

Graph Analysis to Predict Investment

Samhita Karnati

Advisor: Prof. Andrea LaPaugh

Motivation & Goal

- If I were to start a company, and I needed to get funding, is there a more efficient way to find venture capital than just asking everyone or using people I'm connected to?
- Is there a way to know which startups have a higher likelihood of getting funding from particular VCs?

Goal: Use history of funding rounds to determine the probability of getting funded by a particular VC.

Related Work

Previous work can be grouped into two larger categories

1. Network analysis for later-stage financial events
 - a. Mergers & Acquisitions prediction
 - b. IPO prediction
2. Categorical and non-financial features for startup similarity
 - a. What features are common among successful startups?
 - b. Do startups in similar categories (i.e. transportation, healthtech, fintech, etc.) have similar features?

Approach & Implementation

Create a directed graph of money flow between startups and VCs using the crunchbase dataset.

Graph analysis: link-prediction problem

- Jaccard similarity coefficient: analysis of common neighbors

$$jaccard(i, j) = \frac{|i \cap j|}{|i \cup j|}$$

- Adamic-Adar similarity measure: neighbor analysis, but reduce significance of nodes that have lots of neighbors

$$\sum_{z \in \Gamma(x) \cap \Gamma(y)} \frac{1}{\log |\Gamma(z)|}$$

Results & Evaluation

Results have not been evaluated yet, but on cursory look of similarity coefficients, results make sense.

Plan:

- ROC curve: receiver operating characteristic curve
 - Illustrates the performance of a binary classifier system
 - True positive rate against false positive rate at various threshold settings
 - Maximize area under the curve

Future Work

Jaccard similarity and Adamic-Adar similarity measures give the probability that an occur does or does not happen -- with investments, we would also like to know how much money will flow.

Similar VCs invest in similar types of companies -- can money flow be used to predict the space a company is in or other categorical features?