Homophily, Information Asymmetry and Performance in the Angels Market

Buvaneshwaran Venugopal*

November 2017

Job Market Paper

Abstract

Using unique hand-collected data on startups that were seed-funded by individual angel investors, I show that social connections between angels and entrepreneurs, obtained via schools, past employment and ethnicity, positively influence investment decisions of angels, and the subsequent performance of startups. Social connections, irrespective of the ranking of the school or employer in which they were formed, are crucial for obtaining early-stage startup financing particularly in markets with higher information asymmetry. Connected seed-stage startups are more likely to survive longer, raise more series A funds and are more likely to attract venture capital investments than unconnected startups. The estimates of a two-stage selection correction model show that the higher performance of connected startups is because of post-investment influence of angel investors, via better information exchange and coordination.

Keywords: Angel investors, information asymmetry, entrepreneurial finance, social networks
JEL Classification: G24, L14, L26, M13

*C. T. Bauer College of Business, University of Houston; email: bvenugopal@bauer.uh.edu. I thank my advisor Vijay Yerramilli and the members of my dissertation committee, Praveen Kumar and Paul Povel, for their invaluable guidance. I am grateful to Crunchbase and AngelList for giving me access to their respective databases, and to Stony Brook University Data Science Lab and Yahoo! Research for providing me access to their ethnicity-nationality classification algorithm. I also thank Kris Jacobs, Raja Kali, Zack Liu, Kevin Roshak, Manpreet Singh, Scott Smart, Rauli Susmel, and seminar participants at the Annual CIRANO-Sam M. Walton College Workshop on Networks in Trade and Finance, Miami University, San Diego State University, University of Central Florida, University of Houston, and University of Otago for their helpful comments and suggestions. All remaining errors are my own.

Introduction

In this paper, I use hand-collected data on the angel investor market to address two important questions: Can social connections mitigate information asymmetry and influence the two-sided matching of investors with startups? And what is the effect of social connections on startup performance? The main contribution of this paper is to show that social connections do influence the angel-startup matching process and that they have a positive effect on the post-investment performance of early-stage startups. This is important because angel investors are the primary source of external finance for early-stage startups and are vital to the economy. Yet, the literature has paid very little attention to them. This paper helps us in understanding the characteristics of angel investors and their influence on startups.

A large sociology literature shows that shared personal characteristics and backgrounds such as age, gender, education, ethnicity, etc., create a sense of trust and kinship between individuals that shapes network formation in schools and work places (Granovetter (2005) and Currarini et al. (2009)). I focus on three types of social connections between investors and startup founders that are discernible from their professional profiles: schools, employers and ethnicity.

Information asymmetry is high in entrepreneurial financing markets. In these markets, socially connected angels and founders may speak the same language, may have taken a class together, or may have similar professional backgrounds, and hence find it easier to understand and trust each other. Thus, social connections can enable a better flow of information and lead to higher investment activity. I refer to this as the *homophily hypothesis*.

The influence of social connections could extend post investment. Socially connected angels and founders may find it easier to communicate, make one another receptive to suggestions, set common expectations, and thus reduce coordination costs and improve performance (Steiner (1972), Bhagwat (2013), and Hegde and Tumlinson (2014)). I refer to this as the *coordination hypothesis*. On the other hand, social connections can also lead to inefficient monitoring and poor performance via groupthink and social conformity (Janis (1982), Ishii and Xuan (2014), and Gompers et al. (2016)). The empirical evidence from other financial contexts such as boards, venture capitalists and M&A markets on the effect of social connections on performance is mixed.

The story of Yelp Inc. is a good example of the influence social connections have on economic transactions. Yelp started as a simple email circle in 2004 and took off in 2005 after a failed first attempt. The company founded by Jeremy Stoppelman and Russel Simmons, was first funded by Paypal's co-founder, Max Levchin. All three individuals are alumni of the University of Illinois at Urbana-Champaign and worked together at Paypal. In a 2012 interview[1] after Yelp went public, Levchin said that he backed Yelp because he knew that the founders were capable. Stoppelman acknowledged the early support of Levchin, who served as the chairman of the board from its inception to 2015. The founding team's ability is one of the primary factors that investors try to discern when making their investment decisions. Similar to the case of Yelp, we can expect that social connections would be useful in reducing information gaps and improving performance in the early-stage startup financing market.

The early-stage startup financing market is dominated by *angel investors*, who are the primary source of funding for more than 95% of early-stage startups (OECD (2011)). Angel investors[2] are high net worth individuals who invest their personal funds in startups (see Section 1.1 for more details). The angel investment market is an ideal setting to study the effect of social connections on investor-startup matching and the subsequent performance of startups for several reasons. First, it allows me to focus on *individual investors*, who are the decision-makers: I can clearly identify the effect of an individual's social connections on his decisions. In the case of VCs, in contrast, a partner may be in a network with a startup, but he may not be the decision-maker for funding the deal. Hence, it is difficult to identify the effect of an individual's social connections on decisions in the VC market. Second, a VC typically invests in the later stages of a startup's life cycle, which makes it difficult to clearly disentangle the VC's influence on the startup from that of earlier investors. Third, investment selection is endogenous. An investor may use his experience and connections to choose startups with greater potential, which will lead to biased estimates for the investor's influence on performance. In contrast to previous studies, which ignore this issue, I construct *individual level instruments* to correct for the endogeneity, and I appropriately identify

[1] See http://www.sfgate.com/news/article/Yelp-s-Jeremy-Stoppelman-a-profile-3707980.php

[2] Angel investors are accredited investors. Rule 501 of Regulation D defines an accredited investor as: a person with individual net worth or joint net worth (with a spouse) in excess of $1 million, excluding a primary residence; or, a person with annual income in excess of $200,000 or joint income (with a spouse) in excess of $300,000.

the post-investment influence of social connections on startup performance. Fourth, unlike VCs, who invest on behalf of limited partners, there are no agency concerns regarding angel investors since they invest their own funds.

The second main contribution of this paper is the construction of a new dataset on angel investors. Entrepreneurial financing has drawn a lot of attention: First, policy makers have introduced many programs and policies (e.g., the Small Business Investment Act of 1958 and the JOBS Act of 2012) to spur entrepreneurship. Second, the entrepreneurial finance literature has grown large in the last couple of decades and has mainly focused on VC funds. Angels, however, are essential for the entrepreneurship process because they are typically the first to invest in a startup. The angel investor market is also an ideal candidate to test the theories in the financial contracting literature since they typically use setups with a single entrepreneur and investor. In spite of all this, we know very little about angels because structured data was not available.

I collect data from Crunchbase (www.crunchbase.com), which is the largest crowd-sourced database on entrepreneurial activities, and AngelList (www.angel.co), which is the most active fund-raising platform for startups. I use these databases to gather information on angel investors (e.g., investment history, education, employment, etc.) and the startups (e.g., biography of the founders, fund-raising history, etc.) that they seed-funded. The data is further augmented using SEC form D filings, S&P Capital IQ, LinkedIn and Google Trends. A detailed description of the data is provided in section 2. The data includes 9,396 startups with 15,951 founders and seed-funded by 5,417 angels during the period 2005 to 2015.

The effect of social connections on angel-startup matching is consistent with the predictions of the homophily hypothesis. I find that an investment is 23.4% more likely to occur when the angel and startup share a social connection. The probability of matching is strongest when the angel and founder have worked for the same employer during an overlapping time period. Contrary to concerns that this result would be driven by angels and founders who attended elite schools or worked for top employers, I find that connections made at both top and lower ranked schools and employers are equally important. Further, social connections are more important for matching in new product markets, where information asymmetry is higher than established product markets.

I also find that startups with connected angels and founders ("connected startups") are more likely to survive and raise series A funds compared to startups where angels and founders do not share a social connection ("unconnected startups"). However, the better performance of connected startups could be due to the *post-investment influence* of angels (treatment), or due to *pre-investment selection* on information obtained through social connections (selection). As econometricians we do not observe all the factors that influence selection and it is extremely difficult to identify post-investment influence using a fixed effects regression framework. Therefore, I use a *Heckman (1979) selection correction model* to isolate the treatment effect. I use *two instruments* that predict sorting and matching of angels with startups in the selection equation. The instruments are variables that indicate the presence of Crunchbase profile pages for the angel and the startup before the startup's seed-funding date. The presence of a profile page on Crunchbase before a startup's fund-raising date would make it easier for angels and founders to approach each other to form a partnership, especially if they share a social connection. At the same time, these instruments satisfy the exclusion restriction, because the presence or absence of a Crunchbase profile should have not affect the startup's performance (see section 5.2 for more details).

After correcting for selection, I find that connected seed-stage startups are 13.6% more likely to successfully move to the series A stage, and that they raise $0.23 million more in series A funding than their unconnected peers. However, connected startups take longer to reach the series A stage, suggesting that their investors tolerate early delays or failures, consistent with Manso (2011). Further, connected startups are 14.6% more likely to attract a VC investment in the series A stage. The vast resources and certification effects that VCs bring to early-stage startups are invaluable for their continued survival.

This paper contributes to a growing body of research in finance that investigates the effect of social connections on financial transactions.[3] Due to institutional settings, most papers in this literature study the aggregate firm level effect of social connections on financial transactions. In contrast, this paper examines the effect of social linkages between individuals, that is, angels and

[3]For example, the effect social connections has been studied in the context of boards (Chidambaran et al. (2011), Cohen et al. (2012), Engelberg et al. (2013), and Ishii and Xuan (2014)), mutual funds performance (Cohen et al. (2008)), securities analysts recommendations (Cohen et al. (2010)), loan markets (Engelberg et al. (2012) and Lin et al. (2013)) and VC markets (Hegde and Tumlinson (2014) and Gompers et al. (2016)).

founders, on their economic decisions. The extant literature finds mixed evidence regarding the effect of social connections on performance. This is partly because they fail to take into account the assortative matching and endogeneity in partnership selection. In contrast, I explicitly correct for these issues in this paper using a selection correction model with instruments at the individual level, and I find that social connections do improve startup performance.

The entrepreneurial finance literature has focused on venture capitalists (see Da Rin et al. (2013) for a survey) and has ignored the role of angels investors in the economy. This paper contributes to a small but growing part of the literature on angels (Kerr et al. (2014), Bonini et al. (2016), Boulton et al. (2017) Lerner et al. (2017), and Venugopal and Yerramilli (2017)) in the following ways: This paper sheds light on the understudied angels market by bringing in new data, and describing the characteristics and performance of the firms funded by individual angels. Unlike existing papers that study the consequence of an investment using a few Silicon Valley angels or groups, this paper examines the factors that drive the two-sided angel-startup matching using a large sample of individual angels. This paper also contributes to the entrepreneurial finance literature by identifying three social channels that can be used to convey both hard and soft information.

1 Institutional and Theoretical Background

1.1 The Angel Investment Market

Angel investors are wealthy individuals, who are often former entrepreneurs. They have been supporting innovation and startup activity since the second industrial revolution (Lerner (1998) and Lamoreaux et al. (2004)). Unlike VCs, who invest in later-stage startups, angel investors invest in early-stage startups. A startup's life cycle is generally divided into the following stages: pre-seed, seed, series A, series B, series C, etc., and finally exit via failure, acquisition or IPO. Stages till series A are referred to as early stages (Please see appendix for generally accepted definitions of the stage classifications in the industry).

According to the Center for Venture Research at the University of New Hampshire, angels seed-funded more than 99% of the (32,120) early-stage startups in 2013. In dollar terms, the angels market ($24.8 billion) is similar in size to the VC market ($29.6 billion) (Sohl (2015)). The Small Business Administration office estimates that there are more than 200,000 angel investors, who are increasingly organizing themselves into angel groups to finance larger deals (Shane (2012)).

The role of angel investors in the early-stage market is crucial to the economy. Because, early-stage markets allow for experimentation and quick failure, without which the innovation process would stagnate (Kerr et al. (2014)). In addition, Hellmann and Thiele (2015) predicts that a sizable portion of VC market's deal-flow comes from angel investors market. An important distinction between VCs and angels is that the former raise funds from limited partners to invest in startups, while the latter invest their personal funds in startups. This skin-in-the-game setup could alleviate some of the agency conflicts associated with VC funding, and may also motivate angel investors to add value to their investments (Chemmanur and Chen (2014)). Kerr et al. (2014) and Lerner et al. (2017) show that angel investors have a real impact on the firms in which the invest, in terms of exit rates and patenting.

1.2 Theoretical Background and Hypothesis Development

Spence (2002) notes that information gap between counter-parties drives adverse selection, which is an important feature of credit markets. In order to address this issue, banks often collect soft information about the quality of borrowers in addition to hard information (e.g., credit scores) (Petersen and Rajan (2002)). In entrepreneurial financing markets, information gaps are wider; because hard information, such as revenue, product demand, etc., is often unavailable as startups are very young. Even the numbers showcased by founders in pitch meetings are highly subjective and soft in nature.[4] Therefore, the ability of founders and investors to convey information about each other would be invaluable in this market.

Granovetter (2005) argues that social connections can serve as an information channel and influence economic decisions. Below, I hypothesize that social connections could be used to narrow the information gap between investors and founders, which should lead to partnership formation.

[4]Please see `https://www.inc.com/howard-greenstein/how-to-show-market-traction.html` for a guide on expressing market demand (traction).

Further, they should improve startup performance by reducing coordination costs.

Homophily Hypothesis

Individuals often tend to associate and interact with others who have similar characteristics and backgrounds, which is termed as *homophily*. Homophily is the organizing basis of networks and has been documented as the driving force behind group formation in a variety of settings including schools (Currarini et al. (2009)), marriages (Becker (1973)), etc. The sociology literature points to a broad range of characteristics including age, gender, education, social status, etc., as homophily factors that influence partnership formation (see McPherson et al. (2001) for a survey). The homophily behavior is attributed to a sense of trust and familiarity between individuals from a similar background, which has been shown to be important in financial markets (Guiso et al. (2009); Bottazzi et al. (2016)). This is all the more important in opaque and high risk markets such as the early-stage entrepreneurial financing markets. Because, the flow of information, especially soft information, is more likely to happen when a founder-investor pair share homophily characters.

I test three social channels —schools, previous employers and ethnic origin— that may influence information exchange, angels' investment decision and founders' decision to accept funding from certain angels. If social connections do reduce information asymmetry, then an angel who went to same school, worked for the same employer, or belongs to the same ethnic minority as one of the startup's founders is more likely to invest in the startup; and the startup is more likely to accept his investment. I refer to this as the *homophily hypothesis*. Following the same logic, the investment likelihood should increase with the strength of connection, that is, number of channels through which the angel-startup pair is connected. Furthermore, this behavior should be stronger in markets with higher information asymmetry, because social connections could be an effective method to gather metrics on the each other's counter-parties in these markets.

Co-ordination Hypothesis

Social connections that affect the two-sided matching of investors with startups can also impact the performance of startups. On the one hand, social connections can facilitate easier communication between founders and investors, and thereby help them in setting common expectations and reducing coordination costs (Steiner (1972), Cohen et al. (2008), Bhagwat (2013), and Hegde and Tumlinson (2014)). Investors add value to their startups after investment by identifying professional talent, customers, service providers, potential partners, and formulating strategies (Gorman and Sahlman (1989) and Hellmann and Puri (2002)). Social linkages would make it easier for angels to influence founders and make them more receptive to suggestions. Therefore, social connections can lead to higher performance through better communication and post-investment influence. I refer to this as the *coordination hypothesis*. On the other hand, such linkages can lead to inefficient monitoring, reduced innovation and poor performance because of a lack of diverse viewpoints, a tendency for social conformity or groupthink mentality (Janis (1982), Surowiecki (2005), Ishii and Xuan (2014), and Gompers et al. (2016)). Both positive and negative effects of social connections have profound consequences for the survival and performance of early-stage startups. Which effect prevails is an empirical question that is explored in this paper.

2 Data and Key Variable Definitions

2.1 Data Sources

In order to test the hypotheses, we need a sample of startups funded by angel investors, with complete funding histories, and biographies of founders and investors. Even though angel investors are the major source of outside equity for early-stage startups, angel investors are rarely covered by commercial databases. I overcome this challenge by collecting and combining data from a variety of sources. The primary sources are Crunchbase (`www.crunchbase.com`), which is the largest crowd-sourced database on entrepreneurial activity, and AngelList (`angel.co`), which is the most active fund-raising platform for startups.[5] This data is supplemented with data from SEC's notice of exempt offering of securities (Form D), LinkedIn, S&P Capital IQ, Google Trends, and various news websites.

[5] I access the data on Crunchbase and AngelList via their Application Programming Interface (API), which allows us to send requests for data on each investor and start-up using a unique identifier. The output of requests is a JSON (JavaScript Object Notation) file that contains tags for data items such as name, location, role, jobs, etc., that are parsed using a Perl script to form data tables.

Crunchbase and AngelList

Crunchbase, owned by AOL, is a crowd-sourced database that tracks entrepreneurial activity around the globe. Though Crunchbase was started in July 2007, it contains detailed profiles on investors and startups founded as far back as the 1980s. Crunchbase independently verifies the data and has partnerships with various venture capital firms, angels groups, accelerators, etc. to ensure the accuracy of its data. Crunchbase contains more details on early-stage deals compared to similar databases such as PitchBook, Dow Jones Venture Sources, etc.[6]

The data in Crunchbase is organized around collection endpoints as a graph database. I use the "People" endpoint to extract data on founders and angels. Apart from personal details, such as gender, location and education, I collect employment and investment histories, and links to news articles and LinkedIn profiles of founders and investors. Figure 1 provides a snapshot of the data available for Alexis Ohanian, the co-founder of Reddit and the most active angel in 2014.

I use the "Organization" endpoint to extract profiles of start-ups. Although there are some missing variables, for start-ups with complete profile pages, I am able to extract data on the company's founding date, website domain address, location, fund-raising dates, amount of funds raised, stage of each funding round, identity of investors who participated in each financing round, founding team and board members. Figure 2 provides a snapshot of the data available for Uber.

I augment the data using AngelList. Similar to Crunchbase, AngelList also contains data on fund-raising histories of startups, and detailed biography of founders and investors. Even though AngelList's data largely overlaps with that of Crunchbase, the former is better at capturing the dates and amount of funds raised by startups in each round.

2.2 Base Sample Construction

To construct the sample for analysis, I match and deduplicate the data from Crunchbase and AngelList. Since there are no standardized identifiers for startups and investors, I use the following multi-step fuzzy matching (vectorial decomposition) procedure to match the data from the two

[6] Please see https://techcrunch.com/2013/07/23/how-crunchbase-data-compares-to-other-industry-sources/ for a comparison on data coverage by various data providers.

databases[7]: First, lowercase startup names from both sources are used to create a similarity score with a range 0 to 1. Second, the same procedure is used on website domain names to create a similarity score. Third, when available, web addresses of social media platforms such as LinkedIn, Twitter and Facebook are used to create three more similarity scores. Fourth, when at least one of the above five similarity scores is more than 0.8, the matched startups are verified and assigned the same unique identifier in both databases. The remaining startups are considered as unique. Finally, with Crunchbase as the base, I fill in missing information and add new startups (those not covered by Crunchbase) using AngelList data. A similar approach is used to match, deduplicate and identify unique investors from both sources.

I place the following conditions on the bigger matched sample to select startups and investors: First, complete fund-raising data should be available for the startups. Second, the lead investor should be an individual angel investor and should have invested in at least three startups by the end of December 2015. These conditions result in 11,637 startups seed-funded by 3,408 lead angels.[8]

2.3 Additional Data Sources

Funds Raised: The consolidated data from Crunchbase and AngelList contains fund-raising information for only 9,438 of 11,637 startups. Therefore, I turn to SEC's Form D filings to fill-in missing values.[9] Using CIK numbers in the Edgar Company Index file and funding round date, I download Form D filings from SEC's FTP servers. The "Total Amount Sold" field is used to identify funds raised by 631 startups that were previously missing the data. Next, I use Owler, CB Insights and Mattermark to search for fund raising data and obtain information for 509 startups.[10]

[7]The algorithm works as follows: Suppose we have two strings – "Mathew" and "Matthew" – to match. The algorithm breaks the strings into rolling 3 characters (A={"mat", "ath", "the", "hew"}; B={"mat", "att", "tth", "the", "hew"}) and calculates the similarity score $s = \frac{|A \cap B|}{|A \cup B|}$; $0 \leq s \leq 1$.

[8]In cases where there are more than one angel investor in the first seed round, I assign the angel with highest degree centrality as the lead angel investor of the seed round.

[9]As per Regulation D of the Securities Act of 1933, some companies are allowed to offer their securities for sale without having to register with the SEC. This is intended to make access to capital markets possible for small companies that could not bear the costs of a normal SEC registration. Such companies are required to file a Form D with the SEC after making the first sale, which, among other things, contains information on the type of security sold, date of first sale and the amount sold. This data, starting from 2008, is available on SEC's FTP servers.

[10]Mattermark, Owler and CB Insights are databases that are similar to Crunchbase, and sell their data to primarily VCs as a deal prospecting tool. I used their trial subscriptions to fill in missing data when available.

Ethnicity Identification: One of the social channels examined in this paper is coethnicity. I use a name matching algorithm developed by researchers at the Stony Brook University's Data Science Lab and Yahoo! Research (see Ye et al. (2017)) to identify the ethnicity of founders and investors. The program was trained on a set of 74 million (first and last) names from 118 countries to assign probability scores for 39 ethnicities/nationalities. The ethnicity with the highest score is taken as the ethnicity of the person bearing that first and last name. Ye et al. (2017) show that this 39-leaf algorithm classifies names with more precision than any other available method.

I use the high level classification of ethnicities —African, Celtic English, East Asian, East European, Hispanic Jewish, Middle Eastern, Nordic, South Asian, South European and West European— in this study. Following Kerr and Lincoln (2010) and Gompers et al. (2016), African, East Asian, Hispanic, Jewish, Middle Eastern and South Asian are considered as ethnic minorities.

Profiles of Angels and Founders: Biographical information is missing for some angels and founders in both Crunchbase and AngelList. In such cases, I use their LinkedIn and S&P Capital IQ profiles to collect data on their education, employment history and entrepreneurial activities.

2.4 Sample Selection and Variable Definitions

In order to be included in the analysis sample, I require each startup to have data on seed and series A funds raised, location (state), and the identities of founders and investors. I only include startups that were founded in 2005 or later since Google Trends data, which I use to construct the traction measure (see section Appendix A), is available only from 2004. These restrictions result in a sample of 9,396 startups, founded by 15,951 entrepreneurs and seed funded by 5,417 individual angels. The seed rounds were led by 2,655 angels during the period 2005 to 2015. I focus on prior social connections between the seed-stage lead angel and the founders of a startup because the lead angel will be responsible for performing due diligence on the startups.

Social Connection Variables

For each lead angel-startup pair, I create the following binary variables to indicate the presence of a social connection: (i) *Same School* indicates whether the seed-stage lead angel and at least one of the startup's founders attended the same school during an overlapping time period. For example, suppose angel 'A' attended Stanford between 2006 to 2010 and founder 'F' was also in Stanford between 2008 to 2012; then *Same School* for the corresponding angel-startup pair takes a value '1'. (ii) *Same Employer* takes a value one if both the angel investor and founder worked for the same employer during an overlapping time period. The overlapping time period requirement increases the likelihood of the angel and founder actually meeting while with the same school or employer. (iii) *Same Ethnic Minority* takes a value one if the angel and at least one of the founders belong to the same ethnic minority. (iv) *Connected Angel-Startup* takes value one if any of the above three indicators is turned on. In addition, the strength of a lead angel-startup pair's social connection is given by *Connected Depth*, a factor variable that takes values from 0 to 3; 0 indicates the absence of a social connection and 3 indicates that the dyad is connected through all three channels.

For school and employer connections, I create two additional sets of dummy variables based on CrunchBase's ranking of the educational institution and employer.[11] *Same Top School* is a binary variable that indicates whether an angel and at least one of the startup's founders attended the same school that features in the list of top-100 educational institutions on Crunchbase rankings. *Same Bottom School* takes a value one if the angel and founder have attended the same school that is not in the top-100 list. I define *Same Top Employer* and *Same Bottom Employer* along similar lines using Crunchbase's employer rankings.

3 Descriptive Statistics and Univariate Results

3.1 Descriptive Statistics

The analysis sample contains 9,396 startups seed-funded by 2,655 lead angels. Table 1 summarizes the key variables used in this study. The average Traction is 2.97, which suggests that early-stage

[11] Crunchbase rankings are determined by an algorithm that takes into account the number of connections, funding, news articles, M&A activity etc. Please see http://about.crunchbase.com/2016/11/use-crunchbase-rank-trend-score-find-influential-companies-identify-trends/?utm_source=pro_help for a detailed description of the ranking methodology used by Crunchbase. I use Crunhbase rankings rather than Times Higher Education Rankings or Forbes rankings because Crunchbase rankings better reflect the entrepreneurial financing landscape.

companies receive less attention/demand compared to other startups in the same product market. Only 20% of the startups successfully move from seed to series A stage. This shows the high level of failure in the early-stage market as documented by Venugopal and Yerramilli (2017).

Consistent with anecdotal claims, I find that 51% of the lead angels in the sample are past entrepreneurs. In order to measure the skill and quality of investors, I look at the angel's investment history and success. I call an angel's seed investment as successful if he has led the startup from seed to series A stage. An angel's *Seed Success Ratio* in year 't' is the number of successful seed investments as of year 't-1' divided by the total number of seed investments made by the angel as of year 't-1'. The average ratio is 14% suggesting that success is fairly rare in this market.

Table 2 summarizes social connections between lead angels and startups. The top-100 schools and top-100 employers are responsible for churning out 26.9% and 30.7% of the founders, respectively. Similarly, 32.6% and 32.5% of the lead angels are from top-100 schools and top-100 employers, respectively. This is consistent with the level of concentration in educational institutions among founders and investors in VC markets (Gompers et al. (2016) and PitchBook (2017)). Panel D summarizes the key social connection variables. 13% of the startups have a seed-stage lead angel who attended the same school, during an overlapping time period, as the startup's founder. Similarly, 21% of the startups have a lead investor who worked for the same employer as a founder, and 30% of the startups have a ethnic minority connection with their lead angels. Overall, 46% of the startups have at least one social connection with their seed-stage lead angel.

3.2 Univariate Results

In table 3, I present a univariate comparison of connected (A) and unconnected startups (B). The last column reports the *t*-statistic for the difference in the means between the two groups. The comparison of pre-seed variables show that connected startups are younger, have bigger founding teams, and have lower traction compared to unconnected one. However, at the univariate level, there is no difference in the quality of schools and employers of founders between the two groups. Overall, both groups have similar quality founders and seed-funds at the start of their life cycles.

The lead angels of connected startups are more likely to be former entrepreneurs, but have shorter investing experience (2.0 vs. 2.6 years) compared to lead angels of unconnected startups. However, the two lead angel groups are not statistically different on other dimensions such as network centrality, rounds participated, past success, and school and employer quality.

Despite lower traction and similar initial funding, connected startups experience better seed-stage outcomes than the unconnected startups. For example, 23% (17%) of the connected (unconnected) startups successfully move to series A stage, and they raise $0.43 million more in series A funding than unconnected startups.

4 Matching between Angels and Startups

4.1 Empirical Methodology

Startup financing is a two-sided matching process in which both the founder and the investor have to agree to each other's terms (Sørensen (2007)). Moreover, the information asymmetry in these markets make both parties wary of each other (Gompers and Lerner (2001)). In order to examine the factors that affect angel-startup matching, we need a sample of angel-startup pairs containing both actual pairs, for which the angel invested in the startup, and counterfactual pairs, for which the angel could have invested in the startup but did not. There are 9,396 actual lead angel-startup pairs in the sample. The counterfactual pairs are created using the following procedure. First, the startup in each actual lead angel-startup pair is matched with "control" angels who satisfy the following conditions: (i) the angel should be located, or have made investments, or be interested (as disclosed in their profile pages) in the same state as the startup, and (ii) the angel should have made at least one investment in the past 3 years. Second, the angel in each actual lead angel-startup pair is matched with "control" startups that are located in the same state as the angel or located in one of states in the angel's preferred locations list. The constructed sample contains 2,395,651 angel-startup pairs, which is used to estimate the effect of social connections by comparing the actual angel-startup pairs with counter-factual pairs.

I estimate the following linear probability model (LPM) to examine the effect of angel, startup and angel-startup pair characteristics on the likelihood of a match.

$$y_{i,j} = \alpha_0 + \alpha_A A_i + \alpha_S S_j + \alpha_{AS} AS_{ij} + \mu_t + \mu_{ind} + \mu_{loc} + \epsilon_{ij} \tag{1}$$

In the above equation, subscript 'i' denotes the angel, 'j' denotes the startup, 't' denotes the startup's seed-funding year, 'ind' denotes the startup's product market category, and 'loc' denotes the state in which startup 'j' is located. The dependent variable $y_{i,j}$ is $Investment_{i,j}$, a binary variable that takes value one if angel i has invested in startup j (actual pair), and zero otherwise (counterfactual pair). The main independent variables of interest are AS_{ij}, which represent social connections between an angel and startup in each pair. A_i represents angel specific characteristics such as degree centrality, past entrepreneurship, Seed Success Ratio, and the quality of schools attended or past employers. S_j includes startup characteristics such as age, presence of serial entrepreneur, traction, and the quality of schools and past employers of founders. The errors are heteroskedasticity robust and clustered at the product market level.

Since investment decisions can be influenced by the year in which a startup is raising funds (boom or bust years), the startup location, and product markets (hot or cold technologies), I include fixed effects for each of these three variables. The identification of our coefficient of interest, α_{AS}, comes from the variation in angel-startup social connections (AS_{ij}) within a location and market after controlling for angel and startup characteristics. I estimate linear probability models instead of probit model to avoid the incidental parameter problem.

4.2 Effect of Social Connections on Angel-Startup Matching

I report the results of regressions examining the effect of social connections on the two-sided angel-startup matching in table 4. Column (1) shows that the likelihood of an angel investing in a startup increases by 6.1% when the angel and founder have attend the same school during an overlapping time period. This is an economically significant impact considering that the unconditional probability of matching in the sample is 0.41%. A concern with this result is that the matching is driven purely by investors and founders from top schools who may have higher quality projects and significant alumni networks in the startup financing markets. To test this, as discussed in section 2.4, I split the *Same School* variable into two binary variables: *Same Top School* and *Same Bottom School* based on Crunchbase's ranking of educational institutions. Column (2) shows that connections made at both top and bottom schools are equally important for matching.

According to column (3), the probability of matching increases by 28.3% when both the angel and founder have worked for the same employer in the past. Similar to column (2), in column (4) I test whether the employer connection result is driven by past employees of influential companies, such as Google and Apple, who have a strong presence in the entrepreneurial financing markets. The coefficients of both *Same Top Employer* and *Same Bottom Employer* are positive and significant suggesting that irrespective of the employer's quality, past employment connections play a significant role in the matching of investors with startups.

The likelihood of matching increases by 0.7% when angels and founders belong to the same ethnic minority. Column (6) shows the coethnicity effect for each minority group. For example, when both the angel and founder are African, the likelihood is higher by 4.2%; and so on.

I include all three social variables in column (7) to test their relative impact on matching. The coefficient estimates show that *Same Employer* is the most significant channel that affects matching, followed by *Same School* and *Same Ethnic Minority*. The estimate of Same School now is less than half of that in column (1); but the effect of Same Employer is stable across columns (3) and (7). This could be because some angel-founder pairs who attended the same school also ended up working for the same employer at some point in their professional careers. Column (8) suggests that this phenomenon is concentrated at the top-100 schools. The employer channel results suggest that entrepreneurs with employer connections may find it easier to raise seed-funding. The effects of angel and startup controls on the likelihood of matching are reported in table IA.1.

Strength of Social Connections and Angel-Startup Matching: According to homophily hypothesis, social connections have a positive impact on matching. Then, the likelihood should increase with the strength of social connections. I report regressions that test this hypothesis in table 5. Column (1) shows that the likelihood of matching increases with the number of channels through which an angel-startup pair is connected. For example, the likelihood of matching increases by 30.6% when an angel-startup pair is connected through school and employer channels. The likelihood increases by 47.2% when a pair is connected through all three channels.

Connection Depth, a factor variable with four possible discrete values from 0 (no social connection) to 3 (connected through all 3 channels), is an alternate measure of connection strength that does not distinguish between channels. The increasing coefficients in column (2) show that the probability of matching increases with increase in the strength of social connections.

In column (3), I divide education and employer connections into top and bottom categories based on the school and company rankings provided by Crunchbase. The interaction terms in column (4) test the effect of school connections that were carried over to employers at different quality levels. For example, the interaction "Same Top School × Top Employer" shows that the likelihood of matching increases when an angel and founder have attended the same top school and then worked for the same top employer.

Effect of Social Connections on Matching in New Product Markets: Any information channel should become more important when the level of information asymmetry is high. Even though early-stage financing markets are opaque to begin with, financing newly emerging product markets is more challenging, because the angel has to evaluate the prospects of a non-existent product market in addition to evaluating the startup and the founding team. Therefore, it is reasonable to expect that the effect of social connections would be stronger in new product markets.

In order to test this hypothesis I create the dummy variable *New Market* to identify startups that were part of the first 25% of startups that were formed in a product market.[12] In the matching regressions, the coefficient of this variable would capture the likelihood of a seed-stage startup in a new product market getting funded. I report regressions that test the effect of social connections on matching in new product markets in table 6. According to column (1), the presence of a prior social connection between an angel and founder increases the likelihood of matching by 23.4%. The coefficient of New Market in column (2) shows that startups that were formed during the developmental stages of a market are less likely to find a seed-stage investor. However, the interaction between Connected Angel-Startup and New Market is positive and significant,

[12] The number of firms created according to Crunchbase by the end of December 2015 is used as the vantage point to calculate the percentage of startups created in each product market category.

indicating a 8.7% increase in the likelihood of investment when an angel and founder share a social connection. The interactions of New Market in column (3) show that the effect of social connections, especially employer connections, are stronger in new product markets.

Overall, the results in section 4 show that social connections play a vital role in matching angels with startups, and these effects are stronger in new markets where information asymmetry is higher than established markets.

5 Social Connections and Seed-stage Outcomes

About 76% of seed-stage startups fail to reach series A stage (Venugopal and Yerramilli (2017)). Therefore, efficient communication and guidance from investors is crucial for the survival of seed-stage startups. According to the coordination hypothesis, social connections improve communication and coordination between angels and founders, and hence should improve performance. Only 19.83% of 9,396 startups in our sample successfully raised series A funds.[13] I create *Seed-stage Success*, a dummy variable to identify startups in the sample that have successfully moved from seed to series A stage.

5.1 Effect of Social Connections on Seed-stage Success

To test the effect of social connections on a startup's probability to successfully move from seed to series A stage, I use a variation of the equation 1 with *Seed-stage Success* as the dependent variable and additional fixed effects for the startup's lead angel. The regression results are presented in table 7. Column (1) shows that when an angel-startup pair is connected via the school channel, the probability of the startup moving from seed to series A stage increases by 9.1%. Since we have lead angel fixed effects, the *Same School* coefficient captures better performance of connected startups within a angel's portfolio and it is not a mere difference across angels. The effect is positive and significant even after dividing the *Same School* variable into *Same Top School* and *Same Bottom*

[13] All startups in the sample were seed-funded by the end of December 2015. I verified if they have moved from seed to series A stage in April 2017, which allows, at minimum, sixteen months to make the transition. According to CB Insights, a industry analytics leader, investors view startups that did not raise additional funding within 1.5 years as failures. Following this logic, I code startups that did not raise additional funding by April 2017 as failed.

School. This suggests that school connections aid in coordination irrespective of the quality and ranking of the school in which the connection was formed.

The coefficient in column (3) indicates that the probability of a startup's seed success increases by 10.2% when the angel and founder have worked for the same employer in the past. Further, connections formed at top-100 employers have a higher impact on seed success than connections formed at lower ranked employers. Columns (5) shows that coethnicity is also associated with a higher likelihood of a startup moving from seed to series A stage. However, the coefficients of constituent ethnic minorities in column (6) are insignificant because of the relatively small number of observations in each group. The employer channel has the strongest impact on Seed-stage Success, followed by school and ethnic ties. All regressions control for observable angel and startup characteristics. I tabulate their impact on Seed-stage Success in table IA.2.

Strength of Social Connections and Seed-stage Success: If social connections foster post-investment coordination, then an increase in the strength of social connections should increase the probability of a startup's seed-stage success. I present results of regressions that test this hypothesis in table 8. Column (1) reports the impact of all three social channels and their interactions on seed-stage success. An angel-startup pair with both Same School and Same Employer turned on is associated with a 22.8% increase in the likelihood of the startup successfully raising series A funds compared to an unconnected startup. The likelihood of seed-stage success increases by 30.72% when the angel-startup pair is connected via all three social channels. This is a substantial increase in the likelihood of seed-stage success given an unconditional probability of 19.83%.

The coefficients of *Connection Depth* in column (2) shows that the success likelihood increases with the number of channels through which a lead angel-startup pair is connected. In columns (3) to (5), I divide the school and employer connections into top and bottom categories based on Crunchbase rankings. The interaction coefficients imply that an increase in coordination between angels and founders results in higher chances of seed success. The connections formed at top ranked schools and employers contribute (0.09%) more towards a startup's seed-stage success than those formed at lower ranked schools and employers.

Effect of Social Connections on Seed-stage Success in New Product Markets: The guidance and coordination obtained from social connections should be more fruitful in newly emerging product markets than established ones. In table 9, I report regressions that test whether social connections have a higher impact in new markets. Column (1) shows that connected startups are more likely to successfully reach series A stage. The coefficient of New Market in column (2) is positive and insignificant. But, the interaction coefficient is positive and significant, implying that connected early entrants of a product market are more likely to succeed compared to unconnected startups. Column (3) shows that only education and employment channels have a statistically positive effect on Seed-stage Success in new product markets.

Overall, angel-startup social connections are associated with better startup performance. However, these results could also be explained by angel investors' pre-investment selection.

5.2 Correcting Pre-investment Selection Bias

The better performance of connected startups could be due to pre-investment selection of higher quality companies by connected angel investors, who can leverage their social capital to gain an informational advantage, rather than post-investment influence. Moreover, the assortative matching behavior in VC markets (i.e., better quality VCs invest in better quality startups) documented by Sørensen (2007) could be occurring in angels market as well. Therefore, we need to isolate post-investment influence of social connections from unobservable factors, such as knowledge and passion of founders, that may affect both angel-startup matching and the consequent performance.

To disentangle the effects of pre-investment screening from post-investment influence of investors, we need a sample where angels and startups are randomly matched. But section 4 showed that the angel-startup matching is non-random. Therefore, I use a two-stage selection correction model proposed by Heckman (1979) to identify the effect of social connections on seed-stage success. The selection equation coefficient estimates are used to calculate the inverse mills ratio (IMR), a proxy for unobservable factors that affect sorting and matching angels and startups. The IMR is then used in the second-stage to correct for selection bias on the social connection coefficient, and thus identify the post-investment influence of angel investors on startup performance.

To estimate the Heckman (1979) model, we need an instrument that predicts matching of angels with startups, but one that does not directly affect post-investment performance of startups. I use two such instruments: (i) *Angel Profile on CB*, indicates whether the lead angel had a profile on Crunchbase before the startup's first seed round; (ii) *Startup Profile on CB*, indicates whether the startup had a profile on Crunchbase before its first seed round. When an angel has a profile on Crunchbase, it increases the likelihood of entrepreneurs contacting him with a potential investment opportunity. Similarly, a startup's profile on Crunchbase makes it easier for investors to learn about the startup and contact its founder.

The presence of a profile page for the angel or the startup should have no effect of the future startup performance. The profiles could either be created by the concerned angels (or founders) or by Crunchbase's data collection algorithm that monitors various sources to track entrepreneurial activities. Therefore, there is no systematic reason why certain angels or startups would have profile pages before the first seed round date. Moreover, as discussed below, the second stage regression includes startup and angel characteristics to control for time varying factors, and also includes lead angel fixed effects to control for time invariant angel effects. Thus, both instruments would satisfy the exclusion restrictions.

I estimate the following two-stage specification to identify the post-investment influence of angel-startup social connections on seed success and other seed-stage outcomes.

$$\begin{aligned} 1^{st} stage: Investment_{i,j} =& \alpha_0 + \alpha_1 Connected\ Angel\ Startup_{i,j} + \alpha_2 Angel\ Profile\ On\ CB_i \\ &+ \alpha_3 Startup\ Profile\ On\ CB_j + \alpha_A A_i + \alpha_S S_j + \mu_t + \mu_{ind} + \mu_{loc} + \epsilon_{ij} \\ 2^{nd} stage: Outcomes_j =& \beta_0 + \beta_1 Connected\ Angel\ Startup_{i,j} + \beta_2 IMR_{ij} \\ &+ \beta_A A_i + \beta_S S_j + \eta_i + \eta_t + \eta_{ind} + \eta_{loc} + u_j \end{aligned} \tag{2}$$

In the above equation, subscript 'i' denotes the angel, 'j' denotes the startup, 't' denotes the startup's seed-funding year, 'ind' denotes the startup's product market category, and 'loc' denotes the state in which startup 'j' is located. I use the angel-startup (actual and counterfactual pairs) sample that was used in section 4 to estimate the selection equation. Since Crunchbase was founded in 2007, I only include startups that raised seed funding in 2008 or later in this analysis. Column (2) of table 10 reports the first-stage regression with $Investment_{i,j}$ as the dependent variable.

The instruments, *Angel Profile on CB* ($\hat{\beta} = 0.076$) and *Startup Profile on CB* ($\hat{\beta} = 0.051$), are positive and significant (at $p < 0.01$), implying that a Crunchbase profile increases the likelihood of matching between angels and startups. In addition, the joint significance test ($\chi^2 = 863.17$) suggests that the instruments are strong. The parameter estimates in column (2) are used to calculate the IMR for each observation and used as a proxy for unobservables in the second-stage.

Effect of Social Connections on Seed-stage Success

Columns (1) and (3) of table 10 report the linear probability model and the $2nd$-stage of the Heckman (1979) selection correction model. The *Connected Angel-Startup* coefficient in column (3) is an estimate of the post-investment influence of social connections on seed-stage success, after correcting for unobservable factors that affect sorting and matching of angels with startups. The Heckman estimate is larger the LPM estimate (13.6% vs. 8.7%), suggesting that the unobserved variables have a negative impact on startup performance. When combined with the fact that connected startups have lower traction and are more likely to be founded by first-time entrepreneurs, this result shows that angels have a lower threshold for observable quality when they choose socially connected startups. However, the positive and larger *Connected Angel-Startup* coefficient in column (3) shows that in spite of the ex-ante lower quality, connected startups are more likely to be successful due to post-investment influence and coordination between angels and founders.

Effect of Social Connections on Other Seed-stage Outcomes

According to coordination hypothesis, connected startups should perform better than unconnected ones. If so, it should be reflected in the fund-raising efforts and the type of investors a startup is able to attract. In the last four columns of table 10, I test the effect of social connections on other seed-stage outcomes, such as, the amount of funds raised in series A, time taken to reach series A stage, and the ability to attract VC investment.

The coefficient of Connected Angel-Startup in column (4) shows that connected startups raise 12.6% (0.26 million) more in series A funds than unconnected startups. However, column (5) shows that connected startups take 14.1% (about 4 months) more time to reach series A stage than unconnected firms, even though both groups raised similar amount of seed funds.[14] The combined results in columns (4) and (5) suggest that angels are more tolerant with the slower pace and short-term failures of their connected startups. This is a desirable behavior in investors according to the theory in Manso (2011), which states that principals should rely on less performance sensitive contracting schemes in innovative ventures.

The positive impact of social connections also extends to the type of investors startups are able to attract in the subsequent stage. Connected startups are 14.6% more likely attract investments from VC firms in series A stage than unconnected startups. This is a big milestone for seed-stage startups because VCs are capable of facilitating a substantial increase in resources. This result, with the fact that 62% of startups in the sample that reached series A stage attract a VC investment, shows that the angels market is a substantial deal-flow provider for the VC market, that is, angel and VC markets are complementary. However, Hellmann et al. (2015) claim that VC and angels markets are substitutes. The contrasting result could be because Hellmann et al. (2015) use data on Canadian startups that were funded by government programs.

In column (7), I test whether seed-stage angels invite their co-investment and social connections as follow-on investors to their portfolio firms. The dependent variable, *Connected Investor*, takes value one if at least one of the series A investors is connected with the startup's seed-stage lead angel. The positive coefficient shows that connected startups are more likely to receive series A funds from contacts of their seed-stage lead angel. This could be because angels invite their connections in order to bring together investors who are easier to work with.

6 Additional Analysis

Effect of Social Connections on Seed round Valuations

A common concern when investigating the effect of social connections is that the angel might have invested in the startup as a favor to the entrepreneur who is a close friend. In such cases, it is reasonable to expect that the angel would purchase the company's shares at a cheaper price, and

[14] Univariate tests in table 3 and unreported regressions show that social connections have no impact on the amount of seed funds raised by startups.

it would be reflected by the statistical relationship between social connections and the share price.

I use seed valuation of 747 startups in the sample to test the above hypothesis. I use the post money valuations and funds raised in the first seed round to calculate the price for 1% of shares. Using this price as the dependent variable, I report regressions that test the effect of social connections in table IA.3. The results show that there is no significant relationship between the price and any social channel, suggesting that angels pay a fair price to buy into connected startups. It has to be noted that investors with larger networks are able to extract a discount from startups.

Alternative Control Samples

Nearest Neighbor Matched Control Sample: A concern with the analysis performed in the paper is that the control startups used for estimation are not similar to the treated ones. Therefore, I use a nearest neighbor approach to create matched samples and rerun the regressions. I adopt the following procedure to create a matched sample for the two-side matching analysis: For each investment made by an angel (actual pair), I search for at least two startups, in the same location that raised funding in the same or previous year, that are similar in terms of age and pre-seed traction (caliper=0.1). Similarly, for each startup, I search for at least two angels, who have invested or shown interest in the same location and are similar to the actual angel investor in terms of degree centrality and success ratio to create two counterfactual pairs. The matched sample contains 67,664 angel-startup pairs. I follow a similar procedure to construct matched samples to study the effect of social connections on seed-stage outcomes. The regressions that test the effect of social connections on matching and seed-stage outcomes are reported in table IA.4. Overall, the coefficients, though smaller, are qualitatively similar to those in the main analysis.

Startups Funded through AngelList: Another natural concern with the control sample is that the entrepreneur will only reach out to those in his social circle, which means that there is no counterfactual. This is not a big concern here because: (i) such investments are considered as friends and family money which occur pre-seed and are not normally reported, and (ii) this analysis includes only professional investors who have invested in at least three startups. To further address this concern, I redo the analysis using a sample of 1,007 startups that used the

AngelList platform to raise funds and report the results in table IA.5. Entrepreneurs who can secure funding directly may not prefer to fund-raise via AngelList, because listing in AngelList requires significant effort to setup profiles and only accredited investors are allowed to invest via the platform. Table IA.5 shows that social connections play a significant role in angel-startup matching and future performance in this sample as well.

Effect of Social Connections in Small and Large Markets

It reasonable to expect that social connections would play a vital role in smaller markets compared to larger ones; because due to the sheer number of startups created, more information and know-how would be available in larger markets such as the Silicon Valley compared to smaller locations. To test this, I divide the sample into large and small regions based on the total number of startups created in a state between 1990 to 2015. According to Crunchbase, 56.3% of the startups were created in California, New York and Massachusetts, which are designated as large markets and the remaining states are considered as small markets (figure IA.2 shows the distribution of startups in the US). The results in table IA.6 show that social connections are twice as important for matching in low activity regions compared to high activity markets. The last four columns examine the effect of social connections on Seed-stage Success in the two sub-samples. The positive effect of social connections in smaller markets are slightly higher in the small markets compared to the large ones, implying that post-investment influence has more impact in markets with less know-how.

7 Conclusion

I examine the effect of social connections on the two-sided matching of angels with startups and the future performance of startups. For this purpose, I assemble a hand-collected database of 9,396 startups seed-funded by 2,655 lead angels between years 2005 to 2015. I find that an angel and startup are more likely to form a partnership when the angel and the startup's founder attended the same school or worked for the same employer during an overlapping time period, or are coethnic. The likelihood of matching increases with the strength of their social connections. The effect of connections on matching are similar irrespective of the quality of the school or employer at which they were formed. Further, tests show that social connections are more important in markets with higher information asymmetry implying that social connections are a good conduit for information flow.

Since the angel-startup matching process involves endogenous selection and assortative matching, I use a two-stage Heckman (1979) model to identify the post-investment influence of social connections on performance. I use the presence of Crunchbase profile pages for angels and startups before the first seed-funding round as instruments to disentangle post-investment influence of angels from pre-investment selection. The results show that, due to post-investment influence of angels, connected startups more often successfully reach series A stage, raise more funding and are more likely to attract VC investment in subsequent rounds than unconnected startups.

This paper contributes to a debate in the literature that examines the effects of social connections on financial transactions by showing that social connections between entrepreneurs and investors improve performance in the context of early-stage startups. The paper also contributes to a small but growing literature on angel investors by constructing a new database on angel investors, and focusing on individual angel investors and their early-stage startups. Finally, this paper also contributes to the research in labor mobility that investigates the relationship between employee movement and startup creation (Decker et al. (2014) and Jeffers (2017)) by suggesting that employees who leave their jobs to start their own companies are more likely to secure financing from their past colleagues. This is a topic that needs further exploration and I leave it for future research.

References

Becker, G. S. (1973). A Theory of Marriage: Part I. *Journal of Political Economy 81*(4), 813–846.

Bernstein, S., A. Korteweg, and K. Laws (2017). Attracting Early-Stage Investors: Evidence from a Randomized Field Experiment. *Journal of Finance 72*(2), 509–538.

Bhagwat, V. (2013). Manager Networks and Coordination of Effort: Evidence from Venture Capital Syndication. *Working Paper*.

Bonini, S., V. Capizzi, M. Valletta, and P. Zocchi (2016). Groups, Networks and the Business Angels' Investment Process. *Working Paper*.

Bottazzi, L., M. D. Rin, and T. Hellmann (2016). The Importance of Trust for Investment: Evidence from Venture Capital. *Review of Financial Studies 29*(9), 2283–2318.

Boulton, T. J., T. D. Shohfi, and P. Zhu (2017). Angels or Sharks? The Role of Personal Characteristics in Angel Investment Decisions. *Working Paper*.

Chemmanur, T. J. and Z. Chen (2014). Venture Capitalists Versus Angels: The Dynamics of Private Firm Financing Contracts. *Review of Corporate Finance Studies 3*(1-2), 39–86.

Chidambaran, N., S. Kedia, and N. R. Prabhala (2011). CEO-Director Connections and Corporate Fraud.

Cohen, L., A. Frazzini, and C. Malloy (2008). The Small World of Investing: Board Connections and Mutual Fund Returns. *Journal of Political Economy 116*(5), 951–979.

Cohen, L., A. Frazzini, and C. Malloy (2010). Sell-Side School Ties. *Journal of Finance 65*(4), 1409–1437.

Cohen, L., A. Frazzini, and C. J. Malloy (2012). Hiring Cheerleaders: Board Appointments of "Independent" Directors. *Management Science 58*(6), 1039–1058.

Currarini, S., M. O. Jackson, and P. Pin (2009). An Economic Model of Friendship: Homophily, Minorities, and Segregation. *Econometrica 77*(4), 1003–1045.

Da Rin, M., T. F. Hellmann, and M. Puri (2013). A Survey of Venture Capital Research. In G. Constantinides, M. Harris, and R. M. Stulz (Eds.), *Handbook of the Economics of Finance*, Chapter 8, Vol. 2. Amsterdam: Holland, North: Elsevier B.V.

Decker, R., J. Haltiwanger, R. Jarmin, and J. Miranda (2014). The Role of Entrepreneurship in US Job Creation and Economic Dynamism. *Journal of Economic Perspectives 28*(3), 3–24.

Engelberg, J., P. Gao, and C. A. Parsons (2012). Friends with money. *Journal of Financial Economics 103*(1), 169–188.

Engelberg, J., P. Gao, and C. A. Parsons (2013). The Price of a CEO's Rolodex. *Review of Financial Studies 26*(1), 79–114.

Gompers, P. and J. Lerner (2001). The Venture Capital Revolution. *Journal of Economic Perspectives 15*(2), 145–168.

Gompers, P. A. (1995). Optimal Investment, Monitoring, and the Staging of Venture Capital. *Journal of Finance 50*(5), 1461–1489.

Gompers, P. A., V. Mukharlyamov, and Y. Xuan (2016). The Cost of Friendship. *Journal of Financial Economics 119*(3), 626–644.

Gorman, M. and W. a. Sahlman (1989). What Do Venture Capitalists Do? *Journal of Business Venturing 4*, 231–248.

Granovetter, M. (2005). The Impact of Social Structure on Economic Outcomes. *Journal of Economic Perspectives 19*(1), 33–50.

Guiso, L., P. Sapienza, and L. Zingales (2009). Cultural Biases in Economic Exchange? *Quaterly Journal of Economics 124*(3), 1095–1131.

Heckman, J. J. (1979). Sample Selection Bias as a Specification Error. *Econometrica 47*(1), 153–161.

Hegde, D. and J. Tumlinson (2014). Does Social Proximity Enhance Business Partnerships? Theory and Evidence from Ethnicity's Role in U.S. Venture Capital. *Management Science 60*(9), 2355–2380.

Hellmann, T., P. H. Schure, and D. Vo (2015). Angels and Venture Capitalists: Complements or Substitutes? *Saïd Business School Working Paper 2015-02*, 1–48.

Hellmann, T. and V. Thiele (2015). Friends or Foes? The Interrelationship between Angel and Venture Capital Markets. *Journal of Financial Economics 115*(3), 639–653.

Hellmann, T. F. and M. Puri (2002). Venture Capital and the Professionalization of Start-Up Firms: Empirical Evidence. *Journal of Finance LVII*(1), 169–197.

Ishii, J. and Y. Xuan (2014). Acquirer-target social ties and merger outcomes. *Journal of Financial Economics 112*(3), 344–363.

Janis, I. L. (1982). *Groupthink: Psychological Studies of Policy Decisions and Fiascoes* (2nd ed.). Cengage Learning.

Jeffers, J. S. (2017). The Impact of Restricting Labor Mobility on Corporate Investment and Entrepreneurship. *Working Paper*.

Kerr, W. R., J. Lerner, and A. Schoar (2014). The Consequences of Entrepreneurial Finance: Evidence from Angel Financings. *Review of Financial Studies 27*(781), 20–55.

Kerr, W. R. and W. F. Lincoln (2010). The Supply Side of Innovation: H-IB Visa Reforms and U . S . Ethnic Invention. *Journal of Labor Economics 28*(3), 473–508.

Kerr, W. R., R. Nanda, and M. Rhodes-kropf (2014). Entrepreneurship as Experimentation. *Journal of Economic Perspectives 28*(3), 25–48.

Lamoreaux, N., M. Levenstein, and K. Sokoloff (2004). Financing Invention During the Second Industrial Revolution: Cleveland, Ohio, 1870-1920. *NBER Working Paper No. 10923*.

Lerner, J. (1998). Angel financing and public policy: An overview. *Journal of Banking & Finance 22*(6-8), 773–783.

Lerner, J., A. Schoar, S. Sokolinski, and K. Wilson (2017). The Globalization of Angel Investments: Evidence Across Countries. *Journal of Financial Economics Forthcomin*.

Liberti, J. M. and M. A. Petersen (2017). Information: Hard and Soft. *Working Paper*.

Lin, M., N. R. Prabhala, and S. Viswanathan (2013). Judging Borrowers by the Company They Keep: Friendship Networks and Information Asymmetry in Online Peer-to-Peer Lending. *Management Science 59*(1), 17–35.

Manso, G. (2011). Motivating Innovation. *Journal of Finance 66*(5), 1823–1860.

McPherson, M., L. Smith-Lovin, and J. M. Cook (2001). Birds of a Feather: Homophily in Social Networks. *Annual Review of Sociology 27*, 415–444.

OECD (2011). Financing High-Growth Firms: The Role of Angel Investors. Technical report, OECD.

Petersen, M. A. and R. G. Rajan (2002). Does Distance Still Matter? The Information Revolution in Small Business Lending. *Journal of Finance 57*(6), 2533–2570.

PitchBook (2017). Universities Report 2016-2017 Edition. Technical report.

Shane, S. (2012). The Importance of Angel Investing in Financing the Growth of Entrepreneurial Ventures. *Quarterly Journal of Finance 02*(02), 1250009.

Sohl, J. (2015). The Angel Investor Market in 2013: A Return to Seed Investing. Technical Report 603, Center for Venture Research.

Sørensen, M. (2007). How Smart Is Smart Money? a Two Sided Matching Model of Venture Capital. *The Journal of Finance 62*(6), 2725–2762.

Spence, M. (2002). Signaling in Retrospect and the Informational Structure of Markets. *American Economic Review 92*(3), 434–459.

Steiner, I. D. (1972). *Group Process and Productivity (Social Psychological Monograph)*. Academic Press Inc.

Surowiecki, J. (2005). *The Wisdom of Crowds: Why the Many Are Smarter Than the Few and How Collective Wisdom Shapes Business, Economies, Societies and Nations* (2nd ed.). Anchor Books.

Venugopal, B. and V. Yerramilli (2017). How do Investors Accumulate Network Capital? Evidence from Angel Networks. *Working Paper*.

Ye, J., S. Han, Y. Hu, B. Coskun, M. Liu, H. Qin, and S. Skiena (2017). Nationality Classification using Name Embeddings. In *ACM Conference Proceedings*, Washington, DC.

Figure 1 Sample Investor Profile on Crunchbase

The figure below is an excerpt of Alexis Ohanian's (Co-founder of Reddit and most active angel in 2014) profile on Crunchbase.

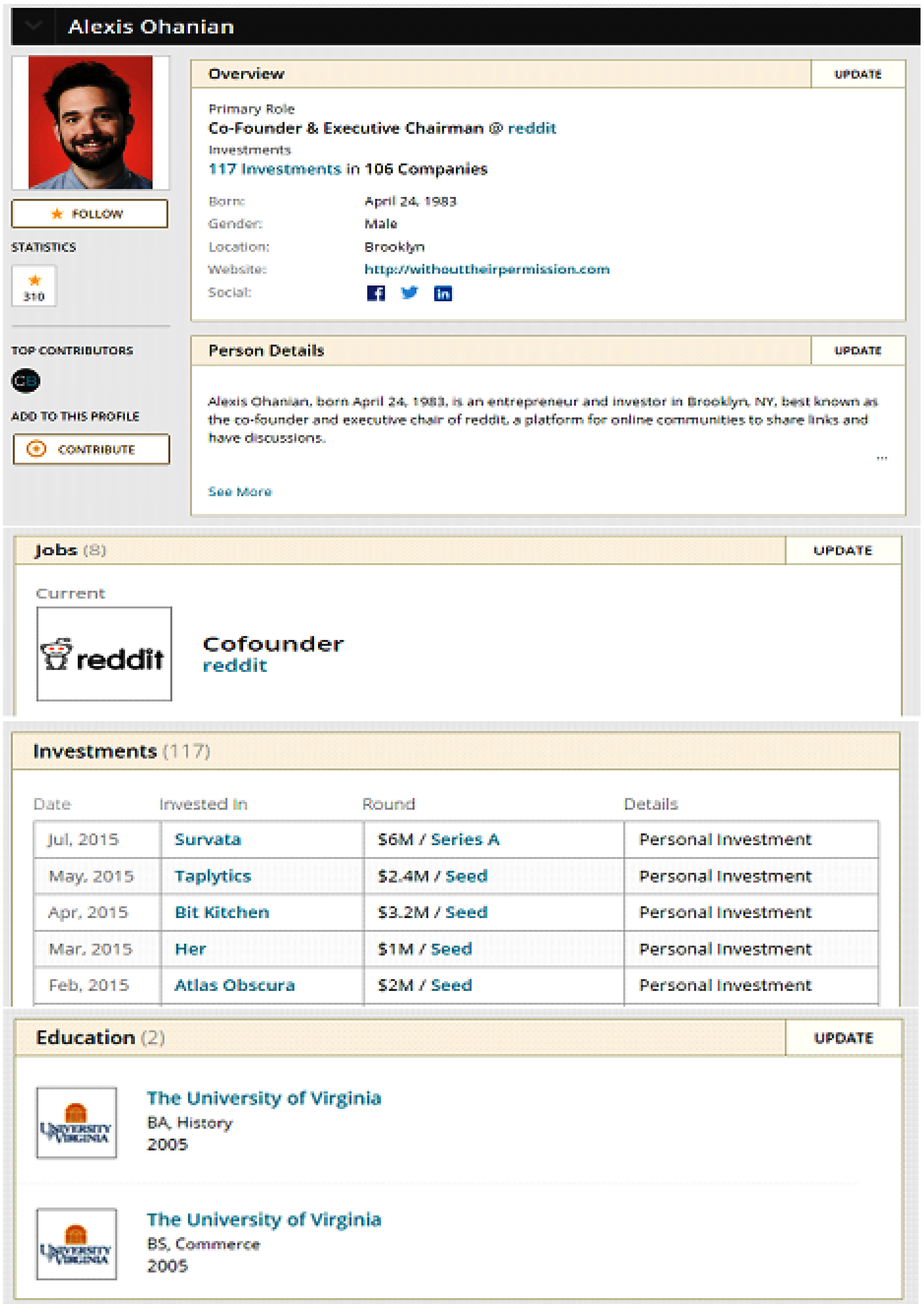


Figure 2 Sample start-up Profile on Crunchbase

The figure below is an excerpt from UBER's profile on Crunchbase.

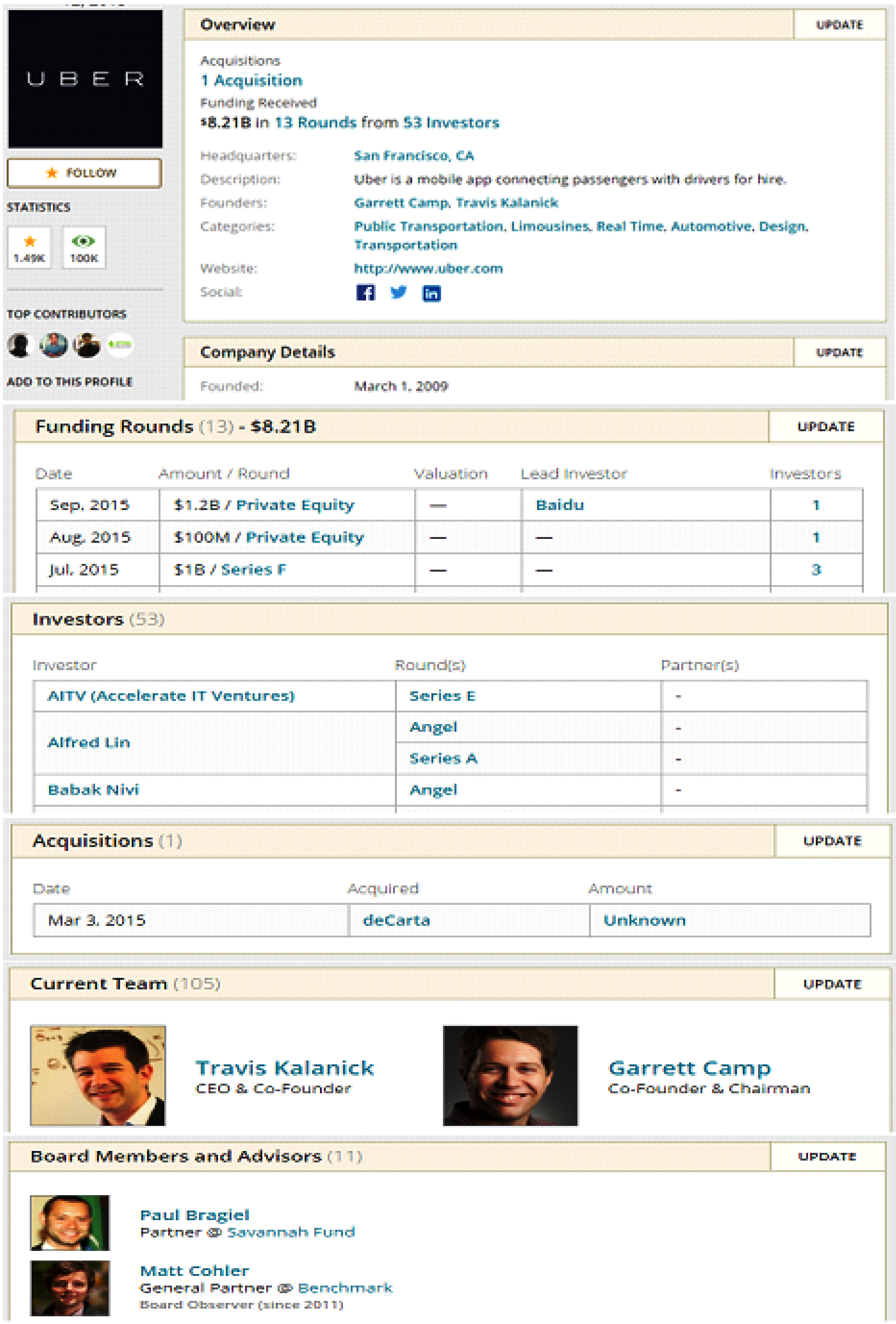


Table 1 Summary statistics

This table reports summary statistics of the key variables for startups in the sample at three different points in time: (i) before startups raised their first seed round (ii) when the startup is in seed-stage, and (iii) after seed-stage, that is, during series A stage. The sample includes 9,396 startups seed funded by 2,655 lead angels during the period 2005 to 2015. All variables are defined in Appendix B.

Variable	Mean	SD	p25	p50	p75	N
Pre-seed Startup Characteristics						
Age at seed	0.97	1.04	0.00	0.67	1.53	9396
No. of Founders	1.87	1.30	1.00	2.00	2.00	9396
Serial Entrepreneur	0.12	0.32	0.00	0.00	0.00	9396
Traction	2.97	2.99	0.60	1.55	5.23	9396
Seed-stage Startup Characteristics						
Seed Funds	0.86	4.87	0.00	0.19	0.75	9396
No. of seed rounds	1.17	0.47	1.00	1.00	1.00	9396
No. of seed investors	1.99	1.78	1.00	1.00	2.00	9396
Lead Angel Characteristics						
Investor Experience	2.33	3.00	1.00	1.00	2.00	9396
Investor Degree	12.68	44.06	1.00	1.00	3.00	9396
Entrepreneur-Investor	0.51	0.50	0.00	1.00	1.00	9396
No. of rounds participated	5.02	13.24	1.00	1.00	2.58	9396
Investor Success Ratio	0.14	0.26	0.00	0.00	0.25	9396
Post-seed Outcomes						
Seed Success	0.20	0.40	0.00	0.00	0.00	9396
Series A Funds	4.24	8.77	0.20	2.00	5.00	1863
Time to Series A	1.47	1.30	0.51	1.17	2.00	1863
VC in Series A	0.62	0.49	0.00	1.00	1.00	1863
No. of Series A rounds	1.19	0.56	1.00	1.00	1.00	1863

Table 2 Summary Statistics: Social Connections between Angels and Entrepreneurs

This table summarizes the social characteristics of 15,951 founders and 2,655 lead angels in the sample. Panel A and B report the top five frequent schools and employers of founders and lead angels in the sample. Panel C shows the ethnic distribution of people in the sample and Panel D summarizes the social connection variables used in this study.

Panel A: Top 5 Common Schools and Employers of Founders

Total no. of founders = 15951

School	Frequency	Previous Employer	Frequency
Stanford University	366	Google	336
Harvard University	186	Microsoft	240
University of California, Berkeley	180	IBM	168
University of Pennsylvania	174	Yahoo	156
University of Southern California	144	Apple	156
% of total Founders	6.58		6.62
% from Top 100 Schools/Employers	26.92		30.71

Panel B: Top 5 Common Schools and Employers of Lead Angels

Total no. of Lead Angels = 2655

School	Frequency	Previous Employer	Frequency
Stanford University	101	Google	69
Harvard University	47	Microsoft	62
University of Pennsylvania	44	Facebook	44
Massachusetts Institute of Technology	43	McKinsey & Company	31
University of Cambridge	28	IBM	29
% of total Lead Angels	9.91		8.85
% from Top 100 Schools/Employers	32.62		32.46

Panel C: Ethnicity of Founders and Lead Angels

Ethnicity	Founders	Lead Angels
African	234	50
Celtic English	6366	1049
East Asian	657	130
East European	654	114
Hispanic	1521	235
Jewish	285	41
Middle East	246	40
Nordic	441	81
South Asian	2772	442
South European	507	77
West European	2268	396
Total	15951	2655

Panel D: Social Connections between Lead Angels and Startups

Variable	Mean	SD	p25	p50	p75	N
Same School	0.13	0.33	0.00	0.00	0.00	9396
Same Employer	0.21	0.41	0.00	0.00	0.00	9396
Same Ethnic Minority	0.30	0.46	0.00	0.00	1.00	9396
Connected Angel-Founder	0.46	0.50	0.00	0.00	1.00	9396

Table 3 Summary statistics: Connected vs. Unconnected Startups

This table reports univariate comparisons of key variables across connected and unconnected startups. Connected startups (A) are startups in which the lead seed-stage angel has a social connection – same school, same employer or same ethnic minority – with at least one of the startup's founder. Unconnected startups (B) are those startups in which the lead seed-stage angel and founders do not share any social connection. The last column provides the t-statistic for the test of difference between the two groups of startups. All variables are defined in Appendix B.

	Connected Startups (A)				Unconnected Startups (B)				t-stat
Variable	Mean	Std	Median	N	Mean	Std	Median	N	(A-B)
Pre-seed Startup Characteristics									
Age at seed	0.89	1.00	0.58	4323	1.04	1.07	0.72	5073	-7.02
No. of Founders	2.11	1.53	2.00	4323	1.67	1.03	1.00	5073	16.06
Serial Entrepreneur	0.07	0.27	0.00	4323	0.16	0.37	0.00	5073	-13.42
Traction	2.89	3.00	1.51	4323	3.04	2.98	1.57	5073	-2.42
Top School	0.28	0.45	0.00	4323	0.27	0.44	0.00	5073	1.08
Top Employer	0.32	0.46	0.00	4323	0.31	0.46	0.00	5073	1.04
Seed-stage Startup Characteristics									
Seed Funds	0.87	6.31	0.15	4323	0.85	3.17	0.20	5073	0.19
No. of seed rounds	1.17	0.46	1.00	4323	1.16	0.47	1.00	5073	1.04
No. of seed investors	2.08	1.83	1.50	4323	1.90	1.73	1.00	5073	4.87
Lead Angel Characteristics									
Investor Experience	2.00	2.49	1.00	4323	2.62	3.35	1.00	5073	-10.27
Investor Degree	11.89	39.96	2.00	4323	13.35	48.92	2.00	5073	-1.59
Entrepreneur-Investor	0.66	0.47	1.00	4323	0.38	0.49	0.00	5073	28.22
No. of rounds participated	4.82	12.36	1.00	4323	5.19	15.72	1.00	5073	-1.28
Investor Success Ratio	0.14	0.28	0.00	4323	0.15	0.29	0.00	5073	-1.69
Top School	0.31	0.46	0.00	4323	0.34	0.48	0.00	5073	-3.10
Top Employer	0.33	0.47	0.00	4323	0.32	0.47	0.00	5073	1.30
Post-seed Outcomes									
Seed Success	0.23	0.42	0.00	4323	0.17	0.38	0.00	5073	7.21
Series A Funds	4.44	4.57	1.70	989	4.01	4.98	2.30	874	1.93
Time to Series A	1.55	1.36	1.21	989	1.40	1.25	1.13	874	2.48
VC in Series A	0.65	0.48	1.00	989	0.59	0.49	1.00	874	2.66
No. of Series A rounds	1.17	0.53	1.00	989	1.21	0.60	1.00	874	-1.52

Table 4 Effect of Social Connections on Angel-Startup Matching

This table reports regressions that test the effect of social connections on angel-startup matching. The dependent variable,*Investment*, is a dummy variable equal to one if the investor made an investment in a startups (actual pairs) and zero otherwise (counter-factual pairs). The independent variables of interest are pairwise social characteristics (education, employment and ethnicity) of angels and startups. All regressions control for observable startup characteristics, such as age, serial founder, traction, and quality of founder's school and employer, and angel characteristics, such as degree centrality, entrepreneur-investor, success ratio, and quality of angel's school and employer. The regressions also include seed year, location and product market fixed effects. Robust standard errors clustered at industry level are reported in parenthesis. ***, **, and * denote statistical significance at 1%, 5% and 10% levels respectively. All variables are defined in Appendix B.

	Investment							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Same School	0.061***						0.027***	
	(0.001)						(0.001)	
Same Top School		0.059***						0.006***
		(0.001)						(0.001)
Same Bottom School		0.066***						0.034***
		(0.004)						(0.004)
Same Employer			0.283***				0.282***	
			(0.001)				(0.001)	
Same Top Employer				0.225***				0.195***
				(0.001)				(0.001)
Same Bottom Employer				0.314***				0.289***
				(0.002)				(0.002)
Same Ethnic Minority					0.007***		0.005***	
					(0.000)		(0.000)	
Both African						0.042***		0.019***
						(0.003)		(0.002)
Both East Asian						0.009***		0.006***
						(0.001)		(0.001)
Both Hispanic						0.008***		0.004***
						(0.000)		(0.000)
Both Jewish						0.016***		0.009***
						(0.001)		(0.001)
Both Middle East						0.021***		0.013***
						(0.002)		(0.002)
Both South Asian						0.004***		0.003***
						(0.000)		(0.000)
Obs.	2395651	2395651	2395651	2395651	2395651	2395651	2395651	2395651
Adj. R^2	0.122	0.129	0.215	0.227	0.121	0.121	0.215	0.227
Location, Prod. Market & Yr. F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Table 5 Effect of Social Connections Strength on Angel-Startup Matching

This table reports regressions that test the effect of social connection strength on angel-startup matching. The dependent variable, *Investment*, is a dummy variable equal to one if the investor made an investment in a startups (actual pairs) and zero otherwise (counter-factual pairs). The independent variables of interest are the interactions of angel-startup social characteristics and Connection Depth. All regressions control for observable startup characteristics, such as age, serial founder, traction, and quality of founder's school and employer, and angel characteristics, such as degree centrality, entrepreneur-investor, success ratio, and quality of angel's school and employer. The regressions also include seed year, location and product market fixed effects. Robust standard errors clustered at industry level are reported in parenthesis. ***, **, and * denote statistical significance at 1%, 5% and 10% levels respectively. All variables are defined in Appendix B.

	Investment				
	(1)	(2)	(3)	(4)	(5)
Same School	0.018***				
	(0.001)				
Same Employer	0.162***				
	(0.002)				
Same Ethnic Minority	0.002***		0.004***	0.004***	0.004***
	(0.000)		(0.000)	(0.000)	(0.000)
Same School × Employer	0.091***				
	(0.005)				
Same School × Ethnic Minority	0.006**				
	(0.003)				
Same Employer × Ethnic Minority	0.048***				
	(0.003)				
Same School × Employer × Ethnic Minority	0.072***				
	(0.009)				
Connection Depth=1		0.023***			
		(0.000)			
Connection Depth=2		0.188***			
		(0.001)			
Connection Depth=3		0.291***			
		(0.006)			
Same Top School			0.006***	0.004***	0.005***
			(0.001)	(0.001)	(0.001)
Same Bottom School			0.034***	0.028***	0.028***
			(0.004)	(0.006)	(0.006)
Same Top Employer			0.195***	0.126***	0.124***
			(0.001)	(0.002)	(0.002)
Same Bottom Employer			0.289***	0.163***	0.161***
			(0.002)	(0.003)	(0.003)
Same Top School × Top Employer				0.093***	0.095***
				(0.009)	(0.010)
Same Top School × Bottom Employer				-0.183***	
				(0.008)	
Same Bottom School × Top Employer				0.081***	
				(0.020)	
Same Bottom School × Bottom Employer					0.086***
					(0.010)
Obs.	2395651	2395651	2395651	2395651	2395651
Adj. R^2	0.227	0.110	0.227	0.236	0.235
Location, Prod. Market, Yr. F.E.	Yes	Yes	Yes	Yes	Yes

Table 6 Effect of Social Connections on Angel-Startup Matching: New vs. Established Markets

This table reports regressions that test the effect of social connections on angel-startup matching in newly product markets. The dependent variable, *Investment*, is a dummy variable equal to one if the investor made an investment in a startups (actual pairs) and zero otherwise (counter-factual pairs). *Connected Angel-Startup* indicates whether the angel and startup have a prior social connection. *New Market* identifies startups in the sample that belong to product markets that emerged in the last three years or those that were the first 25% entrants into a given product market. All regressions control for observable angels and startup characteristics and also include fixed effects for seed year, location and product market. Robust standard errors clustered at industry level are reported in parenthesis. ***, **, and * denote statistical significance at 1%, 5% and 10% levels respectively. All variables are defined in Appendix B.

	Investment		
	(1)	(2)	(3)
Connected Angel-Startup	0.234***	0.182***	
	(0.005)	(0.006)	
Same School			0.023***
			(0.001)
Same Employer			0.215***
			(0.001)
Same Ethnic Minority			0.004***
			(0.000)
New Market		-0.068***	-0.063***
		(0.013)	(0.010)
Connected Angel-Startup × New Market		0.087***	
		(0.008)	
Same School × New Market			0.043*
			(0.022)
Same Employer × New Market			0.091***
			(0.009)
Same Ethnic Minority × New Market			0.019**
			(0.008)
Obs.	2395651	2395651	2395651
Adj. R^2	0.149	0.149	0.228
Location, Prod. Market & Yr. F.E.	Yes	Yes	Yes

Table 7 Effect of Social Connections on Seed-stage Success

This table reports regressions that test the effect of social connections on a startup's seed-stage success. The dependent variable, *Seed-stage Success*, is a dummy variable equal to one if the startup raised series A funds. The independent variables of interest are pairwise social characteristics (education, employment and ethnicity) of angels and startups. All regressions control for observable startup characteristics, such as age, serial founder, traction, and quality of founder's school and employer, and angel characteristics, such as degree centrality, entrepreneur-investor, success ratio, and quality of angel's school and employer. The regressions also include angel, seed year, location and product market fixed effects. Robust standard errors clustered at industry level are reported in parenthesis. ***, **, and * denote statistical significance at 1%, 5% and 10% levels respectively. All variables are defined in Appendix B.

	Seed-stage Success							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Same School	0.091***						0.076**	
	(0.028)						(0.034)	
Same Top School		0.096**						0.080*
		(0.041)						(0.042)
Same Bottom School		0.083**						0.065*
		(0.036)						(0.037)
Same Employer			0.102***				0.127***	
			(0.021)				(0.022)	
Same Top Employer				0.119***				0.131***
				(0.033)				(0.038)
Same Bottom Employer				0.084***				0.101***
				(0.029)				(0.036)
Same Ethnic Minority					0.039**		0.037*	
					(0.019)		(0.019)	
Both African						0.048		-0.042
						(0.094)		(0.094)
Both East Asian						-0.008		0.012
						(0.059)		(0.059)
Both Hispanic						0.001		0.011
						(0.040)		(0.040)
Both Jewish						0.072		0.080
						(0.083)		(0.083)
Both Middle East						-0.085		-0.085
						(0.098)		(0.098)
Both South Asian						0.020		0.025
						(0.032)		(0.032)
Ln(Seed Funds)	0.068***	0.068***	0.067***	0.068***	0.067***	0.067***	0.067***	0.069***
	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)
Obs.	9396	9396	9396	9396	9396	9396	9396	9396
Adj. R^2	0.172	0.170	0.169	0.171	0.169	0.168	0.172	0.172
Location, Prod. Market, Yr. F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Table 8 Effect of Social Connections Strength on Seed-stage Success

This table reports regressions that test the effect of social connection strength on a startup's seed-stage success. The dependent variable, *Seed-stage Success*, is a dummy variable equal to one if the startup raised series A funds. The independent variables of interest are the interactions of angel-startup social characteristics and Connection Depth. All regressions control for observable startup characteristics, such as age, serial founder, traction, and quality of founder's school and employer, and angel characteristics, such as degree centrality, entrepreneur-investor, success ratio, and quality of angel's school and employer. The regressions also include angel, seed year, location and product market fixed effects. Robust standard errors clustered at industry level are reported in parenthesis. ***, **, and * denote statistical significance at 1%, 5% and 10% levels respectively. All variables are defined in Appendix B.

	Seed-stage Success				
	(1)	(2)	(3)	(4)	(5)
Same School	0.071**				
	(0.031)				
Same Employer	0.088***				
	(0.027)				
Same Ethnic Minority	0.027*		0.029*	0.028	0.028
	(0.016)		(0.017)	(0.018)	(0.018)
Same School × Employer	0.069**				
	(0.034)				
Same School × Ethnic Minority	0.028				
	(0.018)				
Same Employer × Ethnic Minority	0.013**				
	(0.006)				
Same School × Employer × Ethnic Minority	0.112**				
	(0.053)				
Connection Depth=1		0.044***			
		(0.018)			
Connection Depth=2		0.079***			
		(0.030)			
Connection Depth=3		0.123**			
		(0.062)			
Same Top School			0.080*	0.067	0.066
			(0.042)	(0.043)	(0.043)
Same Bottom School			0.065*	0.055	0.055
			(0.037)	(0.038)	(0.038)
Same Top Employer			0.131***	0.119***	0.119***
			(0.038)	(0.041)	(0.040)
Same Bottom Employer			0.101***	0.098***	0.098***
			(0.036)	(0.039)	(0.039)
Same Top School × Top Employer				0.077***	0.080***
				(0.031)	(0.030)
Same Top School × Bottom Employer				0.134	
				(0.092)	
Same Bottom School × Top Employer				0.106**	
				(0.048)	
Same Bottom School × Bottom Employer					0.050
					(0.068)
Obs.	9396	9396	9396	9396	9396
Adj. R^2	0.171	0.171	0.171	0.171	0.171
Location, Prod. Market, Yr. F.E.	Yes	Yes	Yes	Yes	Yes

Table 9 Effect of Social Connections on Seed-stage Success: New vs. Established Markets

This table reports regressions that test the effect of social connections on angel-startup matching in newly product markets. The dependent variable, *Seed-stage Success*, is a dummy variable equal to one if the startup raised series A funds. *Connected Angel-Startup* indicates whether the angel and startup have a prior social connection. *New Market* identifies startups in the sample that belong to product markets that emerged in the last three years or those that were the first 25% entrants into a given product market. All regressions control for observable angels and startup characteristics and also include fixed effects for angel, seed year, location and product market. Robust standard errors clustered at industry level are reported in parenthesis. ***, **, and * denote statistical significance at 1%, 5% and 10% levels respectively. All variables are defined in Appendix B.

	Seed-stage Success		
	(1)	(2)	(3)
Connected Angel-Startup	0.095***	0.084***	
	(0.017)	(0.018)	
Same School			0.064**
			(0.031)
Same Employer			0.088***
			(0.023)
Same Ethnic Minority			0.015
			(0.021)
New Market		0.032	0.031
		(0.020)	(0.020)
Connected Angel-Startup × New Market		0.058*	
		(0.033)	
Same School × New Market			0.053*
			(0.031)
Same Employer × New Market			0.073**
			(0.036)
Same Ethnic Minority × New Market			0.022
			(0.032)
Obs.	9396	9396	9396
Adj. R^2	0.143	0.147	0.151
Location, Prod. Market, Yr. F.E.	Yes	Yes	Yes

Table 10 Effect of Angel and Founder Social Connections on Seed-stage Success: Heckman Selection Correction

This table reports results of the two-stage Heckman (1979) model (equation (2)) that attempts to isolate the influence effect of socially connected angels on startups from the selection effect. The selection equation is estimated using the angel-startup (actual and counterfactual) pairs sample with the same controls and fixed effects used in tables 4 and 5. The sample period is limited to 2008-2015 since Crunchbase was started only in July 2007. The instruments included in the first stage are: *Startup Profile on* CB_j, which indicates whether startup *'j'* had a profile in Crunchbase before its first seed funding round, and *Angel Profile on* $CB_{i,j}$, which indicates whether angel *'i'* had a profile on Crunchbase before startup *'j'* raised its first seed round. The Inverse Mills Ratio (IMR) is calculated using the first-stage coefficient estimates and used in the second stage. All regressions include founding year, location and product market fixed effects. Robust standard errors clustered at industry level are reported in parenthesis. ***, **, and * denote statistical significance at 1%, 5% and 10% levels respectively. All variables are defined in Appendix B.

	OLS	2SLS: First stage	2SLS: Second stage				
	(1) Seed Success	(2) Investment	(3) Seed Success	(4) Ln(Series A Funds)	(5) Ln(Time to Series A)	(6) VC in Series A	(7) Connected Investor
Connected Angel-Startup	0.087***	0.112***	0.136***	0.126**	0.141**	0.146*	0.153*
	(0.020)	(0.016)	(0.024)	(0.055)	(0.067)	(0.083)	(0.078)
Ln(Traction)	0.037***	0.002***	0.020*	-0.074	0.025	0.021	0.019
	(0.010)	(0.000)	(0.011)	(0.070)	(0.031)	(0.038)	(0.064)
Ln(Seed Funds)	0.052***		0.093***	0.502***	0.074	-0.044	-0.035
	(0.015)		(0.020)	(0.114)	(0.050)	(0.062)	(0.104)
Ln(Degree)	0.014**	0.000	0.019**	0.068	0.010	0.052**	0.146***
	(0.006)	(0.001)	(0.007)	(0.044)	(0.019)	(0.024)	(0.040)
Seed Success Ratio	0.201***	0.003***	0.166***	-0.052	-0.425***	0.008	-0.047
	(0.031)	(0.001)	(0.036)	(0.186)	(0.082)	(0.102)	(0.171)
Inverse Mills Ratio			-0.082***	-0.131*	0.055*	-0.018	-0.057
			(0.010)	(0.067)	(0.029)	(0.036)	(0.061)
Angel on CB Before Seed		0.076***					
		(0.014)					
Startup on CB Before Seed		0.051***					
		(0.010)					
Obs.	5793	1942292	5793	1167	1167	1167	1167
R^2	0.161	0.397	0.152	0.151	0.294	0.098	0.015
Location, Prod. Market, Yr. F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Appendix A Construction of Startup Traction measure

Investors evaluate startups based on quality of the founding team, viability of the product or service offered, milestones achieved and, especially, traction (Bernstein et al. (2017)). Traction[15], a measure of startup progress or product demand, is firm-specific and soft in nature like most information in this market (Liberti and Petersen (2017)). Since traction data is unavailable to outsiders, I use web-search counts ("Interest over time") from Google Trends as a measure of startup traction. For each startup, I download the entire trends history for the name and primary product of the company. The counts are then normalized to a scale of 0 to 10 within each product market category, where 10 denotes a startup with highest web-search hits in that product market category. This is a reasonable proxy because web-search trends reflect the potential demand for a startup's product, which is the spirit of traction measures used in the industry and academia (Kerr et al. (2014) and Bernstein et al. (2017)).

[15] AngelList founder Naval Ravikant calls traction as "Quantitative evidence of market demand" and one of the most important measure of progress. For example, in case of social networking sites such as Twitter or Facebook one way to measure traction could be average number of users signing-in daily.

Appendix B Variable Definitions

Start-up Financing Stages:

Start-ups raise funds at various stages of their life cycle. Industry participants classify these financing stages as *Seed*, *Series A*, *Series B*, *Series C*, and so on. The academic literature (e.g., see Gompers (1995)) sometimes refers to series A as "early stage," series B as "expansion stage," and series C and beyond as "late stage." The informal definitions of the these stages are as follows:[16]

- *Seed stage:* The purpose of the series seed is for the startup to figure out the product it is building, the market it is in, and the user base. Typically, a seed round helps the company scale to a few employees past the founders and to build and launch an early product.
- *Series A:* Startups that get to this stage have figured out their product and user base, and are trying to establish a viable business model and scale up their operations.
- *Series B:* This stage is all about scaling. Startups that get to this stage have an established product and business model, and are trying to scale up their business model and user base.
- *Series C:* This stage is used by startups to accelerate their growth beyond the Series B stage; e.g., by going international or by making acquisitions. Firms requiring more funds raise them in stages Series D, E, etc.

The startups disclose the financing stage when they raise funds, and this information is reported by Crunchbase and AngelList. Each financing stage may itself involve multiple funding rounds.

Social Connection Variables and Instruments:

- *Same School* is a binary variable that takes value 1 if the (lead) angel and founder attended the same school during an overlapping time period.
- *Same Employer* is a binary variable that takes value 1 if the (lead) angel and founder worked for the same employer during an overlapping time period.
- *Same Ethnic Minority* is a binary variable that takes value 1 if the (lead) angel and founder belong to the same ethnic minority.
- *Connected Angel-Startup* is a binary variable that indicates whether the lead angel and startup founder have a prior social connection.
- *Connection Depth* takes a value '0' if the angel and founder do not share any common homophily characters; takes a value '1' if they share exactly one characteristic; takes a value '2' if they share two characteristics; and finally takes a value 3 if they share all three characteristics.
- *Same Top School* indicates whether the seed-stage lead angel and one of the founders have attended the same top school. I define top schools as those educational institutions that appear in the top 100 spots of Crunchbase rankings.
- *Same Bottom School* indicates whether the seed-stage lead angel and one of the founders have attended the same non-top school. I define non-top or bottom schools as those educational institutions that do not appear in the top 100 spots of Crunchbase rankings.

[16]See `http://blog.eladgil.com/2011/03/how-funding-rounds-differ-seed-series.html` for a more detailed description of these funding stages.

- *Same Top Employer* indicates whether the seed-stage lead angel and one of the founders have worked for the same top employer. I define top employers as those companies that appear in the top 100 spots of Crunchbase rankings.
- *Same Bottom Employer* indicates whether the seed-stage lead angel and one of the founders have worked for the same non-top employer. I define non-top or bottom employer as those companies that do not appear in the top 100 spots of Crunchbase rankings.
- *Startup Profile on* CB_j indicates whether the startup *'j'* had a profile page on Crunchbase before its first seed funding round.
- *Angel Profile on* $CB_{i,j}$ indicated whether angel investor *'i'* had a profile on Crunchbase before the first seed funding round of startup *'j'*.

Startup Characteristics:

- *Age at Seed* is the number of years from the founding date to the first seed round.
- *Serial Entrepreneur* is a binary variable that indicates whether at least one of the founders of the startup is a serial entrepreneur, that is, the founder has started at least one company before a given startup.
- *Traction:* is the "Interest Over Time" Google Trends monthly count normalized on a scale of 0 to 10 with in each product market category.
- *Seed Funds* is the total funds (in $ millions) raised by a start-up in seed rounds.

Angel Investor Characteristics:

- *Degree* $Centrality_t$ is the number of past co-investor connections an angel investor has as of year 't'.
- *Entrepreneur-Investor* indicates whether the angel investor is a past entrepreneur.
- *Seed Success* $Ratio_t$ is the number of successful seed investments as of year 't-1' divided by the total number of seed investments made by an angel as of year 't-1'. A successful seed investment is one that has moved from seed to series A stage.

Matching and Post-seed Outcomes:

- $Investment_{i,j}$ takes a value one when angel i has invested in startup j (actual pair), and zero otherwise (counterfactual pairs).
- *Seed-stage Success* indicates whether a startup successfully moved from seed to series A stage by raised series A funds.
- *Time to Series A* is the number of years between the first seed funding round and the first series A round.
- *Connected Investor* takes a value '1' if at least one of the series A investor is connected with the seed-stage lead angel.
- *VC in Series A* indicates whether at least one VC participated in the first series A round of a startup.

Internet Appendix: Additional Results

Table IA.1 Effect of Angel and Startup Characteristics on Angel-Startup Matching

This table reports regressions that test the effect of angel and startup characteristics on the likelihood of an angel investing in a startup. The dependent variable, *Investment*, is a dummy variable equal to one if the investor made an investment in a startups (actual pairs) and zero otherwise (counter-factual pairs). The independent variables include startup characteristics —such as age, startup traction, and founder school and employer quality— and angel investor characteristics —such as degree centrality, success ratio, and quality of the angel's school and employer. All regressions include seed year, location and product market fixed effects. Robust standard errors clustered at industry level are reported in parenthesis. ***, **, and * denote statistical significance at 1%, 5% and 10% levels respectively. All variables are defined in the Appendix B.

	Investment			
	(1)	(2)	(3)	(4)
Startup Characteristics				
Ln(Age at Seed)	-0.002***	-0.002***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)
Serial Founder	0.001***	0.001***	0.000***	0.000*
	(0.000)	(0.000)	(0.000)	(0.000)
Ln(Traction)	0.001***	0.001***	0.001***	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)
Top School: Founder			0.001***	0.000***
			(0.000)	(0.000)
Top Employer: Founder				0.001***
				(0.000)
Angel Investor Characteristics				
Ln(Degree Centrality)		0.000	0.000	0.000
		(0.000)	(0.000)	(0.000)
Entrepreneur-Investor		0.001***	0.001***	0.001***
		(0.000)	(0.000)	(0.000)
Success Ratio		0.001	0.002**	0.002**
		(0.001)	(0.001)	(0.001)
Top School: Angel			0.001***	0.000***
			(0.000)	(0.000)
Top Employer: Angel				0.001***
				(0.000)
Obs.	2395651	2395651	2395651	2395651
Adj. R^2	0.040	0.040	0.040	0.040
Location, Prod. Market, Yr. F.E.	Yes	Yes	Yes	Yes

Table IA.2 Effect of Angel and Startup Characteristics on Seed-stage Success

This table reports regressions that test the effect of angel and startup characteristics on the likelihood of a startup successfully raising series A funds. The dependent variable, *Seed-stage Success*, is a dummy variable equal to one if the startup raised series A funds. The independent variables include startup characteristics —such as age, startup traction, and founder school and employer quality— and angel investor characteristics —such as degree centrality, success ratio, and quality of the angel's school and employer. All regressions include seed year, location and product market fixed effects. Robust standard errors clustered at industry level are reported in parenthesis. ***, **, and * denote statistical significance at 1%, 5% and 10% levels respectively. All variables are defined in Appendix B.

	Seed-stage Success			
	(1)	(2)	(3)	(4)
Startup Characteristics				
Ln(Age at Seed)	-0.024	-0.043 ***	-0.025	-0.023
	(0.016)	(0.016)	(0.016)	(0.017)
Serial Entrepreneur	0.009	0.006	0.007	0.009
	(0.016)	(0.015)	(0.015)	(0.016)
Ln(Traction)	0.024***	0.027***	0.024***	0.021***
	(0.008)	(0.008)	(0.008)	(0.008)
Top School: Founder			0.058***	0.056***
			(0.020)	(0.022)
Top Employer: Founder				0.083***
				(0.020)
Angel Investor Characteristics				
Ln(Degree)		0.018***	0.012**	0.013**
		(0.005)	(0.005)	(0.006)
Entrepreneur-Investor		0.017	0.012	0.003
		(0.015)	(0.015)	(0.015)
Seed Success Ratio		0.107***	0.109***	0.106***
		(0.026)	(0.026)	(0.026)
Top School: Angel			-0.003	0.013
			(0.024)	(0.026)
Top Employer: Angel				-0.016
				(0.023)
Obs.	9396	9396	9396	9396
Adj. R^2	0.058	0.099	0.109	0.114
Location, Prod. Market, Yr. F.E.	Yes	Yes	Yes	Yes

Table IA.3 Effect of Social Connections on Seed Valuation

This table reports regressions that investigate the effect of social connections on seed round valuations of startups. Seed valuations were available for 747 startups in our sample from AngelList. Assuming that this is the first time founders are distributing equity to outsiders, I calculate the price for 1% of shares using the seed funds raised and post-money valuation provided by AngelList to create the dependent variable *Price for 1% of Shares*, which is the price an angel investor would have paid to purchase 1% of the company. All regressions include fixed effects for seed funding year and product market. Bootstrapped standard errors are reported in parenthesis. ***, **, and * denote statistical significance at 1%, 5% and 10% levels respectively. All variables are defined in Appendix B.

	Ln(1+Price for 1% of Shares)				
	(1)	(2)	(3)	(4)	(5)
Social Connections					
Same School	0.176			0.021	
	(0.318)			(0.332)	
Same Employer		0.219		0.141	
		(0.227)		(0.237)	
Same Ethnic Minority			-0.124	-0.104	
			(0.195)	(0.197)	
Same School × Employer × Ethnic Minority				1.017	
				(1.281)	
Connected Angel-Founder					0.196
					(0.181)
Startup Characteristics					
Ln(Age at Seed)	0.179	0.138	0.164	0.141	0.222
	(0.220)	(0.222)	(0.216)	(0.221)	(0.217)
Serial Entrepreneur	0.073	0.033	0.054	0.067	0.004
	(0.180)	(0.185)	(0.184)	(0.189)	(0.180)
Ln(Traction)	0.002	0.001	0.018	0.008	0.015
	(0.105)	(0.105)	(0.103)	(0.104)	(0.103)
Top School: Founder	0.389	0.426	0.519*	0.457	0.578*
	(0.317)	(0.296)	(0.292)	(0.317)	(0.294)
Top Employer: Founder	0.053	0.007	-0.033	-0.053	0.054
	(0.233)	(0.236)	(0.231)	(0.236)	(0.229)
Angel Investor Characteristics					
Ln(Degree)	-0.301***	-0.305***	-0.277***	-0.295***	-0.239***
	(0.089)	(0.087)	(0.084)	(0.089)	(0.086)
Entrepreneur-Investor	-0.221	-0.240	-0.139	-0.175	-0.050
	(0.177)	(0.175)	(0.166)	(0.181)	(0.172)
Seed Success Ratio	-0.159	-0.161	-0.171	-0.170	-0.171
	(0.249)	(0.248)	(0.245)	(0.246)	(0.244)
Top School: Angel	-0.272	-0.191	-0.324	-0.400	-0.293
	(0.421)	(0.413)	(0.408)	(0.438)	(0.406)
Top Employer: Angel	-0.108	-0.138	-0.002	-0.040	-0.017
	(0.269)	(0.270)	(0.267)	(0.271)	(0.266)
Obs.	747	747	747	747	747
Adj. R^2	0.088	0.091	0.117	0.109	0.118

Table IA.4 Effect of Social Connections on Matching and Seed-stage Outcomes

This table repeats the main analysis of this paper using a nearest-neighbor matched sample. Each angel-startup treated observation is matched with at least two control pairs based on the startup's age, traction, angel's degree and success ratio. All regressions include seed year, location and product market fixed effects. ***, **, and * denote statistical significance at 1%, 5% and 10% levels respectively. All variables are defined in Appendix B.

	(1) Investment	(2) Seed Success	(3) Ln(Series A Funds)	(4) Ln(Time to Series A)	(5) Connected Investor	(6) VC in Series A
Same School	0.013***	0.059**	0.101	0.093*	0.116	0.061*
	(0.001)	(0.029)	(0.063)	(0.055)	(0.085)	(0.035)
Same Employer	0.153***	0.085***	0.196***	0.154***	0.168*	0.089**
	(0.001)	(0.025)	(0.074)	(0.056)	(0.088)	(0.040)
Same Ethnic Minority	0.001***	0.020	0.081	-0.057	0.070	0.038
	(0.000)	(0.021)	(0.062)	(0.065)	(0.093)	(0.042)
Obs.	67664	3954	926	926	926	926
Adj. R^2	0.216	0.123	0.166	0.201	0.097	0.063

Table IA.5 Effect of Social Connections on Startups Funded through AngelList

This table repeats the main analysis of this paper using only those startups that raised funding through AngelList platform. All regressions include product market fixed effects. ***, **, and * denote statistical significance at 1%, 5% and 10% levels respectively. All variables are defined in Appendix B.

	(1) Investment	(2) Seed Success	(3) Ln(Series A Funds)	(4) Ln(Time to Series A)	(5) Connected Investor	(6) VC in Series A
Same School	0.036***	0.061	0.151*	0.174	0.115	0.069
	(0.004)	(0.047)	(0.084)	(0.197)	(0.105)	(0.060)
Same Employer	0.229***	0.261**	0.172***	0.304*	0.171	0.106*
	(0.002)	(0.117)	(0.055)	(0.169)	(0.109)	(0.061)
Same Ethnic Minority	0.003***	0.025	0.065	-0.153	0.107	0.044
	(0.001)	(0.068)	(0.072)	(0.135)	(0.103)	(0.056)
Obs.	202844	1007	297	297	297	297
Adj. R^2	0.262	0.117	0.088	0.111	0.048	0.107

Table IA.6 Effect on Angel-Startup Matching and Seed-stage Success: Large vs. Small Markets

This table reports regressions that test the effect of social connections between a angel-startup dyad on the likelihood investment happening and seed-stage success. In these tests, I divide the sample into high and low activity regions based on the total number of startups founded in each state between 1990 to 2015. The dependent variable in the first four columns is *Investment*, which is a dummy variable equal to one if the investor made an investment in a startups (actual pairs) and zero otherwise (counter-factual pairs). The dependent variable in the last four columns is *Seed-stage Success*, a dummy variable that is equal to one if the startup raised series A funds. The results tabulated below are the same regressions in column (1) of tables 4, 5, 7 and 8 on each subsample. All regressions include seed year, location and product market fixed effects. Robust standard errors clustered at industry level are reported in parenthesis. ***, **, and * denote statistical significance at 1%, 5% and 10% levels respectively. All variables are defined in Appendix B.

	Investment				Seed-stage Success			
	Large Markets		Small Markets		Large Markets		Small Markets	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Same School	0.020***	0.004***	0.041***	0.008*	0.098**	0.075*	0.093**	0.081*
	(0.001)	(0.001)	(0.009)	(0.005)	(0.043)	(0.041)	(0.045)	(0.044)
Same Employer	0.169***	0.099***	0.323***	0.188***	0.093***	0.070*	0.101***	0.075*
	(0.001)	(0.001)	(0.004)	(0.004)	(0.030)	(0.036)	(0.033)	(0.041)
Same Ethnic Minority	0.004***	0.003***	0.012***	0.007***	0.027	0.027	0.022	0.026
	(0.000)	(0.000)	(0.002)	(0.002)	(0.017)	(0.021)	(0.023)	(0.026)
Same School × Employer		0.131***		0.295***		0.060*		0.066*
		(0.004)		(0.021)		(0.034)		(0.036)
Same School × Ethnic Minority		0.009***		0.022		0.034		0.045
		(0.002)		(0.037)		(0.041)		(0.046)
Same Employer × Ethnic Minority		0.102***		0.197***		0.070		0.079
		(0.002)		(0.009)		(0.071)		(0.073)
Same School × Employer × Ethnic Minority		0.016**		0.110**		0.118*		0.128*
		(0.007)		(0.049)		(0.068)		(0.083)
Obs.	1330478	1330478	1065173	1065173	4818	4818	4578	4578
Adj. R^2	0.122	0.136	0.410	0.412	0.160	0.160	0.095	0.094
Location & Prod. Market F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Table IA.7 Effect of Angel and Founder Social Connections on Seed-stage Success: Instrumental Variables

This table reports results of an instrumental variable approach that attempts to isolate the influence effect of socially connected angels on startups from the selection effect. I run variations of the following specification:

$$
\begin{aligned}
1^{st} stage: \quad ConnectedAngelStartup_{i,j} &= \alpha_0 + \alpha_1 AngelProfileOnCB_i + \alpha_2 StartupProfileOnCB_j + \alpha_A A_i + \alpha_S S_j + \mu_t + \mu_{ind} + \mu_{loc} + \epsilon_i j \\
2^{nd} stage: \quad Y_j &= \beta_0 + \beta_1 Connected\widehat{Angel}Startup_{i,j} + \beta_A A_i + \beta_S S_j + \eta_i + \eta_t + \eta_{ind} + \eta_{loc} + u_j
\end{aligned}
$$

The sample period is limited to 2008-2015 since Crunchbase was started only in July 2007. The two instruments are: *Startup Profile on* CB_j, which indicates whether startup *'j'* had a profile in Crunchbase before its first seed funding round, and *Angel Profile on* $CB_{i,j}$, which indicates whether angel *'i'* had a profile on Crunchbase before startup *'j'* raised its first seed round. *Connected Angel-Startup*$_i, j$ indicates whether angel *'i'* and startup *'j'* are socially connected. All regressions include founding year, location and product market fixed effects. Robust standard errors clustered at industry level are reported in parenthesis. ***, **, and * denote statistical significance at 1%, 5% and 10% levels respectively. All variables are defined in Appendix B.

	OLS	IV: First stage	IV: Second stage				
	(1) Seed Success	(2) Connected Angel-Founder	(3) Seed Success	(4) Ln(Series A Funds)	(5) Ln(Time to Series A)	(6) Connected Investor	(7) VC in Series A
Connected Angel-Startup	0.095***		0.141***	0.134***	0.151*	0.176**	0.172*
	(0.017)		(0.043)	(0.021)	(0.081)	(0.072)	(0.099)
Ln(Traction)	0.041***	0.039***	0.068***	-0.005	0.032**	0.010	0.061***
	(0.008)	(0.009)	(0.008)	(0.031)	(0.016)	(0.018)	(0.020)
Ln(Seed Funds)	0.067***	0.047***	0.049***	0.640***	0.155***	-0.048	-0.071**
	(0.014)	(0.016)	(0.015)	(0.048)	(0.031)	(0.030)	(0.034)
Ln(Degree)	0.011*	0.064***	-0.003	0.021	-0.024**	0.122***	0.007
	(0.006)	(0.007)	(0.007)	(0.024)	(0.011)	(0.016)	(0.014)
Seed Success Ratio	0.227***	0.003	0.248***	-0.114	-0.385***	0.033	-0.046
	(0.026)	(0.030)	(0.032)	(0.089)	(0.051)	(0.047)	(0.053)
Angel on CB Before Seed		0.068***					
		(0.020)					
Startup on CB Before Seed		0.049***					
		(0.021)					
Obs.	5793	5793	5793	1167	1167	1167	1167
Adj./ Pseudo R^2	0.170	0.152	0.090	0.186	0.151	0.102	0.082
Location, Prod. Market, Yr. F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Figure IA.1 Social Connections in Big Data Market

The figure below depicts the social connections between entrepreneurs and investors in the Big Data market in the year 2013. The black and gray nodes represent the startups and investors respectively. A red edge implies that at least one of the startup's founder and the investor share a social connection: (i) both went to the same school, (ii) both worked for the same employer in the past, or (iii) both belong to the same ethnic minority.

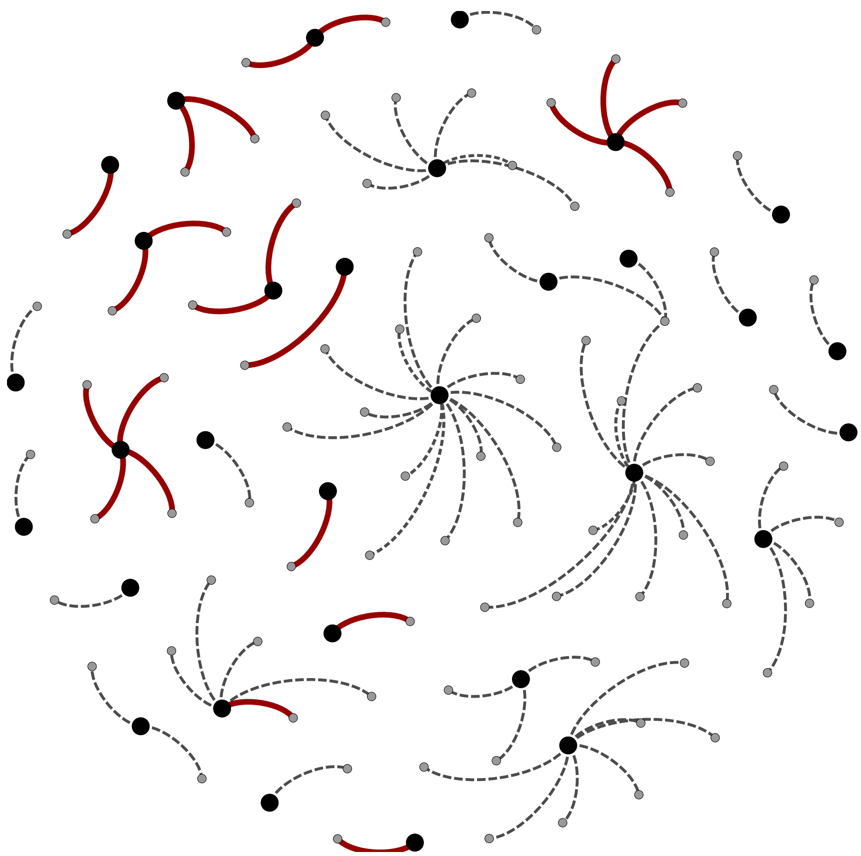
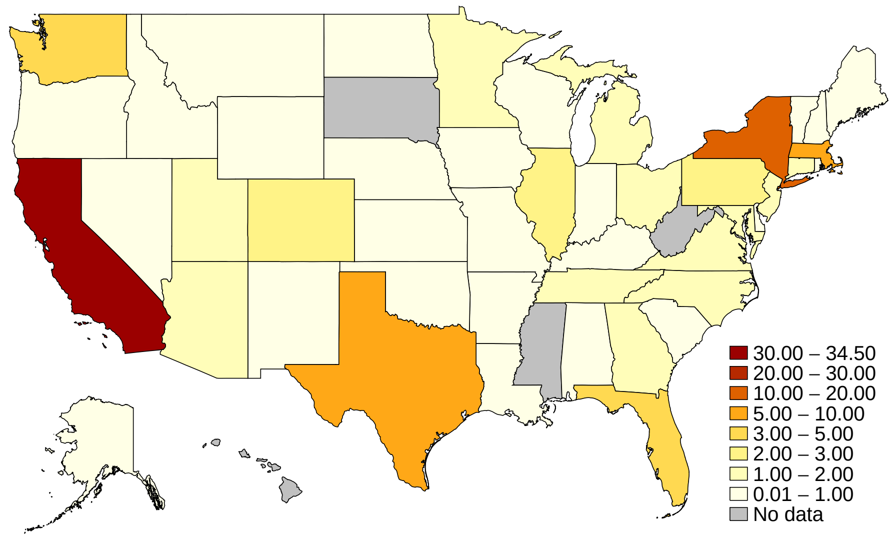


Figure IA.2 Distribution of Startups

This figure shows the proportion of startups in each state. The figure in panel A uses the full Crunchbase dataset to map the percentage of startups located in each state. The figure in panel B was mapped using the 9,396 startups in our sample.

Panel A: Crunchbase Data



Panel B: 9,396 Startups in the paper