A social network comparison of Corporate Venture Capital and Venture Capital firms

Anuj Menta
New York University
New York, USA
menta@nyu.edu

Cole Smith
New York University
New York, USA
css@nyu.edu

Nivedita Ganesh
New York University
New York, USA
nive@nyu.edu

Aneri Dalal
New York University
New York, USA
and301@nyu.edu

***Abstract*—The objective of this study is to draw a quantitative contrast between Corporate Venture Capital (CVC) firms and Venture Capital (VC) firms. We construct a social network where a node is either a CVC firm or a VC firm and an edge exists between two nodes if both organizations have made a mutual investment in a company, at any point in time.**

***Index Terms*—social networks, graph theory, venture capital**

I. Introduction

The purpose of this study is to consider the entire world of start-ups and Venture Capital as a social network and quantitatively compare the Corporate Venture Capital (CVC) firms and general Venture Capital (VC) firms.

II. Data

The data gathering process is a fundamental aspect of any study. We were able to gather all our data from a single source, Crunchbase. At the time of our data collection, there were 17,034 VC firms that were considered for the study. These VC firms were tagged as "Venture Capital" and "Corporate Venture Capital" by Crunchbase and we assume that this classification is accurate.

We then used Networkx, a Python package, to plot the nodes and edges we extracted from Crunchbase in order to create the network that we used throughout the tests. We also used Pandas, another Python package, to read and edit the data we obtained from Crunchbase.

The size of the network made it very challenging to visualize the network and calculate key metrics as a part of the study. Regardless, we managed to extract the necessary parameters at this scale.

The first thing we did after we obtained the data was to build the following visualizations based on the data sets to better understand them.

- Distribution of Companies Founded by Location
- Distribution of Companies by Type Over Time
- Number of Investments Made by VC Firms Over Time

III. Related Works

In-depth studies have been conducted in the context of CVC syndication. Zheng in [1] provides a detailed overview on the importance of social capital among CVCs, and their value as Structural Holes in their networks. Namely, the study finds that the most prominent CVCs act as "social brokers," and tend to syndicate with other prominent CVCs.

In [2], Eugene et al. explains how funding investors often make their investments along social ties, contrary to past research on the subject. They claim that investors are more likely to invest in a particular company if they have a strong social connection with that company. This follows for both direct and indirect relationships. They form their network in a similar way to our study, in that their network is based upon Crunchbase data. Their methodology involves a holistic approach by comparing metrics such as common neighbors, shortest path, Jaccard Coefficient, and others.

IV. Hypotheses

We frame the following hypotheses about the interaction among CVCs and VCs:

Corporate VC firms invest heavily in start-ups working on industries aligned with their parent company as compared to VC firms.

One way to measure this phenomenon is what we call the *Diversity Coefficient* (See V-A), which is structured to quantify the diversity of the type of companies a CVC/VC firm invests in. The higher the diversity coefficient, the less spread out the investments are.

The effective size of Corporate VC firms in a social network of Venture Capitalists is higher than the effective size of VC firms.

With this proposal, we corroborate the findings in [1] that CVC firms exist in their networks as Structural Holes. The likelihood that a node in a network is a Structural Hole can be modeled by the Effective Size of their network (See V-B).

A social network of VC firms is more closely knit as compared to a social network of Corporate VC firms.

The bipartite social network formed by VC firms and start-ups obey the power law (rich get richer phenomena) and exhibits a very high network density (> 0.5).

A network that follows the power law is also called a scale-free network (See V-D). A scale-free network has what we can call "hubs", which are often the highest-degree nodes.

The network of Corporate VC firms exhibits a greater number of Strong Triadic Closures than the network of VC firms.

We measure the propensity for Strong Triadic Closures using a related metric called the *Clustering Coefficient* (See V-E). This coefficient is related to the number of Triadic Closures in the network.

V. Methodology

In order to run our experiments, we source data from Crunchbase using custom parsing programs, which perform a Breadth First Search over the Organizations that Crunchbase tracks. An Organization in this context can be either a VC, CVC, or a start-up. We differentiate a start-up from a CVC/VC by assessing if the Organization contains an "investments" field. Crunchbase does not make a clear distinction between CVCs and VCs, therefore, we use a predefined list of Corporate VCs. We then define the world of CVCs and VCs as a social network graph. Each node in this graph is either a CVC or VC, and each edge between the nodes represents if *both organizations have made a mutual investment in a company, at any point in time.* This network is the basis for the assessment of our hypotheses.

A. Diversity Coefficient

One important observation about a CVC firm is that their investments are heavily focused in companies working in industries which either overlap with interests of their parent company or are very similar.

To quantify that diversity, we define a metric called the *Diversity Coefficient*. The Diversity Coefficient is structured to quantify the diversity of the type of companies a VC firm invests in. The higher the Diversity Coefficient, the less spread out the investments are. The Diversity Coefficient is defined as *the Standard Deviation of the frequency distribution of types of companies a VC firm has invested in.*

B. Effective Size

The *Effective Size* of a network, from a given node, is the number of non-redundant contacts among its immediate contacts. We use Burt's Formula where the Effective Size of node i's network is equal to:

$$\sum_j \left[1 - \sum_q p_{iq} m_{iq}\right], q \neq i, j$$

The summation indicates the redundancy of connections for each of i's contacts, j. This is related to Structural Holes, which are influential nodes that link larger networks. They act as information liaisons.

C. Network Density

The *Network Density* indicates how tightly knit the entire network is. The Network Density is defined as the number of edges in the network over the total number of possible edges in the network. In other words, a very compact network would have a very high Network Density.

D. Power Law

Scale-free networks are networks that follow the *Power Law*. This means that the fraction P(k) of nodes in the network having k connections to other nodes goes for larger values of k as:

$$P(k) \sim k^{-\gamma}$$

E. Strong Triadic Closures

We measure the elicitation of *Strong Triadic Closures* in the CVC/VC network by estimating the network's global Clustering Coefficient. A Strong Triadic Closure is any three nodes for which all nodes are connected to each other by an edge. Similarly, we can count the number of node triplets which are connected, but not strongly connected. Thus, we can define the Clustering Coefficient for a graph as,

$$C = \frac{|T_{strong}|}{|T_{all}|}$$

where T_{strong} is the set of all densely-connected node triplets, and T_{all} is the set of all dense and non-dense node triplets. $|.|$ denotes the cardinality of the set.

F. Network Transitivity

Network Transitivity is the ratio of all triangles over all possible triangles. A possible triangle exists when one person (Fox) knows two people (Fell and Whitehead). So Network Transitivity, like density, expresses how interconnected a graph is in terms of a ratio of actual over possible connections.

VI. Results

A. Diversity Coefficient

After calculating the Diversity Coefficients for all the nodes in the graph, we calculated the average value for CVCs and VCs. According to the formula, a higher Diversity Coefficient means the firms investments are more concentrated. The results can be tabulated as:

Mean Diversity Coefficient	
CVC	6.30
VC	0.37

B. Effective Size

The Effective Size of the network of a VC firm relates to their role in the social network of Venture Capital. The higher the Effective Size of the network, the greater the probability that the VC firm acts as a structural hole in the network. The results below prove that the effective sizes of CVCs are very high compared to the effective sizes of VC firms. The table encapsulates the results for effective network sizes.

Firms with the Largest Effective Size		
Name	CVC or VC	Effective size
Intel Capital	CVC	2872.49
Salesforce Ventures	CVC	1565.54
Tencent Holdings	CVC	1523.37
Qualcomm Ventures	CVC	1322.19
Bain Capital	CVC	1070.52
Matrix Partners	VC	945.13
IDG Capital	VC	944.30

Due to computational constraint, we could not compute the effective size for all the nodes in our network so we restricted the calculation to the top 500 most active VC firms (by the number of investments a firm has made).

C. Network Density

The Network Density is a parameter, which does not tell us a lot about the network in absolute terms but we still wanted to observe the Network Density of the social network of all VC firms.

The Network Density of this network is **0.00044**. While this does not tell us anything about how this parameter could differ for CVC firms, it does tell us that the network is very sparse. In the future, we could watch the parameter change over the years.

D. Network Transitivity

The Network Transitivity of the entire VC network is **0.1089**, which translates into the fact that 10 percent of triads are closed and, as discussed with Network Density (See VI-C), a good follow up would be to calculate this metric as the network changes over time.

We also plotted two separate networks by isolating the nodes that are CVC firms and VC firms. The following are the Network Transitivities of each of those networks. It can be seen that the network of CVC firms follows the Strong Triadic Closure property more than the network of VC firms.

Network Transitivity	
CVC	0.2025
VC	0.1544

We acknowledge that calculating the transitivity separately might not be the most accurate metric to calculate. This is due to the fact that plotting separate networks removes edges mapping from a VC to a CVC (and vice versa) and so the two networks cannot be fully separated from each other and maintain their integrity. Hence the results of calculating the Network Transitivity of the two separate networks should be taken with a grain of salt.

VII. Conclusion

Among all the hypotheses we mention earlier, we were able to validate the following:

- Corporate VC firms invest heavily in start-ups working on industries aligned with their parent company as compared to general VC firms
- The effective size of Corporate VC firms in a social network of Venture Capitalists is higher than the effective size of generic VC firms
- The bipartite social network formed by VC firms and start-ups obey the power law (rich get richer phenomena)
- The network of Corporate VC firms follow Strong Triadic Closure more than the network of VC firms

These remaining hypotheses were invalidated based on the metrics we generated through the study:

- The bipartite social network formed by VC firms and start-ups has a very high network density (¿0.5)
- A social network of VC firms is more closely knit as compared to a social network of Corporate VC firms

VIII. Source Code

All of our Source Code that we used for our analysis can be found at the following Git repository: https://github.com/css459/social-networks-final-project

References

[1] Zheng, Ju Kimberly. *A Social Network Analysis of Corporate Venture Capital Syndication*. 2004. URL: http://hdl.handle.net/10012/854.

[2] Liang Yuxian Eugene and Yuan Soe-Tsyr Daphne. "Predicting investor funding behavior using crunchbase social network features". In: *Internet Research* 26.1 (Jan. 2016), pp. 74–100. ISSN: 1066-2243. DOI: 10.1108/IntR-09-2014-0231. URL: https://doi.org/10.1108/IntR-09-2014-0231.