Analyzing VC Influence on Startup Success

A people-centric network theory approach

by Beth M. Hadley

S.B., Massachusetts Institute of Technology (2015)
Computer Science

Submitted to the
Department of Electrical Engineering and Computer Science
in Partial Fulfillment of the Requirements for the Degree of

Master of Engineering in Electrical Engineering and Computer Science
at the
Massachusetts Institute of Technology
June 2017

Author . . . .
Department of Electrical Engineering and Computer Science
May 22, 2017

Certified by . . . .
Peter Gloor
Research Scientist, MIT Center for Collective Intelligence
Thesis Supervisor

Accepted by . . . .
Christopher J. Terman
Chairman, Masters of Engineering Thesis Committee

Analyzing VC Influence on Startup Success

A people-centric network theory approach

by

Beth M. Hadley

Submitted to the Department of Electrical Engineering and Computer Science
on May 22, 2017, in partial fulfillment of the
requirements for the Degree of
Master of Engineering in Electrical Engineering and Computer Science

Abstract

In this thesis, we study the impact of venture capitalists on startup success using social network analysis. Using multiple sources, we compile a unique dataset of 3199 US-based technology startups and their board members, from which we generate and analyze the interlocking directorates network (formal network) and Twitter network (informal network). We find that startups with more VC board members are more central in the formal network, receive greater funding, have greater annual sales, yet a smaller return-on-investment. We also find that VCs are more central in the Twitter network than non-VCs, have greater Twitter popularity, yet tweet significantly less. Our results indicate that VCs carry a considerable amount of financial and social capital, which they transmit to the startups they invest in, yet their participation leads to lower startup ROI. Additionally, our dataset enabled us to investigate more general questions regarding startup success, including gender diversity on startup boards.

Thesis Supervisor: Peter Gloor
Title: Research Scientist, MIT Center for Collective Intelligence

Acknowledgments

Foremost, I sincerely thank Peter Gloor of MIT's Center for Collective Intelligence for being a tremendously supportive thesis supervisor and mentor on this project. Although when I began I knew very little about social network analysis nor venture capital, Peter never ceased to support and encourage me by sharing his insightful ideas, research expertise, and vast knowledge. He astounded me with his availability and responsiveness - he often replied to my emails just minutes after I had sent them, with specific ideas and responses to my questions. I can't imagine a better thesis advisor than Peter.

I also wish to thank Stephanie Woerner, of MIT's Center for Information Systems Research, for enthusiastically joining this research project and becoming a valuable advisor to my thesis. Stephanie consistently proposed clever ideas and insights that pushed our research questions in new and exciting directions. Her encouragement and enthusiasm were always contagious, and her delicious baked goods always very welcome.

I thank Yuhong Zhou, visiting scholar at the Center for Collective Intelligence, for her partnership on this work, and in general for her friendship. Yuhong provided a unique perspective and helped me to think about this work using a global approach.

I also gratefully acknowledge the developers of the Condor software (Peter Gloor, Marton Makai, Hauke Fuehres, Joao Marcos da Oliveira) for simplifying the creation and analysis of social networks in this thesis. My project would not have been possible without their years of dedication and commitment to Condor.

I also thank Shikha Sharma, MIT Business and Management Librarian, for helping me get started with MIT's business databases.

I'd like to thank my friends, family, and the many MIT professors and staff who supported my decision to take a "gap" year to work abroad - in France and Mongolia - before returning to MIT to complete my MEng. That gap year very much redefined my interests, both personally and professionally, and for that I am tremendously grateful.

I also thank the MIT Computer Science department and specifically professors Rob Miller and Julie Greenberg for extending teaching assistant positions to me for two courses (6.811 Principles and Practice of Assistive Technology and 6.813/6.833 User Interface Design), which provided the funding for my MEng year.

Dedication

I dedicate this thesis to my grandpa, Thomas H. Dooley. Having completed an extremely successful career in corporate finance, Grandpa always encouraged me to diversify my technical education by learning about business, finance, and management. Thank you, Grandpa, for sharing with me your breadth of intellectual curiosity.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The purpose of this research is to use social network analysis techniques to measure the impact of venture capitalists on startup success, by analyzing both formal and informal networks. In addition, the dataset we collected for this purpose allowed us to investigate a number of additional questions, including the gender diversity of board membership, number of founders, geographic location of startup headquarters, industry specialization, and relationship of schooling and social prominence on individual income.

1.1 Motivation

There is much uncertainty involved with early stage startups. Venture capitalists are faced with literally million dollar questions as they seek to evaluate early stage startups, to determine the potential of an investment. Likewise, entrepreneurs are faced with equally valuable questions as they seek venture capitalists from whom to not only gain funding, but also support and mentorship. We hope that this work will provide some answers regarding the VC-startup ecosystem and reveal key trends and indicators of success by studying the formal and informal networks of both venture capitalists and entrepreneurs.

1.2 Our Approach

VC and startup success has been studied from many angles. In this work, we address the question with a people-centric approach, constructing social networks of the board members of each startup. We choose to study the board of directors of a startup, as board directors have a great deal of influence over a startup due to their financial, intellectual, and social capital resources. Furthermore, board members often sit on multiple boards, and therefore may share their resources with multiple startups. The resulting network formed by board membership represents a communication network, through which critical resources (money, knowledge, and opportunities) flow.

We strive to answer questions about startup success and the participation of VCs in this success by constructing two types of networks: a formal network and an informal network. The formal network is an interlocking directorates network composed of people and the boards they sit on, whereas the informal network is a Twitter social network composed of tweets and retweets.

1.3 Research Questions

Fundamentally, we aim to investigate the question of how venture capitalists, in particular through their networking behavior, influence the success of a startup, and how this is manifested in people-centric networks. To answer this broad question, we adopt a network theory approach and investigate questions that we believe shed light on the ultimate question. We first look at the position of successful startups in the formal network, and compare this to the position of startups with many VCs on their boards. Secondly, we construct an informal social network composed of Twitter data, and look at the behavior of venture capitalists in this network. By looking at venture capitalists' influence in these networks, we develop evidence that informs our response to the ultimate research question - how do venture capitalists influence the success of startups.

Furthermore, we extend our analysis by investigating a variety of additional factors that influence startup success. For example, we investigate the correlation the following factors have with startup success: number of founders, female board membership, and tech industry. We also investigate individual success, and ask if someone's location in the Twitter social network is indicative of professional success. By answering our research questions, we provide evidence for various trends of success and failure in the tech startup ecosystem.

1.4 Contributions

In this thesis, we make the following contributions:

1. Demonstrate a novel approach to studying startup success by creating and comparing formal and informal networks based on startup board of directors membership.

2. Collect a substantial dataset from multiple sources (Capital IQ, Crunchbase, OneSource, US Tax Data, and US News & World Report School Rankings) which did not previously exist and which permitted investigation of questions not previously analyzed in this way.

Given our analysis of this dataset, we provide evidence for the following conclusions:

1. Startups and individuals located more centrally in both the formal and informal network are generally more successful - in terms of startup funding, revenue, Twitter popularity, educational background, and personal income.

2. Startups with more VCs on their board tend to receive greater funding, have greater annual sales, but a smaller return-on-investment (defined as revenue $\div$ funding).

3. VCs are significantly more central in both formal and informal networks than non-VCs, and they have greater popularity (defined as the ratio of followers to number of people you follow on Twitter). Interestingly, VCs tweet significantly

less than non-VCs, further proving the point VCs are inherently more "popular" than non-VCs.

4. The number of startup founders is positively correlated with startup success.

5. The number of female board members is negatively correlated with startup funding.

6. A person's educational rank (measured by the prestige of their university) and their social capital (measured by reach-2 in the Twitter network) is correlated with their residential income.[1]

1.5 Outline

In the remaining sections of the Introduction, we provide contextual information and definitions about this thesis. In Chapter 2 we review previous work in this area. In Chapter 3 we summarize the data sources used and the process of constructing the networks. In Chapter 4 we present our analysis of this data and how our results address our research questions regarding the influence of VCs on startups. Chapter 5 discusses some additional research questions we investigated related to startup success. Chapter 6 provides some ideas for future work and concludes the thesis.

1.6 Thesis Scope

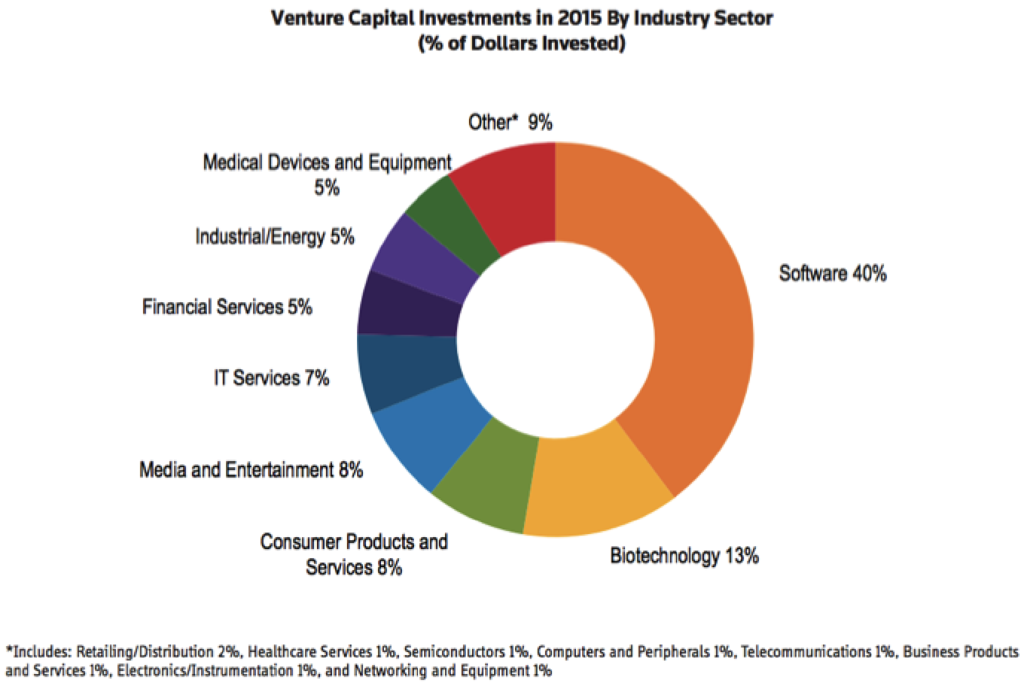
We limited the scope of this study to the US tech startup industry. We describe below specific details regarding the scope limitations we imposed in our study.

Startup: We define a startup as any company founded in 2011 or more recently. We began with a dataset from Capital IQ which contained 3199 startups, out of which 1514 had Crunchbase funding data, from that 525 had OneSource sales data.

Industry: Only startups in the IT/Software domain were investigated. In 2015, software startups received 40% of all dollars invested in startups that year, which is

[1] As defined by the average residential income of the zip code address in which they live. See section 5.3 for details.

by far the largest percentage by sector. Therefore, restricting this study to only IT startups insures both a large dataset and controls for industry effects.



Source: Franklin and Haque [2016]

Figure 1-1: The distribution of venture capital investments in 2015 according to sector. The IT/Software industry accounted for the largest percentage (40%) of venture capital investments in 2015.

Venture Capital Firms: We limit our study to only those firms managing a venture capital investment fund, as opposed to private equity or hedge fund.

Geography: The study is limited to startups with headquarters in the United States.

1.7 Definitions

1.7.1 Network Theory Definitions

Formal Network A network that is constructed based on formal rules and contracts that exist in the real-world. In this thesis, the interlocking directorates network is our formal network.

Informal Network In contrast to a formal network, an informal network is one that is not constructed based on rigid rules and often does not represent strong real-world ties and/or relationships. The informal network is often much more socially structured. In this thesis, the Twitter social network is our informal network.

Interlocking Directorates (ID) Network A network composed of companies (the nodes) and links between two companies where the same director sits on the boards of both companies. This is also referred to as the formal network in this thesis.

ROI (Return-on-investment) Measures the efficacy of a startup in transforming dollars of investment into revenue. Calculated as the ratio between sales and funding. We use sales rather than revenue because the database (OneSource) only provided annual sales data, although we expect sales to be highly correlated with revenue so we do not anticipate this to alter our results.

Social capital The networks of relationships among people who live and work in a particular society, enabling that society to function effectively [Nahapiet and Ghoshal, 1998].

Twitter (Social) Network A network composed of people (the nodes) and a link between two people if one of the people has tweeted about the other person and/or retweeted something that person has tweeted.

VC (Venture Capitalist) A VC, or venture capitalist, is someone who invests private equity in small, early-stage, emerging companies that they deem to have high growth potential. In this thesis, we limit our study to only VCs associated with formal VC firms.

$$\frac{AnnualSales(OneSource)USDmil}{TotalStartupFunding(Crunchbase)USDmil}$$

1.7.2 Network Centrality Measures

Betweenness Centrality The extent to which a vertex lies on paths between other vertices. People with high betweenness may have considerable influence within a network by virtue of their control over information passing between others. Removing them from the network will most strongly affect communications between other vertices because they lie on the largest number of paths taken by messages.

Closeness Centrality The mean distance from a vertex to other vertices. A person with lower mean distance to others (high closeness) are likely able to spread their opinions to others in the community more quickly than someone with higher mean distance, as the information must "travel" farther.

Degree Centrality The number of neighbors a node has. The more neighbors a node has, the greater number of people who will be directly influenced by that person and the information they distribute.

Reach-2 Counts the number of nodes a node can reach in 2 or less steps. Effectively, this is a measure of the strength of your "friend-of-a-friend" network. It measures how many people someone can reach via their friends.

1.8 Research Standards

Throughout this thesis, we compare various sets of data and seek to determine if there is a statistically significant difference and/or correlation between the data sets.

We use t-tests to make a binary conclusion regarding whether or not there is a difference between the mean of two datasets. All t-tests are two tailed Student's t-tests. A t-test is considered significant if the p-value is less than 0.05.

We use correlations to indicate the strength of similarity between two datasets. All correlations are two-tailed Pearson correlations. A correlation is considered significant if it has a p-value less than 0.05. As is typical in research in the domain of managerial science and behavioral economics, it is rare to see correlation coefficients greater than 0.5, and typically correlations of 0.05 - 0.30 are considered interesting, so long as they have a p-value less than 0.05. Furthermore, we do not make any causal claims, as this would require temporal data which we unfortunately had no ability to acquire (see section 6 for a discussion on future work).

Chapter 2

Background and Related Work

2.1 Demonstrated Impact of VCs on Startup Success

Venture capital (VC) is well established as one of the key driving forces in the American entrepreneurial ecosystem. According to the National Venture Capital Association, in 2015 nearly $60 billion in venture capital was deployed across 4,380 deals. More than 30% of those deals were to companies receiving venture investment for the first time [Franklin and Haque, 2016]. That said, studies have not provided consistent conclusions regarding the impact of VC investments on entrepreneurial firms, and whether this impact is a net positive or negative. Although the answer is likely, "it depends", the question is worthy of continued rigorous analysis.

Advocates for the positive influence of VCs claim that VCs serve three main roles to identify and promote successful startups:

1. **Screening**

   VCs choose to invest in high quality companies with promising potential. They are experienced at selecting for certain criteria that predict success, such as technical expertise and founder commitment [Chan, 1983], Amit et al. [1998].

2. **Monitoring**

   VCs track the status of their portfolio companies, comparing investments with market trends and opportunities. They protect the value of their investments by adding credibility and prestige to those companies they invest in [Lerner, 1995], [Kaplan and Strömberg, 2003].

3. **Coaching**

   VCs provide advice and support to their portfolio companies with the intent of improving their chances of success and, in return, the return on their investment. This may include connecting the firm with resources (networking), assisting with recruitment, and providing experience, advice, and mentoring [Hellmann and Puri, 2002], [Hellmann, 2000].

Some researchers advocating for the positive influence of VCs on startup growth have demonstrated the correlation between VC backing and startup success: VC-backed firms have faster growth, faster times-to-market of their products, more patents, higher productivity, and are more likely to go public [Wright and Mike, 1998]. Some studies have also been able to demonstrate a causal relationship between VC involvement and startup success. For example, one study investigated how VCs' on-site involvement with their portfolio companies (as facilitated by the introduction of direct airline routes between VC firms and their portfolio companies) increased the portfolio companies' innovation and success. The study found that the introduction of a new airline route led to a 3.1% increase in the number of patents the portfolio company produces (a measure of "innovation") and an increase of 1.4% probability of having a successful exit (via IPO or acquisition). Although small yet significant, these results do indicate that VC involvement is an important determinant of innovation and success [Bernstein et al., 2015].

Another longitudinal study compared the efficiency gains generated by venture capital (VC) investment in private firms from 1972 to 2000. They compared the productivity and efficiency of 1,881 VC-backed firms with 185,882 non VC-backed firms. They found that the overall efficiency of VC-backed firms is higher than that of non-VC-backed firms at every point in time, and they demonstrated that this was due to both screening (pre-investment) and monitoring/coaching (post-investment). Interestingly, they also demonstrated that companies backed by "high-reputation VC firms" have significantly greater productivity, likely a reflection of the monitoring ability of high-reputation VCs. Lastly, they showed that both screening and monitoring activities of VCs positively affect the probability of a successful exit (IPO or acquisition) [Chemmanur et al., 2008].

That said, reasons and explanations abound which indicate just the opposite, that venture capitalists can and do have a negative influence on startups and their entrepreneurs. Foremost, a large body of research exists regarding conflicts between startup entrepreneurs and VCs. Prior research has identified three main areas of VC-CEO conflicts [Khanin and Turel, 2013]:

Source: Wasserman [2008]

Figure 2-1: Demonstrates the trade-off entrepreneurs must make between financial gains and control over their company. These trade-offs indicate the potential for conflict between entrepreneurs and CEOs.

1. **Conflicts of interests and unfavorable attributions**

   In this conflict type, the interest of the VC/VC firm is directly in conflict with the interest of the entrepreneur/startup. For example, the higher the pre-money valuation (the value of a startup before VC investment), the better for the startup, yet the worse for the VC. Tradeoffs and compromises are often the solution to these conflicts, as illustrated in figure 2-1 (extracted from a Harvard Business Review article) [Wasserman, 2017].

2. **Conflicts of inefficient collaboration**

   In this conflict type, the VC and the entrepreneur fails to work together and does not form a mutually beneficial collaborative partnership. For example, if a VC maintains a high degree of oversight over the startup, the entrepreneur may perceive this as having a low degree of autonomy whereas the VC may simply be wishing to contribute expertise and add value to their investment.

3. **Conflicts of VC-CEO mismatch**

   In this conflict type, the VC and the entrepreneur does not align regarding fundamental characteristics such as personality type, educational background, and/or strategic vision. For example, if the CEO is a "builder", i.e. he is really focused on creating a viable enterprise whereas the VC is an "investor", i.e. primarily focused on achieving fast financial results, conflicts are likely to occur [Perry, 1988].

Although research indicates sources of conflict between founders and VCs abound, the implications of such conflict are not entirely obvious. Conflict does not necessarily have a negative impact on the success of a startup; one study showed how disagreement can be beneficial to venture performance, although conflict as personal friction was negatively associated with performance [Higashide and Birley, 2002]. Nonetheless, at high levels of occurrence and intensity, conflict is generally considered to be costly to those involved [Reve and Stern, 1989]. It can thus be reasoned that in the VC-entrepreneur relationship, conflict is a likely contributing factor to negative outcomes regarding VC investment in startups.

The negative impacts of VC investments have been demonstrated analytically for a number of geographies external to the US, although to the author's knowledge no studies in the US have reached similar conclusions. A study of startups in Singapore demonstrated that the post-IPO operating performance of VC-backed companies was inferior to non VC-backed companies, though they were less underpriced [Wang and et al., 2003]. Another study analyzed startups in China, and found that VC firms failed to enhance the development of startups, and to some extent even exerted a hampering effect on the performance of funded firms [Xi and Su-Sheng., 2016]. A study of the French startup ecosystem from 1996 to 2006 demonstrated that VC financing had a negative impact on firm survival [Pommet, 2017].

There seems to be a growing sentiment emanating from Silicon Valley to be dubious and extremely aware of venture capitalist investment. A number of recent news articles - mostly targeted at entrepreneurs - warns VCs may have a negative or at best neutral or insignificant influence over the success of a startup. In a 2014 article "VC Funding Can Be Bad For Your Start-Up" [Mullins, 2014] published in the Harvard Business Review, the author provides four clear reasons to be wary of VCs:

- Pandering to VCs is a distraction.
- Term sheets and shareholders' agreements can burden you.
- The advice VCs give isn't always that good.
- The stake you keep is small - and tends to get smaller.

Fred Wilson, co-founder at Union Square Ventures, wrote in a blog post in 2013: "The fact is that the amount of money start-ups raise in their seed and Series A rounds is inversely correlated with success. Yes, I mean that. Less money raised leads to more success. That is the data I stare at all the time" [Wilson, 2013]. While academic research remains to prove or disprove his claim, it is inherently intriguing that a venture capitalist would make such a comment, and begs further analytical investigation.

In this thesis, we attempt to disentangle the conflicting messages coming from academia and industry regarding the impact of VCs on startup success. By applying a unique approach to studying the industry, namely a network theory analysis of the board membership network and Twitter social network, we believe our work provides a unique contribution to the literature.

2.2 Networks and Venture Capitalists

In this thesis, we take a unique network theory approach to understanding the impact of VCs on startup success. We discuss below related literature in the domain of social network theory. To the best of our knowledge, we are the first to analyze the question of VC influence through an analysis and comparison between both formal and informal networks.

Since the mid 1980s, social networks and relationships have been demonstrated to be absolutely critical to an entrepreneur's success [Hoang and Antoncic, 2003]. Social networks can take a variety of forms, from people and physical conversations to online personas (via Facebook, LinkedIn, and Twitter) and electronic messages. The value of real-world, physical networking is certainly well-understood and proven [Zimmer, 1986], [Allen et al., 2009]. However, the role of digital networks, and their comparison to physical networks, is less understood, as there is a fundamental difference between strengths of ties in physical networks versus digital ones [De Meo et al., 2014]. In the context of venture capitalists and entrepreneurs, a network composed of venture capitalists and entrepreneurs in the physical world might vary drastically from one in the digital world. In the digital world, a tie, albeit weak, may be created in seconds when a venture capitalists creates a tweet, whereas in the physical world the tie might represent thousands, millions, or even billions of dollars of investment. On the other hand, a simple 140 character tweet could be seen by thousands or millions of people in a single minute, and impact their actions and behaviors on a daily basis. The contrast between these two kinds of networks is not well understood, but better knowledge of their congruency may lead to a better understanding of the influence venture capitalists play on startups, both in the physical world and in the digital.

2.3 Interlocking Directorates

A physical network commonly studied among venture capitalists and entrepreneurs is an "interlocking directorates" network. Companies are represented by nodes in this network, and ties exist between companies if an individual (either a venture capitalist or an entrepreneur) sits on the board of directors of both companies. The presence of interlocking directorate networks was a source of fervent academic study during the 20th century - as concern mounted that monopolistic corporations were supported by interlocking directorate relationships [Mizruchi, 1996]. Studies demonstrated that interlocking networks have correlated with performance of corporations, organizational failure, economic downturns, CEO pay, the sale price of a corporation, stock synchronicity, and prices for a corporation's services. Fundamentally, it is clear that interlocking networks facilitate the flow of information among corporations [Saavedra and et al., 2014]. It is less clear to what extent the practice of venture capitalists sitting on multiple startup boards plays in the success of these startups.

2.4 Social Networking Analysis

In contrast to the interlocking directorates network, a social network is an informal network composed of links between individuals who communicate of their own accord. Studies have demonstrated that social networks can indicate and predict the success of startups. For example, a study analyzing the Twitter network associated with 644 IT startups indicated that the higher the centrality and connectedness of the startup in the social network, the more successful the startup in the real world [Yu and Perotti, 2015]. Another study involving the LinkedIn, Facebook, and e-mail networks of an entrepreneurship community in Boston indicated that the more central actors in the network, the more successful they are. Furthermore, proximity to key people in the network correlates with success [Gloor et al., 2013]. Few studies have analyzed the social networks formed between VCs and entrepreneurs. One study, however, investigated the influence of VCs' social capital on the funding of startups. They constructed a social network based on VC syndication (a link between a startup and a VC if the VC invests in the startup) [Hopp, 2010]. Through the analysis of 1500 funding rounds, they found a positive effect of VCs' social capital on the amount of money that startups receive [Alexy et al., 2012] For this reason, we too expect that those people who are centrally located in our Twitter social network (hence higher social capital) will sit on the boards of more highly-funded startups. To our knowledge, however, no study has analyzed the Twitter social network of people in an interlocking directorates network.

Chapter 3

Data Sources

3.1 Formal Network Data Sources

3.1.1 Constructing the Formal Network Dataset

The process of constructing a dataset to conclusively answer the research questions was a non-trivial task. Foremost, no single data source contained all the data required, so a combination of data sources was required. Secondly, the data being sought was of a sensitive nature, and not commonly nor publicly available.

In summary, the basic[1] dataset we created contains the following fields: Startup Name, Number of Founders, Board Members, Total Funding, and Annual Sales. We also constructed a dataset of individuals, namely the board members: Persons Name, Gender, Board Memberships, Is a VC?. We describe below the datasource used and the process taken to acquire the data fields listed above.

3.1.2 Data Source: Capital IQ

IT Startups, Board Memberships, List of Venture Capitalists

S&P Capital IQ[2] is a financial information platform that was originally designed

[1] Basic meaning the minimal dataset. We compiled a vast amount of additional data on these startups, including for example date founded, headquarters location, investors, industry categories, etc. but these data fields were not used in this analysis.

[2] `https://www.capitaliq.com`

for the investment banking industry. The platform provides detailed content in four major areas: company profiles, industry profiles, transaction profiles, and executive profiles. Over 62,000 public companies and 4.4 million private companies are profiled [Financial, 2016]. The platform provides a convenient interface to generate specific datasets according to search constraints.

Of all the data sources we investigated, Capital IQ proved to have the most complete and easily extractable information regarding startup board memberships. On January 13, 2017 we extracted three datasets from Capital IQ, as seen in table 3.1

Table 3.1: Summarizes the datasets we extracted from Capital IQ, including the size of each unique dataset.

Number of Entries	Description of Dataset	Data Fields Extracted[3]
3199	IT Startups in the US	Board Members, Year Founded, Business Description, Company Status, Current Investors, Number of Employees, Website
6508	Venture Capital Firms	Key Executive, Board Members, and Other Professionals, Address, Website, Business Description, Investments, Fund size, Fund Launch Date.
8474	Board members currently sitting on an IT startup board in the US	Biography, Board Memberships, Primary Company Affiliation, All Company Affiliations, Address, Job Functions, Age, College, Degree, Major

A simple matching was performed to mark board members as venture capitalists if they appear in the venture capital firms dataset. Individual's gender, which was not included as a data field in Capital IQ, was extracted from their biography, with a simple lookup for occurrence of “Ms.” vs. “Mr.” and gendered pronouns.

3.1.3 Constructing the Interlocking Directorates Network

Step 1: People Network

Given the Capital IQ listing of startups and board members, the interlocking directorates network was constructed. Foremost, a network was constructed with 8474 individuals (nodes) and links between individuals who sit on the same startup board. The network consisted of one large central connected component with 412 individuals, surrounded by a sea of smaller connected components. The average individual was connected to 2.94 individuals in the network.

As a preliminary verification of the data, we wanted to ensure that venture capitalists were central within the network. We expected this characteristic because it is well known that VCs obtain control over companies in which they invest by acquiring board seats. Having control over their investment is critical for VCs to attempt at influencing the return on their investment. Influence over decisions made by boards of directors is commonly viewed as among the most effective governance mechanisms in VC investments. In an analysis of 182 startup contracts between VCs and founders, at least one board seat is held by a venture capitalist who invested in the round [Bengtsson, 2011].

To verify VCs were central within our network, we calculated three standard network centrality measures (betweenness, degree, and closeness as described in section 1.7.2) and compared the position of VCs to non-VCs in the network. As expected, VCs were more central. This may be observed visually in figures 3-1 and 3-2, as well as numerically in table 3.2.

Table 3.2: Includes t-tests which demonstrate that VCs are more central in the formal network.

T-Test	VCs	Non-VCs	n	p-value
Mean Degree Centrality	3.4	2.87	1124	6.79e-09
Mean Betweenness Centrality	1662.56	149.8	7350	9.13e-11

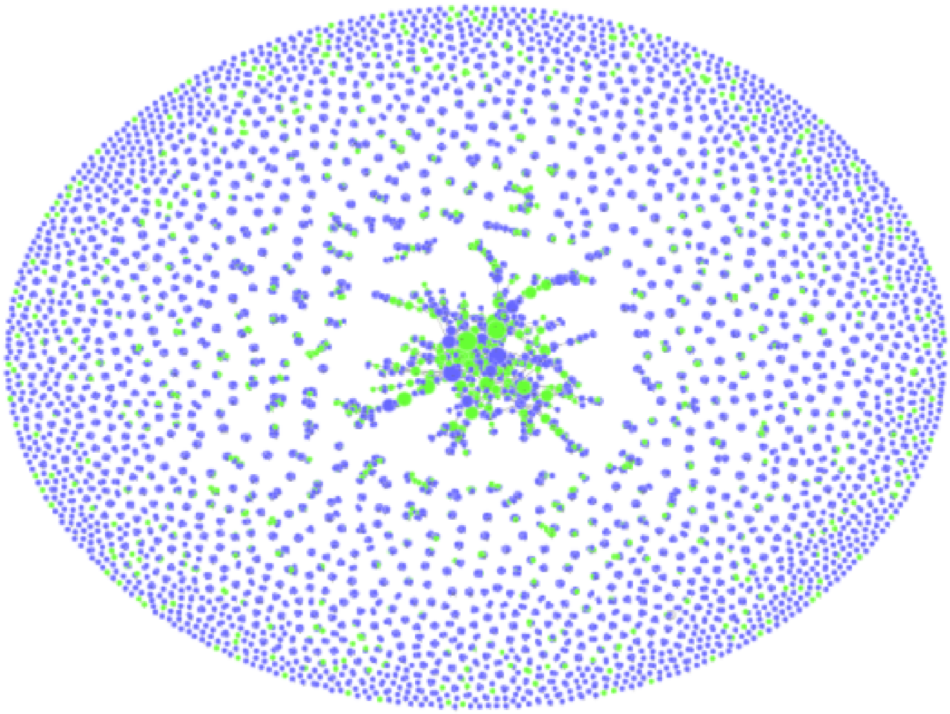


Figure 3-1: The formal interlocking directorates network, containing a single large central connected component surrounded by a sea of disconnected islands.

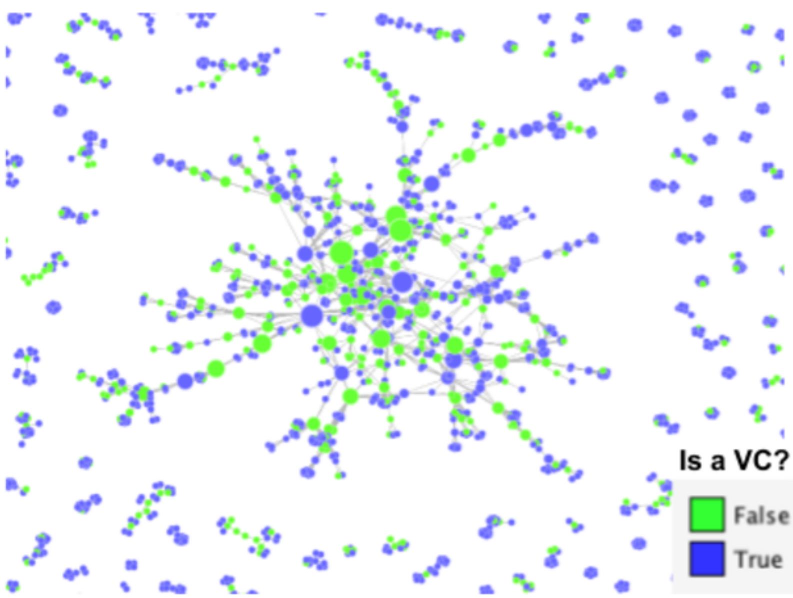


Figure 3-2: A zoom on the central connected component of the interlocking directorates network, where there is a greater presence of VCs. Node size is proportional to betweenness centrality.

VCs not only have on average more connections than non-VCs, but also much higher betweenness centrality. This demonstrates the dataset we collected does indeed conform to our expectations, namely that VCs are more central in the network.

The top 10 people, ranked in order of betweenness centrality, are listed in table 3.3. Eight out of ten of these individuals are VCs, and many are well known in the industry.

Table 3.3: The top 10 people, ranked in order of betweenness, in the interlocking directorates network.

Rank	Name	Betweenness centrality	Degree centrality	Is a VC
1	Stephen Herrod	81023	11	y
2	Mark Leslie	77445	12	y
3	Glenn Solomon	72027	9	n
4	Krishna S. Kolluri	69595	19	y
5	John N. Stewart	66946	14	n
6	Joe Lonsdale	65115	11	y
7	Theodore E. Schlein	64631	14	y
8	Marc L. Andreessen	58736	9	y
9	Karim B. Faris	55864	10	y
10	David B. Aronoff	55540	7	y

Step 2: Startup Interlocking Directorates Network

Given the formal network of 8474 individuals, we then collapsed this network to produce a network of 3199 startups, with links between startups who share at least one board member. This is known as the interlocking directorates network [Mizruchi, 1996]. We calculated three centrality measures on this network: betweenness centrality, degree centrality, and closeness centrality. Listed in table 3.4 are the top 10 startups ranked by betweenness.

Given this network of startups and various centrality measures, we then turned our attention to acquiring the necessary dependent variables to measure startup success.

Table 3.4: The top 10 startups, ranked in order of betweenness, in the interlocking directorates network.

Rank	Startup	Betweenness Centrality	Degree Centrality	Closeness Centrality	Board size
1	Threat Stack, Inc.	79789	22	0.15545	10
2	BetterWorks Systems, Inc.	78697	20	0.15402	9
3	DataGravity, Inc.	66450	17	0.15553	10
4	Turi Inc.	65110	23	0.15435	10
5	Circle Internet Financial, Inc.	59982	15	0.15334	9
6	Nok Nok Labs, Inc.	54984	11	0.15304	10
7	Viptela, Inc.	52513	14	0.15205	10
8	BlueData Software, Inc.	52135	12	0.15164	10
9	QASymphony LLC	50586	11	0.15323	10
10	Wealth Access, Inc.	49946	11	0.15155	10

3.2 The Dependent Variables

One of the fundamental challenges of conducting this kind of study is determining a reliable metric for startup success. Ideally, we would have a dataset containing a listing of startups with their market valuation, in addition to their annual profits and revenue. However, this is practically impossible. Foremost, startups often do not have valuation until they receive funding, and even then this information is rarely disclosed publicly. Startup profits and revenue data is often rarely available, especially for young startups.

In this study, we identify three dependent variables for each startup:

- **Total Funding** The total amount of funding (venture capital or otherwise) that the startup has received since its founding
- **Annual Sales** The startup's most recent annual sales
- **ROI** Return on investment, defined as the ratio of annual sales to total funding

We describe below our process of acquiring total funding data (from Crunchbase) and annual sales data (from OneSource). We also include ROI as a dependent variable because it indicates the efficacy of a startup in transforming dollars of investment into revenue.

3.2.1 Crunchbase: Total Funding

The first dependent variable we sought was the amount of funding each startup received. For this, we turned to Crunchbase[4], an online database containing thorough information primarily about US-based companies and the people (founders, employees, investors) associated. Data is contributed by registered users and approved by the public. Crunchbase is known for its coverage of startups, especially technology startups. It is also commonly used as a data source in academic research, not only for the breadth of the dataset, but also the efficient data extraction interface via its API [Alexy et al., 2012], [Block and Sandner, 2009]. Although in this work we did not gain access to the Crunchbase API, we did subscribe to Crunchbase Pro, which permitted an albeit limited query of the database and download of data into .csv files for analysis.

On January 2, 2017, we extracted from Crunchbase all IT startups founded since 2011 which have non-zero funding amounts. This produced 16,029 results. The following data fields were extracted: Startup Name, Website, Industry Categories, Headquarters, Description, Status, Founded Date, Twitter Handle, Number of Founders, Date of Last Funding, Last Funding Amount, Last Equity Funding Amount, Total Equity Funding Amount, Total Funding Amount, and Number of Investors.

3.2.2 Capital IQ & Crunchbase Dataset Merge

We then merged the datasets from Capital IQ and Crunchbase, by writing an automated matching script based on startup company name. The automated matcher was robust to naming variations such as lower and upper case and suffixes (Inc, Corp, Ltd, etc.). This resulted in a combined dataset of 1514 startups.

3.2.3 OneSource: Annual Revenue

The second dependent variable required in this study was startup revenue/sales. This was certainly the most challenging and least publicly available data field. However, we

[4] `www.crunchbase.com`

found relevant data in the datasource OneSource Global Business Browser, produced by Avention (recently acquired by Dun & Bradstreet[5]). This database includes nearly 50 million company profiles, and collects data from over 100 data sources [Hoovers, 2017]. The database includes a field called "Annual Sales" for each company. This is not exactly revenue, as it neglects long-term revenue, for example from licensing or from patent settlements, although sales is a decent proxy for revenue. To extract this sales data, there was unfortunately no easily automated method so we performed a manual lookup of each startup in the dataset. Of the 1514 startups in the original dataset, 525 were found to have revenue information in OneSource.

Although revenue information is notoriously difficult to acquire for private companies, and harder still for startups, we have reason to believe that the accuracy of the revenue data we collected from OneSource is relatively high. As a check on the accuracy of the revenue dataset, we collected the revenue of 40 startups from the Capital IQ dataset. We ran a simple Pearson's correlation between the revenue data in Capital IQ and the corresponding sales data of startups from OneSource, and found a correlation of r=0.627 (p=1.31E-05). Thus, even with this small sample size of 40 startups we demonstrate the validity of the OneSource sales dataset, which gives us reasonable confidence that the revenue data acquired from OneSource is accurate.

3.2.4 Dependent Variables as Startup Success

We have identified three dependent variables in our study with which we use to approximate startup success. Although we acknowledge that funding and sales are not absolute measures of startup success and/or performance, we make the assertion that startups with higher funding, higher sales, and/or higher ROI are generally considered more successful.

We do claim that there is a positive relationship between startup funding and annual sales. Based on a sample size of 525 startups, we correlated the total funding amount (from Crunchbase) with the annual sales (from OneSource) and obtained a correlation of 0.513 (p=1.17E-36). Therefore, we feel confident making the claim that

[5] `http://www.hoovers.com`

companies with higher funding and higher revenue are more successful.

Furthermore, throughout our analysis we take the logarithm of financial data - funding amount or revenue - to calculate correlations. Taking the log of financial data is standard practice in organizational research. This is convenient, as a log transformation acts as both a transformation to normality and as a variance stabilizing transformation. Fundamentally, it transforms an exponential dataset into a linear dataset, which is appropriate in our analysis as we are applying linear analysis techniques such as Pearson's correlation. Throughout our discussion, when we refer to financial data (funding or sales), we may not explicitly state that we are referring to the log of the financial data, but this is implied.

3.3 Informal Network Data Sources

3.3.1 Constructing the Informal Network

Choice of Social Network

Given our formal interlocking directorates network, we wanted to construct a network of the same people but built on more informal social ties. Whereas the formal network represented social interactions in a very formal context (i.e. the boardroom), this informal network would represent more organic social interactions. We hoped this informal network would provide a unique perspective into the communications among individuals, especially in contrast to the formal network.

A number of sources were considered for constructing such an informal social network, including Facebook, LinkedIn, a survey, and Twitter. Twitter was quickly selected as the optimal choice because:

1. Twitter is the social media platform most extensively used by startups and investors, and broadly used by the business community [Wu et al., 2015].

2. Twitter provides an easy-to-use public API to access its data.

3. Facebook and LinkedIn were simply not viable options because they do not

provide a public API to access data regarding relationships among individuals, and a survey would likely have returned incomplete information and not to the scale we desired.

Matching People with Twitter Handles

Having chosen Twitter, the first task was to lookup the Twitter handles of the 8474 people in our formal network. Unfortunately we did not already have a listing of the Twitter handles of each person, as the data source from which we extracted these people (Capital IQ) did not contain the person's Twitter handle.

We started with our dataset of 8474 people, which included a list of all companies each person is "affiliated with", in addition to the person's biography. From the biography, we extracted the person's nickname, if present. We then wrote a script to query Twitter's search API for users with a matching full name or nickname (including last name). We also extracted people's profiles from Crunchbase, which included an individuals' Twitter handle (although sometimes these pairings were inaccurate).

We then conducted a series of tests to search through the potential Twitter user match(es) in an attempt to only accept those users that represented the individual in our interlocking directorates network. After attempting a variety of strategies, both relaxed and strict, we settled on the following two criteria for accepting a match:

> The Twitter user's description must contain at least one of the companies that the real user is affiliated with
>
> OR
>
> The Twitter user's description contains "vc", "capital", or "partner" and the real user is a venture capitalist.

This resulted in 1,271 matched Twitter handles. To measure the accuracy of this dataset, we manually verified 127 randomly selected people (10% of dataset) and found 87.4% accuracy. We therefore argue that the contribution due to inaccuracy of the dataset is minimized and negligible.

As an added check, we compared the ratio of VCs to non-VCs in the larger Interlocking Directorates (ID) network to the smaller Twitter network. We found that 15% of the ID network are VCs, whereas 23% of the Twitter network are VCs. We argue that these numbers are relatively similar and therefore our Twitter network is an accurate representation of the larger ID network. We do not expect the slight over-representation of VCs in the Twitter network to bias results, especially since we can explain explicitly why this overpopulation occurs: the additional check of keywords "vc", "capital" or "partner" selected specifically for venture capitalists. We experimented with conducting a similar test for "founder", "ceo", and "entrepreneur" but these words we discovered to be much more common in Twitter descriptions and caused a higher rate of incorrect matches. In conclusion, we believe the Twitter dataset of 1,271 Twitter users to be an accurate representation of Twitter users among the 8,474 people in our ID network.

Building the Twitter Network

Given our list of 1,271 twitter handles, we used the Condor tool [6] to automatically generate the tweet and retweet network. This produced a network of 45,521 nodes - the 1,271 people initially in the search combined with any Twitter user that they had either tweeted to (with @user_name) or retweeted. We limited the network construction to only the users' past 100 tweets. This network was constructed on March 23, 2017. Links in this network exist between a person who has tweeted to another person, or a person who has retweeted another person's tweet. This produced 168,326 links as seen in figure 3-3.

[6] `www.galaxyadvisors.com`

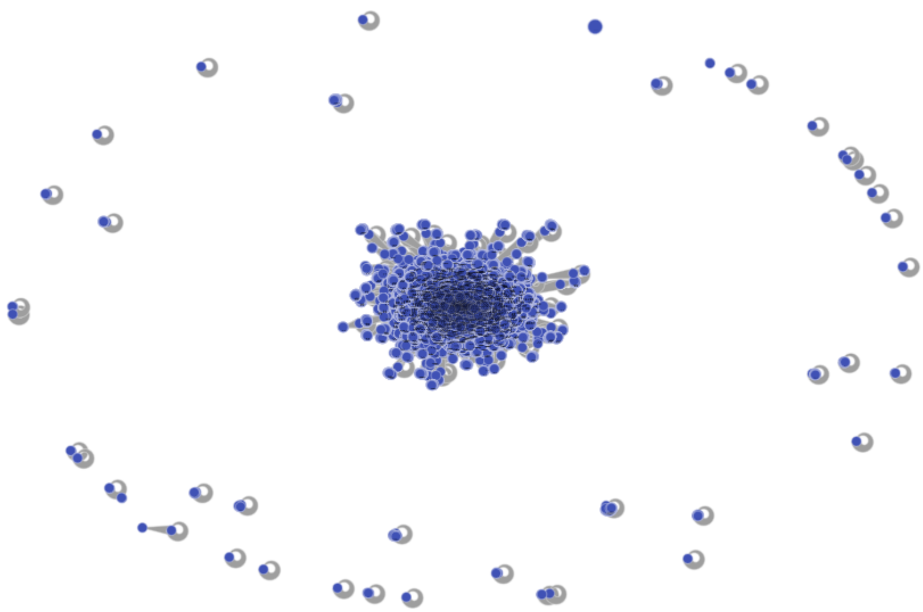


Figure 3-3: The informal twitter network, containing a single large central connected component surrounded by a few satellites of connected islands.

The resulting network was quite highly connected - with a large central connected component surrounded by smaller islands (the image fidelity of the figure is not enough to distinguish the thousands of nodes in the network, but merely presented to provide a visualization of the network). Reasonably, the top Twitter accounts ranked by betweenness were TechCrunch, Forbes, WSJ, realDonaldTrump, nytimes, YouTube, VentureBeat, LinkedIn, POTUS, Inc, FortuneMagazine, Google, WIRED.

Network Centrality Metrics

We analyzed the Twitter network using standard centrality algorithms including betweenness, degree, closeness, and reach-2 (the number of nodes the ego can reach in 2 steps). Additionally, because we had previously constructed our interlocking directorates network from these same individuals, we had centrality metrics (betweenness, closeness, and degree) for each individual from the formal ID network (see section 3.1.3 for a discussion regarding the construction of the interlocking directorates network).

Chapter 4

Analysis of VC Influence on Startup Success

Using the datasets and networks previously described, we conducted an analysis to measure the impact of VCs on startup success. We developed a number of research questions to inform our investigation of this topic.

4.1 Location of Successful Startups in Formal Network

Foremost, in order for us to make any conclusions regarding startup success, we needed to locate successful startups within our networks. Therefore, we investigated the question: are more successful startups more centrally located in the interlocking directorates network? To answer this question, we looked for correlations between the various centrality measures and our three dependent variables (total funding, annual sales, and ROI).

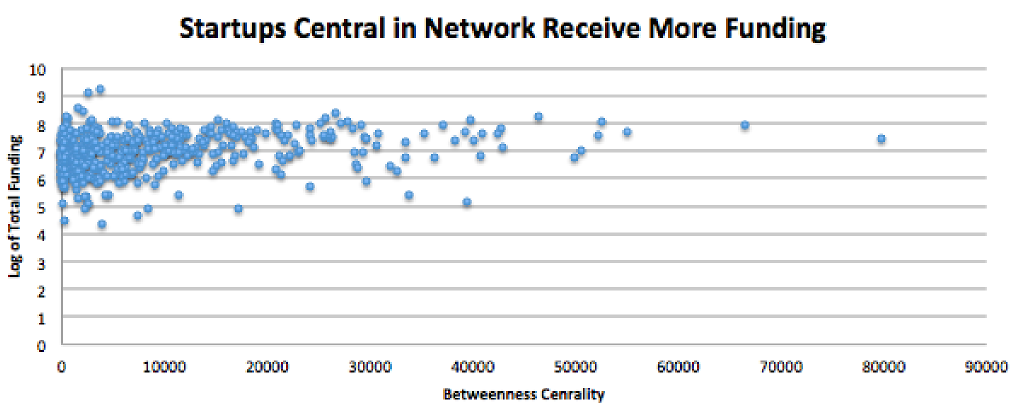
We observed strong positive correlations between total funding and annual sales and all three centrality measures, as seen in Tables 4.1 and 4.2 and in figure 4-1.

Table 4.1: Correlations between centrality measures of formal interlocking directorates network and total startup funding from Crunchbase.

Correlation with:	Log of Total Funding	
	r	p
Betweenness Centrality	0.217	1.51E-17
Degree Centrality	0.312	1.50-35
Closeness Centrality	0.220	4.52E-18
n=1514		

Table 4.2: Correlations between centrality measures in formal interlocking directorates network and log of annual sales from One Source.

Correlation with:	Log of Annual Sales	
	r	p
Betweenness Centrality	0.232	7.33E-08
Degree Centrality	0.298	3.27E-12
Closeness Centrality	0.197	5.17E-06
n=525		



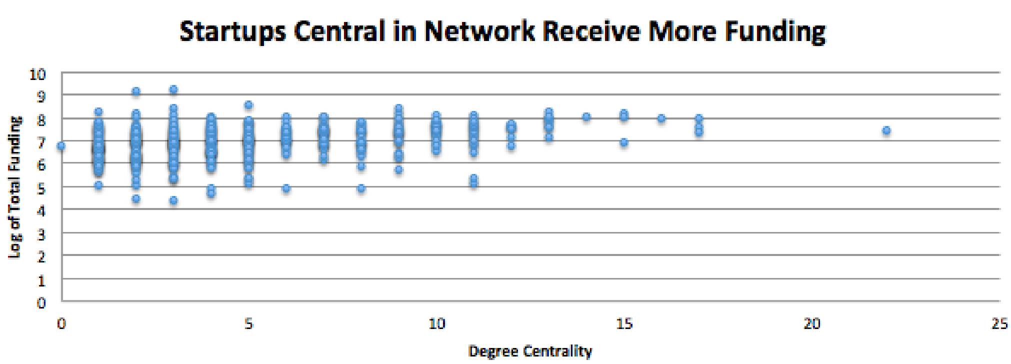


Figure 4-1: This figure demonstrates that both betweenness and degree centrality in the formal interlocking directorates network is correlated with greater startup funding.

Based on these results, we can conclusively say that startups which are more central in the interlocking directorates network tend to receive more startup funding and generate higher sales revenue. In constructing the formal network, we observed that venture capitalists were much more centrally located than non-VCs, as is commonly understood. We therefore proved the hypothesis that startups with more VCs on their board are more successful, in terms of startup funding and annual sales. What was much less clear was the relationship between VC participation on a startup's board and that startup's ROI. To address this question, we studied our dataset of 525 startups, consisting of a total of 1,803 board members (473 of whom are venture capitalists). VC Board membership is distributed throughout the dataset, as seen in figure 4-2.

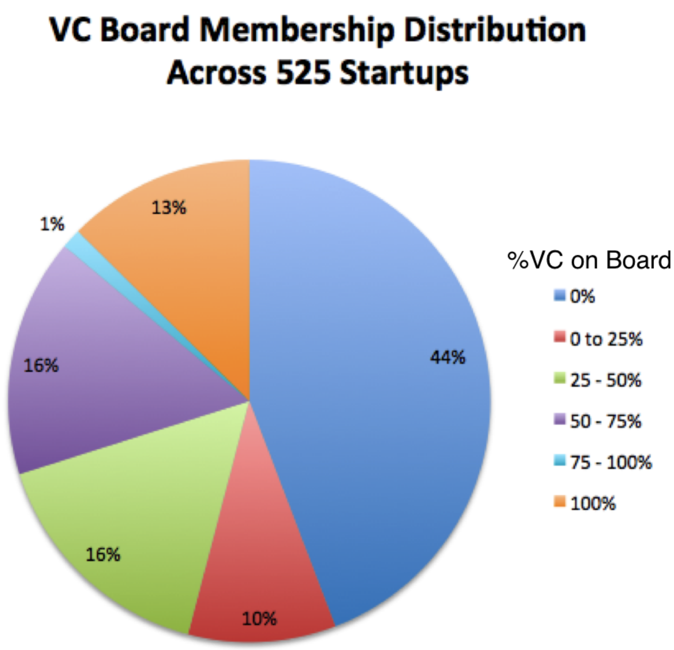


Figure 4-2: Distribution of VC board membership across our dataset of 525 startups.

To answer our research question, we investigated correlations between our three success metrics and VC board membership. We observed positive correlations between VC board membership and total funding (r=0.29, p=6.02E-12, n=525) and between annual sales (r=0.21, p=1.99E-06, n=525). This is not surprising, as venture capital firms which invest in startups often negotiate for board representation, so a correlation between total funding and VC board membership is to be expected. We have already seen a correlation between total funding and annual sales, and this holds true as well. Startups with more VCs on their board tend to have higher annual sales.

We then looked at VC board membership and ROI, defined as the ratio of revenue to funding. Surprisingly, we observed a negative correlation between board membership and ROI (r=-0.10, p=0.02, n=525). We verified these results with a Welch Two Sample t-test, where we broke the dataset into two groups: startups with VCs on their board, and startups without. All t-tests proved statistically significant, as seen in table 4.3.

Table 4.3: T-tests demonstrate that startups with VCs on their board have greater funding and sales, yet lower ROI.

T-Test	Mean (no VCs on Board)	Mean (Nonzero VCs on Board)	p-value
Log of funding	6.679	7.207	9.431e-10
Log of sales	0.4387	0.7639	$<$2.2e-16
ROI	2.15	0.737	0.00273

n=525

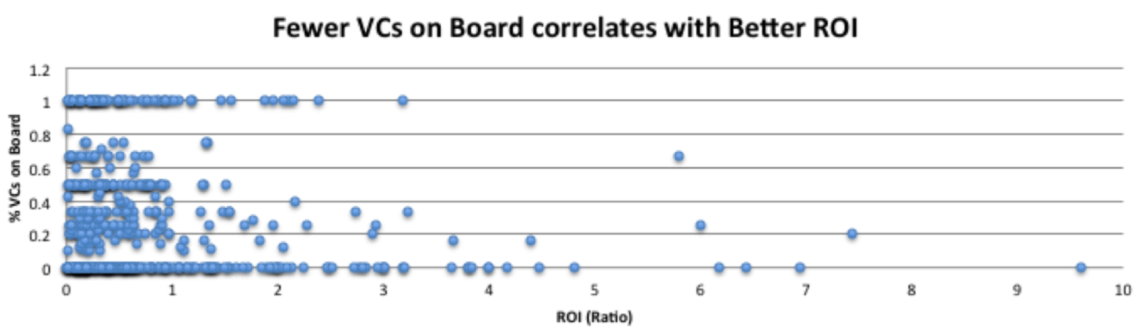


Figure 4-3: Distribution of the percentage of VCs on a starups' board of directors versus the startups' ROI. Fewer VCs on the board correlates with higher ROI.

Turning to the raw data, we discovered that startups with VCs on their board earn on average \$6.81M annually in sales more than startups without VCs on their board, and receive \$15.7M more in funding. However, startups without VCs conclusively experience higher ROI, specifically 191% higher than startups with VCs. When sorted by ROI, the top 20% of startups have on average 18% VC board membership, whereas the bottom 20% of startups have on average 31% VC board membership. This is statistically significant ($p=0.0037$, $n=525$).

Discussion

Given these findings, what does VC membership mean for a startup? We've conclusively demonstrated that more VCs on your board correlates with more funding. This is reasonable, because when VCs commit funding to a startup, it is in their best interest to support that startup's success, not just financially but also by providing advice and opportunities. VCs often request a board seat to gain authority over the startup, and potentially influence decisions that will yield the greatest return on the VC's investment. We've also conclusively demonstrated that more VCs on a startup's board correlates with more sales revenue. There are a variety of explanations for such a correlation. Foremost, startup funding amount and sales amount are correlated, which is logical. The more funding the startup receives, the more resources it has to generate revenue. Additionally, receiving funding is an indication that venture capitalists place trust in a startup and expect a return on their investment, so it is logical to expect such startups to generate greater sales. This said, it is somewhat surprising that VC board memberships correlates negatively with ROI. This is neither good from the startup's perspective nor the venture capitalist's perspective. From the startup's perspective, startup funding is not free money, but rather money that is traded for company equity and potentially decision making power. The more funding a startup receives, the less equity the founders and employees themselves get to keep. Company sales, however, represent money being generated by the company that contribute to a company's valuation. A higher ROI ratio (sales / funding) indicates that the value generated by the startup itself - the founders and the employees - is likely to stay within the startup and not be diluted by external investors such as VCs who "purchased" equity via investment. From the startup's perspective, a low ROI indicates that the startup has taken on a substantial amount of funding but not seen a relative level of sales; and are therefore in the "trough" phase of startup creation (see figure 4-4). We have shown that low ROI correlates with a greater percentage of VCs on the board. It is likely that in this early stage of startup creation, more VCs invest in a startup and gain board membership, yet the cash they pump into the company in the form of investment is not matched in terms of startup revenue generated.

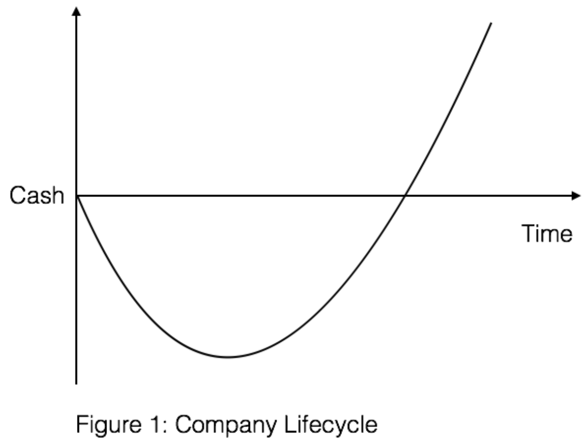


Figure 4-4: The typical company lifecycle in terms of funding, from initial loss to exponential gain.

This is a disappointing realization, especially since VCs typically invest in startups with high growth potential. They expect a rapid return on their investment, and will typically do what it takes to direct a startup down this path to rapid growth. That said, it is well known that venture capitalists invest in a large suite of portfolio startups with the hope of just a small percentage yielding massive returns. According to Dave McClure, a partner at the VC Firm 500 Startups, 50-80% of startups yield no exit or return. 15-25% yield a small return of 2-5 times investment amount. 5-10% of investments might reach a valuation of $100 million with exits yielding 10-20 times investment amount. And unicorns are, of course, extremely rare (<1% reach $1 billion valuations returning 50 times investment amount or more). In summary, McClure concludes that "...most startup investments fail, a few work out ok, and a very tiny few succeed beyond our wildest dreams" [McClure, 2015]. Perhaps this is simply the innate process of VC funding, and our data analysis exposes the inefficiencies of the system. On the other hand, startups with high ROI seem to have less VC members on their board. These startups likely did not receive high amounts of funding, yet are generating a disproportionately large amount of sales. Because they did not receive a large amount of startup funding, they were not in a position to need to accept VCs onto their board. This does seem to lead to the logical conclusion that startups without VCs on their board - and therefore without VC funding - seem to do better in terms of ROI, at least in the first 5 years. This concept has been written about numerous times in a number of entrepreneurship blogs and articles, with the underlying recommendation (as one article put it): "If you are looking towards more measured growth for your startup, want to keep control or you're simply not established yet, you probably want to avoid VCs" [Jee, 2016]. The evidence is clear - avoid VC investment if possible.

4.2 Comparing Formal & Informal Networks

Using our formal network, we demonstrated VC membership on a startup's board is positively correlated with funding and revenue, yet negatively correlated with ROI. This already provides evidence that the formal structure embedded within the VC and tech startup industry enables VCs to associate with successful startups and likely boost their success. However, our formal network was simply a proxy for communication among people in this environment, namely the communication that happens in a boardroom. That said, it is well known that communication happens much more informally than this - and much more frequently. We believe that by analyzing a different, more informal, network of communication, we would likely reach new insights regarding the influence of VCs on the success of startups.

We began our analysis by first comparing the formal network with the informal network, as both networks were composed of the same people. Our approach to analyzing the congruence of these two networks was to use first and foremost a Pearson's correlation between the corresponding centrality measures in each network. We found no significant correlation between closeness nor degree centrality, and although we did find a significant correlation in betweenness centrality ($r=0.093$, $p=.001$, $n=1271$) this is quite low and not representative of a strong correlation. See table 4.4 for details. Note that the interpretation of this correlation as low is distinct from all other correlations we make claims on in this thesis, as other correlations are between real-world variables and a variety of uncontrolled environmental factors, whereas in this correlation we truly are comparing two controlled metrics. A meaningful correlation would need to be much higher than 0.093 to be able to consider the networks significant, and even then we would need to use a different correlation calculation - namely QAP (Quadratic Assignment Procedure) to accurately assess congruence. Therefore, we conclude that these two networks really are not very similar.

Table 4.4: Pearson correlations between centrality measures in the interlocking directorates network and the Twitter network

	Correlation (r)	Significance (p)
Betweenness	0.093	0.001
Closeness	0.025	0.375
Degree	0.000	0.999

n=1271

That said, we wondered if the core networkers - the top 100 most central people - were similar between the two networks. To fairly perform this comparison, we extracted from the interlocking directorates network all those people that did not exist in the Twitter network, leaving only the 1271 people that we have in our Twitter network. We then ranked people by betweenness centrality in each network, and counted the number of people that appeared in the top 100 of each network. We did this for each centrality measure: betweenness, closeness, and degree. Generally, we found relatively little overlap in the "core" networkers in these two networks. By betweenness centrality, 13 people overlapped; by degree, 6 people overlapped (see table 4.5). Overlap due to closeness was inconclusive, as many people in the interlocking directorates network had the same closeness metric. 13% and 6% overlap between the core central members of the two networks represent very little congruence in these networks. Therefore, we can safely conclude that the interlocking directorates network and the Twitter social network are very different networks, governed by different motivating mechanisms.

Although we conclude that the top networkers in the interlocking directorates

Table 4.5: The people who ranked in the top 100 most central people (by betweenness and degree centrality) in both the formal (interlocking directorates) and informal (Twitter) networks.

Rank	Name	Is a VC	ID Rank	Twitter Rank
		Ranked By Betweenness		
1	Jon Sakoda	N	11	66
2	Roger H. Lee	N	30	43
3	Ping Li	Y	9	65
4	David B. Aronoff	Y	7	86
5	David L. M. Sze	Y	71	49
6	John William Gurley	Y	13	2
7	Neeraj Agrawal	Y	77	79
8	James J. Goetz	N	92	22
9	Peter J. Levine	N	8	28
10	Chamath Palihapitiya	Y	16	15
11	Jason M. Lemkin	Y	41	1
12	Byron B. Deeter	N	92	83
13	Martin S. Hauge	N	65	33
		Ranked By Degree		
1	Jon Sakoda	N	68	26
2	Roger H. Lee	N	68	22
3	Peter J. Levine	N	3	31
4	Venkataraman Vishnampet Ganesan	Y	12	95
5	Brian O'Malley	N	22	39
6	Erik Benson	Y	68	71

network are generally different than the top networkers in the Twitter network, we were curious to identify the few individuals who rank in the top 100 in both networks across both measures of centrality - betweenness and degree. Just three such individuals exist: Jon Sakoda, Roger Lee, and Peter Levine, as seen in figure 4-5. These three individuals, all venture capitalists with a long history in the tech startup industry, sit on 12 or more boards (as compared to the average number of boards someone sits on in our dataset which is 3.92) and have an above-median[1] number of Twitter followers (as compared to the median number of followers in our dataset

[1]Median used here rather than mean because of extreme outliers which is typical with Twitter.

which is 1142).

Name	Primary Professional Record	Is a VC	# Current Board Memberships	Twitter Handle	# Twitter Followers	Twitter Description
Jon Sakoda 	New Enterprise Associates	Y	14	@jonsakoda	3113	Entrepreneur turned Venture Capitalist General Partner at New Enterprise Associates
Roger H. Lee 	Battery Ventures	Y	19	@rogerleevc	1433	Software/Consumer investor, ex-entrepreneur, dad. All comments are my own and do not reflect the views of Battery Ventures; nothing here is investment advice.
Peter J. Levine 	Andreessen Horowitz LLC	Y	12	@Peter_Levine	9404	General partner at Andreessen Horowitz (http:// www.a16z.com), lecturer at the Stanford GSB, and an avid ice climber and mountaineer.

Figure 4-5: The 3 people who ranked in the top 100 in both formal and informal networks, ranked by betweenness and degree centrality.

4.3 Venture Capitalists in the Twitter Network

Our comparison of the formal network with the informal network demonstrated that the two are very different networks, with no clear correlation between individuals' position between the two. That said, in our continued attempt to ascertain the influence of venture capitalists on the success of startups, we reasoned that an investigation of venture capitalists' behavior and location in the Twitter network would inform our understanding of their communication patterns. We previously demonstrated that venture capitalists are most central in the formal network, but what about their communication and/or social capital might relate to their influence on a startup?

From our previous study, we found venture capitalists to be centrally located in the interlocking directorates network. This held true for our smaller network composed of only the 1271 people we found Twitter usernames for. We expected venture capitalists to be centrally located in the Twitter network as well.

Using a two-tailed unequal variance t-test, we observed VC members are more central in both the formal interlocking directorates network as well as the informal Twitter network. We measured statistically significant differences in a number of centrality measures among VCs and non-VCs, as seen in table 4.6.

Table 4.6: T-Tests on centrality measures for VCs and non-VCs in formal and informal networks.

Centrality Measure	T-Test p-value	Mean for VCs	Mean for non-VCs
Interlocking Directorates Network			
Betweenness centrality	4.19E-06	4299	615
Degree centrality	2.44E-07	4.34	3.05
Closeness centrality	4.78E-16	0.0001	0.0001
Twitter Network			
Betweenness centrality	0.012	2243375	1980370
Closeness centrality	0.029	0.0036	0.0035
Degree centrality	0.012	65.8	60.2
Reach-2 centrality	2.49E-07	918	671

Once we had demonstrated that VCs are indeed more central in the Twitter network, we were curious to determine what about their Twitter behavior - and by extension their communication behavior - influenced their network centrality.

We conducted t-tests between the VC group and non-VC group for a number of Twitter usage characteristics, including the number of followers the user has, the number of public lists the user is on, the number of tweets the user has made, and the number of people the user is following. To prevent distortion from outliers with extremely many and extremely few followers, we performed our analysis on a truncated mean dataset sample (we sorted the dataset by number of followers, and removed the top and bottom 5%). As a measure of the "popularity" of the user, we took the ratio of the number of users the person is following to number of followers the user has:

$$\frac{AnnualSales(OneSource)USDmil}{TotalStartupFunding(Crunchbase)USDmil}$$

The lower this number, the more followers the user has in proportion to the number of users the user is following, and thus the more "popular" the user.

The results of this analysis is provided in table 4.7. Generally, we discovered that VCs truly are more "popular" than non-VCs. Using as a measure of "popularity" the ratio of the number of people someone is following to the number of followers they have, VCs are significantly more popular than non-VCs (the ratio differs by 23% between the two groups, with a significance of 0.012). Furthermore, VCs have 22% more followers than non-VCs (although this result only has a statistical significance of 0.077). Additionally, VCs appear on 31% more public lists than non VCs.

Table 4.7: T-tests between VCs and non-VCs for a number of Twitter usage characteristics.

Description	T-test	Mean (VCs)	Mean (non VCs)
The number of followers this account currently has.	0.0773	4032	3303
The number of public lists that this user is a member of.	0.0127	194	148
The number of tweets (including retweets) issued by the user.	0.0004	1845	2963
The number of users this account is following.	0.3864	794	681

VCs appear to have greater social reputation on Twitter than non-VCs. Interestingly, VCs tend to tweet less than non-VCs (VCs post/repost 38% less than non-VCs). However, they have a strikingly higher Twitter popularity ratio as compared to non VCs (a higher ratio indicates lower popularity). This means that VCs truly do - at least in the digital social networking space - have a higher social reputation than non VCs.

4.4 Network Centrality and Financial Success

These results lead us to the conclusion that VCs are truly popular people - guardians not only of money but also social status and information. Clearly, this bears implications on the success of the startups they fund and sit on the board of. To this end, we wondered if we could determine a clear correlation between centrality in the Twitter social network and some measure of financial success - either in terms of the individual's personal income and/or the funding of the startup(s) they are affiliated with.

4.4.1 Network Centrality and Income

Foremost, we investigated the correlation between an individual's position in the network and their income, as measured by the average income of their residential zip code (see section 5.3 for a description on acquiring this dataset). We looked at both the formal and informal networks. We found that generally, the more central someone is located in either the formal or informal network, the higher their income (see table 4.8). This was especially true with the measure of reach-2 in the Twitter network (r=0.14, p=0.000, n=1271).

Table 4.8: Correlations between centrality measures and individual income.

Correlation with	Log of Average Income	
	r	p
Interlocking Directorates Network		
Betweenness centrality	-0.0120	0.670
Degree centrality	0.053	0.061
Closeness centrality	0.066	0.018
Twitter Network		
Betweenness centrality	0.079	0.005
Closeness centrality	0.038	0.173
Degree centrality	0.079	0.005
Reach-2 centrality	0.145	0.000

This would imply that VCs, who are generally more central in both networks, tend to have greater income. That said, we did not find a statistically significant difference in income between VCs and non VCs (p=0.25, n=830). Clearly, VCs are not the only individuals in our networks who are generating personal wealth.

Additionally, from our earlier analysis, we demonstrated that VCs have a significantly higher social reputation than non-VCs. We reasoned that perhaps this ratio would correlate with residential income. Indeed, we found just that. We found a significant negative correlation between an individuals' income (on the logarithmic scale) and their ratio of # people they are following / # followers they have. Because this ratio is inversely proportional to social reputation, the greater someone's Twitter social reputation, the greater their income (r=-0.105, p=0.00018, n=1172). Therefore, by transitive reasoning, it can be concluded that VCs are not only socially prominent people - their prominence is rewarded financially.

Our analysis so far enabled us to determine that as individuals VCs are generally more central in both the formal and informal networks, which correlates with greater income and a higher social reputation.

4.4.2 Network Centrality and Startup Funding

Additionally, we investigated whether an individual's position in the social network was indicative of the startup funding this person is affiliated with. We reasoned that those individuals more central in the network - those with higher personal income - would be affiliated with more highly-funded startups.

The Capital IQ database contains a listing of all the companies each individual is affiliated with - either as a board member or employee. If a person was affiliated with a startup we analyzed in our dataset of 1514 startups (from the formal interlocking directorates network), we had a funding amount for that startup. In total, we found funding information for at least one affiliated company for 830 people in our dataset. Less than 10% of these people were affiliated with another startup for which we had funding data, so we decided an accurate and comparable measure would be to take the maximum startup funding of all startups for which we had data for each person.

We looked at the correlation between our centrality measures in the TWitter network and the log of the maximum affiliated startup funding for the 830 people with such data. We found significant positive correlations, as seen in table 4.9

Table 4.9: Correlations between an individual's Twitter centrality measures and an individual's maximum affiliated startup funding.

Correlation with:	Log of Max Funding	
	r	p
Betweenness centrality	0.098	0.005
Closeness centrality	0.063	0.070
Degree centrality	0.109	0.002
Reach-2 centrality	0.244	0.000

n=830

All centrality measures except closeness exhibit a significant positive correlation with the startup funding metric. However, reach-2 displays the largest correlation, at 0.24. A possible explanation for this is that in real life, people are very likely to share leads/opportunities with their close friends (1 degree) and their friends of friends (2 degree). Degree is correlated (r=0.109) with funding, but significantly less correlated than reach-2 (r=0.24). We reason that degree is less strongly correlated than reach-2 because people really do use their friends-of-friends network and don't depend merely on their closest contacts. Using simply degree to indicate funding misses out on the real-world connections that happen due to friends-of-friends. Therefore, the friends-of-friends network (reach-2) is really a much better representation of how information and leads spread throughout the network. It's not who you know that counts, it's who your friends know. This finding is supported by the academic world; in fact, the strength of weak ties (indirect connections, or friends-of-friends) was first presented in Granovetter's seminal work "The Strength of Weak Ties" in 1973 [Granovetter, 1973]. With over 40,000 references to date, in his paper Granovetter presents his theory regarding the prevalence of weak ties and their efficacy in a variety of contexts. Our research appears to affirm that the VC-startup context is yet another domain where weak ties dominate.

Conclusion

In our study of the influence of venture capitalists via formal and informal networks, we have determined that VCs play a central role in the success of a startup. Startups with more VCs as board members are consistently more central in both networks, and startups central in the network receive more funding and have greater revenue. That said, startups with more VCs have lower ROI, which bears implications on the startups' long-term success. We found that people central in the formal network are not necessarily central in the informal network, although VCs in general have greater centrality in both. VCs can be considered the keepers of funding and information, and therefore hold considerable power and influence in the tech startup ecosystem.

Chapter 5

Additional Startup Success Factors

In addition to investigating the influence of venture capitalists on the success of startups, we had the opportunity to investigate more general questions related to startup funding and revenue. These investigations were primarily motivated by having acquired a large dataset that made such analysis possible, however we found several interesting insights in this way.

5.1 Number of Founders

How does the number of founders relate to startup success? To study this question, we began with our largest dataset, which consisted of 1217[1] startups for which we had both funding amount and number of founders (via Crunchbase). In this dataset, the average startup has 2.25 founders. The distribution of the number of founders across startups in this dataset may be seen in figure 5-1.

We observed a significant positive correlation between number of founders and the log of the total funding amount received by the startup (r=0.16, p=7.78E-09, n=1217). This proved to be a linear relationship, as the number of founders grows, the startup receives on average more and more funding (see figure **??**.

[1] This is the dataset of the original 1514 startups matched between Capital IQ and Crunchbase, but of these we only had number of founders data for 1216 startups.

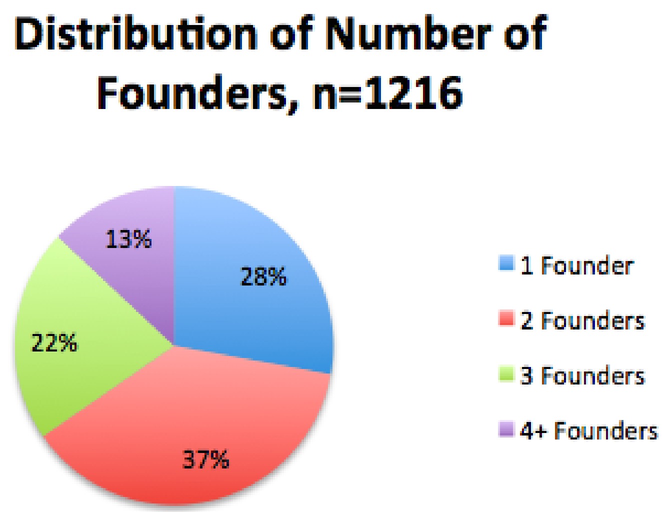


Figure 5-1: Distribution of startups by number of founders.

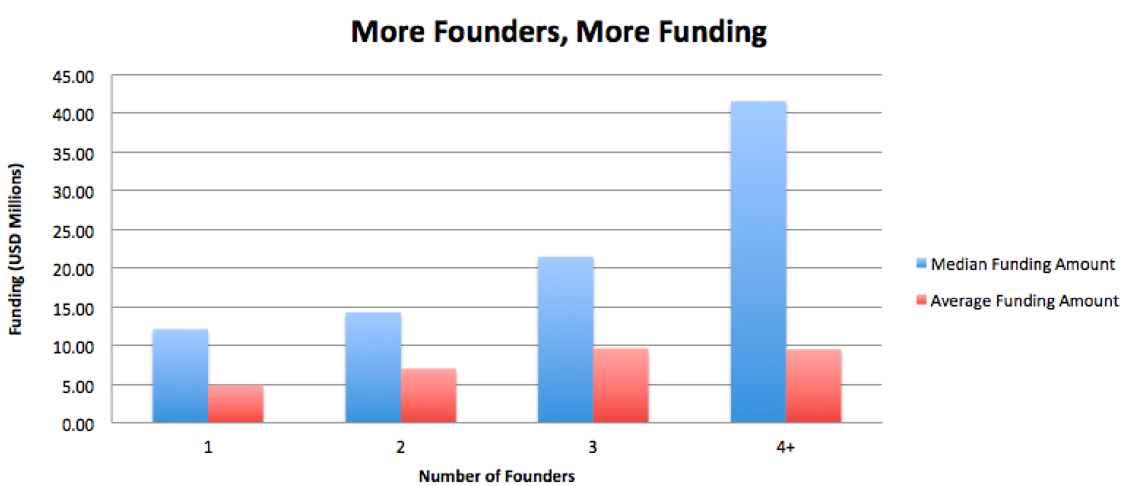
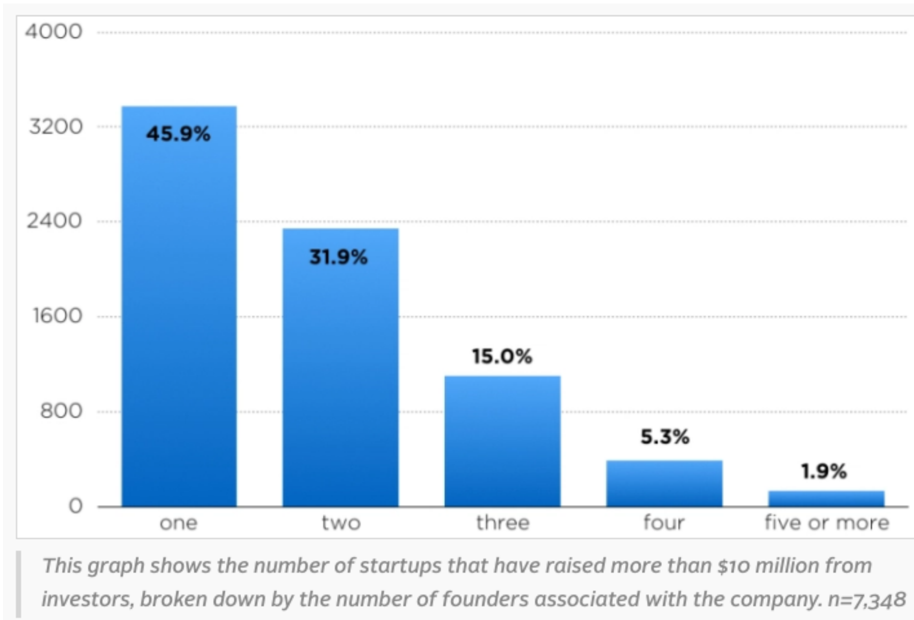


Figure 5-2: A graph showing the number of startup founders versus funding for 1217 startups in our dataset. The more founders, the more funding.

We also conducted the same analysis on our smaller dataset which included startup annual sales information. We began with our dataset of 525 startups, 460 of which contained data on the number of founders. We observed positive correlations between both the log of total funding (r=0.17, p=0.0002, n=460) and log of annual sales (r=0.11, p=0.0200, n=460).

Interestingly, a similar investigation of Crunchbase data led to a strikingly different conclusion. In August 2016, a journalist conducted an analysis of Crunchbase data and found that the number of startups that have raised more than $10 million in funding decreases as the number of founders increases [Kamps, 2016]. This calculation is different than the one we conducted, as Kamps merely counts the raw number of startups with >$10M funding, and classifies according to number of founders.



Source: [Kamps, 2016]

Figure 5-3: A figure from a similar study comparing startup number of founders and funding, provided for comparison.

We conducted the same analysis, starting with our dataset of 2171 startups and filtering by the 458 startups with >$10M funding. We found relatively similar results, with one major exception. We found fewer startups with one founder than two or three, however we did observe the trend that fewer highly funded startups exist as the number of founders increases.

Combined with our earlier analysis demonstrating that the average funding increases with more founders, we make the conclusion that two founders appear to be a nice "sweet spot" in terms of number of founders that actually seem to work in practice. It is interesting that the previous study found more single founder startups with high funding - it is possible this was because they included all companies in Crunchbase, not just 5 year old companies. That said, the article's author does

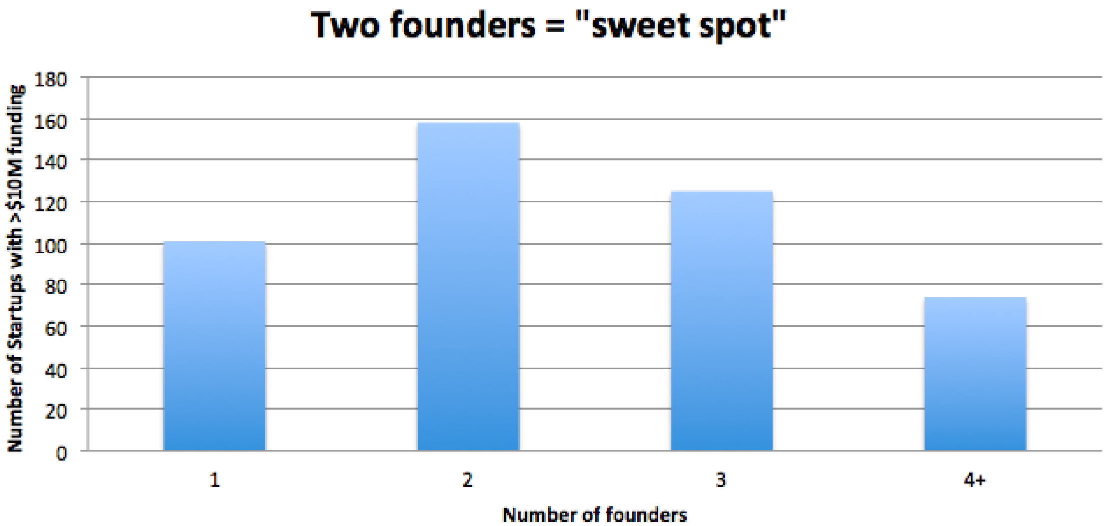


Figure 5-4: Number of founders vs. number of startups with >$10M in funding (enabling direct comparison to previous study).

conclude that "While the above data from Crunchbase suggests that it's possible to raise money and secure an exit as a solo founder, that doesn't mean it's a fantastic idea. In my opinion, building a company with 2-3 co-founders is probably the way to go" [Kamps, 2016]. Our data supports this conclusion. The most commonly highly funded startups are founded by just two people.

5.2 Female Diversity in the Networks

Another research question we found interesting to investigate involved the gender diversity within our networks, both the formal (board membership) and informal (Twitter activity).

5.2.1 Female Board Membership

Foremost, we wondered how female membership on startup boards relate to startup funding, revenue, and ROI. It is well known that females are in the minority in tech, and especially entrepreneurship. We were curious to explore whether their representation correlated with any financial measures of success.

We began with the larger dataset of 1514 startups, consisting of 4,468 total board members, 6.56% of which are female. 84.7% of startups have no females on their boards, 12.5% have one, and just 2.8% have two or more. On average, females represent 6.0% of board members of startups in our dataset. See figure 5-5 for the distribution of female board membership across our dataset.

We found a statistically significant, although small, negative correlation between female board membership and the log of total funding (r=-0.069, p=0.0076, n=1514). We then specifically excluded female venture capital members of the board, hoping to understand the correlation between non-VC female members of the board and funding. 16.7% of the female board members were identified as VCs. Upon looking at just the non-VC female board membership, we found an even stronger negative correlation with funding (r=-0.115, p=6.80E-06, n=1514). This would imply that the female VCs on the board seemed to be contributing a net positive factor to funding, which was indeed the case. The correlation between percentage female VCs on the board and funding is positive (r=0.068, p=0.0077, n=1514).

Table 5.1: T-tests on funding between startups with females and without females on their board (excluding female VCs).

T-tests	1 or more females (Non-VC) n=193	No females (Non-VC) n=1321	p-value
Log (funding)	6.51	6.66	0.006
Funding	$12.8M	$16.5 M	0.115

1 sided unequal variance t-test

Wanting to better understand the correlations between female board membership and startup success, we then conducted the same analysis, but on our smaller dataset for which we also had annual sales data. We eliminated the boards with no females and those entirely composed of females. This left us with 203 startups.

In this dataset, we observed the same negative correlation between female board membership and log of funding (r=-0.21, p=0.00235, n=203). Similar to the larger dataset, this negative correlation was exacerbated with the removal of female VCs (r=-0.365, p=8.42E-0.8, n=203). Female VCs contributed a net positive correlation

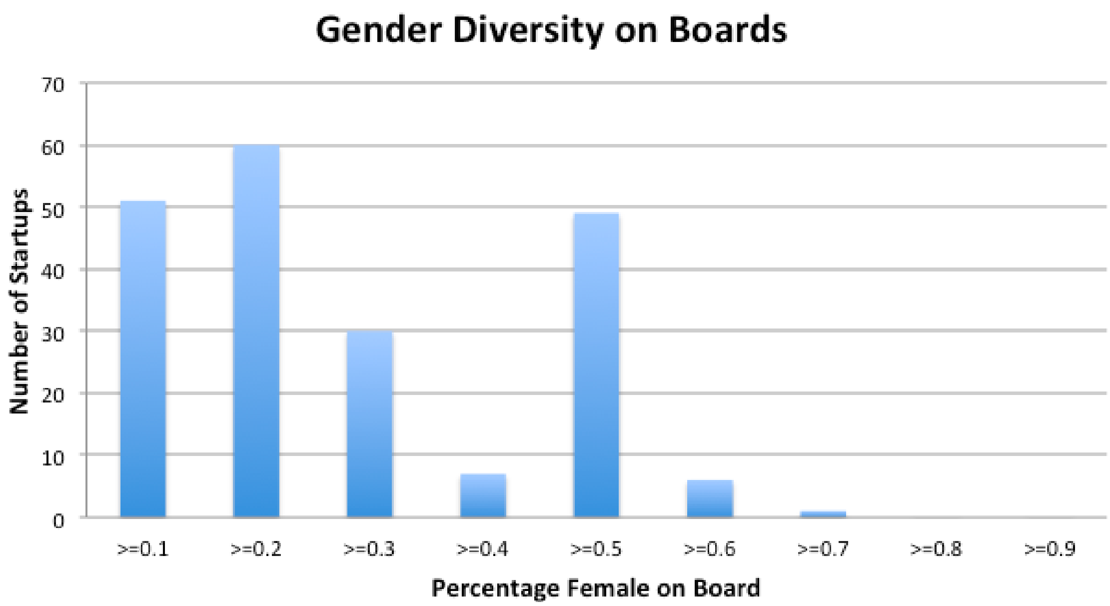


Figure 5-5: Distribution of female board membership on startups boards.

with log of total funding (r=0.208, p=0.00287, n=203). These correlations exhibit the same behavior to the correlations observed with the larger dataset. Fundamentally, greater female board membership is correlated with less total startup funding, and this is especially true for females who are not venture capitalists.

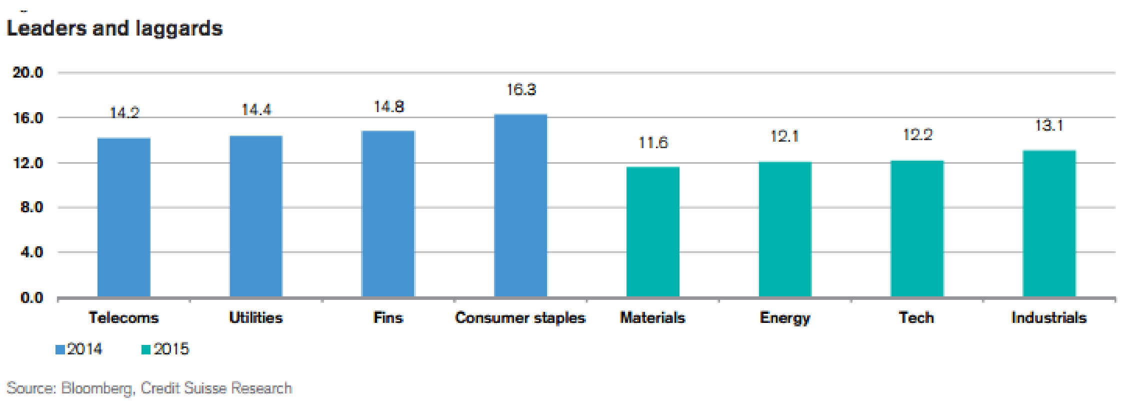
We did not find any statistically significant relationships between female board membership and revenue or ROI. This may be due to the small size of the dataset (n=72), in addition to the fact that the dataset became non-normally distributed once we removed startups for which we did not have funding data.

Based on our finding of a negative correlation between female board membership and funding, we were curious to investigate the position of startups with female board member representation within the interlocking directorates network. We hypothesized that those startups with greater female board membership would be less central in the network, as we had already seen a strong correlation between funding and centrality in the network.

Indeed, we did observe strong negative correlations between a startup's percentage of female board membership and its centrality in the network. (Betweenness centrality: -0.57, Degree centrality: -0.61, Closeness centrality: -0.59, all significant to <0.001).

Discussion

Foremost, there has been considerable research conducted regarding the representation and role of females in the tech startup industry, including their participation on the board. A September 2016 study by the Credit Suisse Research Institute looked at 3,400 public companies worldwide in various industries and found an average female representation of 14.7% at year end 2015. Limiting to just the tech industry brought the number down to 12.2%, which is double the 2010 percentage of 6.8%. Of all the industries, IT was second lowest (just above Materials at 11.6%) regarding female participation on boards, as seen in figure 5-6



Source: [Suisse, 2016]

Figure 5-6: Female board membership (by percentage) across various industries according to a study by Credit Suisse.

In comparison, we found female participation on tech startup boards to be 6%, half what was found by Credit Suisse [Suisse, 2016]. This discrepancy is disheartening, and can be accounted for by the fact that Credit Suisse looked at public tech companies, whereas our dataset was predominantly composed of much younger tech startups. This indicates quite clearly that board membership diversity is not only generally low within the IT industry, but even lower among tech startup boards.

Regarding funding of startups with female board representation, it is known that female founders receive a small percentage of all venture capital funding: according to one study, just 2.7% [Weisul, 2016]. Furthermore, in 2015, only 14% of the startups with at least one woman as an owner were successful in raising funding. This compares to an overall rate of 18% for all companies [Suisse, 2016]. However, to our knowledge, no other research has demonstrated a negative correlation between female board membership and lack of funding. The correlation we demonstrate is alarming. Given the assumption that startups founded by men and women are equally good ideas and equally likely to succeed[2], why is it that startups with more female board members receive less venture capital funding?

We provide evidence for one potential explanation - these startups are not "highly networked" in the interlocking directorates network, and therefore VCs simply have less chance to meet and exchange with them. We provide abundant evidence in this thesis that more highly networked individuals and startups achieve greater success (both in the formal and informal networks). That said, females face many hurdles to become "highly networked" in a male-dominated industry of tech entrepreneurship. A number of factors inhibit female networking, including cultural bias, prejudice, biological female personality traits and tendencies, etc. Furthermore, active attempts to improve a female's network are often relatively unsuccessful. In a 2012 survey conducted of 74 women working in middle management across three organizations, 67% believed networking helped in building their career, yet their networking actions were ineffective in helping them achieve their aims [Vongalis-Macrow, 2012]. Essentially, women claimed that the kinds of actions they identified as critical to networking (including helping others, offering support to others, offering career advice, and supporting the career plans of others) did not showcase their talents or promote their goals [Vongalis-Macrow, 2012]. This may be one explanation for why female-heavy startups remain marginalized in the interlocking directorates network.

The question of course becomes, should anything be done about this, and if so, what?

A growing body of research has demonstrated the advantages of diversity - on

[2] This assumption may very well not be true, especially given the variety of cultural and social factors that inhibit female entrepreneurs. However, we operate under this assumption in order to explore various other factors that may cause the lack of funding awarded to startups with female board members.

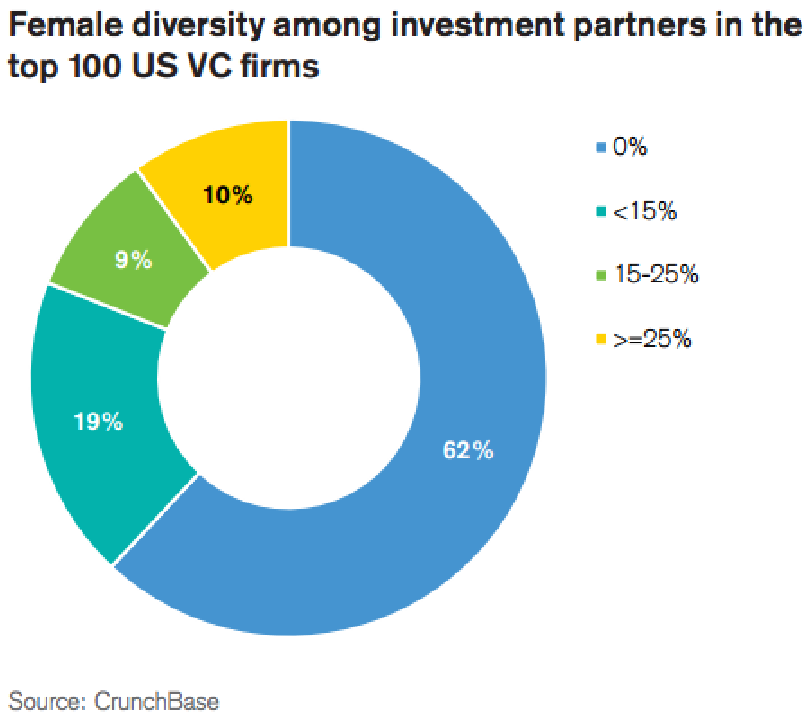
teams, in the boardroom, as startup founders, etc. According to a study by First Round Capital, companies with a woman on the founding team outperform their all-male peers by 63 percent [Capital, 2015]. Furthermore, and more generally, a study on the collective intelligence of groups demonstrated that performance was positively and significantly correlated with proportion of females in the group [Woolley et al., 2010].

That said, research into the success of female-led startups yields concerning results. In a July 2016 article from Harvard Business Review, a study found the following: "If you deme success as an exit from venture capital financing via acquisition or an IPO, female-led startups perform much worse than male-led startups. About 17% of female-led startups successfully exit VC financing whereas 27% of male-led startups do" [Raina, 2016]. The author uses this evidence to refute the claim that VCs hold female-led startups to a higher standard than male-led startups, and therefore are much more selective with allocating funding.

Interestingly, the author offers an alternative explanation that is very much in alignment with some evidence we present in our research regarding the influence of female VCs in terms of startup funding. The author finds that with startups backed by all-male VCs, there is a 25% difference in the exits of female-led and male-led startups. However, when startups are financed by VCs with female partners, this difference is no longer statistically significant.

Similarly, our research indicates that female VC membership appears to "counteract" the negative correlation between female board membership and startup funding. Unfortunately our dataset was too small in order to be able to make any conclusions regarding whether or not female VCs increased funding specifically for startups with female board members, but what we can say more generally is that startups with female VCs on their boards are actually more likely to receive funding according to our research.

So both our research and related research indicate that one solution to the problem of gender diversity in the tech startup industry is increased female VCs. However, the state of gender diversity in the VC industry is, alarmingly, even worse than in the IT startup industry. Among the top 100 VC firms globally in terms of deal flow and size of funds, just 7% of the partners are women and only 38% have at least one female partner [Suisse, 2016], as seen in figure 5-7.



Source: [Suisse, 2016]

Figure 5-7: Female diversity among investment partners in the top 100 US VC firms.

The good news is that VC firms with with female founders tend to invest more in startups with female founders relative to the general average. Furthermore, there is evidence that the number of new firms founded by women is growing much faster than those founded by men and these firms are capturing a growing share of VC funding [Suisse, 2016]. It is a glimmer of hope amidst a sea of dismal data.

Based on our research and review of related research, we propose the following recommendations:

1. Startups with female founders and board members must make a concerted effort to improve their networking position relative to their pears - likely by venturing outside of their comfort zone.

2. VCs should be aware that female-heavy startups consistently receive less fund-

ing, yet there is no evidence to suggest these startups are of less value. Furthermore, given proper VC mentorship (especially from fellow females), these startups are likely to succeed. VCs should make a concerted effort to fund female-heavy startups, as such a strategy is likely to prove rewarding.

3. Females should be encouraged to join venture capital firms - as their participation will likely render the firm more successful as female participation on teams and as mentors has been proven to correlate with more successful outcomes.

5.2.2 Gender and the Twitter Social Network

We saw how gender played a large role in the formal interlocking directorates network - females were significantly less central than males, and therefore were associated with startups of less funding. We therefore wondered if this same tendency appears in the Twitter social network?

Our dataset of 1271 people from which we constructed the Twitter social network contained 97 females (7.6%). This is generally representative of the female representation in the original interlocking directorates network of 8473 board members, which contained 7.5% females.

We first looked at the number of boards each individual in the network sits on. From Capital IQ, we had a listing of all the company boards (not just startup boards) the person currently sits on, and all the boards the person has sat on in the past. Of course, some amount of inaccuracy is inherent in the data, but generally we assume that this inaccuracy is distributed randomly.

On average, females currently sit on 3.05 boards as compared to males who sit on 3.99 boards. Using a two-tailed paired t-test, this is significant ($p=0.015$, $n=1271$). However, in terms of prior board memberships, they sat on 1.35 boards as compared to males who sat on 2.46 boards on average ($p=0.0006$, $n=1271$). Although it is true that females currently sit on nearly one fewer boards than males, this is still slightly better than the past. Today, males sit on 1.3 boards for every 1 board a female sits on, whereas in the past males sat on 1.82 boards for every board a female sat on. Furthermore, the significance of this difference is decreasing, as evidenced by the t-test probability of 0.015 for current board membership versus 0.0006 for past board membership (see table 5.2.

Table 5.2: T-tests on the number of current board memberships and previous board memberships held by women versus men.

T-Tests	p	Mean (Females)	Mean (Males)
Number of current board memberships	0.015	3.052	3.99
Number of prior board memberships	0.0006	1.350	2.46

Although these are just two data points, it certainly appears as though females are becoming more prominent in this highly male-dominated network.

Given this observation, we were curious to investigate the difference in centrality measures between females and males - both in the interlocking directorates network as well as the Twitter network (see table 5.3).

Of the six t-tests we conducted (2 networks x 3 centrality measures) we found only one to be statistically significant; betweenness centrality in the interlocking directorates network (p-value of 0.0007, average=391 (females), 1383 (males). However, this is because a much higher percentage of females have a betweenness of zero as compared to males (93% of females had a betweenness of zero as opposed to only 53% of males). This indicates that generally a greater percentage of females are "marginalized" in the network than males, as more significantly more females than males are not on any shortest paths between two nodes in the network. Hence, information can more efficiently circumvent females more readily than males as it flows through people in the network.

That said, no statistically significant differences were observed between females and males when analyzing the two other centrality metrics (closeness and degree) in the interlocking directorates network. Furthermore, no significant difference was observed between females and males when looking at any of the centrality metrics in the Twitter network.

Furthermore, we looked at a number of other statistics related to Twitter behavior (number of followers, number of friends, number of statues) and none demonstrated a measureable difference between males and females.

Table 5.3: T-tests on a number of statistics between males and females, mostly demonstrating lack of significance.

Metric	T-test p-value
Betweenness centrality (ID network)	0.0007
Degree centrality (ID network)	0.8291
Closeness centrality (ID network)	0.9310
Degree (Twitter network)	0.7255
Reach 2 (Twitter network)	0.5645
followers_count	0.0012
statuses_count	0.1220
friends_count	0.4366
Ratio #following / #followers	0.9532
Normalized Education Rank	0.4366
Normalized Log of Average Income	0.6947
Metric of Educational Prestige and Professional Affluence	0.8555
Max Funding	0.0014

It can be concluded that although females are certainly the minority in this network, and to some extent they are not as central in the network as their male peers, there is no trace of this discrepancy via social media nor in measurable personal outcomes. This we take as a sign of hope that the system is fair and as time progresses and females become more highly represented on boards, they will in turn gain social capital and more of them will join the elite group of highly connected individuals, currently a group dominated by men.

5.3 Influence of Educational Prestige and Social Capital on Income

Because we had constructed a dataset of individuals with their social network, and we had information about their education, namely their university, we were curious to investigate how these variables were related. Specifically, we aimed to investigate the question: does educational prestige and how central someone is in a social network (i.e. Twitter) tell us something significant about their income? To do so, we conducted additional data collection and data preparation, as described below.

Education Level

The Capital IQ database provided a listing of universities that each individual attended. This information appeared to be (surprisingly!) quite complete and accurate. In total, 886 of the 1271 people in our network had at least one university listed. Those with no university listed were excluded from this analysis.

To transform this list of universities into a numerical score of "education prestige", we first mapped each university name to a rank. We used two lists to accomplish this ranking. The first was the U.S. News & World Report ranking of the 2017 Best National Universities[3]. This list ranked the top 231 universities in the USA, selected from a pool of more than 1,800 universities. To include universities from outside the USA, we integrated the 2016 World University Ranking published by the Times Higher Education World University Rankings[4] into our ranking system. Specifically, for all international universities, we identified their position in the World University ranking, and assigned to it the closest more highly ranked US University Ranking equivalent. Careful attention was given to variation in school names (The Ohio State University vs. Ohio State University for example).

For example, the highest ranking school in France (École Normale Supérieure) is ranked 54 on the Times Higher Education World University Ranking. The closest more highly ranked US school on the list is Brown University, which is ranked 14 in the national listing. Therefore, École Normale Supérieure received the ranking 14 (see table 5.4).

This ranking seemed most logical, as most individuals attended US-based schools, so accounting for the outliers who attended non-US schools by mapping their school to US-schools did not introduce an inordinate amount of inaccuracy.

[3] https://www.usnews.com/info/blogs/press-room/articles/2016-09-13/us-news-releases-2017-best-colleges-rankings
https://data.world/education/university-rankings-2017

[4] https://data.world/hhaveliw/world-university-ranking-2016
https://www.timeshighereducation.com/world-university-rankings/methodology-world-university-rankings-2016-2017

Table 5.4: Excerpt from World University Ranking 2016 to illustrate ranking methodology.

Rank	School	Country
51	Brown University	United States of America
52	Australian National University	Australia
53	Technical University of Munich	Germany
54	École Normale Supérieure	France

Ultimately, each person was assigned a "university prestige" ranking based on the most highly ranked university they attended. For example, if they attended the Massachusetts Institute of Technology (ranked 8) for undergrad and Harvard University (ranked 2) for their MBA, they were assigned a ranking of 2. If someone attended no schools on either the national or international listings, they received a rank of 232, which is one rank below the "lowest" ranked university. Finally, these rankings were normalized on a 0 to 1 scale, with 1 being the highest university ranking, and 0 being the lowest. On average, our dataset contained individuals in the upper 30 percentile of educational ranking, with an average normalized university ranking of 0.72 with standard deviation 0.37. We call this the normalized school rank metric.

Individual Income

Although it is impossible to know the income of each individual in our network, we reasoned that an accurate estimate of a person's income would be the average income of the location where they live, according to the US Internal Revenue Service.

The Capital IQ database provided the residential city and zip code of each individual. This information appeared to be (surprisingly!) quite complete and accurate. In total, 1220 of the 1271 people in our network had a valid zip code listed.

To transform zip code to some metric of wealth/affluence, we used zip code data published by the IRS from 2014 [5], which is the most recent release of this dataset. We mapped a person's residential zip code to the average income reported for that zip code. Our dataset contained on average high income zip codes, with an average

[5] https://www.irs.gov/uac/soi-tax-stats-individual-income-tax-statistics-2014-zip-code-data-soi

income of $211.29 K but a large standard deviation ($178.17 K).

We then transformed these numbers into a normalized "professional affluence" ranking on a log scale from 0 to 1, where 1 was the most affluent and 0 the least. This resulted in a dataset of an average 0.72 professional affluence score, with standard deviation 0.10.

Analysis

To answer our research question, we developed a metric to measure the combined educational prestige and social capital of an individual. The metric is based on the normalized school rank (as described in the previous section) as well as the reach-2 metric calculated from the Twitter network. We assert that school rank is a measure of a person's educational prestige, and reach-2 is a measure of their social capital - both in the digital world as well as in the real world. The combined metric is calculated as the product of these two numbers: $schoolRank_{normalized}$ x $reach2_{normalized}$

We looked at the correlation between each of these metrics and the log of average income for each individual, as seen in table 5.5.

Table 5.5: Correlations with average income between school rank, reach-2, and the combined metric. All are significantly correlated.

Correlation with	Log of Average Income	p	n
$schoolRank_{normalized}$	0.076	0.0241	886
$reach2_{normalized}$	0.145	0.000	1220
Educational Prestige and Social Capital Metric	0.18	0.000	865

Clearly, both educational prestige and social capital correlate with income. Interestingly, the combined metric proves to be the most highly correlated with average income.

5.4 Top Startup Categories

The world of tech startups is vast, however amidst the sea of emergent companies, it is well known that certain domains of tech are most popular throughout the community. We investigated what categories of IT startups were most successful in terms of receiving funding. Startups are labeled via Crunchbase into categories, and a startup may have multiple labels. The distribution of IT categories is shown in figure 5-8.

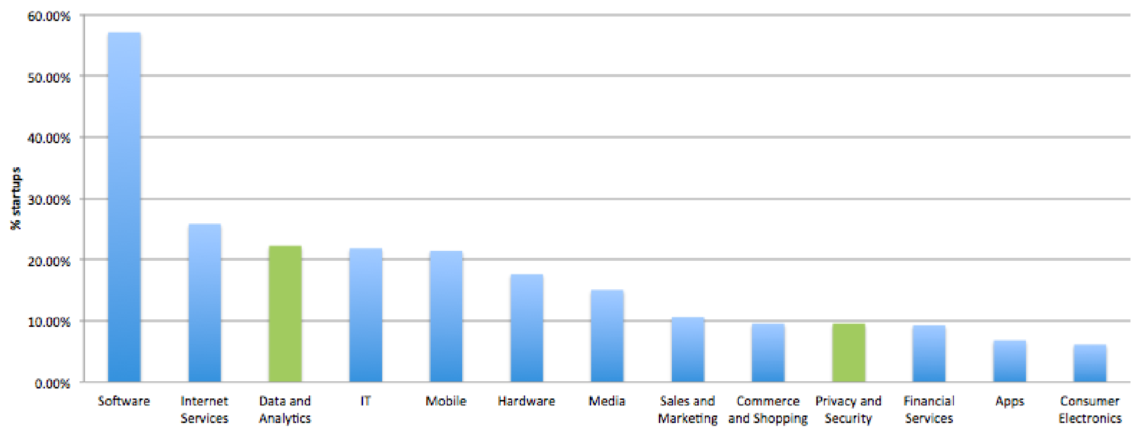


Figure 5-8: Distribution of IT categories among dataset.

We identified the top two categories with the highest correlation with the total funding: Data and Analytics (22.3% of startups) and Privacy and Security (9.6% of startups). Data and Analytics startups receive $1.5M more in funding than the average startup, and Privacy and Security Startups receive $4.7M more in funding than the average startup.

It is well known that two hot fields in technology right now are Data and Analytics and Privacy and Security. Our data certainly reflects this, as startups with those categories do indeed receive a measurably higher amount of funding. We reasoned that these startups would also be more central in the interlocking directorates network, and found that indeed, these two categories are more central (see table 5.6).

Table 5.6: The top two categories with the highest correlations with the log of total funding amount.

Category	# Startups	Log (Total Funding Amount)	Betweenness Centrality	Degree Centrality	Closeness Centrality
Data And Analytics	337	0.14**	0.06*	0.09**	0.07**
Privacy and Security	145	0.13**	0.14**	0.15**	0.11**

$p<0.05$: *, $p<0.01$: **

5.5 Residential Location and Network Centrality

Certain locations in the United States are well known as tech startup hubs - namely Silicon Valley as well as the area of Boston/Cambridge. We wanted to investigate to what extent people from these areas were central in our networks - both the interlocking directorates network as well as the Twitter network.

Foremost, we analyzed the distribution of locations existing in our network of 1271 people. We found that on average 75% of people lived in one of the top 5 most popular locations, ranked in decreasing order of popularity: California, New York, Massachusetts, Washington (the state), and Texas (see figure 5-9). These results indeed indicate our dataset does indeed reflect the natural distribution of people in the startup tech industry.

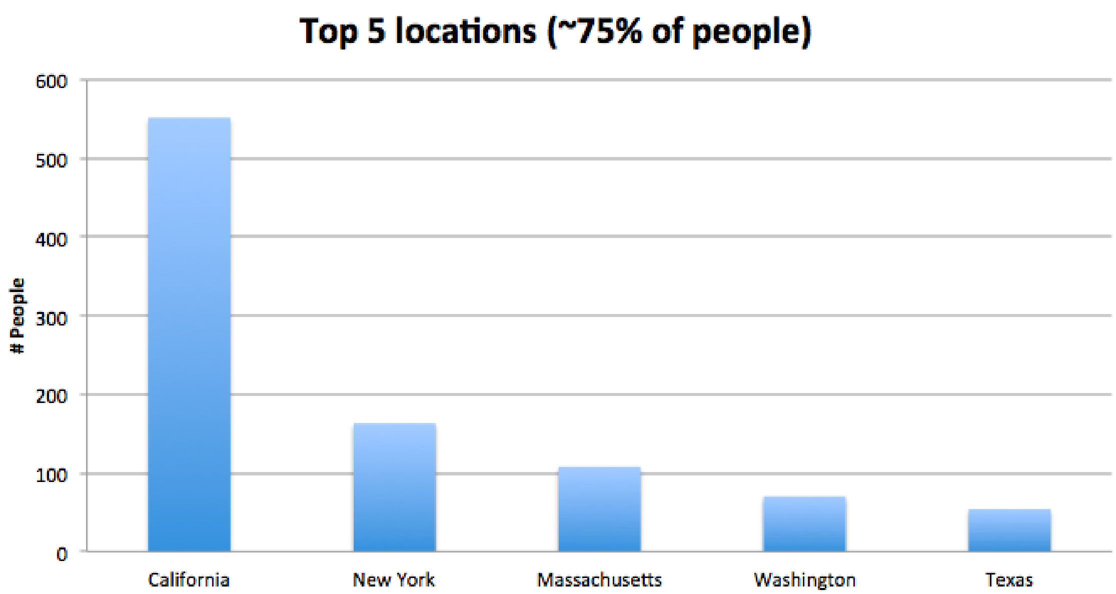


Figure 5-9: The top 5 residential locations of individuals in our dataset.

We next looked at how highly correlated individuals from each state where with centrality in our two networks - the formal interlocking directorates network and the informal Twitter network. Our hypothesis was that those from California would reasonably be most central in the network. In fact, we found significant correlations between California residency and all three centrality measures in the ID network. We also found a significant correlation between California residency and degree centrality in the Twitter network (r=0.088, p=0.0016, n=1271). Furthermore, we found significant correlations between California residency and betweenness and degree centrality in the Twitter social network (see table 5.7). These findings confirm the well known fact - the highly influential people in the tech startup domain - those who are central in both our networks - seem to reside in California.

No other state demonstrated as extreme a correlation with centrality measures. Massachusetts residents correlated with betweenness centrality in the interlocking directorates network (r=0.068, p=0.015, n=1271), but this correlation was the only significant one across the two networks and all centrality measures. Interestingly, New York residents correlated negatively with degree centrality in the interlocking directorates network (r=-0.063, p=0.015, n=1271). We reason that because New York is more of a financial hub than a tech hub, residents of New York would be less likely to be central in our tech-centric interlocking directorates network. No other significant correlations existed between state of residency and centrality measures.

Table 5.7: Pearson correlations on centrality measures of interlocking directorates network and Twitter network between startups from various states.

	r (Betweenness)	p	r (Degree)	p	r (Closeness)	p
Interlocking Directorates Network						
California	0.115**	0.000	0.101**	0.000	0.147**	0.000
Massachusetts	0.068*	0.015	0.037	0.186	0.038	0.178
New York	-0.053	0.059	-0.063*	0.024	-0.050	0.072
Washington	-0.026	0.353	-0.007	0.808	-0.029	0.298
Twitter Network						
California	0.056*	0.045	0.088**	0.002	0.044	0.119
Massachusetts	0.001	0.969	-0.022	0.430	-0.014	0.629
New York	0.006	0.844	0.013	0.646	0.018	0.531
Washington	0.015	0.604	0.026	0.353	-0.012	0.662

n=1271
Note: *p<.05, **p<.01

Chapter 6

Future Work & Conclusion

There are many ways in which the questions studied in this thesis could be further investigated. These are a few ideas for potential extensions on the research presented here.

6.1 Demonstrating Causality

One of the fundamental limitations of our dataset was that it included no temporal data, and thus we were unable to investigate cause-and-effect relationships. We were merely able to observe statistically significant correlations, and hypothesize about these correlations under the assumption that they may indeed be related via causality. Having dependent variable data (startup funding and startup revenue) over time would have facilitated a causal analysis. An interesting question to investigate would be to analyze ROI over the lifetime of a startup, and relate this to VC board participation. Our expectation would be that ROI would increase as VCs became less present on the board over time. Turning to the informal network, it would be interesting to analyze the social network behavior of VCs vs. non-VCs over time, and relate this to the performance of the startups they are involved with. One would expect VCs to gain popularity on Twitter as they become more central in the formal network - and thus become involved with more highly funded startups.

6.2 Construction of a Third Informal Communication network

The two communication networks (formal and informal) studied in this thesis were both proxies for the real-world communications that we assume occur between VCs and entrepreneurs. However, they merely represent one small fraction of the depth of communication that actually occurs between VCs, board members, and startup founders. An interesting study would be to conduct a survey of the individuals in our networks, requesting them to respond to information about their communication patterns on a weekly basis (to whom did they converse, what was the nature of the conversation, etc.) By constructing a third network based on this data, we believe we could develop intriguing insights by comparing to our formal and informal networks.

6.3 Comparing Across Geographies

This study was limited in scope to the United States. However, we reason that the entrepreneurial ecosystem - especially as related to venture capital - varies drastically among geographies worldwide. A number of factors may influence VC behavior from country-to-country, including culture, political structure, regulation, economy, and historical precedent. We would be interested in investigating whether the correlations we observed in the United States appear in various other geographies known to differ drastically according to some of these other factors. In particular, we believe a study comparing the European, North American, and Asian VC-startup ecosystems would be quite interesting.

6.4 Conclusion

In this work, we aimed to better understand the impact of venture capitalists on startup success, by analyzing both formal and informal networks. We used a people-centric network theory approach, constructing a dataset of board members who sit on 3,199 distinct US tech startups. We used this dataset to build two distinct networks: a formal network based on the interlocking directorates (board membership) network, and an informal network based on the Twitter behavior of board members. We also collected financial information regarding these startups, using a variety of sources (Capital IQ, OneSource, and Crunchbase). This dataset enabled us to observe correlations between characteristics in the networks, and ultimately informed our investigation of the research question regarding the impact of venture capitalists on startup success.

We found that startups with VCs as board members are consistently more central in both the formal and informal networks, and startups central in the network receive more funding and have greater revenue. However, startups with more VCs have lower ROI. We discussed the implications of this conclusion on the startup industry, including the growing sentiment of avoiding venture capital funding if possible. We also demonstrated that VCs are most central in both the formal and informal network, and that VCs have greater popularity in the Twitter network, although they tweet significantly less than non-VCs. We discussed how this implies that VCs have a significant influence over the exchange of money and information in these networks.

Additionally, the dataset we collected for the original research question enabled us to explore a number of more general questions, including the gender diversity of board membership, number of founders, geographic location, industry specialization, and relationship of schooling and social prominence on wealth. We show that the number of startup founders is positively correlated with startup success, and we discuss the optimal number of startup founders is likely two. We demonstrate how the number of female board members is negatively correlated with startup funding, and we discuss the implications of this finding to the startup community, and what may be done to address this issue. Finally, we observe that a person's educational rank and their social capital is correlated with their residential income.

We believe our approach to studying the aforementioned topics, namely using a people-centric network theory analysis of startup board membership, provides unique insight and inspires further investigation into the startup ecosystem.

Bibliography

O. Alexy, J. Block, P. Sandner, and A. L. J. Ter Wal. Social capital of venture capitalists and start-up funding. *Small Business Economics*, (39):835–851, 2012.

Thomas J. Allen, Ornit Raz, and Peter A. Gloor. Does geographic clustering still benefit high tech new ventures? the case of the cambridge/boston biotech cluster. *Massachusetts Institute of Technology. Engineering Systems Division*, 2009.

R. Amit, J. Brander, and C. Zott. Why do venture capital firms exist? theory and canadian evidence. *Journal of Business Venturing*, 13:441, 1998.

Ola. Bengtsson. Covenants in venture capital contracts. *Management Science*, 57 (11):1926–1943, 2011. doi: 10.1287/mnsc.1110.1409.

Shai Bernstein, Xavier Giroud, and Richard R. Townsend. The impact of venture capital monitoring. *The Journal of Finance*, 2015.

J. Block and P. Sandner. What is the effect of the financial crisis on venture capital financing? empirical evidence from us internet start-ups. *Venture Capital*, (11): 295–309, 2009.

First Round Capital. 10 year project, 2015. URL `http://10years.firstround.com/`.

Y.S. Chan. On the positive role of financial intermediation in allocation of venture capital in market with imperfect information. *Journal of Finance*, 5(5):1543–1568, 1983.

Thomas J. Chemmanur, Karthik Krishnan, and Debarshi K. Nandy. How does venture capital financing improve efficiency in private firms? a look beneath the surface. *US Census Bureau Center for Economic Studies*, 8(16), 2008.

P. De Meo, E. Ferrara, G. Fiumara, and A. Provetti. On facebook, most ties are weak. *Communications of the ACM*, 57(1):78–84, 2014.

McGraw Hill Financial. The s&p capital iq platform, 2016. URL `http://marketintelligence.spglobal.com/documents/products/SPCIQ_Platform_v2.pdf`.

Bobby Franklin and Maryam Haque. National venture capital association yearbook. 2016.

Peter A. Gloor, Pierre Dorsaz, Hauke Fuehres, and Manfred Vogel. Choosing the right friends-predicting success of startup entrepreneurs and innovators through their online social network structure. *International Journal of Organisational Design and Engineering*, 3(1):67–85, 2013.

Mark S Granovetter. The strength of weak ties. *American journal of sociology*, pages 1360–1380, 1973.

T. Hellmann. Venture capitalists: the coaches of silicon valley. *Stanford University Press*, pages 267–294, 2000.

T. Hellmann and M. Puri. Venture capital and the professionalization of start-up firms: Empirical evidence. *The Journal of Finance*, 57:169–197, 2002. doi: 10.1111/1540-6261.00419.

H. Higashide and S. Birley. The consequences of conflict between the venture capitalist and the entrepreneurial team in the united kingdom from the perspective of the venture capitalist. *Journal of Business Venturing*, 17(1):59–81, 2002.

Ha Hoang and Bostjan Antoncic. Network-based research in entrepreneurship: A critical review. *Journal of Business Venturing*, 18(2):165–187, 2003.

D&B Hoovers. D&b hoovers (avention onesource acquired by dun & bradstreet) solves your sales, marketing, and research data challenges, 2017. URL `http://www.hoovers.com/company-information/onesource-solutions/advantage.html?avention=true`.

C. Hopp. When do venture capitalists collaborate? evidence on the driving forces of venture capital syndication. *Small Business Economics*, 35(4):417–431., 2010.

Charlotte Jee. Why your startup shouldn't go to a venture capital firm for funding, 2016. URL `http://www.techworld.com/startups/why-your-startup-shouldnt-go-vc-for-funding-3643856/`.

Haje Jan Kamps. Breaking a myth: Data shows you don't actually need a co-founder, 2016. URL `https://techcrunch.com/2016/08/26/co-founders-optional/`.

S.N. Kaplan and P. Strömberg. Financial contracting theory meets the real world: An empirical analysis of venture capital contracts. *Review of Economic Studies*, 70(2):281, 2003.

D. Khanin and O. Turel. Conflicts between venture capitalists and ceos of their portfolio companies. *Journal of Small Business Strategy*, 23(1):31–54, 2013.

J. Lerner. Venture capitalists and the oversight of private firms. *Journal of Finance*, 50:301–318, 1995.

Dave McClure. 99 vc problems but a batch ain't one: Why portfolio size matters for returns, 2015. URL https://500hats.com/99-vc-problems-but-a-batch-ain-t-one-why-portfolio-size-matters.

Mark S. Mizruchi. What do interlocks do? an analysis, critique, and assessment of research on interlocking directories. *Annual Review of Sociology*, 22:271, 1996.

John Mullins. Vc funding can be bad for your start-up, 2014. URL https://hbr.org/2014/08/vc-funding-can-be-bad-for-your-start-up.

J. Nahapiet and S. Ghoshal. Social capital, intellectual capital, and the organizational advantage. *Academy of Management Review*, 23(2):242–266, 1998.

Lee Tom. Perry. The capital connection: How relationships between founders and venture capitalists affect innovation in new ventures. *Academy of Management Executive*, 2(3):205–212, 1988. doi: 10.5465/AME.1988.4277257.

Sophie Pommet. The impact of the quality of vc financing and monitoring on the survival of ipo firms. *Managerial Finance*, 43(4):440, 2017.

Sahil Raina. The gender gap in startup success disappears when women fund women, 2016. URL https://hbr.org/2016/07/the-gender-gap-in-startup-success-disappears-when-women-fund-women.

T Reve and L.W Stern. *The relationship between interorganizational form, transaction climate, and economic performance in vertical interfirm dyads*. Lexington Books, Lexington, Massachusetts, 1989.

Serguei Saavedra and et al. Stock fluctuations are correlated and amplified across networks of interlocking directorates. *EPJ Data Science*, 3(1), 2014.

Credit Suisse. The cs gender 3000: The reward for change, 2016. URL https://www.tribunafeminista.org/wordpress/wp-content/uploads/2016/09/csri_gender_3000.pdf.

Athena Vongalis-Macrow. Two ways women can network more effectively, based on research, 2012. URL https://hbr.org/2012/11/two-ways-women-can-network-more.

Clement K. Wang and et al. Effects of venture capitalists' participation in listed companies. *Journal of Banking and Finance*, 27:2015–2034, 2003. doi: 10.1016/S0378-4266(02)00317-5.

N. Wasserman. The throne vs. the kingdom: Founder control and value creation in startups. *Strategy Management Journal*, 38:255–277, 2017. doi: 10.1002/smj.2478.

Noam Wasserman. The trade-off entrepreneurs make, 2008. URL https://hbr.org/2008/02/the-founders-dilemma.

Kimberly Weisul. The new face of funding. *EBSCOhost*, 38(9):76–80, 2016.

Fred Wilson. Maximizing runway can minimize success, 2013. URL `http://avc.com/2013/09/maximizing-runway-can-minimize-success/`.

Anita Williams Woolley, Christopher F. Chabris, Alex Pentland, Nada Hashmi, and Thomas W. Malone. Evidence for a collective intelligence factor in the performance of human groups. *Science*, pages 686–688, 2010.

Robbie Wright and Ken Mike. Venture capital and private equity: A review and synthesis. *Journal of Business Finance & Accounting*, 25(5-6):521–570, 1998.

Andy Wu, Fujie Jin, and Lorin Hitt. Social is the new financial: How startups' social media activities influence funding outcomes. *Working Paper*, 2015.

Yang Xi and Wang Su-Sheng. Do venture capital firms play as supportive mentors or free-riders? evidence from china. *Journal of Commercial Biotechnology*, 22(4): 3–18, 2016.

Yang Yu and Victor Perotti. Startup tribes: Social network ties that support success in new firms. *Americas Conference on Information Systems*, 2015.

Catherine Zimmer. *Entrepreneurship through social networks. The art and science of entrepreneurship*. Ballinger, 1986.