Who Finances Disparate Startups?

S. Katie Moon Paula Suh*

April 29, 2022

(First draft: April 23, 2021)

Abstract

Dozens of mid-sized U.S. cities are fostering startups as regional technology hubs. Using detailed early-stage firm information from Crunchbase, we show such a diminishing industrial agglomeration trend driven by the angel financing. This trend is tied to angel investors' unique portfolio selection of startups that diverges from venture capitals' approach. Specifically, angel investors make geographically concentrated investments with industry diversification, while venture capital investors make industry-concentrated investments with relatively greater geographic diversification. We also show that angel investors' portfolio selection of disparate startups enhances their average portfolio firm performance and plays an important economic role in forming the local entrepreneurial ecosystem.

Keywords: Startup Geographic Location, Startup Business Similarity, Angel Investment, Venture Capital Investment

JEL Classification: M13, L26

---

*Moon is at University of Colorado at Boulder, Leeds School of Business, Campus Box 419, Boulder, CO 80309, USA, [katie.moon@colorado.edu]. Suh is at the University of Georgia, Terry College of Business, 620 S. Lumpkin St., Athens, GA, 30602, [paula.suh@uga.edu]. The authors are grateful to seminar participants at the University of Colorado, University of Georgia, FOM Research Group Virtual Corporate Finance and conference participants at the FMA 2021 meeting and the 2021 Colorado Finance Summit, as well as Hadiye Aslan (discussant), John Barrios, Andrea Buffa, Susan Chaplinsky, Jack He, Yael Hochberg, Pekka Honkanen, Steven Malliaris, Gordon Phillips, Shrihari Santosh, Ilya Strebulaev, Sheridan Titman, Yuri Tserlukevich, Xinxin Wang (discussant), Malcolm Wardlaw, and Xingtan Zhang. An earlier version of this paper was circulated under the title "Startup (Dis)similarity and Types of Early-stage Financing."

1 Introduction

Recent agglomeration literature has greatly expanded into entrepreneurial agglomeration based on the Marshallian spillovers.[1] Earlier studies highlight that the entrepreneurial agglomeration is highly localized and benefits from common input sharing, quality of matching in local labor markets, and knowledge spillovers (Jaffe, Trajtenberg, and Henderson, 1993; Rosenthal and Strange, 2008; Kolympiris, Kalaitzandonakes, and Miller, 2011). Entrepreneurial finance is also thought to play an important role in the agglomeration of innovation and entrepreneurship as spatial proximity helps with screening of ventures, monitoring and advising of portfolio firms, and thus mitigating information asymmetry and moral hazard (Hochberg, Ljungqvist, and Lu, 2007, 2010; Chen, Gompers, Kovner, and Lerner, 2010). These agglomeration benefits still strongly persist, such that Silicon Valley has continuously been the nation's leading technology hub. However, we are seeing a gradual declustering of firms especially in high-tech industries with many startups.[2]

The recent declustering of startups is in part driven by high-tech firms' decreased reliance on the proximity to physical resources. For example, Silicon Valley's own products and services, such as cloud computing, video-conferencing, and online collaboration, have made it possible to exploit the traditional benefits of geographic knowledge spillover possible in ventures anywhere, where living costs and competition are lower (The Economist 2018). More importantly, the difficulty of accessing traditional financing and, to some extent, venture capital (VC) financing as VCs tend to locate near crowded tech hubs, raises relative benefits of launching business in areas where other early-stage financing might be accessible.

In this paper, we empirically document whether entrepreneurship is geographically declustering and examine how the new startup dynamics affect entrepreneurial location choice, access to early-stage financing, and success of firms as a result. To this end, we offer an al-

[1]See for example, Glaeser, Kerr, and Ponzetto (2010); Carlino and Kerr (2015); Chatterji, Glaeser, and Kerr (2014) among others.

[2]The Economist (September 1, 2018), "Silicon Valley is changing, and its lead over other tech hubs narrowing."

ternative perspective on a trade-off between agglomeration and access-to-financing benefits. In doing so, we highlight the rising role of angel investors, who are geographically spread over the nation, in paving the way for the geographic dispersion and industry declustering of new startups.

We use the startup firm data from Crunchbase, the leading platform for superb information on early-stage firms for the sample period from 2007 to 2018. The information provided by the database includes founding years, geographic locations, detailed industry classifications, early-stage financing rounds, and founder and team members. The data is gathered from investment firms that submit monthly portfolio updates to Crunchbase, community contributors, artificial intelligence and machine learning of the online information, and Crunchbase's own data team. The Crunchbase database not only provides the most extensive data on startup firms with a greater focus on high-tech sectors but also has the distinct advantage of identifying their business characteristics in great detail over other databases.

Particularly, Crunchbase tags firms with multiple industry classifications composed of 47 broad industry groups and 742 detailed industries. Using these detailed industry classifications, we create an industry vector of 742 elements for each firm and compare those industry vectors across firms within a Metropolitan Statistical Area (MSA). To examine industry declustering in an MSA, we develop a cosine-similarity measure of these industry vectors. It is important to note that the similarity measure is time varying as the geographic peers change over time due to the new entries of firms. This is an important component of externality in this measure that later serves as an identification strategy for casual interpretations.

We first document two stylized facts on similarity. First, we find that startup firms geographically spread throughout the country over time. Although a few big MSAs, such as San Francisco, New York, and Los Angeles, are still leading in terms of the total numbers of firms and new startup foundations, we note that other MSAs, such as Boston, Chicago, Washington DC, Miami, Austin, Dallas, Atlanta, and Boulder, have emerged as new startup hubs.[3] Second, we show that MSA similarity scores have decreased significantly over time. These results overall are consistent with the interpretation that industry clustering that is first introduced by Porter (1990) began to retrogress more recently. The literature thus far highlights the unrealistic cost of living and the heightened competition for talent (Kerr and Robert-Nicoud, 2020) in the areas of severe business clustering as possible reasons for the industry declustering. In this paper, we attempt to examine other driving forces by focusing on the financing channel.

We further examine how a firm's similarity to potential peers in a given MSA affects the firm's location choice. We find that in general a firm is more likely to enter into a market when the potential peers of the market is similar to itself based on our similarity measure. This overall effect is consistent with industry clustering. However, the effect has been reversed over time and becomes negative in the more recent period. We also find that the negative impact of potential similarity of a market on the firm entry decision to the market is particularly stronger when the market has a stronger angel-investor presence than any other early-stage investors including VC.

Next, we examine how a firm's business similarity to geographic peers is associated with the type of early-stage financing. We expect the startup similarity to geographic peers to be associated with angel financing relative to VC financing, based on our entry model results above. Our main analyses show that startup firms are more likely to receive angel financing when they are dissimilar to their geographic peers. A one standard deviation *decrease* in a firm's similarity to its geographic peers is related to a 27.9% increase in the unconditional likelihood of receiving angel financing in the subsequent year. In contrast, a one standard deviation *increase* of the similarity relates to a 16.9% increase in the unconditional likelihood of receiving VC funding in the subsequent year. We take into account that angel financing and VC financing can be sequential for some firms and thus consider only the first funding

[3]We find our list of MSAs with a high startup concentration very similar and consistent with that of Kerr and Robert-Nicoud (2020).

rounds of all firms or subsets of firms that exclusively receive angel fundings or VC fundings only. We find the results are robust to using these refined samples. Also, these effects are likely causal because changes in our measure of similarity to geographic peers are exogenously given beyond the firm's initial entry decision depending on the industry characteristics of newly entered of firms each year.

We propose that the relation between business similarity to geographic peers and types of funding can be attributed to the mechanism based on unique portfolio preferences of angel investors. Unlike VC investors who specialize in a small number of industries and use various control mechanisms to reduce portfolio risks, angel investors' small scale and diverse expertise and experiences may raise coordination frictions between them and their portfolio firms. Therefore, we predict that angel investors are more likely to create portfolio by investing in geographically close firms to facilitate interaction with them but by diversifying over portfolio-firm industries within the close proximity.

To test this prediction, we examine investor-level portfolios using the same Crunchbase data. We find that portfolio preferences are significantly different between angel investors and VC investors. More specifically, angel investors' portfolios are associated with greater industry dissimilarity but smaller geographic distance. On the other hand, VC investors maintain their portfolios under higher business similarity but are less likely to restrict their portfolio firms based on geographic distance. More importantly, we find that angels with funded firms in close proximity but diverse in industry also tend to have higher success rates as measured by exits through IPOs or acquisitions. Notably, this relationship is not observed for any other types of investors. This final result suggests that a regional business declustering of firms allow angel investors to form an optimal portfolio, which in turn, improves the success of startups funded by angels.

Overall, the results support the idea that firms that are dissimilar to geographic peers are more likely to be considered favorably by angel investors. This mechanism may incentivize new entrepreneurs to start business in a market with less industry clustering especially when angel investors' presence is greater in the market. Considering a significant increase in the number of angel investors (Figure 1), who are widespread throughout the country, we ascribe our evidence of geographic dispersion and industry declustring of new startups to spreading angel financing.

Our paper relates to the three areas in the finance and economics literature. First, our paper adds to the entrepreneurial agglomeration literature. The recent agglomeration literature expands on the intellectual spillover (Marshall, 1920) and shows that entrepreneurial activities are much more localized than other economic forces linked to agglomeration (Rosenthal and Strange, 2008; Glaeser et al., 2010; Kolympiris et al., 2011; Carlino and Kerr, 2015; Chatterji et al., 2014). Particularly, VC has been extensively studied as an important factor of entrepreneurial agglomeration as VC tends to locate closely to tech hubs and attract new startups in proximity (Chen et al., 2010; Hochberg et al., 2007, 2010). We complement the earlier studies and expand the literature by documenting a gradual declustering of startup firms over time while the traditional agglomeration benefits still exist. We link this phenomenon with the emergence of angel investors as the rising source of early-stage financing, which attracts startups away from more expensive and competitive tech hubs and trades off the traditional agglomeration benefits with funding opportunities.

Second, our paper is closely related to studies on angel investors. Studies on angel investors are relatively underdeveloped due to the data limitations on angel financing, compared to the extensive work on VC financing (Hellmann, Schure, and Vo, 2021; Lindsey and Stein, 2019). The previous studies on angel investors focus on the real effects of angel financing and success of angel-financed startups. Lindsey and Stein (2019) show that a reduction in a market's pool of angel investors is negatively associated with firm entry into the market and local employment. In contrast, Denes, Howell, Mezzanotti, Wang, and Xu (2020) find that angel investor tax credits have no significant effect on entrepreneurial activity. Our study may reconcile the gap between the two studies to some extent by showing that the effectiveness of angel financing, in part, depends on the startup clustering dynamics and similarity of firms in the area. Furthermore, our paper complements earlier studies of angel investors by showing that the local startup dynamics and similarity are important factors in angels' funding decisions (Bernstein, Korteweg, and Laws, 2017) and successful exits (Kerr, Lerner, and Schoar, 2014).

Lastly, our paper broadly connects to the literature on early-stage financing and financial contracting. We consider the preferences of angel investors in creating and diversifying their investment portfolios and show that the industry diversification motive of angel investors is closely related to business declustring. These investment preferences of angel investors that are summarized as geographic concentration and industry diversification are contrary to those of VC investors who prefer industry specialization (Sorenson and Stuart (2001); Hochberg et al. (2007); Chen, Gompers, Kovner, and Lerner (2009)). The difference in investment preferences highlights potential limitations of angel investors' financial contracts, where the use of risk-sharing mechanisms in VC contracts, such as syndication and staged financing (Lerner (1994); Lerner (1995); Kaplan and Stromberg (2001); Cornellii and Yosha (2003)) is limited.

Our paper proceeds as follows. In Section 2, we discuss our data and sample selection. Section 3 discusses our measure of business similarity to geographic peers and presents the results on the business similarity and firm entry decisions. Section 4 examines the relation between the business similarity and types of early-stage financing. Section 5 presents the results on entrepreneurial outcomes. Section 6 concludes.

2 Data and Sample

2.1 Crunchbase Data

We collect our main data on startup firms and their funding information from Crunchbase. Crunchbase is a crowdsource platform started in 2007 and provides information about private and public companies and investors. Its parent company is TechCrunch which is an online newspaper focusing on high-tech and startup companies. TechCrunch originally operated Crunchbase as its online encyclopedia for startup company information. Although the initial data relied on web searching and scraping, now Crunchbase collects data directly from its wide network of partners covering companies, executives, entrepreneurs, and investors. Furthermore, Crunchbase maintains high quality data through machine learning algorithms to validate data for accuracy and anomalies daily and ensures capturing notable funding rounds, acquisitions, and exits through following over 2,000 of the top news publications. Lastly, Crunchbase covers the most comprehensive set of startup firms, especially high-tech startups. However, the coverage is not just restricted to the well-known innovation hubs in California as shown in the Appendix Table A1.

Crunchbase data has been used by academics in finance for research in early-stage financing. Kaplan and Lerner (2016) introduce Crunchbase as the best known source of venture capital financing. Further, the use of the data has expanded since covering details of all types of early-stage financing beyond venture capital financing. Davis, Morse, and Wang (2020) use detailed startup financing types, venture debt in particular, from Crunchbase to study startup firms' capital structure decisions and successes. Ewens, Nanda, and Rhodes-Kropf (2018) use the startup team member characteristics for examining VC portfolio strategies. Wang (2018) uses the Crunchbase data for tracking entrepreneurship decisions. In this paper, we further make extensive use of Crunchbase for the detailed industry classifications to create our unique business similarity measure. We also benefit from the detailed location information of firms and investors and funding information by different investor types from Crunchbase.

We access the snapshot of information on 753,938 organizations on Crunchbase as of April, 2019 with unique Crunchbase identifiers (uuid). The organizations are further categorized into company, investor, and school. We keep organizations whose primary role is "company." Each company receives a unique Crunchbase identifier (uuid), and all information such as investors and funding rounds is categorized into a node that also receives a unique identifier. We utilize these identifiers to connect firms to their funding investors and rounds information. Besides, Crunchbase also provide detailed information on exits including IPOs and acquisitions and founder and team member information.

Despite the fact that Crunchbase provides the most comprehensive data of early-stage startup companies, one may be concerned with the possibility that firms on Crunchbase (i.e., self-select to be listed on Crunchbase) are systematically different. However, for this selection issue to explain our main results, the bias has to be correlated both with the business similarity and financing choice. This seems unlikely. First, the similarity measure is a dynamic measure affected not only by the firm's choice but also by the neighboring firms' choices. Hence, a firm's choice to be on Crunchbase does not seem to correlate with our similarity measure in any clear way. Second, being on Crunchbase does not guarantee an access to financing or any specific type of financing. We find that 18% of firms in our sample have at least one funding record, and the funding type covers a wide spectrum from seed money to post-IPO debt financing.

2.2 Sample Selection

We restrict our sample to firms that have non-missing headquarter location, valid zip code, and founding year. For our analyses, we also limit the age of a startup firm up to ten years since its founding year.[4] It is important to note that Crunchbase includes not only startups but also public firms. We consider those public firms as geographic peers to startup firms and do not exclude them based on the age cutoff when calculating the business similarity. We exclude observations from the sample after firms close their business or exit through an IPO. After all these screens, our final sample comes down to 119,605 unique firms founded between 2007 and 2018.[5] We extend the data with investor and funding information, which

[4]Some private firms in our sample are very old. For example, Scovill Fasteners was founded in 1802, and thus skews the age distribution to the right. Hence, we use the median age at the IPO of 2,067 exiting firms in our sample, which is ten, as the age cutoff.

[5]Although Crunchbase provides information on founding activities even before 1990, we intentionally restrict our sample to the period after 2007 when Crunchbase launched in July of 2007. Our results are

results in about 1.5 million firm-year observations. About 18% of the sample firms have at least one funding record (denoted funding sample) and corresponding funding investor information.

Table 1 describes firm age and funding information for the funding sample. The average (median) age of firms in our full sample or funding sample is 5.5 (6). Firms in the funding sample have on average about two funding records. The average age at which firms receive funding is 3.07, whereas the age at receiving the first funding record is slightly younger at 2.48.

[Insert Table 1 Here]

We then summarize funding data from angel and VC investors separately in detail. A smaller fraction of firms in the funding sample receive funding from angel investors than from VCs (19% vs. 59%). Firms are younger (2.35 vs. 3.37 years-old) at the time they receive angel financing than VC financing. The statistics on funding amounts and geographical distance between funded firms and corresponding investors are consistent with earlier studies (MIT Entrepreneurship Center, 2000). The median size of angel financing is smaller than that of VC financing ($875,000 vs. $4.3 million), whereas the angel investors tend to fund firms that are in significantly closer geographic proximity than do VC investors (88 miles vs. 427 miles).

We further describe angel investors from our sample more in detail in Appendix Table A3. There are 7,125 unique angel investors that have funded firms in our sample. The angel investors are dominated by male investors and make on average 3.69 investments during our sample period. Angel investors hold on average 1.51 academic degrees, and 24% and 5% of them hold MBA and PhD degrees, respectively. About half of the angel investors (i.e., 3,279 out of 7,125) have founded at least one company, and conditional on having entrepreneurial experience, they have founded on average 1.67 companies. Also, angel investors hold on average about two advising roles for entities.

robust to including the data before 2007.

3 Business Similarity

3.1 Business Similarity Measure

One unique feature of Crunchbase is its industry classification of each firm. Since a large proportion of Crunchbase firms are startups, they do not use conventional industry classifications such as SIC or NAICS code. Instead, Crunchbase has created more technology-oriented industry classifications of its own. There are 46 broad classification, which covers anything from Consumer Goods to Information Technology, and total 742 finer sub-classifications. See Appendix A for the 742 industries and their groups. For example, Information Technology further breaks down into Cloud Data Services, Cyber Security, Data Integration, Sales Automation, Video Conferencing, etc.. Thus, Crunchbase industry classifications provide very detailed industry coverage from conventional industries, such as consumer goods, to high-technology industries that better capture wide variety of business within startup companies. Firms can have multiple categories that describe their business at the time of creating company profile and are allowed to update them over time.[6] In our sample, firms on average report three categories (median of 2.73). Only about 10% of firms report more than 5 categories.

Using this big advantage of 742 detailed industry classifications provided by Crunchbase, we create a business similarity measure for each firm-year. We first create a vector of 742 elements for each firm where each element corresponds to one of the 742 industries. Each element receives 1 for the reported category and 0 otherwise. Then, we use these vectors to compute the cosine similarity score of a pair of a given firm and its geographic peer. Suppose that there are N firms in an MSA in a year. A given firm i's industry vector can be represented by a vector of v_i with element h $(1 \leq h \leq 742)$ being one if firm i's industry classifications by Crunchbase include the category h and zero otherwise. Then, we compute

[6]In Appendix Table A2, we provide a list of top 20 popular categories chosen by firms over time. The list seems to ensure that the Crunchbase categories depict industry trends in a timely manner. For example, whereas Software has always been the most popular category, Artificial Intelligence, Blockchain, FinTech, and Machine Learning have become rising categories in most recent year.

the firm-to-firm business similarity using the vectors of v_i and v_j for a pair of firm i and firm j based on the product cosine similarity:

$$Pairwise\ Business\ Similarity_{i,j} = (v_i \cdot v_j). \tag{1}$$

The pairwise business similarity ranges from zero to one. We end up with $N-1$ pairwise business similarity scores for each firm and then take the average of them to denote a firm's similarity to its geographic peers within MSA for a given year. We compute the similarity scores for each startup company from 2007, the year in which Crunchbase was found, but consider all firms founded before 2007 as existing geographic peers. As new firms are added to an MSA the similarity measure evolves over time for both the existing and new firms in the data. The economic interpretation of the similarity variable is that it serves as a proxy for how close a firm is to its geographic peers within MSA in terms of Crunchbase industry classifications; higher values of similarity suggest that the firm clusters closely with its peers within MSA.

We begin by documenting the geographic dispersion of firms and industry declustering of new startups over time. Figure 2 displays a series of heat-maps of both the number of firms on the left column (panels (a), (c), and (e)) and the business similarity on the right column (panels (b), (d), and (f)) by state and over time. First, we find that startup firms become geographically dispersed over time. In 2007, a large number of our sample firms are located in California, Texas, and New York. These states continue to attract new startups in the next decade, but the number of startup firms also starts increasing in Illinois, Florida, and Massachusetts in 2012 and continue to grow in 2018. Second, we also see that the business similarity of firms decrease over time but not necessarily only in the states where the number of firms grows significantly.

[Insert Figure 2 Here]

Further, Figure 3 shows clearly decreasing trends of business similarity in all 416 MSAs and in a few selected MSAs that are known to have extensive startup activities including San Francisco, New York, and Los Angeles. We note that this trend is not just driven by these few selected MSAs but widely observed. Although the numbers of firms have been steadily increasing across the county and in the notable MSA areas, we find that similarity scores move up and down over time, indicating that the increase in the number of firms mechanically drives a decrease in similarity scores.

[Insert Figure 3 Here]

The decreasing business similarity is numerically summarized in Table 2. In the beginning of our sample period, the average (median) MSA-level similarity score is 0.0412 (0.0314). Over the decade, the MSA-level similarity scores decrease to the average (median) score of 0.0372 (0.0297). This is a 9.59 percentage point decrease from the value in 2007. This stylized fact also holds when we expand out our sample period back to 1990 using Crunchbase's backdated firm data. This trend does not seem to be driven mechanically by increasing number of firms over time nor by firms choosing to locate in MSAs with greater number of investors. Overall, the average of similarity scores in the full sample is 0.0384 with the standard deviation of 0.033 from approximately 1.5 million firm-year observations.

[Insert Table 2 Here]

3.2 Similarity and Entry Decision

In this section, we examine firms' entry decisions to a specific MSA and implications of potential business similarity on firms' final choice of location. We expect that firms are likely to consider its business similarity to local incumbent firms at the time in the area as an important factor in choosing a location to launch its business. We thus examine firms' entry decisions by focusing on business similarity and also considering other local dynamics that are known to affect location choices.

To explore firms' entry decisions, we consider a cross-sectional data at the time of the entry year t and compute a firm's similarity to incumbents in each of 416 MSAs in the year prior to the entry. Hence, each firm has 416 cross-sectional observations. Our dependent variable, $Entry_{i,m,t}$, is an indicator, which is equal to one if the MSA m is chosen by entering firm i at entry year t and zero otherwise. We also compute firm i's potential business similarity scores to all incumbent firms in MSA m at entry year t. Then, for a pair of firm i and MSA m, we take the average of $Pairwise\ Business\ Similarity_{i,j}$ defined in Equation (1) where v_i is the industry vector of firm i and v_j is that of an incumbent firm j in MSA m. We denote the average business similarity for pairs of firm i and all incumbent firms in MSA m by $Similarity_{i,m,t}$. Incumbent firm data are lagged one year at time $t-1$ relative to firm i's entry year t. We specifically estimate the following regression:

$$Entry_{i,m,t} = \alpha + \beta Similarity_{i,m,t} + \gamma \Gamma_{m,t-1} + \epsilon_{i,m,t}, \tag{2}$$

where $\Gamma_{m,t}$ is a set of MSA-level control variables in the year prior to firm i's entry year t, including the number of incumbent firms, number of investors, net job creations from Business Dynamics Statistics, and per capita personal income from Bureau of Economic Analysis. We also control for local innovation ecosystem dynamics measures from Andrews, Fazio, Guzman, Liu, and Stern (2019). Depending on specifications, we include MSA and entry-year fixed effects separately or MSA-by-year fixed effects. The results are presented in Table 3. In columns (1) through (3) and (5), we include MSA and entry-year fixed effects separately to examine the effects of MSA factors in firm entry decisions. In columns (4) and (6), we include MSA-by-year fixed effects, and thus all those MSA factors are subsumed. Standard errors are clustered by MSAs to account for time-series correlation within MSAs.

[Insert Table 3 Here]

Column (1) of Table 3 shows that firms are more likely to enter an MSA if its business is more similar to those of incumbent firms in that MSA. The coefficient on *Similarity* is positive and statistically significant at the 1% level. This relationship is economically significant. A one standard deviation increase in similarity (0.037) is associated with an increase in the probability of entry by 0.03 percentage point, which is equivalent to a 12% increase from the unconditional probability of 0.24 percentage points.[7] Other than the similarity measure, we do not find that MSA characteristics of the number of incumbent firms, number of investors, net job creations, and per capita personal income matter for the entry decision. In column (2), we additionally include three innovation ecosystem measures. Entrepreneurship Quality Index (EQI) is the average quality of startups in an area, RECPI is the expected number of successful growth of startups in an area, and Regional Ecosystem Acceleration Index (REAI) is the expected number of growth events, such as an IPO or acquisition at a meaningful valuation. We find that, of the three innovation ecosystem measures, RECPI strongly factors into firms' entry decisions. We note that the coefficient estimate for *Similarity* becomes even slightly greater with the inclusion of innovation ecosystem measures.[8] Overall, in the first two columns of analyses in Table 3, we find that the business similarity, RECPI, and the number of local investors are important factors in firm entry decisions.

Next, in columns (3) through (6), we further investigate whether the business similarity effect on firms' entry decisions change over time by interacting the similarity measure with indicator variables for time periods. The last two columns (5) and (6) additionally control for the founder's degree location to rule out the possibility that the results are driven by founders who simply stay in the area close to where they received their degrees rather than choosing a location.[9] Columns (3) and (5) include MSA and year fixed effects, and columns (4) and (6) include most stringent MSA-by-year fixed effects. The positive effect

[7]The probability of a random choice of one MSA out of the total 416 MSAs is 0.0024. The standard deviation of *Similarity* in this analysis, 0.037, is different from that in Table 2 because we consider all potential similarity scores of a firm to all incumbent firms in both chosen and unchosen MSA locations for this analysis.

[8]The number of observations decreases when local economic variables and innovation ecosystem measures are included as control variables. The former is due to non-matching MSAs, and the latter is due to the fact that the innovation ecosystem data ends in 2016.

[9]Founder degree location information is populated for about a quarter of our sample firms in data. This reduces the number of observations in regressions in columns (4) and (8), which include the founder degree location control.

of similarity shown in the previous results progressively weakens over time, especially in the more recent period from 2015 to 2017. The magnitude of the estimated coefficient for $Period[2015-17] \times Similarity$ (-0.111) is as large as that for $Similarity$ alone (0.129) in column (3), indicating that the positive effect of similarity on firm entry decision is almost muted in the recent years. The coefficients in columns (5) and (6) with founder degree location control are larger in magnitude although the statistical significance weakens slightly.

[Insert Figure 4 Here]

Importantly, we find that the negative significant effects are much stronger in MSAs where the ratio of angel investors to VC investors is greater than the year's median, to the extent that the aggregated effect of similarity becomes negative eventually. The coefficient plot in Figure 4 reinstates this result more precisely by each year. In the beginning of our sample period in 2007, the triple interaction-term coefficient is slightly negative but statistically insignificant. Over time, however, the coefficient becomes more negative and statistically significant starting in 2013. These results are particularly important for our main analysis later where we examine the relation between business similarity and funding types. The fact that dissimilarity plays a significant role in firm entry decisions when the presence of angel investors is relatively strong foreshadows one of our main empirical findings that business dissimilarity enhances the likelihood of angel financing.

3.3 Identification for Angel Investor Presence

In this section, we explore the possibility of angel investors as one of the driving forces of the geographic spread and business declustering of new startups. Based on our results in the previous section, strong presence of angel investors is significantly associated with a decrease in business clustering. Following Denes et al. (2020), who find that angel investor tax credits increase angel investments, we exploit the staggered introductions and expiration of angel tax credit programs in the U.S. states during our sample. The angel tax credit program data are from Denes et al. (2020), and we use specifically states maximum angel tax credit percentages. We repeat the regression in Equation (2) but replace the main variable of interest with the interaction between *Similarity* and either an indicator variable for a state-level angel tax credit program or the maximum angel tax credit percentage in a given year. Table 4 presents the results. Columns (1) to (4) correspond to the regressions where the indicator for angel tax credits ($\mathbb{1}$(Angel Tax)) is used, and columns (5) to (8) correspond to regressions where the continuous variable of maximum angel tax credit percentages is used. Depending on specifications, we include MSA and entry-year fixed effects separately or MSA-by-year fixed effects. Standard errors are clustered by MSAs to account for time-series correlation within MSAs

[Insert Table 4 Here]

We first find in column (1) that a state's introduction of angel tax credit programs magnifies the negative effect of *Similarity* on a firm's entry to an MSA in the state. When the state has an angel tax credit program, a one standard deviation decrease in similarity (0.037) is associated with an increase in the probability of entry by approximately 0.02 percentage point, or a 8.5% increase from the unconditional probability of a random entry. In column (2), the effect becomes much stronger when the presence of angel investors is strong in the MSA. The estimated coefficient of $\mathbb{1}$(Angel Tax) $\times$ High Angel $\times$ *Similarity* translates into a 38.6% increase in the likelihood of an entry with a one standard deviation decrease in similarity where the MSA has both strong presence of angel investors and state-wide angel tax credit programs. Column (3) reinforces the results in column (2) by replacing the separate MSA and year fixed effects with the MSA-by-year fixed effects. Column (4) additionally controls for the founder's degree location to rule out the possibility that the results are driven by founders who simply stay in the area close to where they received their degrees rather than choosing a location.[10] We find the effects in columns (3) and (4) remain

[10]The reduction in the number of observations for columns including the founder degree location is due to missing information.

robust and consistent. We note that angel tax credit itself is not significantly associated with firm entry decisions.

Using a continuous variable of angel tax credit percentages in the next four columns, we find consistent results. Column (5) shows that a one standard deviation decrease in similarity and a 10% increase in state-wide angel tax credit percentages increase the likelihood of a startup entry to the MSA by 2.4%. In column (6), the effect is about four times as large as that in column (5) when the MSA has strong presence of angel investors. We also find that our results are intact when including the MSA-by-year fixed effects and founder degree location control in columns (7) and (8), respectively.

Overall, the results in Sections 3.2 and 3.3 show strong evidence that angel investors significantly affect startup entries to the locations in which they are unique relative to local peers in terms of business similarity. We address the concern that local economic factors (potentially omitted variables) may impact dissimilarity and the angel investors' presence in a specific location simultaneously by using a stringent regression specification with MSA and year fixed effects and also exploiting staggered implementations of angel tax credit polices in a difference-in-differences approach. Importantly, these results serve as motivating economic insights that some special features of angel financing may benefit startup firms that differentiate themselves from local peers–i.e., angel investors' positive role for promoting local entrepreneurship of unique characteristics.

4 Similarity and Financing

4.1 Funding Types

In this section, we examine how business similarity affects startup firms' access to financing by funding investor types. Each investor (whether an organization or a person) in Crunchbase reports one or more investor type. We group the 22 detailed investor types into three groups broadly: angel, venture capital, and others. Angels include fund providers of "preseed", "seed", and "angel" investments, and VC include fund providers of any "series-X" rounds. Others include investment banks, hedge funds, pension funds, private equities, and accelerators.

Figure 5 first provides some insights into the relation between business similarity and the access to differing funding types. Overall, we find from the figures that VC financing is the most frequent financing type for startup firms in our sample. Figure (a) shows that the probability of receiving angel funding is higher when an MSA has lower similarity among firms. Conversely, the probability of receiving VC funding is slightly higher when an MSA has high similarity among firms. Similarly, as shown in Figure (b), the fraction of funding provided by angel investors is also higher in the low-similarity MSA group, whereas the fraction of funding provided by VC investor is higher in the high-similarity MSA group.

Next, we further investigate these findings in regression analyses. Specifically, we consider the following regression specification:

$$Funding_{i,t} = \alpha + \beta Similarity_{i,m,t} + \gamma Ln(firms)_{m,t} + \epsilon_{i,t}, \tag{3}$$

where $Funding_{i,t}$ is one of the dependent variables that we consider including an indicator variable for funding received from one type of investor in a year, the average fraction of funding received from the type of investor weighed by the size of each funding round in a year, the fraction of the total amount of funding received from the type of investor in a year, and the total amount of funding in million dollars received from the type of investor in a year. Similar to the previous analyses, our main variable of interest is $Similarity_{i,m,t}$, which is the average business similarity for pairs of firm i and all incumbent firms in MSA m. The specification includes the log number of firms, G Index and EG Index in MSA m at year t as control variables. A firm's total funding amount received in a year is also controlled when the dependent variable is the total amount of funding in million dollars received from a given type of investor. Firm and year fixed effects are also included, and standard errors are clustered by MSAs. Table 5 present the results. Columns (1) to (3) estimate the regression only using firm-year observations with any reported funding in the Crunchbase data (i.e., conditional on receiving at least one funding), and columns (4) to (6) use our full sample by replacing observations with no funding data from the Crunchbase with zeros.

[Insert Table 5 Here]

In Panel A of Table 5, we focus on the funding provided by angel investors. In column (1), we find that a decrease in business similarity increases a firm's access to angel financing. The coefficient estimate for *Similarity* implies that a one standard deviation (0.033) drop in similarity scores increases the probability of receiving funding from angel investors by 23 percentage points. Column (2) shows that the total amount raised from angel investors as a fraction of all funding amount in a given year increases by about 13.6 percentage points for a one standard deviation decrease in business similarity. We also find in column (3) that the total funding amount from angel investors increases by about $66,000 for a one standard deviation decrease in business similarity. When we consider unconditional funding results in the next four columns, results are consistent although the magnitudes of the effects are smaller. We find that a one standard deviation decrease in similarity scores implies an increase of 1.2 percentage points in the likelihood of angel funding, 0.7 percentage points in the fraction of the angel funding amount in the total funding amount, and $4,884 in the total funding amount. All estimated coefficients of *Similarity* are statistically significant at the 1% level except for column (3).

On the other hand, when we consider VC funding in Panel B of Table 5, we find strikingly opposite results to those with angel financing. That is, a decrease in business similarity *decreases* firms' access to VC financing. In columns (1) and (2), we find that an increase in similarity scores is significantly associated with an increase of VC funding. For example, in column (1), the coefficient estimate for *Similarity* implies that a one-standard deviation increase in similarity scores increases the probability of receiving VC funding by 7.8 percentage points. The results in column (2) shows that the amount of funding raised from VC investors as a fraction of the total funding amount increases by 12.3 percentage points for a one-standard deviation increase in similarity scores. Also in column (3), the total funding amount received from VC investors increases by $5.1 million dollars, when a firm's similarity score to local peers increase by a one standard deviation. However, the unconditional VC funding results in columns (4) to (6) become statistically weaker for most but for the total funding amount result in column (6). The total VC funding amount increases by $227,000 for one-standard deviation increase in similarity scores.

One may be concerned about some unobserved economic shock of an MSA that attracts new startups, resulting in a local composition of more dissimilar firms, and increases the local angel funding opportunities at the same time. However, the essence of our main results is that any external shocks that disturb the business similarity of an MSA including changes in economic conditions *deferentially* affect the types of financing of the firms that become more dissimilar vs. similar in the MSA with the arrivals of such shocks. Therefore, those external shocks rather help us exploit a meaningful variation in the changes in similarity scores across firms within the MSA. On the other hand, one may have a good reason to be concerned about the possibility that the entry of a group of firms with certain firm-level characteristics both attracting different businesses into the same MSA and simultaneously increasing their own likelihood of getting angel fundings relative to VC fundings. Although such potential firm-level characteristic that casts broad impact over the entire MSA is hardly known in this context, we specifically mitigate such concern in Table A4 by excluding all entry year observations and allowing the results to be driven only by subsequent external disturbances in the similarity scores. Our main results remain robust and consistent.

4.2 Robustness

One may be concerned that angel funding and VC funding are sequential for some firms and thus VC funding generally follows by angel funding. To address this concern and focus on the initial funding opportunity sets, we now examine how business similarity affects the type of the first funding by restricting the sample to only the very first funding observations. We repeat the regression in column (1) of Panels A and B of Table 5 with the first-funding sample and report the results in Table 6.

[Insert Table 6 Here]

The dependent variables in Table 6 are all indicator variables for the first funding that is received from a given investor type. The odd columns include the funding year and founding year fixed effects, and the even columns include funded firms' MSA and investors' MSA fixed effects additionally. In the first four columns, we consider all first funding observations, while, in the next four columns, we exclude funding observations that are received from multiple types of investors. In columns (1) and (2), we find that the coefficient estimates for *Similarity* are negative and significant in predicting angel funding as the initial funding of a firm. These results indicate that a lower business similarity score of a firm to its local peers increases the probability that the first funding is received from angel investors. In contrast, columns (3) and (4) show that the coefficient estimates for *Similarity* are positive and significant in predicting initial VC funding. These results are consistent with a higher business similarity score of a firm to its local peers increase the probability that the first funding is received from VC investors. The magnitude of the effect, for example in columns (1) and (3), implies that a one standard deviation increase (0.033) in similarity scores decreases the probability of receiving angel funding as the first funding by about 1.5 percentage point but increases the probability of receiving VC funding as the first funding by about 3.2 percentage points. When we reduce possible noise from the mixed types of investors by excluding first funding observations from multiple investor types, the corresponding results in columns (5) and (7) strengthen to a decrease of 1.7 percentage points in the likelihood of the first angel funding and an increase of 3.5 percentage points in the likelihood of the first VC funding, respectively.

Alternatively, one may be concerned that our main results are driven by the correlation between the similarity and funded firm quality, which affects the demand for certain type of early-stage financing. Specifically, one may be concerned that the low similarity of firms in an MSA translates into low competition in an MSA, therefore indicating a lack of high quality firms that would have chosen to receive funding from VC otherwise. We rule out this possibility in Table 7.

[Insert Table 7 Here]

The dependent variable is EQI, which measures the MSA-level entrepreneurship quality, re-scaled by multiplying the index by 100. We regress the EQI on MSA-level similarity and find that high MSA-level similarity actually lowers the regional entrepreneurship quality, even after controlling for a number of MSA-level economic variables and entrepreneurial indices, such as entrepreneurship cohort potential index and regional ecosystem acceleration index. The coefficients range from -0.054 to -0.083, strengthening with more control variables. The coefficient of -0.083 translates into a 6% decrease in the entrepreneurial quality for a one standard deviation increase in MSA similarity. This result has two implications. First, our main result is not driven by low quality firms choosing to receive angel funding over VC funding. Second, given that the EQI is measured as the probability of observing growth outcome, i.e. IPO success, conditional on firms' characteristics at birth, the negative correlation between MSA-level similarity and EQI is likely to convey fiercer entrepreneurial environment for survival when regional similarity is high.

We also consider two alternative measures of similarity in Table 8, Soft Similarity and Dynamic Similarity. First in Panel A, we replace our primary similarity measure with another similarity score that additionally takes into account of correlations among finer subclassifications within 46 broad industry groups, based on the soft-cosine similarity calculation technique introduced by Sidorov, Gelbukh, Gómez-Adorno, and Pinto (2014). Instead of treating all 742 elements in an industry vector as completely unrelated, we adjust the elements under the same broad industry group to be treated equally more similar.[11] In

[11]The soft-cosine similarity technique is used in machine learning and natural language processing where text sentences are expressed in a non-orthogonal basis. For example, "Hi" and "Hello" are synonyms, and thus the angle between two basis vectors of "Hi" and "Hello" is set to be non-orthogonal.

Panel B, we next consider potential changes of industry vectors over time. It is possible that firms may voluntarily update their industry classifications as their businesses are expanding, contracting, or pivoting. To address this possibility, we use multiple Crunchbase data dumps that have been acquired in different years including 2016, 2018, 2019, and 2020 and create industry vectors that change over time.[12]

[Insert Table 8 Here]

The results in both panels of Table 8 show that our main results on funding types are consistent when we consider correlations among sub-classifications under a broad industry group and potential updates of industry classifications over time. Specifically, the results using the soft-cosine similarity are almost the same to those using our primary similarity measure, although the magnitude of the effect is weaker as one second to third. The results using the dynamic similarity are insignificant for some VC funding measures, while the results for angel funding continue to hold.

Lastly, we consider a broader spatial unit using Combined Statistical Area (CSA)[13] to rule out the possibility that our results are predominantly driven by the smaller MSAs, where angel investors may mechanically appear to provide more funding given their preference for geographically closer funded firms. We re-run our main regressions in Table 5 using subsample of observations belonging to CSAs. In Table A5, we find that our main results are consistent and hold robust to using the broader spatial unit. Collectively, our main results provide strong evidence that angel investors and VC investors have the opposite preferences with respect to business similarity. That is, angel investors prefer firms that stand out from their local peers as unique businesses, while VC investors are more likely to finance firms with similar businesses to those of their local peers.

[12]We thank Ilya Strebulaev for graciously sharing the Crunchbase data dumps in earlier years. When there is no Crunchbase data dump for a specific year, we use the data from the closest prior year (e.g., using the data in the 2016 dump for the 2017 industry vectors). For the years that have no prior-year data, we use the earliest possible data (i.e., using the data in the 2016 dump for years from 2007 to 2015).

[13]Combined Statistical Areas represent groupings of Metropolitan and Micropolitan Statistical Areas (in any combination) and can be characterized as representing larger regions that reflect broader social and economic interactions.

4.3 Mechanism: Investor Portfolio Preferences

Thus far, we show that business dissimilarity facilitates the access to angel financing but reduces the access to VC financing. In this section, we propose economic explanations to the stark contrast between fundings from the two types of investors. Our explanations focus on different organizational characteristics of angel investors and venture capitalists that constitute two very important investors in start-ups and dynamic substitutes for each other (Hellmann et al., 2021).

Angel investors are high net-worth individuals who are accredited investors defined by the Securities and Exchange Commission Rule 501. Hence, different from professional VCs that might be geographically concentrated in specific locations (i.e., their own industry clustering), private angel investors are more likely to be widely dispersed.[14] Also, angel investors tend to invest in companies in close proximity to their home as MIT Entrepreneurship Center (2000) and Kauffman Foundation (2002) show in their reports that most active angel investors investigate opportunities and memberships in their local areas and do not invest in opportunities outside a 1-2 hours driving distance from home. Furthermore, angel investors are more diverse individuals with widely varying investment styles and behaviors, experiences, personality, and networks even among the experienced angel investors and thus tend to consider investment opportunity sets based on their personal expertises (MIT Entrepreneurship Center, 2000).

On the other hand, VCs are characterized by specializing in a particular industry (Sorenson and Stuart (2001), Hochberg et al. (2007), Hochberg et al. (2010)). Therefore, to diversify the portfolio risks coming from the industry concentration, they use a few important control mechanisms, such as convertible securities, syndications, and staging of capital infusion (Gompers, 1995), in addition to contractual provisions (Kaplan and Stromberg, 2001) and monitoring through board seats (Lerner, 1995). Individual angel investors have limitations

[14]Wharton Entrepreneurship and Angel Capital Association (2017) identified that 63% of angel investors were located outside California, New York, and Boston, with a sizable presence in Great Lakes, Southeast, and Mid-Atlantic.

on the full exploitation of these control mechanisms, which makes the large organizational differences from VCs. For example, Wong (2002) finds that there are fewer staged financing, where most consecutive rounds do not have angel investors as follow on investors (especially when the subsequent financing round involves VC participation), that only 0.59 board seats are added in angel investment rounds on average, compared to 1.12 seats for VC rounds, and that contractual provisions, such as first refusal provisions, contingent equity stakes, and full ratcheting protections, are less common in angel fundings.

We also find that angel investors and VCs are significantly different in investment frequencies. Appendix Figure A1 (left panel) shows that during our sample period, the median number of investments made by angel investors is about two investments per every ten year, while VC investors have made more frequent investments at six per every ten year. This indicates that when angel investors make their funding decisions, it is likely that they comprehensively consider the project characteristics including geographic distance and business similarity over the years rather than promptly evaluating the flow of new deals as they come in as VC investors are known to do so.

Based on these stylized facts on angel and VC investors' characteristics, we provide a simple theoretical model in Appendix B that illustrates our proposed mechanism of investors' differential preferences for investment portfolios. In our model, we first assume that the investor is risk-neutral. The search costs of the investor for projects include the term that captures her cost of efforts to overcome the physical distance between herself and an invested firm and the knowledge gap between the industry of her expertise and the industry of the invested firm. In this setup, the investor's relative tolerance for geographical distance and knowledge gap plays a significant role in explaining the trade-off of a search between geographic proximity and business expertise constraints. Based on the prediction of the model, we expect that angel investors with strong preference for local investments opportunities (e.g., within 1-2 hours driving distance from home) will search for a project with higher tolerance toward knowledge gap from dissimilar projects. As such, angel investors tolerate dissimilar startups despite their costs to learn about unfamiliar businesses. Second, we relax our assumption of the risk-neutral investor and allow the investor to be risk averse. In this case, angel investors even more prefer dissimilar projects to lower their portfolio variance, because project payoffs are correlated according to their business similarity. In both cases, we find that dissimilar startup firms are more likely to be added to angel investors' existing portfolios than similar firms, all else equal.

On the other hand, the same incentive to maintain an investment portfolio that consists of geographically concentrated funded firms is significantly weaker for VCs because they specialize in a narrow industry with higher tolerance for geographic distance. As Sorenson and Stuart (2001) show that VCs that build axial positions tend to invest more frequently in spatially distant companies, VCs tend to reach out to targets in relatively distance.

We examine these intuitions in our data and find the results consistent with these underlying channels. We present the results in Table 9. In Panel A of Table 9, we first examine individual investors' portfolio preferences. In columns (1) and (2), we consider business similarity of funded firms in each investor's portfolio. The funded firm business similarity in a portfolio is the average firm business similarity to all other firms in the portfolio. This test is cross-sectional across investors, and thus the sample consists of one observation per investor. If an investor has only one funded firm, we drop the investor from the analysis, as the business similarity with other funded firms can not be computed in this case. In column (1), we consider all three types of investors, angel, VC, and other investors with the other type as a control group. In column (2), we only contrast angel investors and VCs with VCs as a control group.

[**Insert Table 9 Here**]

Column (1) shows that VCs have strong preference to fund companies with similar industry classifications relative to both the angel and other types of investors. In column (2) when we compare angels and VCs, we find that funded firms in angel investors' portfolios are relatively more dissimilar than those in VC portfolios. The coefficient estimate for *Angel* in column (2) indicates that our business similarity score is lower by 1.8 percentage points, or 19.4% from the unconditional mean of portfolio-firm similarity at 0.093.

We repeat analogous tests in columns (3) and (4) by replacing the dependent variable with the average geographical distance between a given investor and the funded firms in its investment portfolio. Both columns (3) and (4) show that angel investors have significantly stronger preference to maintain geographical proximity to their funded firms than that of VCs or other types of investors. For example, the coefficient estimate for *Angel* in column (2) implies that the distance between angel investors and their portfolio firms is about 107 miles closer, or 14.7% shorter compared to the distance between VCs and their portfolio firms.

In Panel B of Table 9, we consider funding-level observations instead to examine business similarity and geographic distance between existing portfolio firms and newly funded firms. Consistent with the results in Panel A, we find in columns (1) and (2) that angel investors strongly prefer dissimilar firms in their investment portfolios relative to VCs and other types of investors. Similarly, we continue to find in columns (3) and (4) that angel investors fund new firms when they are in close proximity to the existing funded firms in their portfolios. For example, in column (4) angel investors choose to invest in a firm that is approximately 39 miles closer to the funded firms in their existing portfolios relative to VC investors.

Surprisingly, the heterogeneity in portfolio preference described and shown in Table 9 is evident within the angel investor group and appears to be a function of investment size in general. In Figure 6, we plot the trade-off between geographic distance and industry diversification by angel investment sizes (represented by color). We find that the preference for geographic proximity over industry specialization decreases monotonically with the angel investment size and increasingly resembles the VC portfolio preferences when the angel investment size grows sufficiently large.[15] This shift in the portfolio preference for industry specialization seems to be consistent with the greater rent extraction and resource realloca-

[15] When the angel investment size is too small (brightest yellow plot), i.e. only two funded firms in a portfolio, the diversification benefit may be limited.

tion efficiency benefits of large investments for more specialized portfolios shown by Fulghieri and Sevilir (2009).

Overall, results in Table 9 show strong evidence that angel investors and VC investors have contrasting portfolio preferences; Angel investors prefer firms that are dissimilar each other in their portfolios but in close proximity geographically, while VC investors seek for industry similarity without focusing on the geographic proximity. This stark contrast in portfolio preferences is likely due to the difference in the extent to which other risk-sharing control mechanisms are feasible or available to them. The evidence in this section supports the conclusion that business similarity plays an important role in determining the types of early-stage financing and reinstates that the easier access to angel financing with the increased dissimilarity may contribute to industry declustering of entrepreneurship across the country.

5 Entrepreneurial Outcome

In this section, we further investigate investment outcomes of angel investors focusing on whether the close-distance and diverse-industry investment strategies lead to more successful exits of their investments. Specifically, we examine the likelihood of a funded firm's exit through an IPO or acquisition and whether an angel investment is followed by a subsequent investment round. We estimate the following regression specification for this analysis:

$$Outcome_j = \alpha + \beta_1 Angel_j + \beta_2 Close\&Dissimilar_j + \beta_3 Angel_j \times Close\&Dissimilar_j + \gamma\Gamma_j + \epsilon_j, \quad (4)$$

where $Outcome_j$ is one of the dependent variables that we consider including an indicator variable for a subsequent funding after investor j makes a funding, the average IPO exit rate among all funded firms by investor j, and the the average successful exit rate including both IPO and acquisitions exist among all funded firms by investor j. Our main variable of interest is $Close\&Dissimilar_j$ which is an indicator variable if investor j's investment portfolio is identified to follow both the close-distance strategy and the diverse-industry strategy. An investor's portfolio is considered to follow the close-distance strategy if the average distance among the funded firms in the investor's portfolio is below the median of all portfolios of investors of the same type. Likewise, an investor's portfolio is considered to follow the diverse-industry strategy if the overall similarity score among the funded firms in the investor's portfolio is below the median of all portfolios of investors of the same type. We then consider the interaction between $Close\&Dissimilar_j$ and the angel investor indicator ($Angel$) to examine whether the close-distance strategy and the diverse-industry strategy only concerns angel investment strategies. Γ_j is a set of control variables at the investor level including the log number of total investments made by investor j, the log of investing years since investor j made her first investment, and the log of the average funding amount made by investor j. Table 10 present the results.

[Insert Table 10 Here]

Columns (1) and (2) report the regression results for subsequent investments. In Column (1), we examine the close-distance strategy and the diverse-industry strategy separately, and in Column (2) we consider the intersection between the two strategies combined as in $Close\&Dissimilar$. First, we note that the angel investor indicator is positive and significant in predicting subsequent funding rounds. This indicates that angel funding is more likely to come before any other funding types, which is discussed in the previous section regarding possible sequential investments between angel and VC. More importantly, we find that the interaction term of $Angel$ with $Dissimilar$ is significantly positive at the 1% level, implying that the diverse-industry strategy has a positive effect on the likelihood of an interim success of a funded firm by an angel manifested as the incidence of the next funding round. The economic effect translates into a 13.0 percentage point increase in the likelihood that a funded firm in the angel investor's portfolio receives the next round of funding. However, we do not find that the close-distance strategy has a significant effect on the likelihood of a subsequent funding. In Column (2), when we consider the close-distance strategy and the diverse-industry strategy together, we find the effect of the combined strategy has a similar effect on the likelihood of a subsequent funding round. The economic magnitude of the combined-strategy effect is estimated at an 11.6 percentage point increase, or 17% from the unconditional mean.

In Columns (3) and (4), we report results from the analogous tests by replacing the dependent variable with the fraction of firms with eventual IPO exits in the investor's portfolio. We note that angel investors in general have significantly lower IPO exit rates relative to other types of investors. Unlike the previous results for the subsequent funding in Column (1), we find each of the close-distance strategy and the diverse-industry strategy has a significantly positive effect on the IPO exit rate of an angel investor's portfolio. The estimated effects are 3.4 percentage point increases in the portfolio's IPO exit rate for both the close-distance strategy and the diverse-industry strategy. When we combine the two strategy in Column (4), the effect becomes an 3.2 percentage point increase in the IPO exit rate for angle investors. We find consistent, if not stronger, results by additionally considering a funded firm's being acquired as a successful outcome in Columns (5) and (6). Angel investors' successful exit rates, both in IPO and acquisition exits, increase by 3.1 and 7.0 percentage points by following the close-distance strategy and the diverse-industry strategy, respectively. When the two strategies are combined in Column (6), the increase in successful exits is estimates at 5.9 percentage points, or 27% from the unconditional mean exit rate.[16]

Also, we note that these strategies, separately or combined, do not increase the likelihood of successful outcomes for non-angel investors, the majority of which is VC investors. Therefore, results in this section show strong evidence that the close-distance and diverse-industry strategy indeed is an optimal investment strategy for angel investors only. Overall, these results are consistent with the interpretation that angel investors by diversifying their portfolios over the dimension of industry classifications (with maintaining close geographical proximity to funded firms) can optimally include the net-present-value positive investments

[16]There may be a concern that investors with only one investment may drive our results. However, our similarity score cannot be computed when there is only one investment, and thus the case falls into $Dissimilar = 0$. We also find consistent results after excluding investors with only one investment.

from their investment opportunity sets. Without forgoing optimal investment opportunities with proper diversification, we find angel investors can significantly improve their successful exit rates.[17]

6 Conclusion

In this paper, we introduce a new measure for startup business similarity using detailed industry classifications available from the Crunchbase data. This measure allows us to examine time-varying and across-region firm similarity and its effects on the types of early-stage financing. We find that increasingly more startup firms tend to enter a regional market when the local peers are more dissimilar to them. This unique trend appears to be associated with local angel investment opportunities as our evidence shows that dissimilar startup firms are more likely to be funded by angel investors.

We propose a possible mechanism for this finding based on different diversification preferences of angel investors. Angel investors make geographically concentrated investments with industry diversification, while venture capital investors make industry-concentrated investments with geographic diversification. Finally, we show that successful exit rates of angel investors significantly increase when they stick to these investment preferences. Collectively, our results offer a new insight that angel finance has an important implications for startup firms' entry decisions.

[17]Consistent with this interpretation, the average number of investments is significantly greater at 3.69 for angel investors with the combine close-distance and diverse-industry strategy, relative to 3.05 for angel investors with no such strategy.

References

Andrews, Raymond, Catherine Fazio, Jorge Guzman, Yupeng Liu, and Scott Stern, 2019, The startup cartography project: Measuring and mapping entrepreneurial ecosystems Working Paper.

Bernstein, Shai, Arthur Korteweg, and Kevin Laws, 2017, Attracting early-stage investors: Evidence from a randomized field experiment, *Journal of Finance* 62, 509–538.

Carlino, Gerald, and William Kerr, 2015, Agglomeration and innovation 5, edited by Gilles Duranton, J. Vernon Henderson, and William Strange, 349–404, Handbook of Regional and Urban Economics.

Chatterji, Aaron, Edward Glaeser, and William Kerr, 2014, Clusters of entrepreneurship and innovation 14, In: Lerner, J., Stern, S. (Eds.), Innovation Policy and the Economy.

Chemmanur, Thomas J, and Zhaohui Chen, 2014, Venture capitalists versus angels: The dynamics of private firm financing contracts, *The Review of Corporate Finance Studies* 3, 39–86.

Chen, Henry, Paul Gompers, Anna Kovner, and Josh Lerner, 2009, Buy local? the geography of successful and unsuccessful venture capital expansion Working paper.

Chen, Henry, Paul Gompers, Anna Kovner, and Josh Lerner, 2010, Buy local? the geography of successful and unsuccessful venture capital expansion, *Journal of Urban Economics* 67, 90–102.

Cornellii, Francesca, and Oved Yosha, 2003, Stage financing and the role of convertible securities, *Review of Economic Studies* 70, 1–32.

Davis, Jesse, Adair Morse, and Xinxin Wang, 2020, The leveraging of silicon valley Working Paper.

Denes, Matthew, Sabrina Howell, Filippo Mezzanotti, Xinxin Wang, and Ting Xu, 2020, Investor tax credits and entrepreneurship: Evidence from us states Working Paper, Carnegie Mellon University.

Ewens, Michael, Ramana Nanda, and Matthew Rhodes-Kropf, 2018, Cost of experimentation and the evolution of venture capital, *Journal of Financial Economics* 128, 422–442.

Fulghieri, Paolo, and Merih Sevilir, 2009, Size and focus of a venture capitalist's portfolio, *The Review of Financial Studies* 22, 4643–4680.

Glaeser, Edward, William Kerr, and Giacomo Ponzetto, 2010, Clusters of entrepreneurship, *Journal of Urban Economics* 67, 150–168.

Gompers, Paul, 1995, Optimal investment, monitoring, and the staging of venture capital, *Journal of Finance* 50, 1461–1489.

Hellmann, Thomas, and Veikko Thiele, 2015, Friends or foes? the interrelationship between angel and venture capital markets, *Journal of Financial Economics* 115, 639–653.

Hellmann, Thomas F, Paul Schure, and Dan Vo, 2021, Angels and venture capitalists: substitutes or complements?, *Journal of Financial Economics* forthcoming.

Hochberg, Yael, Alexander Ljungqvist, and Yang Lu, 2007, Whom you know matters: Venture capital networks and investment performance, *Journal of Finances* 62, 251–302.

Hochberg, Yael, Alexander Ljungqvist, and Yang Lu, 2010, Networking as a barrier to entry and the competitive supply of venture capital, *Journal of Finances* 65, 829–859.

Jaffe, Adam, Manuel Trajtenberg, and Rebecca Henderson, 1993, Geographic localization of knowledge spillovers as evidenced by patent citations, *Quarterly Journal of Economics* 108, 577–598.

Kaplan, Steven, and Josh Lerner, 2016, Venture capital data: Opportunities and challenges NBER Working Paper.

Kaplan, Steven, and Per Stromberg, 2001, Venture capitalists as principals: Contracting, screening, and monitoring, *American Economic Review* 91, 426–430.

Kauffman Foundation, 2002, Business angel investing groups growing in north america Kauffman Foundation.

Kerr, William, Josh Lerner, and Antoinette Schoar, 2014, The consequences of entrepreneurial finance: Evidence from angel financing, *Review of Financial Studies* 27, 20–55.

Kerr, William, and Frederic Robert-Nicoud, 2020, Tech clusters, *NBER Working Paper* 1–26.

Kolympiris, Christos, Nicholas Kalaitzandonakes, and Douglas Miller, 2011, Spatial collocation and venture capital in the us biotechnology industry, *Research Policy* 40, 1188–1199.

Lerner, Joshua, 1994, The syndication of venture capital investments, *Management Science* 23, 1–14.

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

Lindsey, Laura, and Luke Stein, 2019, Angels, entrepreneurship, and employment dynamics: Evidence from investor accreditation rules Working paper.

Marshall, Alfred, 1920, Principles of economics .

MIT Entrepreneurship Center, 2000, Venture support systems project: Angel investors MIT Entrepreneurship Center.

Porter, Michael E, 1990, The competitive advantage of nations, *Harvard business review* 68, 73–93.

Rosenthal, Stuart, and William Strange, 2008, The attenuation of human capital spillovers, *Journal of Urban Economics* 64, 373–389.

Sidorov, Grigori, Alexander Gelbukh, Helena Gómez-Adorno, and David Pinto, 2014, Soft similarity and soft cosine measure: Similarity of features in vector space model, *Computación y Sistemas* 18, 491–504.

Sorenson, Olav, and Toby Stuart, 2001, Syndication networks and the spatial distribution of venture capital investments, *American Journal of Sociology* 106, 1546–1588.

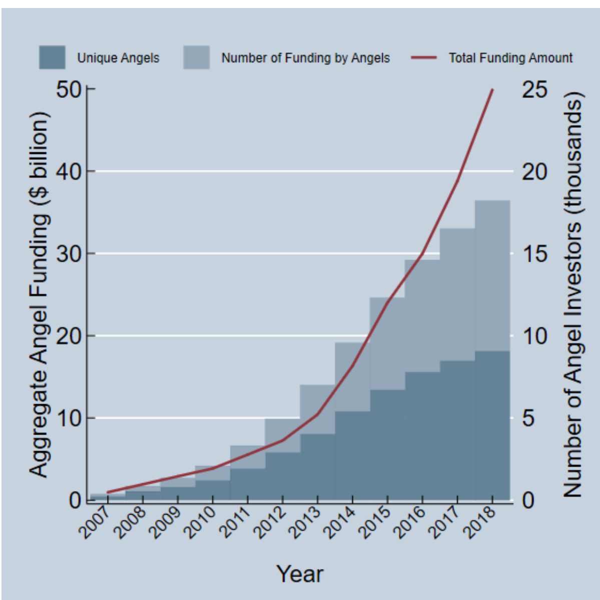
Wang, Xinxin, 2018, Catering innovation: Entrepreneurship and the acquisition market Working Paper.

Wharton Entrepreneurship and Angel Capital Association, 2017, The american angel Wharton Entrepreneurship and Angel Capital Association.

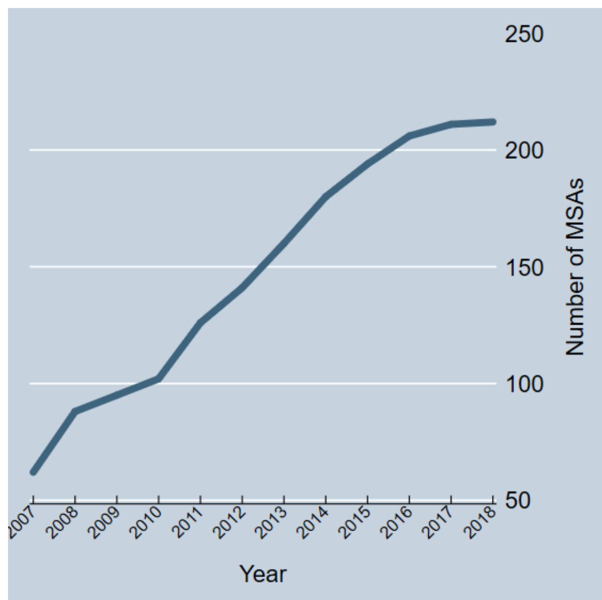
Wong, Andrew, 2002, Angel finance: The other venture capital Working paper.

Figure 1: Growth of Angel Investors and Angel Funding Over Time

The figure shows the cumulative angel funding amount and number of angle investors in Crunchbase over our sample period 2007-2018. In Panel (a), the darker blue bar represents the cumulative number of unique angel investors, and the lighter blue bar represents the cumulative number of funding financed by our sample angel investors. The red line shows the cumulative total dollar amount of angel funding in $ billion. In Panel (b), we show the number of MSAs with at least one Crunchbase angel investor over time. The total number of MSAs is 416.



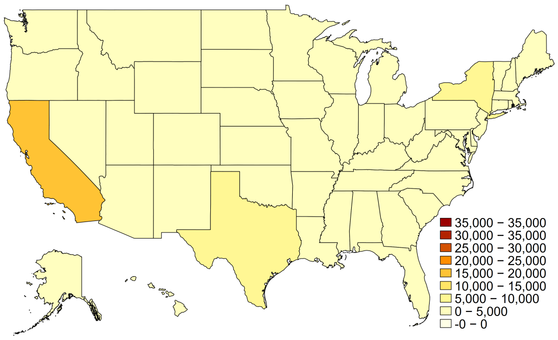
(a) Cumulative Aggregate Funding Amount and Number of Angels



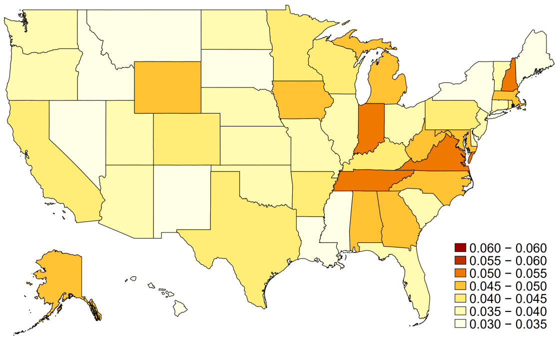
(b) Number of MSAs with Angel Investor Presence

Figure 2: Changes in Firm Concentrations and Similarity by State

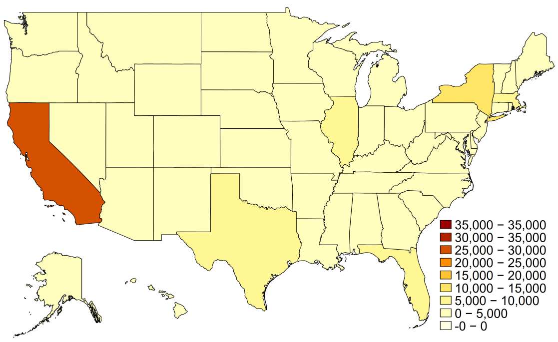
The figure shows changes in the similarity measure and the number of firms by state over the last three decades. Each map in panel (a), (c), and (e) is a snapshot of state-level average number of firms in given year, and each map in panel (b), (d), and (f) is a snapshot of state-level average similarity in given year. For variable definitions and further details of their construction, see Appendix C.



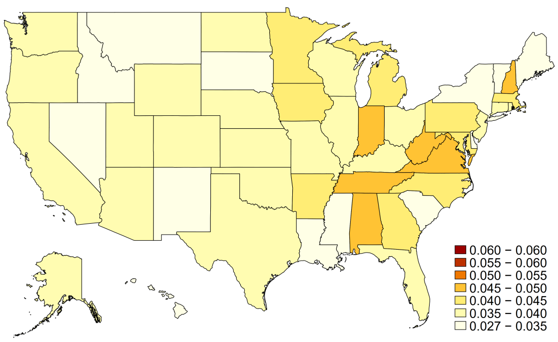
(a) Number of firms, 2007



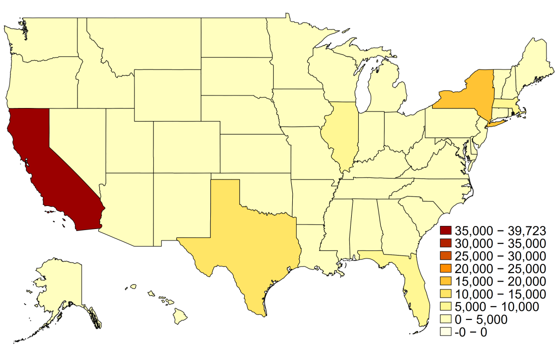
(b) Average similarity, 2007



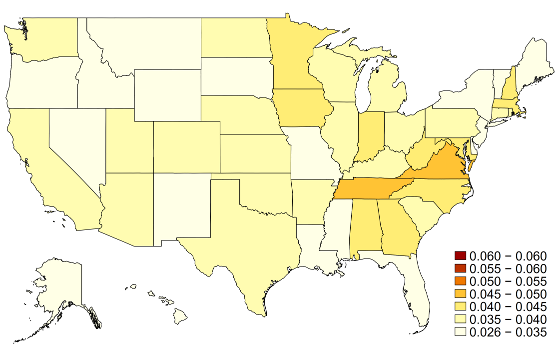
(c) Number of firms, 2012



(d) Average similarity, 2012



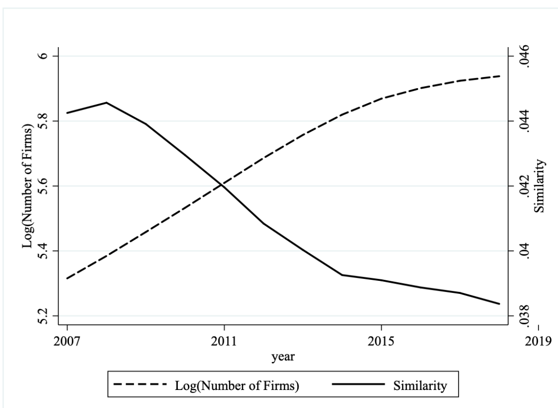
(e) Number of firms, 2018



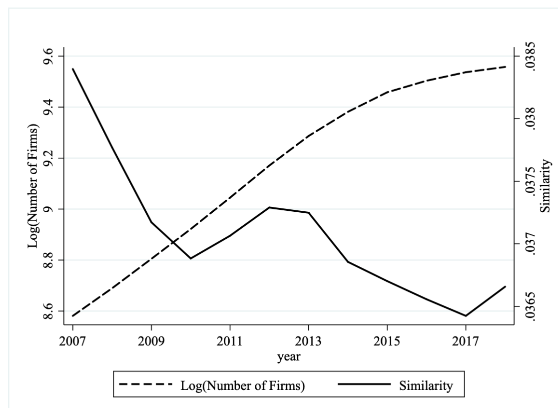
(f) Average similarity, 2018

Figure 3: Similarity over Time

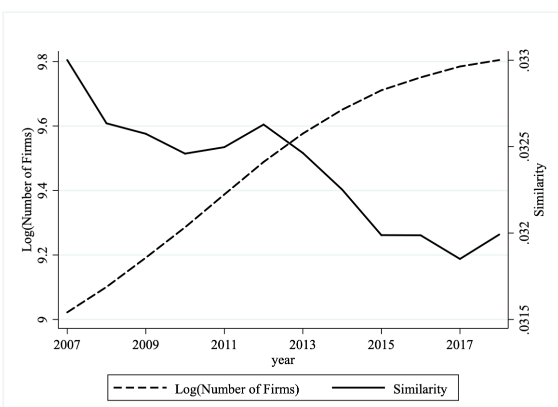
The figure shows time trends of firm similarity and the number of firms for the entire MSAs overall in (a) and San Francisco, New York, Los Angeles MSAs that are the top 3 active areas for startup activities in (b), (c), and (d), respectively. The solid line is our similarity measure. The short dash line is the total number of firms. For variable definitions and further details of their construction, see Appendix C.



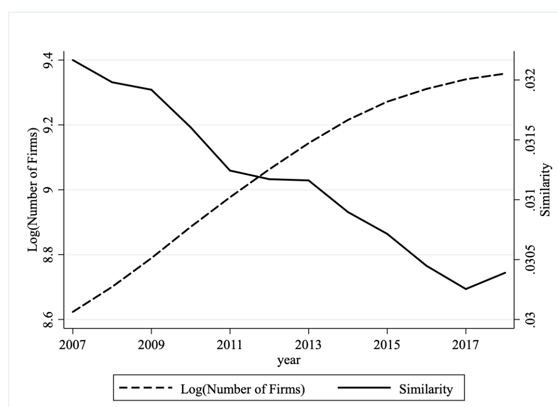
(a) All MSAs



(b) San Francisco MSA



(c) New York MSA



(d) Los Angeles MSA

Figure 4: Similarity, Angel Investor, and Entry, over Time

The figure shows the effects of firm similarity and MSA percentage of angel investors on entry decisions. We consider the regression of firm entry in an MSA on similarity, high_angel, year dummies, and their interactions terms. *Entry* is an indicator variable equal to 1 if a given MSA is an entering location and 0 otherwise. *Similarity* is the similarity measure for all entry firm-MSA pairs, measured using existing firms in the year prior to an entry. *HighAngel* is an indicator variable equal to one if the MSA's percentage of angel investors is greater than the year median and zero otherwise. Each bar represents the magnitude of the coefficient estimates for the triple interaction terms among *year*, *HighAngel*, and *Similarity*. A capped spike shows the 90 percent confidence interval. For variable definitions and further details of their construction, see Appendix C.

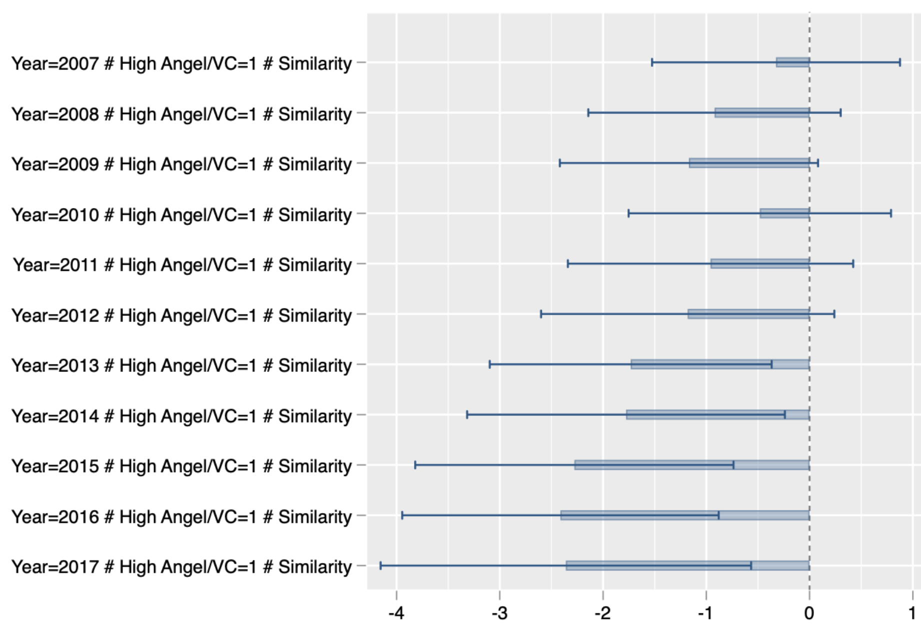
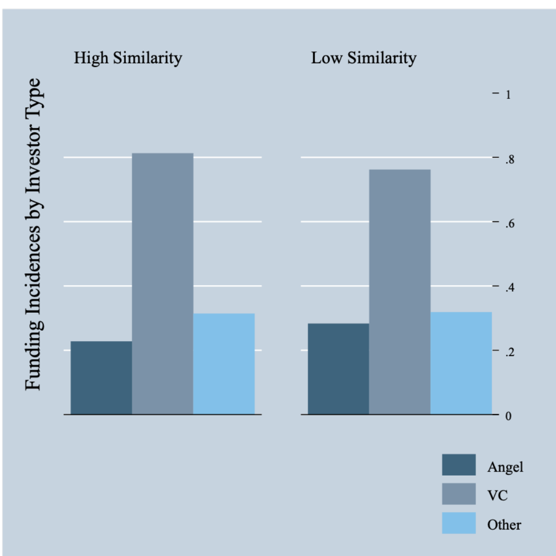
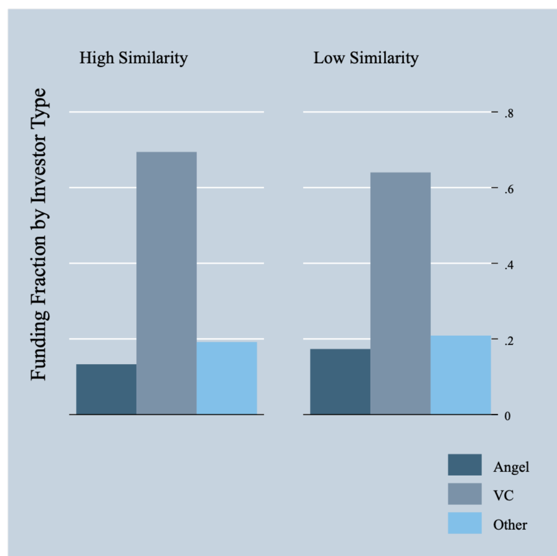


Figure 5: Funding by Investor Types in High vs. Low Similarity MSAs

The figures show the firm-year level funding investor types (Panel (a)) and the average fraction of funding raised from each type of investors (Panel (b)) by the similarity measure. The figures are constructed using firms in our sample that have any funding records. Panel (a) shows that the probability of a firm raising *angel* funding is *higher* in low similarity MSAs than in high similarity MSAs. In contrast, the probability of a firm raising *vc* funding is *lower* in low similarity MSAs than in high similarity MSAs. Panel (b) shows that the fraction of funding amount raised from *angel* investors is *larger* for firms in low similarity MSAs than in high similarity MSAs. In contrast, the fraction of funding amount raised from *vc* investors is *smaller* for firms in low similarity MSAs than in high similarity MSAs. For variable definitions and further details of their construction, see Appendix C.



(a) Funding by investor type



(b) Fraction of funding by investor type

Figure 6: Heterogeneity in Portfolio Strategy by Angel Investment Size

The figure shows the angel investor portfolio strategy by investment size. *Diversification* measures the business similarity among an angel's portfolio firms. *Distance* measures the average distance between an angel and its portfolio firms. Each bubble represents angel investment size decile, where the tenth decile denotes largest investment size. The color of the bubbles denote the scale of angel investment sizes, where the color gets closer to blue as the investment size grows. For variable definitions and further details of their construction, see Appendix C.

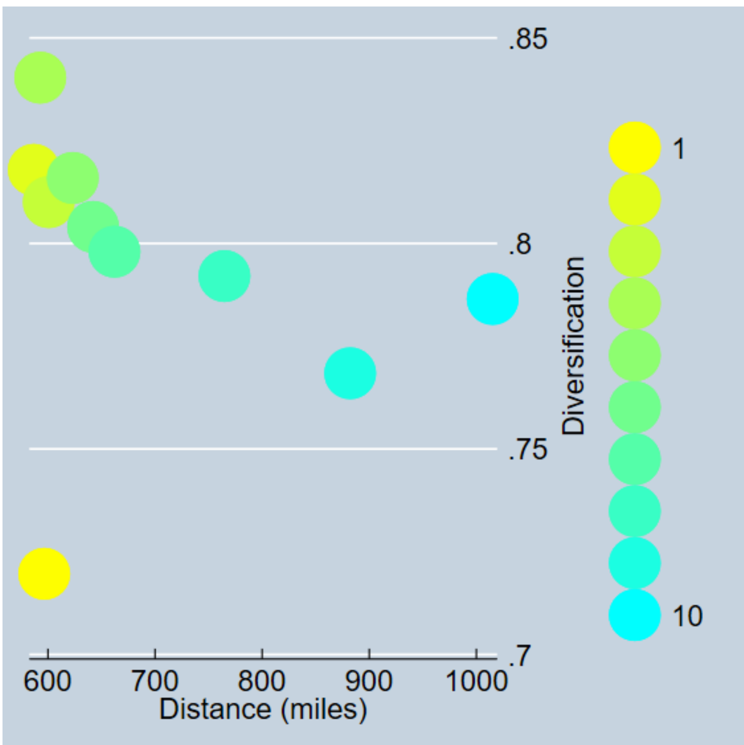


Table 1: Summary Statistics of Crunchbase Variables

The full sample for the analyses consists of 119,605 unique firms between 2007-2018. The top panel presents firm-level summary statistics on the funding sample, which consists of 18,451 firms that report at least one funding round between 2007-2018 (except for $All firm age$). The middle and bottom panel present descriptive statistics on funding by investor types using the funding sample. There are 47,121 funding round-investor type observations. $Fraction\, of\, funding$ is the fraction of funding amount provided by the investor type on a given funding round. $Distance\, to\, investors$ is the distance measured in miles between the funded firm and the given type of funding investor. For variable definitions and further details of their construction, see Appendix C.

	mean	sd	min	p50	max	N
Firm-level						
All firm age (snapshot in 2018)	5.51	2.73	0	6	10	73,822
Funded firm age (snapshot in 2018)	5.53	2.50	0	5	10	13,087
Number of funding rounds	1.90	1.33	1	1	29	18,451
Age at funding	3.07	2.49	0	3	10	18,451
Age at first funding	2.48	2.48	0	2	10	18,451
Funding-level: Angel						
Angel (dummy)	0.19	0.39	0	0	1	47,121
Funded firm age	2.35	2.13	0	2	10	9,052
Funded firm age at first funding	1.75	1.91	0	1	10	5,828
Fraction of first funding	0.22	0.41	0	0	1	26,469
Number of investors	2.07	2.14	1	1	61	9,052
Funding amount ('000s)	2,427.93	9,469.87	0	875	499,505	9,052
Fraction of funding	0.43	0.33	0	0	1	9,052
Distance to investors (miles)	549.79	791.05	0	88	3,294	8,774
Funding-level: VC						
VC (dummy)	0.59	0.49	0	1	1	47,121
Funded firm age	3.37	2.59	0	3	10	27,694
Funded firm age at first funding	2.53	2.46	0	2	10	14,742
Fraction of first funding	0.56	0.50	0	1	1	26,469
Number of investors	2.41	1.74	1	2	31	27,694
Funding amount ('000s)	11,459.04	45,980.61	1	4,286	4,620,000	27,694
Fraction of funding	0.52	0.33	0	1	1	27,694
Distance to investors (miles)	724.35	809.61	0	427	4,972	27,428

Table 2: Summary Statistics of Business Similarity

The table summarizes our similarity measure over the entire sample period and separately over the three decades. Similarity is defined at the firm-year level and is measured as the average of all pairwise cosine similarity scores between the firm and each of all other firms in a given MSA using Crunchbase categories/industries as vectors. The number of firms is measured at the MSA-level. For variable definitions and further details of their construction, see Appendix C.

	Similarity			Number of firms			
	mean	sd	p50	mean	sd	p50	N
2007	0.0412	0.0372	0.0314	2,636	2,563	1,850	84,034
2008	0.0406	0.0365	0.0312	2,879	2,791	2,008	90,040
2009	0.0400	0.0354	0.0310	3,175	3,079	2,262	97,155
2010	0.0395	0.0345	0.0309	3,519	3,415	2,480	104,939
2011	0.0390	0.0337	0.0309	3,922	3,820	2,654	113,300
2012	0.0386	0.0330	0.0308	4,373	4,271	2,876	122,329
2013	0.0382	0.0324	0.0307	4,827	4,718	3,170	131,341
2014	0.0378	0.0319	0.0303	5,246	5,128	3,365	139,816
2015	0.0375	0.0316	0.0300	5,616	5,489	3,561	146,873
2016	0.0373	0.0314	0.0298	5,867	5,736	3,699	151,704
2017	0.0372	0.0313	0.0297	6,077	5,945	3,784	155,566
2018	0.0372	0.0314	0.0297	6,204	6,076	3,835	157,756
2007-2010	0.0403	0.0359	0.0311	3,079	3,024	2,165	376,168
2011-2014	0.0384	0.0327	0.0306	4,631	4,575	2,987	506,786
2015-2018	0.0373	0.0314	0.0298	5,947	5,826	3,699	611,899
Full sample	0.0384	0.0330	0.0304	4,779	4,958	3,131	1,494,853

Table 3: Entry Decision

The table presents cross-sectional regressions of entry decision on similarity. The sample is firm-MSA-entry year level observations for the 81,718 entering firms between 2007-2018. *Entry* is an indicator variable equal to 1 if a given MSA is an entering location and 0 otherwise. *Similarity* is the similarity measure for all entry firm-MSA pairs, measured using existing firms in the year prior to an entry. The control variables are measured in the year prior to the entry year. Column (1) controls for the local economic and column (2) additionally controls for the entrepreneurial ecosystem (Andrews et al., 2019) effects. Columns (3) through (6) examine the similarity effect on entry decision over time. Columns (5) and (6) additionally controls for the founder's degree city. $Ln(Investors)$ is the log (1+ number of CB investors). $Ln(Firms)$, $Ln(Net\ Jobs)$, and $Ln(Personal\ Income)$ are the log of total number of firms, net job creation, and per capita personal income (thousands), respectively. *Founder same city* is an indicator variable equal to 1 if the founder received an academic degree in the same city as the founded company and 0 otherwise. $High\ Angel/VC$ is an indicator variable equal to one if the MSA's ratio of angel investors to VC investors is greater than the year median and zero otherwise. *Period* is an indicator variable for the respective time periods. All standard errors are clustered at the MSA-level. For variable definitions and further details of their construction, see Appendix C.

	Entry × 100					
	(1)	(2)	(3)	(4)	(5)	(6)
Similarity	0.794***	0.808***	0.129***	0.127***	0.201***	0.179***
	(0.190)	(0.194)	(0.029)	(0.029)	(0.049)	(0.049)
Ln(Firms)	0.072	0.061	0.070		0.276**	
	(0.092)	(0.150)	(0.063)		(0.117)	
Ln(Investors)	0.003	-0.003	0.011		-0.007	
	(0.026)	(0.021)	(0.029)		(0.019)	
Ln(Net Jobs)	0.002	0.001	0.001		0.001	
	(0.002)	(0.003)	(0.002)		(0.002)	
Ln(Personal Income)	0.159*	-0.023	0.179		-0.296	
	(0.083)	(0.095)	(0.117)		(0.181)	
Entrepreneurship Quality Index		7.884				
		(14.454)				
Regional Entrepreneurship Cohort Potential Index		0.077***				
		(0.019)				
Regional Ecosystem Acceleration Index		0.000				
		(0.000)				
Founder same city					0.004	0.004
					(0.071)	(0.071)
Period[2009-11] × Similarity			-0.012	-0.007	-0.065	-0.046
			(0.040)	(0.041)	(0.073)	(0.075)
Period[2012-14] × Similarity			-0.047	-0.051*	-0.150***	-0.122**
			(0.030)	(0.030)	(0.053)	(0.054)
Period[2015-17] × Similarity			-0.111***	-0.112***	-0.170***	-0.156***
			(0.035)	(0.035)	(0.059)	(0.058)
Period[2009-11] × High Angel/VC × Similarity			-0.621	-0.723	-1.992	-1.632
			(0.836)	(0.804)	(2.071)	(1.889)
Period[2012-14] × High Angel/VC × Similarity			-1.563*	-1.652**	-2.827	-2.579
			(0.848)	(0.831)	(1.889)	(1.635)
Period[2015-17] × High Angel/VC × Similarity			-2.571***	-2.623***	-4.171*	-3.776*
			(0.957)	(0.958)	(2.160)	(1.983)
Observations	17662847	9648875	17662847	17662847	5423263	5423263
Adjusted R^2	0.049	0.046	0.049	0.049	0.081	0.082
MSA FE	Y	Y	Y	N	Y	N
Year FE	Y	Y	Y	N	Y	N
MSA-by-Year FE	N	N	N	Y	N	Y

Standard errors in parentheses
***p<0.01, **p<0.05, *p<0.1

Table 4: Angel Tax Credit and Entry Decision

The table presents cross-sectional regressions of entry decision on similarity. The sample is firm-MSA-entry year level observations for the 81,718 entering firms between 2007-2018. *Entry* is an indicator variable equal to 1 if a given MSA is an entering location and 0 otherwise. *Similarity* is the similarity measure for all entry firm-MSA pairs, measured using existing firms in the year prior to an entry. The control variables are measured in the year prior to the entry year. Columns (1) through (3) controls for the local economic and entrepreneurial ecosystem (Andrews et al., 2020) effects. Columns (4) through (6) examine the similarity effect on entry decision over time. *Founder same city* is an indicator variable equal to 1 if the founder received an academic degree in the same city as the founded company and 0 otherwise. $\mathbb{1}(Angel\ Tax)$ is an indicator variable equal to 1 if the firm is located in a state with angel tax credit program and 0 otherwise. *Angel Tax* is is a continuous variable of angel tax credit percentages. $Ln(Investors)$ is the log (1+ number of CB investors). $Ln(Firms)$, $Ln(NetJobs)$, and $Ln(PersonalIncome)$ are the log of total number of firms, net job creation, and per capita personal income (thousands), respectively. $High\ Angel/VC$ is an indicator variable equal to one if the MSA's ratio of angel investors to VC investors is greater than the year median and zero otherwise. *Period* is an indicator variable for the respective time periods. All standard errors are clustered at the MSA-level. For variable definitions and further details of their construction, see Appendix C.

	Entry × 100							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Similarity	1.004***	0.099***	0.093***	0.091***	0.978***	0.105***	0.099***	0.101***
	(0.292)	(0.024)	(0.023)	(0.033)	(0.276)	(0.023)	(0.023)	(0.033)
Founder same city				0.004				0.004
				(0.071)				(0.071)
Ln(Firms)	0.078	0.062			0.084	0.069		
	(0.093)	(0.075)			(0.095)	(0.076)		
Ln(Investors)	0.003	0.011			0.003	0.011		
	(0.026)	(0.039)			(0.026)	(0.039)		
Ln(Net Jobs)	0.002	0.002			0.002	0.002		
	(0.002)	(0.002)			(0.002)	(0.002)		
Ln(Personal Income)	0.145*	0.142*			0.143*	0.141*		
	(0.081)	(0.082)			(0.080)	(0.082)		
$\mathbb{1}$(Angel Tax)	-0.010	-0.011						
	(0.013)	(0.007)						
$\mathbb{1}$(Angel Tax) × Similarity	-0.554*	-0.008	-0.003	0.020				
	(0.294)	(0.034)	(0.034)	(0.047)				
$\mathbb{1}$(Angel Tax) × High Angel/VC × Similarity		-2.507**	-2.490**	-3.121**				
		(1.112)	(1.127)	(1.394)				
Angel Tax					-0.028	-0.029		
					(0.037)	(0.019)		
Angel Tax × Similarity					-1.587*	-0.075	-0.064	-0.013
					(0.819)	(0.088)	(0.089)	(0.127)
Angel Tax × High Angel/VC × Similarity						-6.874**	-6.816**	-8.618**
						(3.066)	(3.103)	(3.818)
Observations	17662847	17662847	17662847	5423263	17662847	17662847	17662847	5423263
Adjusted R^2	0.049	0.049	0.049	0.082	0.049	0.049	0.049	0.082
MSA FE	Y	Y	N	N	Y	Y	N	N
Year FE	Y	Y	N	N	Y	Y	N	N
MSA-by-Year FE	N	N	Y	Y	N	N	Y	Y

Standard errors in parentheses
***p<0.01, **p<0.05, *p<0.1

Table 5: Similarity and Funding Types

The table shows results from the firm-year level regressions of funding characteristics on similarity. The main variable of interest is *Similarity*. In each panel, columns (1)-(3) only use observations with any reported funding, and columns (4)-(6) uses all sample observations by replacing no-funding observations with zeros. $\mathbb{1}(InvestorType)$ is a dummy variable that is equal to one if funding is received from the respective investor type and zero otherwise. *Frac. Funding* is the fraction of the total funding amounts received from the given investor type in a year. $\$Funding$ is the total amount of funding in millions from the given investor type in a year. $\$TotalFunding$ is the total funding amount in millions received in a year. All specifications control for G and EG Indexes. Standard errors are clustered at MSA-level. For variable definitions and further details of their construction, see Appendix C.

Panel A: Angel Funding

	(1)	(2)	(3)	(4)	(5)	(6)
	$\mathbb{1}$(Investor Type)	Frac. Funding	$Funding	$\mathbb{1}$(Investor Type)	Frac. Funding	$Funding
Similarity	-7.119***	-4.127***	-1.987*	-0.359***	-0.235***	-0.148***
	(1.179)	(0.634)	(1.110)	(0.129)	(0.077)	(0.054)
Ln(Age)	-0.076***	-0.086***	-0.161***	0.002	-0.001	-0.001*
	(0.020)	(0.016)	(0.024)	(0.001)	(0.001)	(0.001)
Ln(Firms)	-0.641***	-0.372***	-0.426***	-0.057***	-0.033***	-0.026***
	(0.116)	(0.061)	(0.082)	(0.013)	(0.008)	(0.005)
$ Total Funding			-0.000***			-0.000
			(0.000)			(0.000)
Firm FE	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y
Observations	21,178	21,178	21,620	762,299	762,299	762,299
Adjusted R^2	0.372	0.442	0.155	0.123	0.101	0.093

Panel B: VC Funding

	(1)	(2)	(3)	(4)	(5)	(6)
	$\mathbb{1}$(Investor Type)	Frac. Funding	$Funding	$\mathbb{1}$(Investor Type)	Frac. Funding	$Funding
Similarity	2.370**	3.730***	153.967**	-0.144	-0.054	6.883**
	(1.182)	(1.257)	(62.249)	(0.181)	(0.157)	(2.835)
Ln(Age)	0.140***	0.189***	4.481***	0.028***	0.026***	0.288***
	(0.017)	(0.027)	(0.749)	(0.004)	(0.004)	(0.082)
Ln(Firms)	-0.190**	0.009	35.102***	-0.107***	-0.088***	0.465
	(0.091)	(0.077)	(5.409)	(0.023)	(0.019)	(0.310)
$ Total Funding			0.215***			0.224***
			(0.018)			(0.012)
Firm FE	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y
Observations	21,178	21,178	21,620	762,299	762,299	762,299
Adjusted R^2	0.437	0.469	0.248	0.218	0.207	0.378

Standard errors in parentheses
***p<0.01, **p<0.05, *p<0.1

Table 6: Similarity and First Funding Type

The table shows the results from the firm-year level regressions of funding characteristics on similarity. The sample is restricted to first funding observations only. Columns (1)-(4) include all first funding observations, whereas columns (5)-(8) exclude firms that have the first funding funded by multiple types of investors. The main variable of interest is *Similarity*. $\mathbb{1}(Angel)$ and $\mathbb{1}(VC)$ are dummy variables that are equal to one if funding is received from the respective investor type and zero otherwise. Note that the first funding may be funded by more than one type of investors, in which case we cannot include investor-level FE. All specifications control for G and EG Indexes. Standard errors are clustered at funded firm MSA-level. For variable definitions and further details of their construction, see Appendix C.

	All first funding				First funding by single investor type			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	$\mathbb{1}$(Angel)		$\mathbb{1}$(VC)		$\mathbb{1}$(Angel)		$\mathbb{1}$(VC)	
Similarity	-0.457***	-0.286***	0.962***	0.850***	-0.521***	-0.399***	1.075***	0.766***
	(0.114)	(0.110)	(0.139)	(0.129)	(0.105)	(0.074)	(0.171)	(0.143)
Ln(Firms)	0.019***	-0.043	0.044***	-0.044	-0.004	0.001	0.038***	-0.078
	(0.005)	(0.047)	(0.008)	(0.067)	(0.005)	(0.042)	(0.008)	(0.070)
Funding year FE	Y	Y	Y	Y	Y	Y	Y	Y
Founding year FE	Y	Y	Y	Y	Y	Y	Y	Y
Firm MSA FE	N	Y	N	Y	N	Y	N	Y
Investor MSA FE	N	N	N	N	N	Y	N	Y
Observations	20,941	20,879	20,941	20,879	15,419	15,269	15,419	15,269
Adjusted R^2	0.038	0.055	0.028	0.063	0.014	0.108	0.037	0.153

Standard errors in parentheses
***p<0.01, **p<0.05, *p<0.1

Table 7: Similarity and MSA-level Entrepreneurship Quality

The table presents MSA-year-level regression of Entrepreneurship Quality Index on MSA similarity. The dependent variable is MSA-level Entrepreneurship Quality Index measure from Andrews et al. (2019), which is available until 2013. Standard errors are clustered at MSA-level. For variable definitions and further details of their construction, see Appendix C.

	(1)	(2)	(3)
	EQI×100	EQI×100	EQI×100
MSA Similarity	-0.054**	-0.059**	-0.083**
	(0.024)	(0.026)	(0.041)
Ln(Firms)		-0.008	-0.064*
		(0.014)	(0.038)
Ln(Investors)		0.002	-0.001
		(0.002)	(0.003)
Ln(Personal Income)		0.034***	0.003
		(0.013)	(0.011)
Regional Entrepreneurship Cohort Potential Index			0.004***
			(0.000)
Regional Ecosystem Acceleration Index			-0.000***
			(0.000)
MSA FE	Y	Y	Y
Year FE	Y	Y	Y
Observations	3536	3372	2054
Adjusted R^2	0.838	0.838	0.778

Standard errors in parentheses
***p<0.01, **p<0.05, *p<0.1

Table 8: Robustness: Soft-cosine Similarity and Dynamic Similarity

The table shows results from the firm-year level regressions of funding characteristics on soft-cosine similarity scores in Panel A and dynamic similarity scores in Panel B. *SoftSimilarity* a cosine similarity measure that takes account of correlations among finer sub-classifications within 46 broad industry groups. The measure is based on the soft cosine similarity computation method introduced by Sidorov et al. (2014) adjusting the elements under the same broad industry group to be treated equally more similar in our setting. *DynamicSimilarity* is a cosine similarity measure using changing industry vectors over time using multiple Crunchbase data dumps that have been acquired in 2016, 2018, 2019, and 2020. When there is no Crunchbase data dump for a specific year, we use the data from the closest prior year. For the years that have no prior-year data, we use the earliest possible data. In each panel, columns (1)-(3) only use observations with any reported funding, and columns (4)-(6) uses all sample observations by replacing no-funding observations with zeros. $\mathbb{1}(InvestorType)$ is a dummy variable that is equal to one if funding is received from the respective investor type and zero otherwise. *Frac. Funding* is the fraction of the total funding amounts received from the given investor type in a year. $\$Funding$ is the total amount of funding in millions from the given investor type in a year. $\$TotalFunding$ is the total funding amount in millions received in a year. All specifications control for G and EG Indexes. Standard errors are clustered at MSA-level. For variable definitions and further details of their construction, see Appendix C.

Panel A: Using Soft-cosine Similarity

	(1) $\mathbb{1}$(Investor Type)	(2) Frac. Funding	(3) \$Funding	(4) $\mathbb{1}$(Investor Type)	(5) Frac. Funding	(6) \$Funding
Angel Funding						
Soft Similarity	-3.691***	-2.317***	-1.470***	-0.096***	-0.064***	-0.036***
	(0.537)	(0.339)	(0.456)	(0.021)	(0.012)	(0.010)
Observations	18682	18682	19138	1396889	1396889	1396889
Adjusted R^2	0.386	0.455	0.141	0.160	0.140	0.115
VC Funding						
Soft Similarity	1.704***	2.690***	169.129***	-0.082	-0.059	1.860*
	(0.619)	(0.760)	(49.208)	(0.053)	(0.049)	(0.992)
Observations	18682	18682	19138	1396889	1396889	1396889
Adjusted R^2	0.445	0.469	0.402	0.223	0.211	0.269
Firm FE	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y

Panel B: Using Dynamic Similarity

	(1) $\mathbb{1}$(Investor Type)	(2) Frac. Funding	(3) \$Funding	(4) $\mathbb{1}$(Investor Type)	(5) Frac. Funding	(6) \$Funding
Angel Funding						
Dynamic Similarity	-0.908**	-0.376	-1.116*	-0.041**	-0.021**	-0.032**
	(0.448)	(0.238)	(0.610)	(0.018)	(0.010)	(0.014)
Observations	28157	28157	28709	1988051	1988051	1988051
Adjusted R^2	0.346	0.411	0.118	0.091	0.067	0.097
VC Funding						
Dynamic Similarity	0.334	0.617	105.345**	-0.036	-0.025	2.944**
	(0.353)	(0.472)	(43.773)	(0.039)	(0.035)	(1.328)
Observations	28157	28157	28709	1988051	1988051	1988051
Adjusted R^2	0.401	0.418	0.281	0.172	0.159	0.293
Firm FE	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y

Standard errors in parentheses
***p<0.01, **p<0.05, *p<0.1

Table 9: Angel vs. VC Portfolio Preferences

The table shows results from the regressions of investor portfolio characteristics on investor type. Regression in Panels A and B are at the aggregate investor level and each funding level, respectively. We consider within-investor industry category similarities in columns (1) and (2) and the geographic distance in columns (3) and (4). For those measures, we consider all fundings during the entire sample period by each investor in Panel A and previous fundings by each investor before making a given funding in a given year in Panel B. The sample in columns (1) and (3) consists of all investor types, and the sample in columns (2) and (4) consists of only vc and angel investors. *Angel* and *VC* are indicator variables. $\$Funding$ is the average funding amount, $Ln(Investments)$ is the log of the total number of investments, and $Ln(InvestingYears)$ is the log of the years since the first investment. $Ln(Investments)$ and $\$Funding$ in Panel B only consider previous investments before a given funding year. All standard errors are clustered at investor MSA-level. For variable definitions and further details of their construction, see Appendix C.

Panel A: Investor-level

	(1)	(2)	(3)	(4)
	Industry Similarity		Geographic Distance	
Angel	0.009	-0.018*	-93.914**	-106.908***
	(0.011)	(0.011)	(42.571)	(32.192)
VC	0.031***		20.988	
	(0.011)		(25.017)	
Ln(Investments)	-0.054***	-0.058***	-7.638	-8.484
	(0.005)	(0.006)	(7.573)	(7.351)
Ln(Investing Years)	-0.019	-0.055***	-79.014	-77.079
	(0.018)	(0.016)	(54.991)	(67.447)
$Funding	0.000	0.001**	4.360***	5.020***
	(0.000)	(0.000)	(0.571)	(1.196)
Observations	7,185	6,070	11,853	10,384
Adjusted R^2	0.076	0.081	0.068	0.060
Investor MSA FE	Y	Y	Y	Y

Panel B: Funding-level

	(1)	(2)	(3)	(4)
	Industry Similarity		Geographic Distance	
Angel	-0.018***	-0.018***	-154.580***	-52.453***
	(0.005)	(0.004)	(33.119)	(15.073)
VC	0.002		-108.195***	
	(0.004)		(27.449)	
Ln(Investments)	-0.012***	-0.013***	25.426***	26.962***
	(0.002)	(0.001)	(8.354)	(7.934)
Ln(Investing Years)	0.013***	0.010***	79.431***	68.528***
	(0.005)	(0.002)	(16.851)	(19.850)
$Funding	0.000	0.000**	3.535***	2.264**
	(0.000)	(0.000)	(0.774)	(0.989)
Observations	45,158	39,326	44,445	38,732
Adjusted R^2	0.571	0.582	0.251	0.255
Firm FE	Y	Y	Y	Y
Funding year FE	Y	Y	Y	Y
Investor MSA FE	Y	Y	Y	Y
Funded MSA FE	Y	Y	Y	Y

Standard errors in parentheses
***p<0.01, **p<0.05, *p<0.1

Table 10: Outcome of Investments

The table shows results from the regressions of successful outcomes on investor portfolio characteristics. Regression are at the aggregate investor level with one observation per investor. We consider within-investor industry category similarities and the geographic distance as portfolio characteristics. For those measures, we consider all fundings made by each investor during the entire sample period. *Angel* is an indicator variable for angel investors. $Ln(Investments)$ is the log of the total number of investments by a given investor, $Ln(InvestingYears)$ is the log of the years since the first investment of a given investor, and $\$Funding$ is the average funding amount made by a given investor. All standard errors are clustered at investor MSA-level. For variable definitions and further details of their construction, see Appendix C.

	(1)	(2)	(3)	(4)	(5)	(6)
	$\mathbb{1}$(Subsequent Funding)		Exit Rate - IPO		Exit Rate - IPO/Acq.	
Angel	0.081***	0.095***	-0.068***	-0.045***	-0.051***	-0.023**
	(0.009)	(0.007)	(0.005)	(0.006)	(0.011)	(0.010)
Close	-0.004		-0.034***		-0.034***	
	(0.010)		(0.005)		(0.012)	
Close × Angel	-0.024		0.034***		0.031***	
	(0.019)		(0.005)		(0.010)	
Dissimilar	-0.074***		-0.043***		-0.061***	
	(0.009)		(0.005)		(0.008)	
Dissimilar × Angel	0.130***		0.034***		0.070***	
	(0.011)		(0.005)		(0.012)	
Close&Dissimilar		-0.051***		-0.039***		-0.056***
		(0.009)		(0.006)		(0.009)
Close&Dissimilar × Angel		0.116***		0.032***		0.059***
		(0.018)		(0.006)		(0.016)
Ln(Investments)	0.173***	0.173***	0.001	-0.001	-0.007*	-0.009***
	(0.008)	(0.008)	(0.002)	(0.002)	(0.004)	(0.003)
Ln(Investing Years)	-0.120***	-0.117***	-0.116***	-0.114***	-0.676***	-0.673***
	(0.040)	(0.040)	(0.010)	(0.010)	(0.028)	(0.028)
$Funding	-0.000**	-0.000***	0.000***	0.000***	0.000*	0.000*
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Mean of the Dep. Var.	0.665		0.041		0.221	
Observations	12554	12554	12554	12554	12554	12554
Adjusted R^2	0.216	0.213	0.077	0.066	0.106	0.103
Investor MSA FE	Y	Y	Y	Y	Y	Y

Standard errors in parentheses
***p<0.01, **p<0.05, *p<0.1

Figure A1: Funding Round Timing and Count by Investor Type

The figure shows the comparison of the frequency (left) and the number (right) of funding rounds by angel and VC investors within our sample period between 2007-2018.

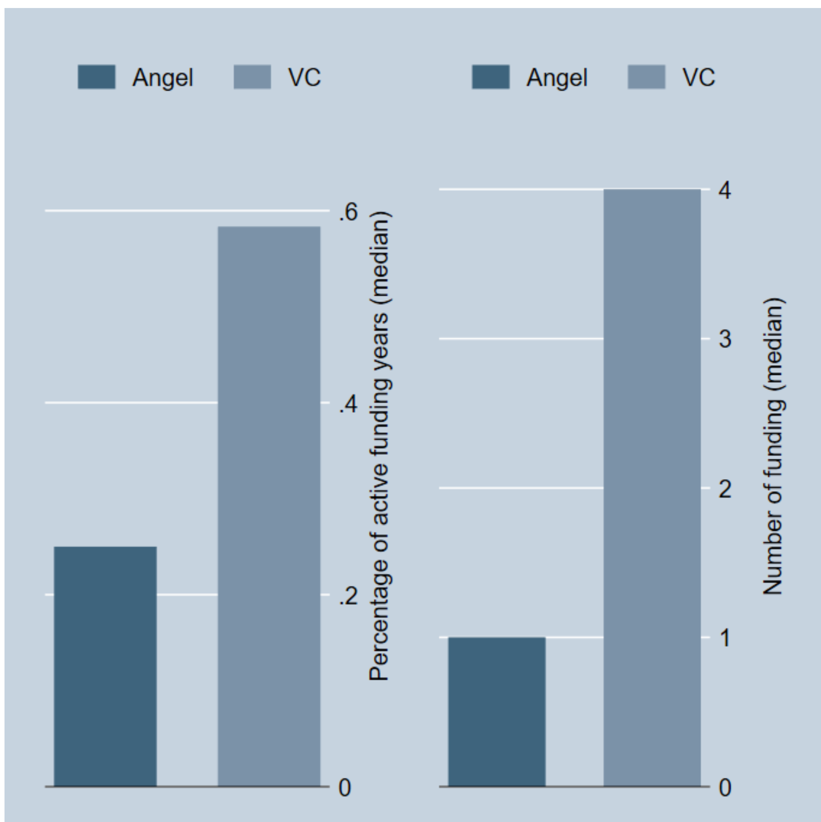


Table A1: Geographic Distribution of Crunchbase Firms

Sample Firms				All Firms in Crunchbase			
State	Count	%	Cum%	State	Count	%	Cum%
California	40,591	25.4	25.4	California	62,332	25.6	25.6
New York	16,910	10.6	35.9	New York	26,890	11.1	36.7
Texas	10,713	6.7	42.6	Texas	16,038	6.6	43.3
Florida	8,270	5.2	47.8	Florida	12,406	5.1	48.4
Massachusetts	7,846	4.9	52.7	Massachusetts	11,449	4.7	53.1
Illinois	6,500	4.1	56.8	Illinois	10,065	4.1	57.2
Pennsylvania	4,655	2.9	59.7	Washington	7,041	2.9	60.1
Washington	4,528	2.8	62.5	Pennsylvania	6,865	2.8	62.9
Colorado	4,409	2.8	65.3	Colorado	6,399	2.6	65.6
Georgia	4,147	2.6	67.8	Georgia	6,329	2.6	68.2
New Jersey	4,137	2.6	70.4	New Jersey	5,979	2.5	70.6
Virginia	4,008	2.5	72.9	Virginia	5,703	2.3	73
Ohio	3,341	2.1	75	Ohio	4,954	2	75
North Carolina	3,058	1.9	76.9	North Carolina	4,435	1.8	76.8
Maryland	2,734	1.7	78.6	Michigan	3,953	1.6	78.5
Michigan	2,629	1.6	80.3	Maryland	3,919	1.6	80.1
Arizona	2,544	1.6	81.9	Arizona	3,886	1.6	81.7
Minnesota	2,321	1.5	83.3	Minnesota	3,453	1.4	83.1
Connecticut	1,944	1.2	84.5	Utah	2,993	1.2	84.3
Utah	1,942	1.2	85.8	Tennessee	2,990	1.2	85.5
Tennessee	1,932	1.2	87	District of Columbia	2,888	1.2	86.7
Oregon	1,847	1.2	88.1	Oregon	2,882	1.2	87.9
Wisconsin	1,709	1.1	89.2	Connecticut	2,736	1.1	89
Missouri	1,655	1	90.2	Wisconsin	2,600	1.1	90.1
District of Columbia	1,649	1	91.2	Missouri	2,556	1.1	91.2
Indiana	1,536	1	92.2	Indiana	2,318	1	92.1
Nevada	1,428	0.9	93.1	Nevada	2,217	0.9	93
South Carolina	913	0.6	93.7	South Carolina	1,541	0.6	93.7
Delaware	806	0.5	94.2	Kentucky	1,270	0.5	94.2
Kansas	796	0.5	94.7	Alabama	1,168	0.5	94.7
Kentucky	784	0.5	95.2	Louisiana	1,166	0.5	95.1
Alabama	740	0.5	95.6	Delaware	1,140	0.5	95.6
Louisiana	704	0.4	96.1	Kansas	1,132	0.5	96.1
Oklahoma	683	0.4	96.5	Oklahoma	1,099	0.5	96.5
New Hampshire	672	0.4	96.9	New Hampshire	969	0.4	96.9
Iowa	625	0.4	97.3	Iowa	896	0.4	97.3
Nebraska	532	0.3	97.6	Nebraska	804	0.3	97.6
Arkansas	525	0.3	98	Arkansas	768	0.3	97.9
Idaho	415	0.3	98.2	Idaho	630	0.3	98.2
Rhode Island	398	0.2	98.5	New Mexico	617	0.3	98.5
New Mexico	395	0.2	98.7	Rhode Island	582	0.2	98.7
Maine	361	0.2	98.9	Maine	559	0.2	98.9
Hawaii	288	0.2	99.1	Hawaii	466	0.2	99.1
Vermont	273	0.2	99.3	Vermont	381	0.2	99.3
Montana	235	0.1	99.4	Montana	341	0.1	99.4
Mississippi	204	0.1	99.6	Mississippi	328	0.1	99.5
Wyoming	179	0.1	99.7	Wyoming	258	0.1	99.7
South Dakota	149	0.1	99.8	North Dakota	248	0.1	99.8
North Dakota	140	0.1	99.9	South Dakota	233	0.1	99.8
West Virginia	118	0.1	99.9	West Virginia	193	0.1	99.9
Alaska	105	0.1	100	Alaska	172	0.1	100
Total	160,023			Total	243,237		

Table A2: Crunchbase Category Trends

The table presents trends of Crunchbase startup categories over time between 2007-2018. The rank is determined by computing the percentage of firms reporting a given category in each year. There are 742 Crunchbase categories, and firms can report multiple categories. The trends in category choices by firms is dominantly driven by new firms added into the data (also see dynamic similarity measure and related discussion in Section 4.2).

	Categories		
Rank	2007	2012	2018
1	Software	Software	Information Technology
2	Information Technology	Mobile	Software
3	Health Care	E-Commerce	Internet
4	Advertising	Information Technology	Health Care
5	Consulting	Health Care	Artificial Intelligence
6	Internet	Internet	E-Commerce
7	E-Commerce	Advertising	SaaS
8	Biotechnology	Social Media	Blockchain
9	Mobile	Enterprise Software	Financial Services
10	Enterprise Software	Education	Consulting
11	Manufacturing	Consulting	FinTech
12	Medical	Analytics	Machine Learning
13	Social Media	SaaS	Advertising
14	Education	Apps	Real Estate
15	Video	Biotechnology	Mobile Apps
16	Marketing	Medical	Education
17	Financial Services	Big Data	Marketing
18	SaaS	Marketing	Cryptocurrency
19	Analytics	Fashion	Mobile
20	Web Development	Finance	Apps

Table A3: Angel Investors

The table presents summary statistics on angel investor characteristics. There are total 7,125 unique angel investors who funded our sample firms. Panel A describes angel investors' funding amount and demographics. Panel B describes angel investors' experience in entrepreneurship. Panel C describes angel investors' reported jobs and advising positions.

	Mean	Sd	Min	Median	Max	N
Panel A: Funding and demographics						
Funding amount ('000s)	1,383.61	8,984.19	1.00	333.33	499,505.13	6,608
Seed funding amount ('000s)	343.38	520.50	1.00	206.02	10,000.00	4,206
Gender	0.92	0.27	0.00	1.00	1.00	7,094
Number of investments	3.69	8.60	1.00	1.00	243.00	7,125
Number of academic degrees	1.51	0.70	1.00	1.00	6.00	2,836
MBA (indicator)	0.24	0.43	0.00	0.00	1.00	2,836
Ph.D (indicator)	0.05	0.23	0.00	0.00	1.00	2,836
Panel B: Entrepreneurship						
Number of founded entities	1.67	0.98	1.00	1.00	8.00	3,279
% of founded entities that went public	4.88	19.17	0.00	0.00	100.00	3,279
% of founded entities that are company	84.74	32.91	0.00	100.00	100.00	3,279
% of founded entities that are investor	15.20	32.87	0.00	0.00	100.00	3,279
Panel C: Jobs						
Number of jobs (all)	6.28	7.26	1.00	4.00	67.00	6,631
Number of jobs (current)	3.63	4.48	0.00	2.00	56.00	6,631
% of jobs working as an employee	65.94	34.90	0.00	75.00	100.00	6,631
% of jobs working as a board member	26.19	27.78	0.00	20.00	100.00	6,631
% of jobs working as an advisor	2.73	9.31	0.00	0.00	85.25	6,631
% of jobs working as an executive	5.14	13.43	0.00	0.00	100.00	6,631
% of employers that are companies	69.26	40.15	0.00	100.00	100.00	6,631
% of employers that are investors	12.88	25.18	0.00	0.00	100.00	6,631
% of employers that are public companies	14.43	24.15	0.00	0.00	100.00	6,631
Number of advising roles	1.99	1.35	1.00	1.00	9.00	2,651
% of advising entities that are companies	88.56	26.72	0.00	100.00	100.00	2,651
% of advising entities that are investors	10.47	25.93	0.00	0.00	100.00	2,651
% of advising entities that are public companies	7.10	22.11	0.00	0.00	100.00	2,651

Table A4: Similarity and Funding Types - Dropping the entry year

The table shows results re-estimating the regressions in Table 5 after dropping the first year observations to mitigate concerns with selection. The main variable of interest is *Similarity*. In each panel, columns (1)-(3) only use observations with any reported funding, and columns (4)-(6) uses all sample observations by replacing no-funding observations with zeros. $\mathbb{1}(Investor\ Type)$ is a dummy variable that is equal to one if funding is received from the respective investor type and zero otherwise. $Frac.\ Funding$ is the fraction of the total funding amounts received from the given investor type in a year. $\$Funding$ is the total amount of funding in millions from the given investor type in a year. $\$Total\ Funding$ is the total funding amount in millions received in a year. All specifications control for G and EG Indexes. Standard errors are clustered at MSA-level. For variable definitions and further details of their construction, see Appendix C.

Panel A: Angel Funding

	(1)	(2)	(3)	(4)	(5)	(6)
	𝟙(Investor Type)	Frac. Funding	$Funding	𝟙(Investor Type)	Frac. Funding	$Funding
Similarity	-6.627***	-3.851***	-2.193**	-0.381***	-0.227***	-0.176***
	(1.128)	(0.634)	(0.876)	(0.138)	(0.075)	(0.057)
Ln(Age)	-0.254***	-0.149***	-0.303***	-0.022***	-0.013***	-0.015***
	(0.047)	(0.036)	(0.047)	(0.007)	(0.004)	(0.005)
Ln(Firms)	-0.564***	-0.327***	-0.333***	-0.062***	-0.032***	-0.025***
	(0.105)	(0.061)	(0.069)	(0.016)	(0.008)	(0.006)
$ Total Funding			-0.000			-0.000
			(0.000)			(0.000)
Firm FE	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y
Observations	18,155	18,155	18,544	678,838	678,838	678,838
Adjusted R^2	0.390	0.449	0.183	0.141	0.113	0.093

Panel B: VC Funding

	(1)	(2)	(3)	(4)	(5)	(6)
	𝟙(Investor Type)	Frac. Funding	$Funding	𝟙(Investor Type)	Frac. Funding	$Funding
Similarity	1.986	2.439*	154.733**	-0.240	-0.147	5.959**
	(1.316)	(1.267)	(69.626)	(0.248)	(0.224)	(2.745)
Ln(Age)	0.128***	0.273***	9.477***	-0.002	0.005**	0.470***
	(0.027)	(0.049)	(1.965)	(0.004)	(0.002)	(0.152)
Ln(Firms)	-0.170*	-0.052	31.736***	-0.139***	-0.119***	0.147
	(0.093)	(0.085)	(4.424)	(0.033)	(0.027)	(0.221)
$ Total Funding			0.205***			0.219***
			(0.019)			(0.012)
Firm FE	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y
Observations	18,155	18,155	18,544	678,838	678,838	678,838
Adjusted R^2	0.432	0.470	0.243	0.237	0.224	0.383

Standard errors in parentheses
***p<0.01, **p<0.05, *p<0.1

Table A5: Similarity and Funding Types - Combined Statistical Area

The table shows results re-estimating the regressions in Table 5, keeping only the combined statistical areas. There are total 149 CSAs matched to 245 of 416 MSAs in our sample. The main variable of interest is *Similarity*. In each panel, columns (1)-(3) only use observations with any reported funding, and columns (4)-(6) uses all sample observations by replacing no-funding observations with zeros. $\mathbb{1}(Investor\ Type)$ is a dummy variable that is equal to one if funding is received from the respective investor type and zero otherwise. *Frac. Funding* is the fraction of the total funding amounts received from the given investor type in a year. $*Funding* is the total amount of funding in millions from the given investor type in a year. $*Total Funding* is the total funding amount in millions received in a year. All specifications control for G and EG Indexes. Standard errors are clustered at MSA-level. For variable definitions and further details of their construction, see Appendix C.

Panel A: Angel Funding

	(1)	(2)	(3)	(4)	(5)	(6)
	$\mathbb{1}$(Investor Type)	Frac. Funding	$Funding	$\mathbb{1}$(Investor Type)	Frac. Funding	$Funding
Similarity	-7.479***	-4.351***	-2.098*	-0.471***	-0.300***	-0.174**
	(1.105)	(0.616)	(1.181)	(0.174)	(0.103)	(0.071)
Ln(Age)	-0.073***	-0.078***	-0.164***	0.002**	-0.001	-0.002*
	(0.024)	(0.018)	(0.026)	(0.001)	(0.001)	(0.001)
Ln(Firms)	-0.685***	-0.402***	-0.428***	-0.068***	-0.038***	-0.030***
	(0.123)	(0.059)	(0.082)	(0.011)	(0.006)	(0.004)
$ Total Funding			-0.000***			-0.000 ***
			(0.000)			(0.000)
Firm FE	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y
Observations	18,624	18,624	18,987	608,155	608,155	608,155
Adjusted R^2	0.372	0.440	0.165	0.124	0.103	0.103

Panel B: VC Funding

	(1)	(2)	(3)	(4)	(5)	(6)
	$\mathbb{1}$(Investor Type)	Frac. Funding	$Funding	$\mathbb{1}$(Investor Type)	Frac. Funding	$Funding
Similarity	3.175***	4.519***	153.021**	-0.149	-0.033	9.694**
	(1.117)	(1.172)	(72.015)	(0.252)	(0.218)	(3.802)
Ln(Age)	0.136***	0.185***	4.688***	0.031***	0.028***	0.320***
	(0.018)	(0.031)	(0.791)	(0.005)	(0.005)	(0.096)
Ln(Firms)	-0.189*	0.032	37.852***	-0.122***	-0.099***	0.581*
	(0.102)	(0.078)	(3.653)	(0.021)	(0.018)	(0.305)
$ Total Funding			0.208***			0.220***
			(0.011)			(0.010)
Firm FE	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y
Observations	18,624	18,624	18,987	608,155	608,155	608,155
Adjusted R^2	0.433	0.470	0.231	0.220	0.210	0.368

Standard errors in parentheses
***p<0.01, **p<0.05, *p<0.1

Appendix A Crunchbase Industry Classification

Crunchbase maintain its company data using more than 700 Industries and 47 Industry Groups. See https://support.crunchbase.com/hc/en-us/articles/360043146954. Industry Groups are broader subjects that encompass multiple industries. Industries are more specific market segments. Company profiles can belong to multiple industries and industry groups.

Industry Group	Industries
Administrative Services	Archiving Service, Call Center, Collection Agency, College Recruiting, Courier Service, Debt Collections, Delivery, Document Preparation, Employee Benefits, Extermination Service, Facilities Support Services, Housekeeping Service, Human Resources, Knowledge Management, Office Administration, Packaging Services, Physical Security, Project Management, Staffing Agency, Trade Shows, Virtual Workforce
Advertising	Ad Exchange, Ad Network, Ad Retargeting, Ad Server, Ad Targeting, Advertising, Advertising Platforms, Affiliate Marketing, Local Advertising, Mobile Advertising, Outdoor Advertising, SEM, Social Media Advertising, Video Advertising
Agriculture and Farming	Agriculture, AgTech, Animal Feed, Aquaculture, Equestrian, Farming, Forestry, Horticulture, Hydroponics, Livestock
Apps	App Discovery, Apps, Consumer Applications, Enterprise Applications, Mobile Apps, Reading Apps, Web Apps
Artificial Intelligence	Artificial Intelligence, Intelligent Systems, Machine Learning, Natural Language Processing, Predictive Analytics
Biotechnology	Bioinformatics, Biometrics, Biopharma, Biotechnology, Genetics, Life Science, Neuroscience, Quantified Self
Clothing and Apparel	Fashion, Laundry and Dry-cleaning, Lingerie, Shoes
Commerce and Shopping	Auctions, Classifieds, Collectibles, Consumer Reviews, Coupons, E-Commerce, E-Commerce Platforms, Flash Sale, Gift, Gift Card, Gift Exchange, Gift Registry, Group Buying, Local Shopping, Made to Order, Marketplace, Online Auctions, Personalization, Point of Sale, Price Comparison, Rental, Retail, Retail Technology, Shopping, Shopping Mall, Social Shopping, Sporting Goods, Vending and Concessions, Virtual Goods, Wholesale
Community and Lifestyle	Adult, Baby, Cannabis, Children, Communities, Dating, Elderly, Family, Funerals, Humanitarian, Leisure, LGBT, Lifestyle, Men's, Online Forums, Parenting, Pet, Private Social Networking, Professional Networking, Q&A, Religion, Retirement, Sex Industry, Sex Tech, Social, Social Entrepreneurship, Teenagers, Virtual World, Wedding, Women's, Young Adults
Consumer Electronics	Computer, Consumer Electronics, Drones, Electronics, Google Glass, Mobile Devices, Nintendo, Playstation, Roku, Smart Home, Wearables, Windows Phone, Xbox
Consumer Goods	Beauty, Comics, Consumer Goods, Cosmetics, DIY, Drones, Eyewear, Fast-Moving Consumer Goods, Flowers, Furniture, Green Consumer Goods, Handmade, Jewelry, Lingerie, Shoes, Tobacco, Toys
Content and Publishing	Blogging Platforms, Content Delivery Network, Content Discovery, Content Syndication, Creative Agency, DRM, EBooks, Journalism, News, Photo Editing, Photo Sharing, Photography, Printing, Publishing, Social Bookmarking, Video Editing, Video Streaming
Data and Analytics	A/B Testing, Analytics, Application Performance Management, Artificial Intelligence, Big Data, Bioinformatics, Biometrics, Business Intelligence, Consumer Research, Data Integration, Data Mining, Data Visualization, Database, Facial Recognition, Geospatial, Image Recognition, Intelligent Systems, Location Based Services, Machine Learning, Market Research, Natural Language Processing, Predictive Analytics, Product Research, Quantified Self, Speech Recognition, Test and Measurement, Text Analytics, Usability Testing
Design	CAD, Consumer Research, Data Visualization, Fashion, Graphic Design, Human Computer Interaction, Industrial Design, Interior Design, Market Research, Mechanical Design, Product Design, Product Research, Usability Testing, UX Design, Web Design
Education	Alumni, Charter Schools, College Recruiting, Continuing Education, Corporate Training, E-Learning, EdTech, Education, Edutainment, Higher Education, Language Learning, MOOC, Music Education, Personal Development, Primary Education, Secondary Education, Skill Assessment, STEM Education, Textbook, Training, Tutoring, Vocational Education
Energy	Battery, Biofuel, Biomass Energy, Clean Energy, Electrical Distribution, Energy, Energy Efficiency, Energy Management, Energy Storage, Fossil Fuels, Fuel, Fuel Cell, Oil and Gas, Power Grid, Renewable Energy, Solar, Wind Energy
Events	Concerts, Event Management, Event Promotion, Events, Nightclubs, Nightlife, Reservations, Ticketing, Wedding

Financial Services	Accounting, Angel Investment, Asset Management, Auto Insurance, Banking, Bitcoin, Commercial Insurance, Commercial Lending, Consumer Lending, Credit, Credit Bureau, Credit Cards, Crowdfunding, Cryptocurrency, Debit Cards, Debt Collections, Finance, Financial Exchanges, Financial Services, FinTech, Fraud Detection, Funding Platform, Gift Card, Health Insurance, Hedge Funds, Impact Investing, Incubators, Insurance, InsurTech, Leasing, Lending, Life Insurance, Micro Lending, Mobile Payments, Payments, Personal Finance, Prediction Markets, Property Insurance, Real Estate Investment, Stock Exchanges, Trading Platform, Transaction Processing, Venture Capital, Virtual Currency, Wealth Management
Food and Beverage	Bakery, Brewing, Cannabis, Catering, Coffee, Confectionery, Cooking, Craft Beer, Dietary Supplements, Distillery, Farmers Market, Food and Beverage, Food Delivery, Food Processing, Food Trucks, Fruit, Grocery, Nutrition, Organic Food, Recipes, Restaurants, Seafood, Snack Food, Tea, Tobacco, Wine And Spirits, Winery
Gaming	Casual Games, Console Games, Contests, Fantasy Sports, Gambling, Gamification, Gaming, MMO Games, Online Games, PC Games, Serious Games, Video Games
Government and Military	CivicTech, Government, GovTech, Law Enforcement, Military, National Security, Politics, Public Safety, Social Assistance
Hardware	3D Technology, Application Specific Integrated Circuit (ASIC), Augmented Reality, Cloud Infrastructure, Communication Hardware, Communications Infrastructure, Computer, Computer Vision, Consumer Electronics, Data Center, Data Center Automation, Data Storage, Drone Management, Drones, DSP, Electronic Design Automation (EDA), Electronics, Embedded Systems, Field-Programmable Gate Array (FPGA), Flash Storage, Google Glass, GPS, GPU, Hardware, Industrial Design, Laser, Lighting, Mechanical Design, Mobile Devices, Network Hardware, NFC, Nintendo, Optical Communication, Playstation, Private Cloud, Retail Technology, RFID, RISC, Robotics, Roku, Satellite Communication, Semiconductor, Sensor, Sex Tech, Telecommunications, Video Conferencing, Virtual Reality, Virtualization, Wearables, Windows Phone, Wireless, Xbox
Health Care	Alternative Medicine, Assisted Living, Assistive Technology, Biopharma, Cannabis, Child Care, Clinical Trials, Cosmetic Surgery, Dental, Diabetes, Dietary Supplements, Elder Care, Electronic Health Record (EHR), Emergency Medicine, Employee Benefits, Fertility, First Aid, Funerals, Genetics, Health Care, Health Diagnostics, Home Health Care, Hospital, Medical, Medical Device, mHealth, Nursing and Residential Care, Nutraceutical, Nutrition, Outpatient Care, Personal Health, Pharmaceutical, Psychology, Rehabilitation, Therapeutics, Veterinary, Wellness
Information Technology	Business Information Systems, CivicTech, Cloud Data Services, Cloud Management, Cloud Security, CMS, Contact Management, CRM, Cyber Security, Data Center, Data Center Automation, Data Integration, Data Mining, Data Visualization, Document Management, E-Signature, Email, GovTech, Identity Management, Information and Communications Technology (ICT), Information Services, Information Technology, Intrusion Detection, IT Infrastructure, IT Management, Management Information Systems, Messaging, Military, Network Security, Penetration Testing, Private Cloud, Reputation, Sales Automation, Scheduling, Social CRM, Spam Filtering, Technical Support, Unified Communications, Video Chat, Video Conferencing, Virtualization, VoIP
Internet Services	Cloud Computing, Cloud Data Services, Cloud Infrastructure, Cloud Management, Cloud Storage, Darknet, Domain Registrar, E-Commerce Platforms, Ediscovery, Email, Internet, Internet of Things, ISP, Location Based Services, Messaging, Music Streaming, Online Forums, Online Portals, Private Cloud, Product Search, Search Engine, SEM, Semantic Search, Semantic Web, SEO, SMS, Social Media, Social Media Management, Social Network, Unified Communications, Vertical Search, Video Chat, Video Conferencing, Visual Search, VoIP, Web Browsers, Web Hosting
Lending and Investments	Angel Investment, Banking, Commercial Lending, Consumer Lending, Credit, Credit Cards, Financial Exchanges, Funding Platform, Hedge Funds, Impact Investing, Incubators, Micro Lending, Stock Exchanges, Trading Platform, Venture Capital
Manufacturing	3D Printing, Advanced Materials, Foundries, Industrial, Industrial Automation, Industrial Engineering, Industrial Manufacturing, Machinery Manufacturing, Manufacturing, Paper Manufacturing, Plastics and Rubber Manufacturing, Textiles, Wood Processing
Media and Entertainment	Advice, Animation, Art, Audio, Audiobooks, Blogging Platforms, Broadcasting, Celebrity, Concerts, Content, Content Creators, Content Discovery, Content Syndication, Creative Agency, Digital Entertainment, Digital Media, DRM, EBooks, Edutainment, Event Management, Event Promotion, Events, Film, Film Distribution, Film Production, Guides, In-Flight Entertainment, Independent Music, Internet Radio, Journalism, Media and Entertainment, Motion Capture, Music, Music Education, Music Label, Music Streaming, Music Venues, Musical Instruments, News, Nightclubs, Nightlife, Performing Arts, Photo Editing, Photo Sharing, Photography, Podcast, Printing, Publishing, Reservations, Social Media, Social News, Theatre, Ticketing, TV, TV Production, Video, Video Editing, Video on Demand, Video Streaming, Virtual World
Messaging and Telecommunications	Email, Meeting Software, Messaging, SMS, Unified Communications, Video Chat, Video Conferencing, VoIP, Wired Telecommunications

Mobile	Android, Google Glass, iOS, mHealth, Mobile, Mobile Apps, Mobile Devices, Mobile Payments, Windows Phone, Wireless
Music and Audio	Audio, Audiobooks, Independent Music, Internet Radio, Music, Music Education, Music Label, Music Streaming, Musical Instruments, Podcast
Natural Resources	Biofuel, Biomass Energy, Fossil Fuels, Mineral, Mining, Mining Technology, Natural Resources, Oil and Gas, Precious Metals, Solar, Timber, Water, Wind Energy
Navigation and Mapping	Geospatial, GPS, Indoor Positioning, Location Based Services, Mapping Services, Navigation
Other	#REF!
Payments	Billing, Bitcoin, Credit Cards, Cryptocurrency, Debit Cards, Fraud Detection, Mobile Payments, Payments, Transaction Processing, Virtual Currency
Platforms	Android, Facebook, Google, Google Glass, iOS, Linux, macOS, Nintendo, Operating Systems, Playstation, Roku, Tizen, Twitter, WebOS, Windows, Windows Phone, Xbox
Privacy and Security	Cloud Security, Corrections Facilities, Cyber Security, DRM, E-Signature, Fraud Detection, Homeland Security, Identity Management, Intrusion Detection, Law Enforcement, Network Security, Penetration Testing, Physical Security, Privacy, Security
Professional Services	Accounting, Business Development, Career Planning, Compliance, Consulting, Customer Service, Employment, Environmental Consulting, Field Support, Freelance, Intellectual Property, Innovation Management, Legal, Legal Tech, Management Consulting, Outsourcing, Professional Networking, Quality Assurance, Recruiting, Risk Management, Social Recruiting, Translation Service
Real Estate	Architecture, Building Maintenance, Building Material, Commercial Real Estate, Construction, Coworking, Facility Management, Fast-Moving Consumer Goods, Green Building, Home and Garden, Home Decor, Home Improvement, Home Renovation, Home Services, Interior Design, Janitorial Service, Landscaping, Property Development, Property Management, Real Estate, Real Estate Investment, Rental Property, Residential, Self-Storage, Smart Building, Smart Cities, Smart Home, Timeshare, Vacation Rental
Sales and Marketing	Advertising, Affiliate Marketing, App Discovery, App Marketing, Brand Marketing, Cause Marketing, Content Marketing, CRM, Digital Marketing, Digital Signage, Direct Marketing, Direct Sales, Email Marketing, Lead Generation, Lead Management, Local, Local Advertising, Local Business, Loyalty Programs, Marketing, Marketing Automation, Mobile Advertising, Multi-level Marketing, Outdoor Advertising, Personal Branding, Public Relations, Sales, Sales Automation, SEM, SEO, Social CRM, Social Media Advertising, Social Media Management, Social Media Marketing, Sponsorship, Video Advertising
Science and Engineering	Advanced Materials, Aerospace, Artificial Intelligence, Bioinformatics, Biometrics, Biopharma, Biotechnology, Chemical, Chemical Engineering, Civil Engineering, Embedded Systems, Environmental Engineering, Human Computer Interaction, Industrial Automation, Industrial Engineering, Intelligent Systems, Laser, Life Science, Marine Technology, Mechanical Engineering, Nanotechnology, Neuroscience, Nuclear, Quantum Computing, Robotics, Semiconductor, Software Engineering, STEM Education
Software	3D Technology, Android, App Discovery, Application Performance Management, Apps, Artificial Intelligence, Augmented Reality, Billing, Bitcoin, Browser Extensions, CAD, Cloud Computing, Cloud Management, CMS, Computer Vision, Consumer Applications, Consumer Software, Contact Management, CRM, Cryptocurrency, Data Center Automation, Data Integration, Data Storage, Data Visualization, Database, Developer APIs, Developer Platform, Developer Tools, Document Management, Drone Management, E-Learning, EdTech, Electronic Design Automation (EDA), Embedded Software, Embedded Systems, Enterprise Applications, Enterprise Resource Planning (ERP), Enterprise Software, Facial Recognition, File Sharing, IaaS, Image Recognition, iOS, Linux, Machine Learning, macOS, Marketing Automation, Meeting Software, Mobile Apps, Mobile Payments, MOOC, Natural Language Processing, Open Source, Operating Systems, PaaS, Predictive Analytics, Presentation Software, Presentations, Private Cloud, Productivity Tools, QR Codes, Reading Apps, Retail Technology, Robotics, SaaS, Sales Automation, Scheduling, Sex Tech, Simulation, SNS, Social CRM, Software, Software Engineering, Speech Recognition, Task Management, Text Analytics, Transaction Processing, Video Conferencing, Virtual Assistant, Virtual Currency, Virtual Desktop, Virtual Goods, Virtual Reality, Virtual World, Virtualization, Web Apps, Web Browsers, Web Development
Sports	American Football, Baseball, Basketball, Boating, Cricket, Cycling, Diving, eSports, Fantasy Sports, Fitness, Golf, Hockey, Hunting, Outdoors, Racing, Recreation, Rugby, Sailing, Skiing, Soccer, Sporting Goods, Sports, Surfing, Swimming, Table Tennis, Tennis, Ultimate Frisbee, Volley Ball
Sustainability	Biofuel, Biomass Energy, Clean Energy, CleanTech, Energy Efficiency, Environmental Engineering, Green Building, Green Consumer Goods, GreenTech, Natural Resources, Organic, Pollution Control, Recycling, Renewable Energy, Solar, Sustainability, Waste Management, Water Purification, Wind Energy

Transportation	Air Transportation, Automotive, Autonomous Vehicles, Car Sharing, Courier Service, Delivery Service, Electric Vehicle, Ferry Service, Fleet Management, Food Delivery, Freight Service, Last Mile Transportation, Limousine Service, Logistics, Marine Transportation, Parking, Ports and Harbors, Procurement, Public Transportation, Railroad, Recreational Vehicles, Ride Sharing, Same Day Delivery, Shipping, Shipping Broker, Space Travel, Supply Chain Management, Taxi Service, Transportation, Warehousing, Water Transportation
Travel and Tourism	Adventure Travel, Amusement Park and Arcade, Business Travel, Casino, Hospitality, Hotel, Museums and Historical Sites, Parks, Resorts, Timeshare, Tour Operator, Tourism, Travel, Travel Accommodations, Travel Agency, Vacation Rental
Video	Animation, Broadcasting, Film, Film Distribution, Film Production, Motion Capture, TV, TV Production, Video, Video Editing, Video on Demand, Video Streaming

Appendix B Theoretical Framework

The appendix shows an underlying theoretical framework on which our proposed mechanism for investor portfolio preference is based. The model shows how an investor with a geographic restriction creates her investment portfolio with startup firms. The physical position of the investor is denoted as X and the business similarity of her portfolio is S. We assume that the investor is risk neutral and that there is no strategic interaction among investors (Fulghieri and Sevilir, 2009), moral hazard (Chemmanur and Chen, 2014), and switching investor types between VCs or angels (Hellmann and Thiele, 2015).

The objective function for the investor includes the term that captures her cost of efforts to overcome the physical distance between herself and an invested firm and the knowledge gap between the industry of her expertise and the industry of the invested firm. Using a quadratic function to make the cost convex, the cost term for each project (firm) i is

$$-\{\alpha_f (X_i - X_f)^2 + \beta_f (S_i - S_f)^2\},$$

where (X_f, S_f) are the geographic position and the industry expertise of the investor, and (X_i, S_i) are the geographic position and the industry of firm i for funding. The investor cares about the relative importance between geographic distance α_f and business similarity β_f. Instead of considering a maximization of expected net payoff, we simplify the setup by directly assuming a binding constraint that the investor must spend a fixed budget on her portfolio with as many projects around (X_f, S_f) as possible.

Suppose projects are uniformly distributed in both X and S dimensions so that in each $dS \cdot dX$ area there is an equal number of projects. Then, it can be normalized to one per unit of the area. If the projects are identical except for the distance and expertise dimensions, the search-area boundary of the investor with some positive ς of the maximum cost can be written as

$$\alpha_f (X_i - X_f)^2 + \beta_f (S_i - S_f)^2 < \varsigma,$$

with the ellipse around (X_f, S_f),

$$\frac{(X_i - X_f)^2}{\varsigma/\alpha_f} + \frac{(S_i - S_f)^2}{\varsigma/\beta_f} = 1,$$

meaning that the investor searches for close and familiar projects near her location. The diameters of the ellipse are $A = \sqrt{\varsigma/\alpha_f}$ and $B = \sqrt{\varsigma/\beta_f}$, and the area inside the ellipse is

$$\pi AB = \pi \frac{\varsigma}{\sqrt{\beta_f \alpha_f}}.$$

Then, the number of projects to be financed is the same as the number of projects inside the search area, which is proportional to $(\beta_f \alpha_f)^{-1/2}$. The number of financed projects by the investor is greater when the preferences for the distance and similarity dimensions are the same as $\beta_f = \alpha_f$ (a circle). If the number of projects in any investor's portfolio is fixed as C, we can therefore derive the following condition:

$$\frac{\varsigma}{\beta_f \alpha_f} = C$$

When α_f becomes large, the geographic distance becomes more important than business expertise as for angel investor relative to VC. It follows from the above condition that β_f has to become proportionally smaller to keep $\beta_f \alpha_f$ constant. In other words, the investor with large α_f must learn about dissimilar projects when the physical distance cannot be decreased. Visually, if the search area ellipse is flatter in one dimension, it must become longer in another dimension to keep the search area the same.

The investor's tolerance for X or S dimension along the search boundary can be found by implicit differentiation. That is,

$$\frac{d(\delta X)}{d(\delta S)} = -\frac{2\beta_f}{2\alpha_f},$$

is her tolerance for δX on the boundary when the constraint in S dimension is relaxed by δS. We note that this tolerance measure is negative indicating that when the business dissimilarity of financed projects increases, the search must be conducted closer in geographic distance and vise versa. The magnitude of this offsetting effect is proportional to the investor's relative sensitivity $\frac{\beta_f}{\alpha_f}$.

Next, we relax the assumption on the risk-neutral investor (*i.e.,* the investor is risk averse). It is reasonable to assume that project payoffs are correlated according to their business similarity. Then, the covariance between projects i and j becomes:

$$cov(\tilde{\theta}_i, \tilde{\theta}_j) = \mathbb{E}(\tilde{\theta}_i \tilde{\theta}_j) - \mathbb{E}(\tilde{\theta}_i)\mathbb{E}(\tilde{\theta}_j) \propto \frac{k}{|S_i - S_j|}$$

with the assumption of $\mathbb{E}(\tilde{\theta}_i) = \mathbb{E}\theta =$ constant for all i and $k >= 0$. The covariance also can be negative if we allow the negative correlation between the projects indicating a hedging effect. The variance of the portfolio of N projects with each project's cost of c_i is

$$\begin{aligned} var\left[\frac{1}{N}\sum_i(\tilde{\theta}_i - c_i)\right] &= \frac{var(\tilde{\theta}_i)}{N} + \frac{1}{N^2}\sum_{i,j} cov(\tilde{\theta}_i, \tilde{\theta}_j) \\ &= \frac{var(\tilde{\theta}_i)}{N} + \frac{(N-1)!}{N^2}\left[\frac{k}{|S_i - S_j|} - [\mathbb{E}\theta]^2\right] \end{aligned}$$

From the above, we note that the portfolio variance decreases with the dissimilarity between projects ($|S_i - S_j|$) and also with the total number of projects in the portfolio (N), while it increases with the individual project variance ($var(\tilde{\theta}_i)$). Therefore, risk-averse angel investors should take many projects and preferably different projects to diversify given the higher tolerance for S dimension.[1]

[1]It is also possible to allow that angel investors are more risk averse than VCs. If so, angel investors' preference for dissimilar projects becomes even stronger.

Appendix C Variable Definition

Variable Name	Definition
Similarity	A firm-MSA-year level business similarity measure computed by taking the average of cosine similarity scores between the focal firm and each of the rest of firms in a given MSA-year using 742 Crunchbase industry categories.
Industry Similarity	A within-investor industry similarity of portfolio firms.
Angel	An indicator variable equal to one if given investor's Crunchbase investor type includes angel and zero otherwise.
VC	An indicator variable equal to one if given investor's Crunchbase investor type includes venture capital and zero otherwise.
Ln(Firms)	The natural logarithm of the number of firms in Crunchbase.
Ln(Investors)	The natural logarithm of the number of investors in Crunchbase.
Ln(Net Jobs)	The natural logarithm of the number of net job created in an MSA from the Business Dynamics Statistics (BDS) from the public Census data.
Ln(Personal Income)	The natural logarithm of per capita personal income obtained from the Bureau of Economic Analysis.
Entrepreneurship Quality Index (EQI)	MSA-year level average entrepreneurial quality obtained from Andrews et al. (2019).
Regional Entrepreneurship Cohort Potential Index (RECPI)	MSA-year level expected number of growth events given the start-up characteristics of a cohort at birth obtained from Andrews et al. (2019).
Regional Ecosystem Acceleration Index (REAI)	MSA-year level ratio of the realized growth events to expected growth events (RECPI) obtained from Andrews et al. (2019).
Angel Tax	An indicator variable is equal to one if a firm resides in a state where there is angel tax credit (Denes et al., 2020). *AngelTax* is a continuous variable of angel tax credit percentages.
Investor Type	An indicator variable equal to one if funding is raised from a given investor type and zero otherwise.
%Investor Type	The average fraction of the funding amounts received from the given investor type, weighted by the size of each funding round in a year.
%Funding	The fraction of the total funding amounts received from the given investor type in a year.
$ Total Funding	The total funding amount received in a year (in millions).
Geographic Distance	A distance in mile between an investor and funded firm.
Ln(Investments)	The natural logarithm of the total number of investments of a given investor.
Ln(Investing Years)	The natural logarithm of the number of years since the first investment record.
Subsequent Funding	An indicator variable equal to one if a given firm has record of any type of subsequent funding.
Exit Rate-IPO (Acq)	The average IPO (Acquisition) exit rate among all funded firms for given investor.
Close	An indicator variable equal to one if an investor's portfolio firms consist of below median average distance between the investor and portfolio firms and zero otherwise.
Dissimilar	An indicator variable equal to one if an investor's portfolio consists of firms with below median average similarity and zero otherwise.
Close&Dissimilar	An indicator variable in the intersection of *Close* and *Dissimilar*