not always aligned. Investment firms often search for projects with outsize returns—if they can find just one of these "unicorn" firms, out of twenty, this can be a successful strategy for the investor. On the other hand, entrepreneurs would like to become profitable and stay in business, and might prefer to be supported by an investor with more consistent, rather than dispersed, outcomes for its startup companies.

   The "startup_companies.csv" file indicates what state each company is in with regard to profitability. Companies can be generating revenue, out of business, or fully profitable, among other outcomes.

   So far, we have used regression to predict numerical outcomes. It is also possible to use regression to predict categorical outcomes as well. We will use a "multinomial logit" to predict the likelihood of a company being in any particular state. Consider four states:

   (a) Ramp-up: in clinical trials or startup phase

   (b) Generating revenue: generating revenue (profitable or not profitable)

   (c) Profitable: fully profitable

   (d) Failed: out of business or bankrupt

Set up the multinomial logit using the command `multinom` from the "nnet" package. This can be executed with

`model = multinom(startup_state` $\sim$ `diversification + status + diversification*status, data)`

This command does not estimate p-values for statistical significance on its own, so instead these can be estimated by first computing the z-scores of the coefficients:

`z = summary(model)`$coefficients/summary(model)$`standard.errors`

and then conducting the significance test:

```
(1 - pnorm(abs(z), 0, 1)) * 2
```

If the value returned is below 0.05, then we can conclude that the predictor was significant.

What does this model suggest about the benefits or drawbacks of diversification and status for startup firms, as opposed to investment firms?