JOURNAL OF FINANCIAL AND QUANTITATIVE ANALYSIS Vol. 55, No. 2, Mar. 2020, pp. 621–651
COPYRIGHT 2019, MICHAEL G. FOSTER SCHOOL OF BUSINESS, UNIVERSITY OF WASHINGTON, SEATTLE, WA 98195
doi:10.1017/S002210901900005X

https://doi.org/10.1017/S002210901900005X Published online by Cambridge University Press

Venture Capital Communities

Amit Bubna, Sanjiv R. Das, and Nagpurnanand Prabhala*

Abstract

Although venture capitalists (VCs) can choose from thousands of potential syndicate partners, many co-syndicate with small groups of preferred partners. We term these groups "VC communities." We apply computational methods from the physical sciences to 3 decades of syndication data to identify these communities. We find that communities comprise VCs that are similar in age, connectedness, and functional style but undifferentiated in spatial location. Machine-learning tools classify communities into 3 groups roughly ordered by their age and reach. Community VC financing is associated with faster maturation and greater innovation, especially for early-stage firms without an innovation history.

I. Introduction

Venture capitalists (VCs) raise capital from wealthy individuals and institutional investors and invest it in young firms with promising upsides. According to the National Venture Capital Association, as of Dec. 2016, there were 898 venture capitalists and 1,562 funds in the United States, with $333.5 billion in assets under management. The top 5 successes of VC investing, Alphabet, Amazon, Apple, Facebook, and Microsoft, have 543,443 employees and have a combined market capitalization of $2.28 trillion.

Firms financed by VCs tend to be young and risky and have unproven business models. VCs use many strategies to manage the risks and resource demands

*Bubna, amitbubna@gmail.com, Cornerstone Research; Das, srdas@scu.edu, Santa Clara University Leavey School of Business; and Prabhala (corresponding author), prabhala@jhu.edu, Johns Hopkins University and University of Maryland College Park. We are very grateful to Jarrad Harford (the editor) and an anonymous referee for extensive and thoughtful feedback. We thank Alexandre Baptista, David Feldman, Jiekun Huang, Ozgur Ince, Vladimir Ivanov, Pete Kyle, Josh Lerner, Laura Lindsey, Vojislav Maksimovic, Robert Marquez, Manju Puri, Krishna Ramaswamy, Rajdeep Singh, Richard Smith, Anjan Thakor, Susan Woodward, and Bernard Yeung for their comments. We also thank participants at the 2011 ISB CAF, 2011 Financial Intermediation Research Society, 2013 Midwest Finance Association, 2011 World Private Equity, and 2012 T. A. Pai Management Institute (TAPMI) conferences for their helpful comments; and we thank the seminar participants at BlackRock Inc., University of Florida, George Washington University, Georgia State University, Georgia Tech University, the Indian School of Business, Northwestern University, the University of Maryland, College Park, National University of Singapore, the Bay Area R User Group, Rutgers University, Tulane University, the University of New South Wales, and Villanova University for helpful comments.

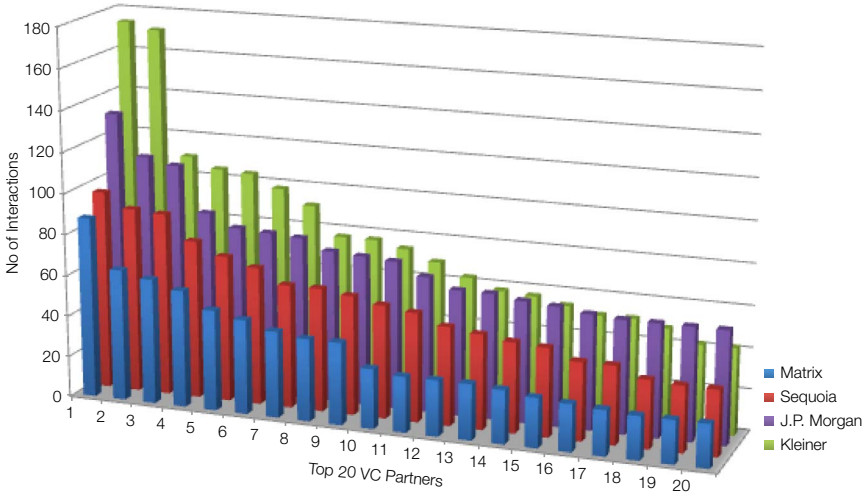
placed by such investments. Perhaps the most common strategy is to *syndicate* deals, or coinvest in portfolio firms in partnership with other VC firms. Virtually all VCs enter into syndications. Only 5% of VCs never syndicate, and these are small, peripheral firms.[1]

We study the VC syndication process. In the research in this area, the unit of study is typically a pair of VCs that join together to syndicate.[2] Our study departs from this focus by examining *groups* of VCs that we call "VC communities." Formally, communities are groups of VCs with the property that group members have greater propensities to enter into syndicate partnerships with each other than with others outside the group.

To motivate the community construct, consider Figure 1, which plots the histogram for the partners of 4 large VCs. A thick left tail indicates that VCs strongly prefer some partners more than others. In our sample, JP Morgan's preferred partners include Kleiner Perkins, Oak Investment Partners, and the Mayfield Fund.[3]

FIGURE 1
Distribution of the Number of Interactions of 4 Top Firms with Their Top 20 Collaborators

Figure 1 displays the distribution of the number of partners for the 4 venture capital firms with the greatest number of deals in our sample.



[1]Other strategies for managing the investment process include rigorous ex ante screening, ex post monitoring, and advising on strategy (Gorman and Sahlman (1989), Hellmann and Puri (2002)). VCs also incorporate security design features, such as priority, staged financing and contracting over control rights and governance (Cornelli and Yosha (2003), Neher (1999), Kaplan and Stromberg (2003), Robinson and Stuart (2007), and Robinson and Sensoy (2011)). See Da Rin, Hellmann, and Puri (2013) for a survey of the academic literature on venture capital.

[2]See Lerner (1994), Cestone, Lerner, and White (2006), Bhagwat (2013), Gompers, Mukharlyamov, and Xuan (2012), and Hochberg, Lindsey, and Westerfield (2015).

[3]Anecdotal evidence for such partner preferences includes the comment by Fred Wilson of Union Square Ventures, a prominent investor in major social networking sites such as Tumblr, Twitter, and Zynga. He says, "there are probably five or ten VCs who I have worked with frequently in my career and I know very well and love to work with. It's not hard to figure out who they are" (http://www.avc.com/a_vc/2009/03/coinvestors.html).

https://doi.org/10.1017/S002210901900005X Published online by Cambridge University Press

https://doi.org/10.1017/S002210901900005X Published online by Cambridge University Press

Figure 1 also shows a thin and long tail. Thus, community VCs do co-syndicate outside their preferred-partner group but infrequently so. Alternatively, the preference for community partners is probabilistic, not deterministic.

Our study has two aims. Our first objective is to introduce the concept of a VC community. We discuss the economic forces that lead to community formation and discuss the underlying mathematics. We show that communities are a solution to a well-defined, although difficult-to-solve, mathematical problem and discuss the unusual computational approach involved in identifying communities. We implement the methods to identify communities from a large sample of VC syndications. We then address several questions of economic interest. For example, we examine the nature of community VCs and the types of realized communities, and we investigate the association between community VC financing and two economic outcomes: the maturation of portfolio firms and their innovation activities.

Our first goal is to detect communities from observed syndication data. Formally, a community is a group of VCs whose members are statistically more likely to co-syndicate with each other than with other VCs outside the group. Detecting communities is like a cluster-identification problem, but it is more complex. The complexity arises because, unlike standard clustering problems, we impose few restrictions on the cluster structure. For example, we do not fix the number of clusters, the size of each cluster, and the number of VCs that belong or do not belong in the clusters. We optimize over these features using a nonstandard maximand, the propensity of VCs to cluster relative to what is expected by chance. The additional features result in a reasonably realistic representation of syndication but also result in a far more complex problem that requires nonstandard solution techniques (Fortunato (2009), Pons and Latapy (2005)).

Our data comprise U.S. VC syndications between 1980 and 2010. We detect communities based on VC financings over 5-year periods with rolling windows moving forward 1 year at a time. A portfolio firm enters our sample only once in each 5-year period. On average, 11% of the VCs in each 5-year period belong to communities. The median community has 8 members. A VC firm's community status is stable. A VC belonging to a community has an 88% probability of belonging to a community 5 years later. We find a pecking order among VCs. Community VCs are the largest, oldest, and most active VCs and tend to finance large firms. VCs that are neither in communities nor partnered with those in communities tend to be small, young VCs. VCs that only partner with community VCs but do not themselves belong to communities display intermediate characteristics. We also consider the composition of VC communities. We ask whether VCs within communities are similar or dissimilar to each other in terms of several characteristics. Dissimilarity indicates that VCs prefer partners with complementary characteristics, whereas similarity is consistent with theories of contracting with private and manipulable signals or homophilous preferences.[4] We find that communities tend to comprise VCs that are similar to each other.

[4] See Cestone et al. (2006), McPherson, Smith-Lovin, and Cook (2001), and Hochberg et al. (2015). However, Hochberg et al. point out that in the case of syndicate pairs, similarity in partner characteristics is necessary but not sufficient to conclude birds-of-a-feather homophily.

https://doi.org/10.1017/S002210901900005X Published online by Cambridge University Press

Perhaps a more interesting question concerns the differences *between* realized communities. We use machine-learning tools to understand these differences. Over the sample period, there are 413 communities. Unsupervised learning methods partition communities into three types. Type 1 communities are what we call "young" or "specialist" communities that comprise young, small VC firms that tend to have low overall centrality (i.e., fewer connections to the overall VC industry). These VCs have concentrated portfolios by region, type of firm, and industry, and they typically finance small firms. A second cluster includes "mature" communities that tend to include the VCs that are the oldest, the largest, and the most networked. The financiers in this cluster tend to fund large firms across multiple industries. The third cluster includes VCs with characteristics somewhere between those of clusters 1 and 2. These groups indicate a pecking order among communities going from clusters that pool large, mature VCs to those that pool emerging VCs. Supervised learning methods identify the characteristics that discriminate between the clusters and show a 93.9% correct classification rate.

We then examine the association between sourcing financing from community VCs and economic outcomes. We note that these outcomes could reflect ex ante selection effects or the ex post real impact of community VCs. We do not disentangle these two sources of skill.[5] The first outcome is patenting, which reflects the production of innovation. Sourcing financing from community VCs is associated with an increase of between 2% and 6% in the number of patents and a 5%–17% increase in citations per patent. These effects display heterogeneity. For example, they are more economically and statistically significant for early-stage portfolio firms with no prior innovation that likely face more ambiguity, so community membership helps more. Because communities can be viewed as alliances with soft boundaries and implicit contracts, the results support the view that alliances foster innovation (Gonzalez-Uribe (2014), Seru (2014)).

A second outcome is exit, which reflects the maturation of portfolio firms. Firms may exit via initial public offerings (IPOs), which is a signature of success, or through mergers and acquisitions (M&As), which reflect a mix of both failures and success stories, such as Microsoft's acquisition of Skype. We estimate a model in which exits are only through IPOs, a model in which exits are through M&As or IPOs, and a model in which M&As are competing hazards for IPO exits. Financing rounds in which capital is sourced from community VCs are 9% more likely to successfully exit via IPO or acquisition and do so sooner by approximately 13%.

We proceed as follows: Section II formally defines communities and develops the related optimization problem and solution methods. Section III discusses the data. In Section IV, we describe the implementation of community-detection methods on the VC data and discuss the baseline results. Section V examines the attributes of VCs within communities and also the different types of realized communities identified in our study. In Section VI, we examine the detected communities using supervised and unsupervised machine-learning techniques that elicit

[5] As Sorensen (2007) illustrates, the techniques to disentangle these two effects are challenging even when modeling pairwise syndications. The structural modeling of communities, which are groups derived from pairwise ties, is even more difficult. We leave this for future work.

the structure of these communities and their differences. Section VII studies the relation between communities and two economic outcomes, innovation and exit. Section VIII concludes and suggests directions for future research. The Appendix gives some sample code and defines the variables used in our analysis.

II. Community Definition and Detection

A community is essentially a group of VCs that have strong propensities to join each other in syndicates. In this section, we explain why such groups could form and then turn to their mathematical definition and their detection from pairwise syndication data.

A. Economic Intuition

Economic theory explains why VCs cluster into preferred-partner groups. Goldfarb, Kirsch, and Miller (2007) and Sorensen (2008) argue that VC investing is a complex process that involves unfamiliar terrain. Thus, investing skills must often be acquired through a learning-by-doing approach. Using familiar partners can aid such learning, for example, through greater familiarity with partner decision-making norms and processes (Gertler (1995), Porter (2000)) or through greater trust, reciprocity, and social capital (Granovetter (1985), Gulati (1995), Guiso, Sapienza, and Zingales (2004), Bottazzi, Da Rin, and Hellmann (2016)) that can help in incomplete contracting situations (Grossman and Hart (1986), Hart and Moore (1990)).

However, syndicating *exclusively* with a small number of VCs is not necessarily optimal. Although exclusivity generates more social capital, it can generate groupthink and limit access to new knowledge (Granovetter (1973)). As Uzzi (1997) argues, something in between where agents have a few strong ties but also other weaker ties is likely optimal. This is, of course, the definition of a VC community.

B. Community Mathematics

We suppress the time-period subscripts t for compactness. Suppose that there are N VCs and K communities, with the kth community comprising n_k VCs; then $N_k = \sum_{k=1}^{K} n_k$ VCs belong to communities, and $N - N_k$ do not. Let c_k denote community k and $\delta_{ij}(c_k)$ be an indicator variable equal to 1 when both VC i and VC j belong to community k, and 0 otherwise.

Community detection chooses the number of groups K, the size of each group n_k, and the set of indicator variables $\delta_{ij}(c_k)$ to maximize *modularity*, Q, where

$$Q = \frac{1}{2m} \sum_{k} \sum_{i,j} \left[a_{ij} - \frac{d_i \times d_j}{2m} \right] \times \delta_{ij}(c_k), \tag{1}$$

where the dummy variable $\delta_{ij}(c_k)$ equals 1 when i and j belong to community c_k, and 0 otherwise. The first term in square brackets in equation (1) is the observed number of syndications between VCs i and j, and the second term is the number

https://doi.org/10.1017/S002210901900005X Published online by Cambridge University Press

https://doi.org/10.1017/S002210901900005X Published online by Cambridge University Press

of syndications between the two that would occur by chance.[6] Modularity is the sum of this difference across all communities, or the difference between the actual incidence of in-community syndications minus its expected value across all communities. When the modularity is high, the number of actual syndication ties between the VCs in communities exceeds what is expected by chance, whereas low modularity indicates that syndications are less likely relative to chance.

The modularity-maximization problem is computationally complex. A brute-force method, which involves searching over all possible partitions of the set of VCs into groups, is computationally infeasible because of the large number of ways in which groups of varying sizes can be formed. Such a problem has no known analytic solutions beyond tiny systems. We use the walk-trap method, which initiates simultaneous random walks at several nodes and takes random steps from each. Communities are clusters from which the random walks fail to exit within a fixed number of steps (Fortunato (2009), Pons and Latapy (2005)). The intuition is that if a trial cluster is a community, a step away from any VC in the walk-trap algorithm should lead to another VC within the same cluster. Thus, when multiple steps lead to VCs within the same cluster, this cluster is likely to be a community.

C. Communities as Social Network Metrics

Community membership is a type of social network metric. In this regard, it is related to but very different from social network metrics such as centrality that are now widely used in the finance literature (Hochberg, Ljungqvist, and Lu (2007), Fracassi and Tate (2012), Engelberg, Gao, and Parsons (2013)).[7]

Social network metrics start with one input, the adjacency matrix that lists the pairwise ties between individuals. However, there is little that is common between the construct of community and measures of influence such as centrality. Operationally, the two are constructed in very different ways. Community detection solves the modularity-optimization problem in equation (1). Centrality is essentially a count or weighted count of the number of connections. The economic intuitions underlying the two constructs are also different. Community detection identifies a set of different-sized groups of VCs and VCs that do not belong to groups. Centrality is an individual attribute capturing a VC's influence or reach. The two are fundamentally different concepts.[8]

A related question is that of the advantage of analyzing communities, which are groups derived from pairwise ties, rather than pairwise ties themselves.[9]

[6]The factor of 2 in the second term reflects that the ties between i and j and the ties between j and i should not be double counted.

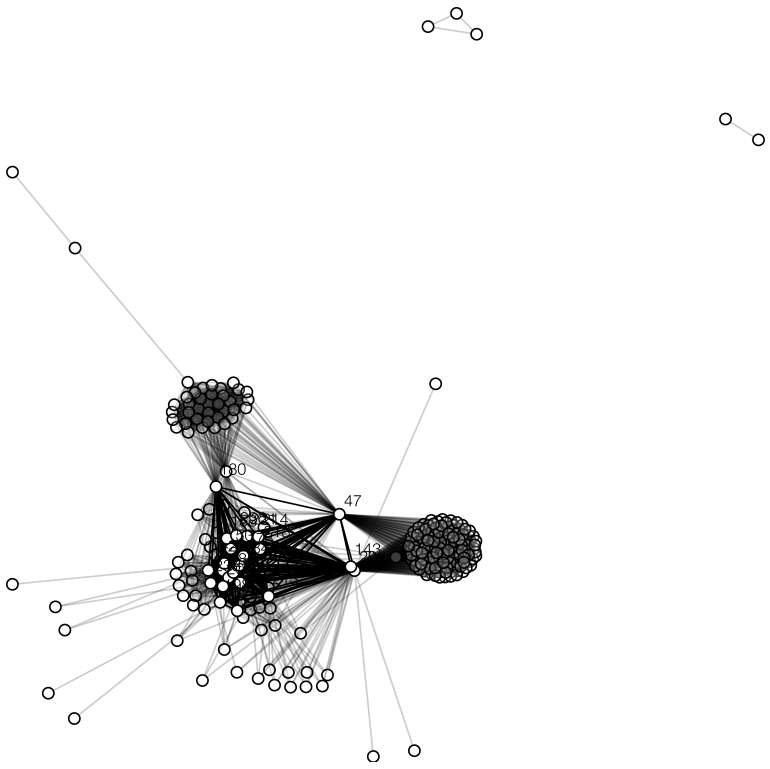
[7]For a textbook treatment of these metrics, see, for example, http://faculty.ucr.edu/~hanneman/nettext/.

[8]For further discussion of these distinctions, see Sections 7.1, 7.2, and 11.6 of Newman (2010). Applications of community detection include, for instance, work on identifying politicians who vote together (Porter, Mucha, Newman, and Friend (2007)), product word groups (Hoberg and Phillips (2010)), and collaboration networks (Newman (2001)). Consider also Figure 2, reproduced from Burdick et al. (2011). The three banks with high centrality are Citigroup, JP Morgan, and Bank of America. These banks do not belong to communities, which are in the left and right nodes of Figure 2.

[9]Work on pairwise ties and their relation to economic outcomes includes Cohen, Frazzini, and Malloy (2010), (2012), Hwang and Kim (2009), Chidambaran, Kedia, and Prabhala (2010), and in

FIGURE 2
Communities and Centrality in Bank Co-Lending Networks

Figure 2 depicts the network structure in the syndicated loans. The network shows 3 banks that are central to the network (banks 47, 143, and 180) but that are outside communities, which are banks that tend to co-syndicate together and are represented as clusters in the figure. See Burdick et al. (2011).



We make two points. One is a practical econometric point. Given that VCs number in the thousands, for even *one* realized pairwise tie, there are millions of counterfactual unrealized ones. Moreover, the specification of pairwise ties likely features complex dependence patterns. For example, each new choice of a syndicate partner likely depends on past interactions between a VC and all other VCs, the outcomes from the interactions, and the nature of the new financing opportunity on hand. These forces create a complex set of temporal and cross-sectional dependence patterns and errors that are not captured by simple independent and identically distributed (IID) logit-type specifications.[10] Community detection turns the pairwise approach on its head. It takes as its primitive the groups formed from the

the VC context, Bhagwat (2013) and Gompers et al. (2012). See also Bengtsson and Hsu (2010) and Hegde and Tumlinson (2014).

[10]Simpler network-formation models are difficult to estimate (e.g., Currarini, Jackson, and Pin (2012)). A structural model for community formation can perhaps be derived from the theory of "preferential attachment" (Barabasi and Albert (1999)), which argues that well-connected nodes attract connections to even more nodes. Adding the feature that familiarity aids learning and thus increases the likelihood of working together potentially creates a model for community formation.

https://doi.org/10.1017/S002210901900005X Published online by Cambridge University Press

pairwise ties and inverts these revealed preferences of VCs to understand the tie formation.

Our second point is economic. Community detection is not just an alternative estimation technique for tie formation. It identifies groups. Understanding groups and characterizing them are topics of economic interest in their own right.

III. Data

Data on venture capital syndications from 1980 to 2010 come from Thomson's Venture Economics (or VentureXpert) database. We start in 1980 because it is the year in which the VC industry was institutionalized (Gompers and Lerner (2001)). We end in 2010 to allow sufficient time to observe investment outcomes. We follow Du (2011) and Tian (2011) in addressing data issues regarding a given financing round appearing multiple times or being split up. We start with an initial sample of 10,677 unique VC firms. To concentrate solely on investments by U.S.-based VC funds, we refine the initial sample from VentureXpert by dropping non-U.S. VCs and buyout funds, which reduces the VC sample to 5,458 VCs. We limit our analysis to VC firms' U.S. portfolio companies, which further reduces the VC sample to 5,300. We include all deals in which there is at least one identified investor that is not an individual investor (e.g., an angel or management). This step reduces the sample to 3,926 VC firms. We exclude VCs that have only one financing round. The final sample has 3,275 unique VC firms.

We do *not* exclude deals that involve institutions such as subsidiaries of financial institutions and technology-transfer offices of universities. Although these investors have different incentive schemes, there are two reasons to keep them in the sample. First, the deals they finance involve traditional institutional VCs. Second, each syndicated deal, whether between institutional VCs or by institutional VCs with others, offers VCs an opportunity to interact and learn through the deal or transitively through deals with partners of noninstitutional VCs. Interaction and learning are the essence of community formation.

Table 1 shows that our sample includes 3,275 unique VC firms. On average, a VC invests in 24 portfolio firms and 56 rounds. Each round involves investments of $4.25 million. VC headquarters are spread across 153 Metropolitan Statistical Areas (MSAs) in our sample, with a mean (median) of approximately 20 (4) VCs per MSA. California (CA) and Massachusetts (MA) account for approximately 35% of the VCs' headquarters.[11] The mean (median) amount of capital raised by a VC equals $472 million ($60 million). Close to three-quarters of the rounds are syndications, and approximately one-third are flagged as early stage. A VC's age at investment is close to 11 years.

Data on patenting, a measure of innovation activity, come from the National Bureau of Economic Research (NBER) patent database (Hall, Jaffe, and Trajtenberg 2001). The data set includes 3 million utility patents granted by the U.S. Patent and Trademark Office from 1976 to 2006. There are approximately 218,000 unique assignees, of which 113,000 are located in the United States and account for approximately 1.4 million patents. To match the data with the VC database,

[11] We do not include satellite offices in the analysis.

https://doi.org/10.1017/S002210901900005X Published online by Cambridge University Press

TABLE 1
VCs in Our Sample

Table 1 provides descriptive statistics for the 3,275 unique U.S.-based venture capitalists (VCs) in our database over the entire 31-year period, from 1980 to 2010. Data are from Venture Economics and exclude non-U.S. investments, angel investors, and VC firms that focus on buyouts. We report the number of rounds of financing (NO_OF_ROUNDS) and the count of portfolio companies a VC invests in (NO_OF_COMPANIES). INVESTMENT_PER_ROUND is the amount a VC invests in a round. %_DEALS_SYNDICATED is the number of a VC's syndicated rounds as a percentage of all rounds that a VC invested in. %_EARLY_STAGE_DEALS is the number of a VC's investment rounds classified by Venture Economics as early stage as of the round-financing date, as a percentage of all Venture Economics deals for the VC between 1980 and 2010. VC_AUM is the sum of the capital under management of a VC in all funds that invested during 1980–2010. TOTAL_INVESTMENT is the sum of a VC's investments over this time period. AGE is defined as the difference in the year of the VC's last investment in the period 1980–2010 and the VC firm's founding date. NO_OF_VC_FIRMS_PER_MSA is the total number of unique VCs headquartered in a metropolitan statistical area (MSA). CA/MA_VC is the fraction of all VCs that are headquartered in either California or Massachusetts.

Variables	Mean	Median	No. of Obs.
NO_OF_ROUNDS	56.11	11.00	3,275
NO_OF_COMPANIES	24.14	8.00	3,275
INVESTMENT_PER_ROUND ($millions)	4.25	2.74	3,252
%_DEALS_SYNDICATED	0.73	0.80	3,275
%_EARLY_STAGE_DEALS	0.30	0.27	3,275
VC_AUM ($millions)	472.17	60.00	1,859
TOTAL_INVESTMENT ($millions)	285.26	32.73	3,252
AGE	11.50	9.00	3,275
NO_OF_VC_FIRMS_PER_MSA	19.95	4.00	153
CA/MA_VC	0.35	0.00	3,275

we standardize names and match by both names of assignees and location. If there is no exact name match, we use a fuzzy matching algorithm that we refine through manual checks. To account for name changes due to mergers, we use the pre- and post-merger names of firms if provided in VentureXpert and examine additional data from the BusinessWeek Web site. We have approximately 7,000 unique name matches that account for approximately 100,000 unique patents.

VC firms can exit through M&As or through IPOs. The IPO data come from Thomson Financial's SDC Platinum. We match portfolio firms by their Committee on Uniform Securities Identification Procedures (CUSIP) identifiers, cross-check the matches against actual names, and further hand-match the names with those in the VentureXpert database. In our sample, 2,545 ventures in our sample exit via IPOs. We obtain M&A data from Thomson Financial's SDC M&A database. There are 5,106 exits via mergers in our sample.

IV. Baseline Results

A. Community Detection

We implement the community-detection algorithm on syndications drawn from 5-year overlapping windows. These windows allow sufficient time to identify preferences for particular partners and avoid excessively long periods containing stale information. We require a minimum community size of 5 members and that the end-to-end diameter not exceed one-fourth that of the entire network.[12]

[12] We choose 5 as a minimum size for a community. In making this choice, we note that the median number of VCs per syndicate is 2. We choose the minimum community size to be more than 2 so that communities represent something other than a syndicate. At the same time, we do not set the minimum size so large that this empirical choice drives the results. The choice of 5 as the minimum community size also returns roughly the same number of communities regardless of the step size used in implementing the walk-trap algorithm. The diameter is the longest of the shortest paths from one node to another within a community. Its restriction is not ex post binding in our data set.

https://doi.org/10.1017/S002210901900005X Published online by Cambridge University Press

We count multiple links among VCs with any portfolio firm just once in each 5-year window.

Table 2 shows that there are 7–25 communities in each 5-year window. Over the whole sample period, the median community has 8 members, and on average, approximately 11% of the active VCs belong to communities. Figure 3 shows that both the number of community VCs and the percentage of VCs classified as community VCs vary over time. To check whether the time-series variation is correlated with market conditions, we identify hot and cold periods as years when the total VC investment deflated by the gross domestic product (GDP) price deflator is in the top and bottom quartiles of the full-period distribution, respectively. We find that the number of community VCs is significantly greater in hot periods (157 VCs on average) than in cold periods (81 VCs on average).[13]

TABLE 2
Stability of Community Status

Table 2 reports the number of venture capitalists (No. of VCs); the number of communities (No. of Comm.); the number of VCs that belong to communities (No. of Comm. VCs); and the fraction of VCs that remain in a community after 1, 3, and 5 years from the initial window. The numbers in parentheses are for the case when only first-round VCs are considered.

				% In Communities After		
Window	No. of VCs	No. of Comm.	No. of Comm. VCs	1 Year	3 Years	5 Years
1980–1984	814	11 (7)	87 (50)	1 (1)	0.98 (1)	0.93 (0.92)
1981–1985	906	10 (10)	92 (65)	1 (1)	0.97 (0.97)	0.93 (0.92)
1982–1986	959	10 (11)	101 (76)	1 (1)	0.97 (0.91)	0.92 (0.86)
1983–1987	969	15 (11)	136 (81)	0.99 (0.94)	0.96 (0.89)	0.90 (0.84)
1984–1988	968	15 (7)	139 (62)	0.98 (0.98)	0.93 (0.98)	0.81 (0.92)
1985–1989	936	18 (5)	147 (55)	0.99 (1)	0.93 (1)	0.82 (0.93)
1986–1990	889	11 (7)	112 (60)	0.99 (1)	0.94 (0.92)	0.88 (0.92)
1987–1991	828	14 (3)	112 (27)	0.93 (1)	0.89 (0.96)	0.85 (0.89)
1988–1992	772	8 (4)	83 (29)	0.98 (0.97)	0.96 (0.93)	0.89 (0.86)
1989–1993	702	7 (6)	73 (40)	1 (1)	0.90 (0.98)	0.85 (0.90)
1990–1994	673	9 (4)	71 (29)	0.99 (1)	0.94 (0.97)	0.93 (0.93)
1991–1995	773	11 (3)	88 (24)	0.97 (0.96)	0.92 (0.96)	0.89 (0.96)
1992–1996	905	11 (5)	101 (44)	0.96 (1)	0.92 (0.98)	0.94 (1)
1993–1997	1,060	10 (5)	107 (41)	0.98 (1)	0.99 (1)	0.95 (1)
1994–1998	1,269	13 (8)	129 (60)	1 (1)	0.98 (0.98)	0.95 (0.98)
1995–1999	1,614	18 (4)	168 (46)	0.99 (0.98)	0.96 (0.98)	0.93 (0.96)
1996–2000	2,039	18 (7)	248 (75)	0.99 (1)	0.98 (1)	0.86 (0.92)
1997–2001	2,156	23 (8)	294 (82)	0.99 (1)	0.96 (0.98)	0.81 (0.93)
1998–2002	2,205	22 (12)	297 (101)	0.99 (0.99)	0.88 (0.97)	0.76 (0.95)
1999–2003	2,166	24 (13)	307 (94)	0.98 (1)	0.86 (0.97)	0.80 (0.93)
2000–2004	2,140	19 (12)	257 (80)	0.96 (1)	0.88 (0.94)	0.84 (0.91)
2001–2005	1,941	25 (6)	232 (39)	1 (1)	0.96 (0.97)	0.91 (0.97)
2002–2006	1,852	19 (7)	186 (40)	0.99 (1)	0.98 (1)	
2003–2007	1,847	17 (10)	198 (60)	0.99 (0.98)	0.98 (0.98)	
2004–2008	1,844	15 (9)	194 (54)	0.99 (1)		
2005–2009	1,801	21 (6)	202 (44)	1 (1)		
2006–2010	1,742	19 (8)	207 (58)			

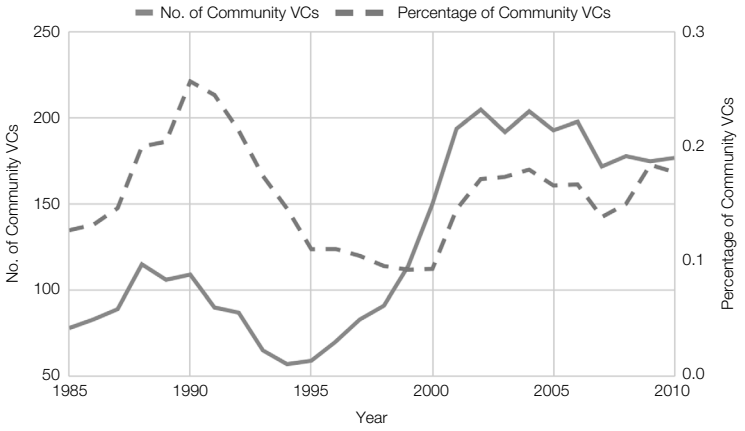
In Table 2, we also assess whether a VC that belongs to a community in one 5-year period $[t, t+4]$ belongs to *some* community the next 5-year period $[t+1, t+5]$; 99% of community VCs retain community status in this test. We also consider longer horizons of $[t, t+4]$, $[t+3, t+7]$, and $[t+5, t+9]$. Community status is relatively stable over these longer horizons as well. For instance, 88% of community VCs remain so 5 years later. We find similar stability when

[13] The regression coefficient for the percentage of community VCs is lower in hot periods than in cold periods. The results are available from the authors.

https://doi.org/10.1017/S002210901900005X Published online by Cambridge University Press

FIGURE 3
Number and Percentage of Community VCs Each Year

Figure 3 shows the time series of the number and percentage of venture capitalists (VCs) that belong to communities in our sample.



communities are detected using first-round investments only. These results are reported in parentheses in Table 2.[14]

Table 3 shows the classification of VCs each year as those belonging to communities, those syndicating with community VCs but that are not community VCs themselves, and VCs that are neither community VCs nor syndicate with community VCs. We find a distinct pecking order among the three categories. Community VCs are more active, make larger investments, are more likely to syndicate, have greater amounts of assets under management, are better networked, and invest greater amounts in portfolio firms with more diverse portfolios. Noncommunity VCs that syndicate with community VCs are more low profile compared with community VCs. For instance, they are smaller and less networked than community VCs. At the bottom of the pecking order are those that are not community VCs and do not co-syndicate with community VCs.

B. Descriptive Statistics

Table 4 shows round-level descriptive statistics that are classified by whether a round involves financing by a community VC or not. Of the 73,414 rounds, 32,547 (approximately 44%) are community rounds, and these account for 65% of the proceeds. Of the 35,088 syndicated rounds, 22,812 (or 65%) are

[14] An additional stability measure is whether VCs tend to belong to the *same community* in successive periods. The problem is that there is no identifiable "one" community but multiple communities that represent reconfigurations of VCs into new community groups in each successive period. One approach is to use the "Jaccard" index. For two sets A and B, the Jaccard index is a number in $(0, 1)$ equal to the ratio of the size of the intersection set $A \cap B$ to the size of the union set $A \cup B$. We compute the average Jaccard index between all communities in a given 5-year period with all communities in the next succeeding 5-year period. We cannot reject the hypothesis that the Jaccard index is similar to an index constructed using simulated communities.

https://doi.org/10.1017/S002210901900005X Published online by Cambridge University Press

TABLE 3
Community Connections among VCs

Table 3 provides characteristics of 3 categories of venture capitalists (VCs) in any given year: those that are neither community VCs nor have syndicated with a community VC (Never Comm.), those VCs that are not community VCs but have syndicated with at least one community VC (Only Partner with Comm.), and VCs that are community VCs (Comm.). For any given VC category, the characteristics are based on their activities in the previous 5-year window that are averaged across all the years. We report the number of rounds of financing (NO_OF_ROUNDS) and the count of portfolio companies a VC invests in (NO_OF_COMPANIES). INVESTMENT_PER_ROUND is the amount a VC invests in a round. %_DEALS_SYNDICATED is the number of a VC's syndicated rounds as a percentage of all rounds that a VC invested in during the previous 5-year window. %_EARLY_STAGE_DEALS is the number of a VC's investment rounds classified by Venture Economics as early stage, as a percentage of all deals for the VC during the previous 5-year window. AVERAGE_VC_AUM is the average of the sum of the capital under management of a VC in all funds during the previous 5-year window for the subset of VCs relevant to the column. TOTAL_INVESTMENT is the sum of a VC's investments during the previous 5-year window. AGE is defined as the difference in the year of the VC's last investment during the previous 5-year window and the VC firm's founding date. CENTRALITY is based on each VC's eigenvector centrality, determined in the 5-year window. For the remaining attributes, we calculate the Herfindahl–Hirschman Index (HHI) as the sum of the squared share in each subcategory of the attribute. INDUSTRY_HHI is the Herfindahl index based on the percentage of a VC's deals in each industry, whereas STAGE_HHI is the Herfindahl index based on the percentage of deals in each stage of investment. COMPANY_REGION_HHI is the Herfindahl index based on the percentage of deals in each geographic region.

	Never Comm.	Only Partner with Comm.	Comm.
NO_OF_ROUNDS	11.75	28.00	83.82
NO_OF_COMPANIES	7.71	16.29	45.70
INVESTMENT_PER_ROUND ($millions)	3.99	4.11	5.60
%_DEALS_SYNDICATED	0.64	0.78	0.86
%_EARLY_STAGE_DEALS	0.29	0.29	0.24
AVERAGE_VC_AUM ($millions)	113.61	175.99	599.99
TOTAL_INVESTMENT ($millions)	47.82	126.91	562.09
AGE	8.55	8.49	12.83
CENTRALITY	0.01	0.04	0.19
INDUSTRY_HHI	0.52	0.41	0.29
STAGE_HHI	0.60	0.49	0.37
COMPANY_REGION_HHI	0.61	0.48	0.37

community rounds. VentureXpert classifies approximately a third of the rounds as "early stage," and 41% of these are community rounds. In total, 35,414 deals, or close to one-half of the investment rounds, are in the California and Massachusetts (CA/MA) geographical clusters, reflecting concentration patterns of VC investments and their representation in VC databases (Kaplan, Sensoy, and Stromberg (2002)).

Panel B of Table 4 shows the industry classifications based on the 10 categories reported in VentureXpert. The software industry accounts for the largest share of financing rounds in our sample, followed by Internet firms, medical or health firms, and communications and media firms. However, community VC financing is the least likely for consumer product or industrial businesses, which are the relatively less risky and complex financing deals. This finding implies that VCs are less likely to lean on preferred partners in deals with less ambiguity (Cestone et al. (2006)).

Panel C of Table 4 shows other characteristics that are associated with community rounds. These rounds involve significantly greater investment on average and involve more VC firms in syndicated and nonsyndicated rounds, in early versus late rounds, and in initial versus later rounds. Panel D gives information on exits, which we analyze and discuss later. We find that community VCs have greater engagement in the later stages of the financing life cycle, which is consistent with the view that repeated interactions are a source of soft information exchange in communities. The percentage of community VC financings in round 1 is 13.2%, whereas for noncommunity VCs, it is roughly double at 26.6%. The

https://doi.org/10.1017/S002210901900005X Published online by Cambridge University Press

TABLE 4

Descriptive Statistics for 73,414 Rounds in 26,995 Unique Portfolio Firms from 1985 to 2010

Table 4 provides descriptive statistics for 73,414 rounds in 26,995 unique portfolio firms from 1985 to 2010. A round is a community round if at least one venture capitalist (VC) firm participating in it comes from a VC community. Communities are detected using a walk-trap algorithm applied to syndicated deals over 5-year windows rolled forward 1 year at a time. The sample comprises VC deals obtained from Venture Economics but excludes non-U.S. investments, angel investors, and VC firms focusing on buyouts. Industry classifications are as per Venture Economics. Exit data are obtained by matching with the Thomson Financial Initial Public Offering (IPO) and Mergers and Acquisitions (M&A) databases. Data on follow-on funding are restricted to rounds financed up to 2009.

Variable	Total	Community Round	Not Community Round
Panel A. Counts by Round			
No. of deals	73,414	32,547	40,867
Round 1	23,253	7,232	16,021
Round 2	15,771	6,773	8,998
Round 3	11,237	5,626	5,611
SYNDICATED	35,088	22,812	12,276
EARLY_STAGE	22,923	9,399	13,524
VC_GEOGRAPHICAL_CLUSTER	35,414	21,074	14,340
Rounds with			
VC_GEOGRAPHICAL_CLUSTER	44,560	28,052	16,508
CORPORATE_VC	8,593	5,084	3,509
FI_VC	10,391	5,720	4,671
Panel B. Percentage by Venture Economics Industry			
Biotech	7.6	3.7	3.8
Commu&Media	10.3	5.2	5.2
Hardware	5.1	2.6	2.6
Software	21.6	10.1	11.5
Semiconductor, Electricals	7.4	3.9	3.6
Consumer Products	5.0	1.3	3.7
Industrial, Energy	4.9	1.4	3.6
Internet	17.7	7.9	9.8
Medical	13.1	6.5	6.6
Others	7.2	1.8	5.4
Panel C. Round Statistics			
Proceeds ($millions)	19 (8)	27 (15)	12 (4)
No. of VCs	2.09 (1)	2.83 (2)	1.50 (1)
In syndicated rounds	3.34 (3)	3.69 (3)	2.69 (2)
In early-stage rounds	1.89 (1)	2.49 (2)	1.47 (1)
In round 1	1.54 (1)	2.01 (2)	1.32 (1)
In round 2	1.46 (1)	1.63 (2)	1.35 (1)
In round 3	2.03 (2)	2.27 (2)	1.80 (2)
Panel D. Exit			
Rounds with			
IPO exits	7,816	4,307	3,509
M&A exits	15,660	8,049	7,611
Follow-on funding	45,866	22,419	23,447

Kolmogorov–Smirnov test for differences has a p-value of 0.03. Figure 4 depicts the result visually. Community VCs show a thicker mass in earlier rounds.

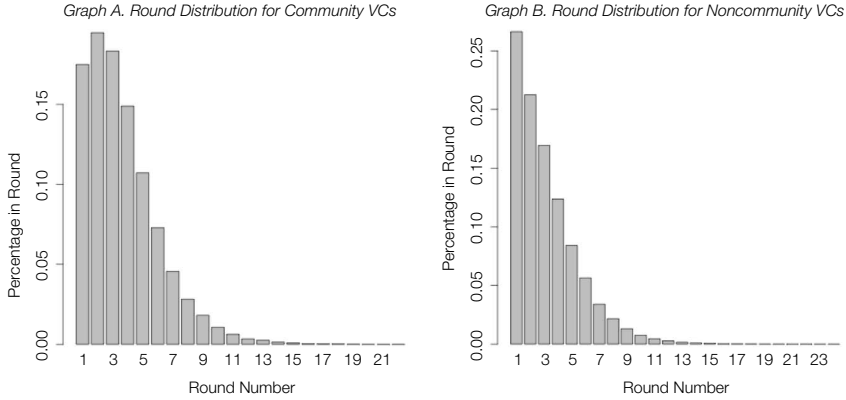
We provide a further description of community VC rounds by examining rounds with at least two community VCs and rounds with at least two VCs from the same community. Of the community rounds, over 40% have more than one community VC, and over 16% have at least two VCs from the same community. The incidence of both types of multiple-community VC participation increases in later financing rounds. The majority (65%) of early-stage rounds do not have multiple-community VCs. Only 15% of early-stage rounds have at least two VCs from the same community. The same-community multiple-VC rounds tend to be larger deals with more syndicate members. Multiple-community VC participation in a round is associated with a greater likelihood of exit. In 44%

https://doi.org/10.1017/S002210901900005X Published online by Cambridge University Press

FIGURE 4

Distribution of Repeated Interactions for Community and Noncommunity VCs

As shown in Figure 4, venture capitalists (VCs) in communities have a greater proportion of cofinanced rounds in rounds after the first round than VCs that are not in communities. The percentage of community VC financings in round 1 is 13.2%, whereas for noncommunity VCs, it is roughly double at 26.6%. The standard deviation of the percentage of round 1 financings for all 390 communities formed over the entire sample is 8%. Community VCs (Graph A) show a thicker mass in earlier rounds than noncommunity VCs (Graph B). A Kolmogorov–Smirnov test for the difference in the 2 distributions shows a highly significant D-statistic, with a p-value of 0.0003.



of rounds in which there are multiple-community VCs (50% with multiple VCs from the same community), we see exit, compared with 34% for rounds with a single-community VC.[15]

V. VC Community Composition

A. Approach

In this section, we analyze the characteristics of VC communities. The first 3 attributes capture a VC's reach and influence. Following Hsu (2004), we consider a VC's age, which is the difference between the VC's last investment in year t and the VC's founding year. We additionally consider the assets under management (AUM) of the VC and, following Hochberg et al. (2015), the VC's eigenvector centrality. Older VCs and those with greater AUM and centrality are more influential.

A second set of attributes reflects a VC's investing style. We follow the literature in considering three major dimensions of VC style: industry, stage, and geography (Sorenson and Stuart (2001), Chen, Gompers, Kovner, and Lerner (2010), and Lerner, Hardymon, and Leamon (2007)). Stage and industry styles are based on past investment patterns that are reported in VentureXpert. For geography, we consider the locations of both portfolio firms and the VC headquarters. These attributes capture region-specific information that is gleaned by due diligence on portfolio firms or by the top management located at the VC headquarters.[16]

[15]The results are available from the authors. We note that regressions based on the alternative definitions of community rounds give results similar to those reported in Section VII.

[16]Anecdotal evidence is consistent with these dimensions of functional styles. For example, SV Angel says it is a "seed-stage" fund with "tentacles into New York media and the advertising world" (http://techcrunch.com/2011/05/24/sv-angel-partners-with-lerer-ventures-to-cross-syndicate-valleynyc-deals).

https://doi.org/10.1017/S002210901900005X Published online by Cambridge University Press

https://doi.org/10.1017/S002210901900005X Published online by Cambridge University Press

To assess whether community VCs are similar to each other, we estimate the within-community variation in several characteristics. For continuous variables, the within-community variation of a characteristic is its standard deviation across community members. For discrete variables, our similarity measure is the standard deviation of the Herfindahl–Hirschman index (HHI) for the category across all VCs in a community. The HHIs use equal weights. Value weights produce similar results.[17] For VC headquarters, we consider the proportion of all VCs in a community that share headquarters.

We consider a more stringent version of style similarity that not only conditions on how VCs distribute capital but also the specific sectors they allocate it to. For example, HHI dispersion will be 0 if there are 10 VCs, 5 with 100% focus in software and 5 with 100% focus in biotech, but these VCs are clearly different. We compute the sector-specific similarity of VCs as the standard deviation of the proportions invested by VCs relative to the aggregate of all VCs in the community.[18]

For inferring significance, we benchmark the results relative to those for simulated communities. We conduct round-level simulations to generate null distributions as follows: We condition simulations on 3 features of rounds, namely, industry, stage, and region of portfolio firm. We draw VCs for a given financing round without replacement from VCs with at least one investment matching all 3 round features, and occasionally, 2 criteria if there is not a match on all 3. We match the number of VCs drawn for a financing round to the actual syndicate size.

B. Results

Table 5 shows the characteristics of community VCs and those of simulated communities. The average age and size of community VCs are not different, but community VCs are more networked and have a greater spread of capital over the industry, stage, and geography of portfolio firms compared with simulated communities.

In Table 6, we analyze the similarities of attributes within communities. The estimates in Panel A show that communities are more homogeneous in terms of age and centrality than in simulated communities. The results in Panel B show that VCs in communities also display greater similarity in functional style along the dimensions of industry, stage, and location.[19] The lower variation in the HHI within realized communities compared with that in simulated communities shows that generalist (specialist) VCs tend to draw other generalist (specialist) VCs as partners. The rows below the HHI statistics in Panel B show similarity even with the more stringent HHI definition. Panels C and D show that the differences in ownership form or in geographies are not significant.

[17] Both measures are reasonable. The key resource in a VC is a partner's time, which first-order scales by the number of investments, whereas the capital at risk is proportional to the proceeds invested in a firm.

[18] Let the proportion of capital in bucket j for VC i in community k be f_{ijk}. We compute $\sigma_{jk} = \sum_{i=1}^{n}(f_{ijk} - \overline{f_{jk}})^2)/(n-1)$.

[19] The reported location results use the 14-region classification given by VentureXpert that reflects homogeneous operational clusters. Other classifications based on state and MSA give similar results.

TABLE 5
Characteristics of Same-Community VCs

Table 5 compares the key community characteristics with those of simulated communities generated by choosing venture capitalists (VCs) that have invested at least once in the same industry, stage, and region as the portfolio firm. For each community (and simulated community), we generate the mean of the characteristic and present the average value across communities. AGE uses the number of years between a VC's last investment in a 5-year window and the founding year of the VC firm. AVERAGE_VC_AUM is the average AUM of all funds of the VC firm in the 5-year period for the relevant subset, which is either a community or a simulated community. CENTRALITY is based on each VC's eigenvector centrality determined for each 5-year rolling window. For the remaining attributes, we calculate the Herfindahl–Hirschman Index (HHI) as the sum of the squared share in each subcategory of the attribute. INDUSTRY_HHI is the Herfindahl index based on the percentage of a community VC's deals in each industry, whereas STAGE_HHI is the Herfindahl index based on the percentage of deals in each stage of investment. COMPANY_REGION_HHI is the Herfindahl index based on the percentage of deals in each geographic region. In unreported tests, we see similar results when we use HHI based on the amount invested. The industry, stage, and geographic region classifications are those provided by Venture Economics. The last column shows the *p*-values from testing the equality of the means of the community and bootstrapped community characteristics. *, **, and *** denote 10%, 5%, and 1% significance, respectively.

	Community	Simulated Community	*p*-Value
Reach and Influence			
AGE	12.00	12.13	0.25
AVERAGE_VC_AUM	461.99	476.71	0.25
CENTRALITY	0.15	0.14	0.05**
Style			
INDUSTRY_HHI	0.34	0.40	0.01***
STAGE_HHI	0.41	0.43	0.01***
COMPANY_REGION_HHI	0.40	0.44	0.01***

One issue is that closeness in style within communities could be driven by similarities between VCs in unpopulated or sparsely populated cells rather than buckets with significant deal flow. Benchmarking with simulations mitigates such concerns, but we also consider a test with a higher bar. In each 5-year period, we consider only the top 4 industries, top 2 stages, and top 3 geographic areas for each VC. Table 7 shows that even in this test, VCs within communities exhibit less variation in each of the buckets.[20]

We next use the simulation data to examine the differences *between* communities. One possibility is community specialization. For instance, clean energy investments may require specialized technical skills or regulatory knowledge, and cancer drug development may call for special knowledge of new therapeutic protocols. If VCs with similar specialties tend to partner with each other, we should see *differences between* and *similarity within* communities. To understand intercommunity differences, we test whether the distances between community centroids is greater than what we observe in simulated communities. Table 8 gives the results. Other than for VC age, we find little evidence that there is more variation in the distances between communities than in simulated ones. Communities do not appear to seek locations to achieve differentiation. We also find that the Panel B results show that community VCs are less concentrated along various style dimensions, and in Panel C, communities are not geographically more concentrated relative to the simulated communities. Thus, communities do not appear to be related to the possible anticompetitive effects from more densely networked VC markets found in Hochberg, Ljungqvist, and Lu (2010).

[20]The top industries, stages, and geography of interest change over time. For instance, the consumer products industry is in the top 4 in the early 1980s, but the Internet industry replaces it after the 1990s. The cosine similarity measures as in Hoberg and Phillips (2010) give similar results.

https://doi.org/10.1017/S002210901900005X Published online by Cambridge University Press

TABLE 6
Similarity of Within-Community VCs

Table 6 presents the variation in key attributes (Panels A and B), the mean geographic location Herfindahl–Hirschman index (HHI) (Panel C), and the ownership HHI (Panel D) of venture capitalists (VCs) within communities and compares these to those of simulated communities generated by choosing VCs that have invested at least once in the same industry, stage, and region as the portfolio firm. We calculate the HHI as the sum of the squared deviation of each subcategory of the attribute. AGE uses the number of years between a VC's last investment in a 5-year rolling window and the founding year of the VC firm. AVERAGE_VC_AUM is the average AUM of all funds of the VC firm in the 5-year period for the relevant subset, which is either a community or a simulated community. CENTRALITY is based on each VC's eigenvector centrality determined for each 5-year rolling window. INDUSTRY_HHI, STAGE_HHI, and COMPANY_REGION_HHI are based on the percentage of a VC's deals in each of the 10 industries, each of the 5 stages, and each of the 14 U.S. geographic regions, respectively, as classified by Venture Economics. In Panels A and B, variations in reach attributes and HHI attributes, respectively, are the standard deviation of each attribute of a community's VC, averaged across all communities. Variation in each attribute in Panel B measures the mean (across all communities) of the sum of the squared deviation in each subcategory (e.g., industry *j*) of each attribute (e.g., industry) averaged across all subcategories and all within-community VCs. Panel C uses alternative geographic location variables, from the most granular (Metropolitan Statistical Area (MSA)) to the least granular (region) and calculates the geographic HHI of a community's VCs, averaged across all communities. Panel D calculates the ownership HHI of a community's VCs, averaged across all communities. The last column shows the *p*-values testing the equality of the means of the community and simulated community characteristics. *, **, and *** denote 10%, 5%, and 1% significance, respectively.

	Community	Simulated Community	*p*-Value
Panel A. Variation in Reach Attributes			
AGE	7.56	7.87	0.01***
AVERAGE_VC_AUM	527.19	492.45	0.1*
CENTRALITY	0.10	0.12	0.01***
Panel B. Variation in Functional Styles			
INDUSTRY_HHI	0.16	0.20	0.01***
STAGE_HHI	0.13	0.17	0.01***
COMPANY_REGION_HHI	0.18	0.21	0.01***
INDUSTRY_VARIATION	0.11	0.13	0.01***
STAGE_VARIATION	0.08	0.11	0.01***
COMPANY_REGION_VARIATION	0.10	0.12	0.01***
Panel C. Mean of Community Geographic HHI			
VC_MSA_HHI	0.28	0.29	0.25
VC_STATE_HHI	0.44	0.43	0.25
VC_REGION_HHI	0.40	0.40	0.25
Panel D. Mean of Community Ownership HHI			
VC_OWNERSHIP_HHI	0.60	0.59	0.25

VI. Classifying Realized Communities

In this section, we use machine-learning tools from the data sciences to assess the nature of the realized communities in our sample. We use both unsupervised and supervised machine-learning algorithms. Unsupervised machine learning refers to classification methods in which the target classification, in this case the number and nature of clusters of communities, is not known. We use the unsupervised learning results to inform the supervised learning methods that uncover the variables that are most significant in explaining the clustering patterns of the 413 realized communities in our sample.

A. Unsupervised Learning

We use a hierarchical clustering algorithm to identify the structure of the realized communities. Hierarchical clustering is a bottom-up agglomerative approach that generates an entire hierarchy of possible clustering schemes and chooses one that best represents the data. The input is the set of 413 realized communities that we detected over our entire sample period and a vector of characteristics that we

https://doi.org/10.1017/S002210901900005X Published online by Cambridge University Press

TABLE 7
Functional Expertise Similarity of Within-Community VC

In Table 7 we present the mean (across all communities) of the sum of the squared deviations of a venture capitalist's (VC's) share of deals in some subcategories (based on number of deals in a 5-year rolling window in each of the top 4 industries, top 2 stages, and top 4 company regions, with the remaining share of investment comprising the last subcategory in each). We compare these values to those of simulated communities generated by choosing VCs that have invested at least once in the same industry, stage, and region as the portfolio company. The last column shows the *p*-values from testing the equality of the means of the community and simulated community characteristics. *, **, and *** denote 10%, 5%, and 1% significance, respectively.

	Community	Simulated Community	*p*-Value
Industry Rank			
1	0.14	0.16	0.01***
2	0.12	0.15	0.01***
3	0.11	0.12	0.01***
4	0.09	0.11	0.01***
5 = others	0.16	0.17	0.01***
Stage Rank			
1	0.16	0.19	0.01***
2	0.15	0.17	0.01***
3 = others	0.16	0.17	0.01***
Company Region Rank			
1	0.18	0.21	0.01***
2	0.12	0.12	0.25
3	0.10	0.11	0.05**
4	0.08	0.08	0.05**
5 = others	0.15	0.17	0.01***

TABLE 8
Similarity across Communities

Table 8 presents the across-community variation in the (average) key venture capitalist (VC) attributes (Panels A–B), in the geographic location Herfindahl–Hirschman index (HHI) (Panel C), and in the ownership HHI (Panel D) of VCs within communities and compares these to those of simulated communities generated by choosing VCs that have invested at least once in the same industry, stage, and region as the portfolio company. The HHI is the sum of the squared deviations of each subcategory of the average of an attribute within a community. INDUSTRY_HHI, STAGE_HHI, and COMPANY_REGION_HHI are based on the percentage of a VC's deals in each of the 10 industries, each of the 5 stages, and each of the 14 U.S. geographic regions, respectively, as classified by Venture Economics. Panels A and B show the mean across-community standard deviation of the average community attribute in each 5-year rolling window. Panel C uses alternative geographic location variables, from the most granular (Metropolitan Statistical Area (MSA)) to the least granular (region) and calculates the standard deviation of the geographic HHI of communities in each 5-year window, averaged across all such windows. Panel D calculates the standard deviation of the ownership HHI of communities in each 5-year window, averaged across all such windows. The last column shows the *p*-values from testing the equality of the means of the community and simulated community characteristics. *, **, and *** denote 10%, 5%, and 1% significance, respectively.

	Community	Simulated Community	*p*-Value
Panel A. Variation in Reach Attributes			
AGE	3.36	3.54	0.10*
AVERAGE_VC_AUM	263.23	271.38	0.50
CENTRALITY	0.07	0.07	0.25
Panel B. Variation in Functional Styles			
INDUSTRY_HHI	0.08	0.09	0.25
STAGE_HHI	0.04	0.04	0.50
COMPANY_REGION_HHI	0.13	0.13	0.50
Panel C. Variation of Community Geographic HHI			
VC_MSA_HHI	0.12	0.13	0.25
VC_STATE_HHI	0.20	0.19	0.25
VC_REGION_HHI	0.17	0.17	0.50
Panel D. Variation of Community Ownership HHI			
VC_OWNERSHIP_HHI	0.18	0.18	0.50

https://doi.org/10.1017/S002210901900005X Published online by Cambridge University Press

https://doi.org/10.1017/S002210901900005X Published online by Cambridge University Press

used to classify the communities. We standardize the vector of characteristics so that it has a 0 mean and a unit standard deviation. As we choose more clusters, the ratio of between-group variation to within-group variation must improve as a mathematical necessity. In our application, this improvement is small after 3 clusters.[21]

Panel A of Table 9 reports the results. Communities separate into 3 types that display a clear pecking order. Cluster 1 (denoted "Young") features young (average age = 8.15 years), less networked venture capitalists (centrality = 0.10) that tend to finance small portfolio firms ($204 million) and form small communities (size = 7.70). These communities tend to be more locally focused than the other communities, as indicated by the higher geography HHI of 0.33 for this cluster relative to others. Cluster 2 (denoted "Mature") comprises the older and more networked VCs. The average age of VCs in this cluster is 15.99 years, approximately double that for the communities in cluster 1. This cluster also finances the largest firms, whose average size of $970 million is over four times the average size of $204 million for the portfolio firms in cluster 1. The VCs in this cluster are more spread out across industries, as indicated by the lowest industry HHI among the 3 clusters, but there is not the same pattern in the stage or region HHIs. Cluster 3 is, in most respects, in between clusters 1 and 2.

TABLE 9

Unsupervised and Supervised Learning on Communities

In unsupervised learning, shown in Panel A of Table 9, we carry out a clustering of communities by characteristics. We use a hierarchical clustering algorithm to generate 3 clusters across all communities. The input variables are standardized to unit standard deviations. Characteristics of the clusters are shown, revealing a pecking order across communities. Supervised learning is then undertaken on the estimated clusters. Panel B shows the confusion matrix derived from a predictive analysis done by using the random-forest algorithm, which is an ensemble extension of the classification and regression trees (CART) algorithm. The random-forest algorithm was also used to assess the most important variables in the classification, in terms of the ability to improve on Gini impurity. These results are shown in the last column of Panel A.

Panel A. Cluster Characteristics

Variable	Cluster 1 (Young)	Cluster 2 (Mature)	Cluster 3 (Established)	Mean % Decrease in Gini
COMMUNITY_SIZE	7.70	18.77	8.73	31.79
AGE	8.15	15.99	13.77	77.20
AVERAGE_VC_AUM ($millions)	203.81	970.12	434.94	46.24
CENTRALITY	0.10	0.21	0.15	42.01
INDUSTRY_HHI	0.24	0.18	0.24	14.86
STAGE_HHI	0.35	0.31	0.29	31.78
COMPANY_REGION_HHI	0.33	0.29	0.22	23.24

Panel B. Confusion Matrix

		Predicted		
		Cluster 1	Cluster 2	Cluster 3
Actual	Cluster 1	157	0	10
	Cluster 2	7	88	4
	Cluster 3	6	4	143

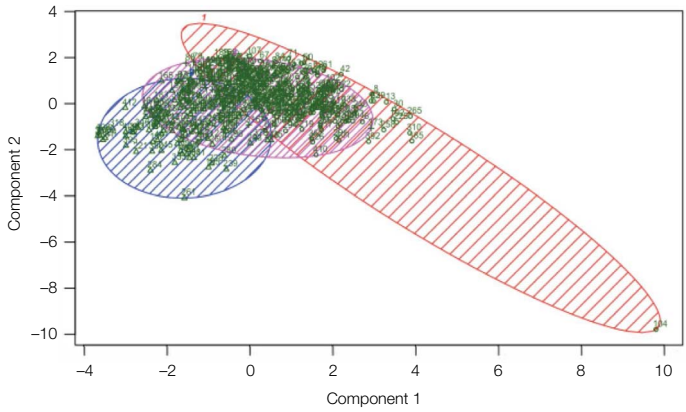
A principal-components analysis shows that the first 2 components of the characteristics account for 53% of the variation across clusters. The top 3 variables that differentiate communities are, in order of importance, age, firm size,

[21] We implement clustering by using the `hclust` function in the `stats` package in the `R` programming language.

and centrality. Similar-characteristic VCs appear to be more likely to engage with each other and cluster into communities. Figure 5 shows the clusters and the community numbers (from 1 to 413) along the axes that represent the first 2 principal components. We see a hierarchy of communities, going from ones that comprise emerging VCs to ones that comprise very seasoned ones.

FIGURE 5
VC Clusters

In Figure 5, we estimate a principal-components analysis of the venture capitalist (VC) clusters described in Table 9. The first 2 components account for 53.27% of the point variability across clusters. We depict the placement of the VCs in the sample in the principal-component space in this figure. The top 3 variables that differentiate communities are, in order of importance, VC age, firm size, and centrality.



B. Supervised Learning

We turn to the results from supervised learning. Given the 3 clusters of communities (which is the supervisory input into the learning algorithm), we run a "classification tree" (CART) algorithm to assess the variables that are salient in deciding the cluster each community is located in. This algorithm uses a process known as recursive partitioning (Breiman (1984), Cook and Goldman (1984)) to generate a decision tree that drives the clusters into which each community is assigned. The results are consistent with those from unsupervised learning, as we discuss next.

The CART procedure works as follows: At the top of the tree, at the root node, the algorithm experiments with an independent variable and a cutoff and bifurcates the sample of communities into 2 based on whether this variable is smaller or greater than a cutoff value. For example, if the VC's age is the partitioning variable and the cutoff for generating 2 partitions is 10 years, communities are divided into 2 groups, one where the average age is less than 10 years and another where it is greater than 10 years. The algorithm searches over all explanatory variables and cutoffs to generate a choice that minimizes the average within-group variation. This search results in 2 child nodes of the root node that contain subgroups of the sample.

https://doi.org/10.1017/S002210901900005X Published online by Cambridge University Press

https://doi.org/10.1017/S002210901900005X Published online by Cambridge University Press

The learning algorithm repeats the treatment of picking variables and cutoffs sequentially by applying them to the child nodes. The process continues recursively down the tree until the reduction in within-group variation is small. The stopping rule is usually modulated with a complexity (error or loss) parameter, which is analogous to an R^2. We stop if the improvement is less than 0.01. We see which cluster is most represented by the chosen group of variables, and then we assign these last (leaf) nodes to that single cluster. The entire process effectively creates a calibrated decision tree. The procedure can be used to classify other communities out of sample if needed.

For robustness, we also use a random-forest algorithm (Ho (1995), (1998)), which generates hundreds of CART trees based on a randomly selected subset of variables from those used for clustering. (If all the variables are used, then the algorithm generates the same CART tree every time. By selecting a random subset of the explanatory variables, the random forest algorithm creates many different decision trees.) In our implementation, we generated 500 different random CART trees with only 2 randomly chosen candidate bifurcation variables at each branching node in the decision tree. Any new community is classified by all n trees, where each tree decides which cluster a community belongs to. The community is assigned to the cluster category that is the modal choice among the n trees. This type of machine learning, where different models essentially vote on classification, is known as "ensemble" learning. Panel B in Table 9 gives the results from the random-forest approach. Given the actual classification of the three clusters of communities and the predicted classifications from the supervised learning algorithm, we can compute the percentage of communities that are correctly classified. Of the 413 communities, the algorithm classifies 388 (93.9%) into the correct cluster.[22]

We next compute the reduction in the Gini impurity coefficient G for each variable to determine which of the clustering variables are the most important in classification. The Gini impurity measures the dispersion in a set. Suppose that the members of the set fall into K categories; then, if f_k denotes the fraction of members in the set in each category k, $G = \sum_{k=1}^{K} f_k(1 - f_k)$. For example, if 60 elements are classified into 3 categories with 10, 10, and 40 members, then $G = 0.50$.[23] Now consider a variable that splits the elements into 2 groups with numbers in each category as follows: $\{10, 10, 0\}$ and $\{0, 0, 40\}$. The first group has $G_1 = 1/2$, and for the second group, it is $G_2 = 0$, so the average Gini impurity is now $1/4$ because of the bifurcating variable. Hence, the percentage reduction in Gini impurity is 25%. Table 9 reports the Gini impurity results. They confirm that VC age, portfolio firm size, and centrality have the highest discriminatory power relative to the other variables when ranked by the reduction in Gini impurity. The importance of the variables roughly corresponds to what we find from the cluster analysis in Section VI.A.

[22] The data sciences literature refers to the actual versus predicted classification matrix as the "confusion" matrix. Better classification algorithms produce more observations on the diagonal of the confusion matrix.

[23] $G = (10 \times 50)/(60^2) + (10 \times 50)/(60^2) + (40 \times 20)/(60^2) = 0.50$. Note that $0 < G < 1$. The lowest value is attained with a single group, whereas the greatest value occurs when membership is dispersed across a very large number of categories.

https://doi.org/10.1017/S002210901900005X Published online by Cambridge University Press

VII. Investing Outcomes

The two economic outcomes we study are innovation (patenting) and exit. A community round is one with at least one community-based VC financing the portfolio firm. We analyze these outcomes at the level of a financing round. Thus, output from the 1985 investment year comes from a community round if at least one VC financing this syndication deal belongs to a community inferred from syndications from 1980 to 1984.[24]

A. Patenting

The scale of innovation is the natural logarithm of 1 plus the number of patents granted. The quality of innovation is the natural logarithm of 1 plus the average number of forward citation counts per patent (Hall, Jaffe, and Trajtenberg (2005)). To handle the truncation in the number of years for which we can count citations for patents granted toward the end of the sample period, we use the estimated shape of the citation-lag distribution as a correction (Hall et al. (2001)).[25] Following Seru (2014), we subtract from the patenting or citation measures the average for the technology class and year.

Panel A of Table 10 shows that on average, a round has 0.36 patent applications and 3.68 citations. Community rounds score higher than noncommunity rounds. In Panel B, we adjust both of the patenting metrics for the average innovation in each cohort (by application year and technology class) and find that the extent of innovation in rounds with a community VC exceeds that in rounds

TABLE 10
Descriptive Statistics for Innovation Outcomes from 1985 to 2006

Table 10 provides mean values for the innovation variables. The unit of observation is company × round. TOTAL_PATENTS is the number of patents applied for in the year of the financing round. AVERAGE_CITATIONS is the total number of citations per patent that is corrected for truncation bias using Hall et al. (2001) (as described in the text). Panel A considers all company × rounds. In Panel B, we adjust both innovation measures for the average innovation in the same cohort (based on application year and technology class) to which the patent belongs. Panel C uses the subsample where TOTAL_PATENTS is positive. The last column tests for the equality of means of innovation outcomes for community and noncommunity rounds. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Variable	Total	Community Round	Not Community Round	Mean Equality
Panel A. All Observations				
TOTAL_PATENTS	0.36	0.46	0.29	***
AVERAGE_CITATIONS	3.68	4.99	2.66	***
Panel B. All Observations, Cohort-Adjusted Innovation Variables				
TOTAL_PATENTS	0.06	0.08	0.05	***
AVERAGE_CITATIONS	0.25	0.34	0.18	***
Panel C. Cohort-Adjusted Innovation Variables if TOTAL_PATENTS >0				
TOTAL PATENTS	0.48	0.49	0.46	**
AVERAGE CITATIONS	2.01	2.17	1.81	***

[24]The main results are based on interactions between VCs regardless of which round the interactions take place in, of course, counting each round only once in a 5-year period. However, for robustness, we conduct our analyses using communities detected based only on first-round deals. The results are qualitatively similar.

[25]We also obtain similar results if we follow Bernstein, Giroud, and Townsend (2016) and use 3-year forward citation rates. A slight difference is that we count the data from the application rather than the grant date used in their study.

https://doi.org/10.1017/S002210901900005X Published online by Cambridge University Press

without a community VC. Because many portfolio firms in VentureXpert are not active in patenting (only 16% of the sample firms file patent applications), Panel C also reports data on financing rounds with at least one application for a patent. Our result, that community rounds are more innovative than noncommunity rounds, continues to hold.

We estimate the regression models next. The dependent variable is the relevant patenting measure. The key independent variable of interest is the community dummy variable. We report results for the full sample and subsamples to understand heterogeneity. We include several controls. One control is whether a round is syndicated or not (Brander, Amit and Antweiler (2002)), so the community coefficient is net of whether the round is syndicated or not. Other controls include financing stage; whether the headquarters are in California or Massachusetts or not (Chen et al. (2010)); and VC experience and skill (Krishnan and Masulis (2011)), as determined through AUM, centrality, the rate at which a VC takes its firms public, and the average VC age as of the year before the financing round. We control for whether VCs are arms of financial institutions or corporate VC investors and whether the VCs have early-stage and industry focus through the fractions of firms in these focus areas. All models include year and industry fixed effects. Table 11 reports the results.

In the baseline estimates for the amount of patenting in columns 1 and 2 of Table 11, the coefficient for community rounds is positive and significant. Controlling for past patenting in column 3 preserves these results. Column 4 includes all other controls. The community coefficient is positive but no longer significant. In column 5, we find that the community coefficient is positive and significant at the 10% level in the full specification for citations. Among the control variables, early-stage firms patent less, but when they do, they have more citations. More citations may reflect the high risk and greater tail outcomes in early-stage investments. California and Massachusetts firms have positive and significant coefficients, as do better-networked VCs, as reflected in centrality. The coefficients are positive for VCs that have a greater focus on the industry of the portfolio firm.

The heterogeneity results are interesting. In Table 12, we sort the sample based on whether the portfolio firm being financed is early stage or not. We further classify firms based on whether they have prior patenting or not. We find that the participation of a community VC is associated with a 1.6% greater volume of patenting and 5.3% greater citations for early-stage ventures that do not have a prior history of patenting. Sourcing funding from a community VC is reliably associated with innovation when the portfolio firms are early stage and have no patenting history.

B. Exits

Our final tests examine the relation between financing sourced from a community VC and exiting via IPOs or M&As. The literature agrees that IPOs are markers of successful portfolio firm maturation (Hochberg, Ljungqvist, and Lu (2007), Kaplan et al. (2002), Maats et al. (2008)). However, exits via M&As are divided between failed financings and successful ones, such as Skype, which was acquired by Microsoft. There is no consensus on how best to deal with this issue

TABLE 11
Innovation and VC Communities

Table 11 reports the ordinary least squares (OLS) estimates for innovation activity between 1985 and 2006. The dependent variable in specifications 1–4 is the natural logarithm of (1 plus) the number of patents that a portfolio firm applied for in the financing year. In specification 5, the dependent variable is the natural logarithm of (1 plus) the per-patent citation count corrected for truncation bias. Both total patents and average citations are adjusted for average innovation in the same cohort (based on application year and technology class). The observations are at the company × round level. See the Appendix for a description of the independent variables. Standard errors are clustered at the portfolio firm level. *t*-statistics are in parentheses. All specifications are overall significant at the 1% level. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

	TOTAL_PATENTS				AVERAGE_CITATIONS
	1	2	3	4	5
COMMUNITY	0.022*** (12.31)	0.021*** (12.41)	0.014*** (9.40)	0.003 (1.31)	0.008* (1.80)
PAST_INNOVATION			0.085*** (30.08)	0.084*** (28.71)	0.196*** (29.52)
EARLY_STAGE				−0.003** (−2.12)	0.024*** (5.94)
COMPANY_GEOGRAPHICAL_CLUSTER				0.012*** (6.36)	0.033*** (7.79)
AUM_ROUND				0.001 (0.94)	−0.002 (−1.48)
CORPORATE_VC				0.004 (1.55)	0.013** (2.06)
FI_VC				0.002 (1.10)	−0.003 (−0.56)
SYNDICATED				0.006*** (4.04)	0.020*** (5.77)
IPO_RATE				0.003 (0.66)	0.006 (0.59)
CENTRALITY				0.020*** (2.78)	0.076*** (4.40)
VC_GEOGRAPHICAL_CLUSTER				0.001 (0.77)	0.003 (0.69)
EXPERIENCE				0.000 (0.17)	0.003 (1.11)
EARLY_STAGE_FOCUS				0.003 (0.52)	0.020 (1.63)
INDUSTRY_FOCUS				0.013*** (2.58)	0.022* (1.80)
No. of obs.	58,126	58,126	58,126	53,000	53,000
Year fixed effects	No	Yes	Yes	Yes	Yes
Industry fixed effects	No	Yes	Yes	Yes	Yes
Adj. R^2	0.005	0.087	0.206	0.209	0.176

in large-scale VC studies such as ours.[26] We follow the VC literature in modeling only exits via IPO (Gompers et al. (2012)) and also by using multiple measures of exit, as described in the following discussion.

We refer to Panel D of Table 4 for univariate evidence on exit. Our analysis is at the level of a financing round. We find that 23,476 (32%) of the financing rounds feature exit through IPOs or M&As. IPOs account for 11%, or a third, of these. In community rounds, 13% exit through IPOs, and 25% exit through M&As, compared with 9% and 19% for noncommunity rounds, respectively. We

[26]Coding every M&A exit requires an exhaustive search for news stories that are not always available because the firms involved are private and not widely followed in the press. The news stories often do not let us sharply discriminate between failures and successes.

https://doi.org/10.1017/S002210901900005X Published online by Cambridge University Press

TABLE 12
Innovation and VC Communities, Financing Stage

Table 12 reports the ordinary least squares (OLS) estimates for innovation activity between 1985 and 2006. The dependent variable in specifications 1–3 is the natural logarithm of (1 plus) the number of patents that a portfolio firm applied for in the financing year. In specifications 4–6, the dependent variable is the natural logarithm of (1 plus) the per-patent citation count that is corrected for truncation bias. Both total patents and average citations are adjusted for average innovation in the same cohort (based on application year and technology class). The key variable of interest is EARLY_STAGE × COMMUNITY. Specifications 2 and 5 (3 and 6) are based on subsamples of observations where there are no patent applications (at least one patent application) in the 3 years prior to the financing round. The observations are at the company × round level. We include all control variables in Table 11 but do not list them for brevity. See the Appendix for a description of the independent variables. Standard errors are clustered at the portfolio firm level. *t*-statistics are in parentheses. All specifications are overall significant at the 1% level. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

	TOTAL_PATENTS			AVERAGE_CITATIONS		
	All	Prior Innovations		All	Prior Innovations	
		No	Yes		No	Yes
	1	2	3	4	5	6
COMMUNITY	0.002	−0.001	0.009	0.003	−0.012	0.030
	(1.03)	(−0.41)	(1.02)	(0.66)	(−1.39)	(1.64)
EARLY_STAGE	−0.004**	0.027***	−0.022**	0.016***	0.116***	0.026
	(−2.16)	(6.78)	(−2.20)	(3.48)	(7.70)	(1.14)
EARLY_STAGE x COMMUNITY	0.001	0.016***	0.013	0.018**	0.053**	0.056*
	(0.42)	(2.77)	(1.00)	(2.40)	(2.52)	(1.67)
No. of obs.	53,000	16,337	10,245	53,000	16,337	10,245
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Industry fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R^2	0.209	0.081	0.140	0.176	0.083	0.127

find similar patterns when considering exits classified by the number of portfolio firms rather than the number of financing rounds.

Table 13 reports the estimates of multivariate models. The independent variable of interest is the dummy variable for sourcing financing from a community VC. We use the same controls as in the analysis of patenting in Section VII.A. For brevity, we do not report these coefficients. Errors are clustered at the portfolio firm level. Specification 1 in Table 13 is a Cox proportional hazards model in which success is an exit through either M&A or an IPO. We report Cox

TABLE 13
Time to Exit

Specification 1 in Table 13 reports the estimates of a Cox proportional hazards model. The dependent variable is the number of days from financing to the earlier of the date of exit or Apr. 10, 2013. Specifications 2–4 report estimates of a competing hazards model where the event of interest is exit only through an initial public offering (IPO) (specification 2) or an IPO or follow-on financing (specifications 3 and 4) after round 1 or 2. A merger is the competing risk in the competing hazards models. The sample comprises venture capitalist (VC) deals obtained from Venture Economics but excludes non-U.S. investments, angel investors, and VC firms focusing on buyouts. All specifications include year and industry fixed effects (FEs), as well as control variables from Table 11, whose coefficients are not reported for brevity. See the Appendix for a description of the independent variables. Both the specifications are overall significant at 1%. *t*-statistics in parentheses are clustered at the portfolio firm level. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

	Cox Model	Competing Hazards		
		IPO Only	IPO or Round 1	IPO or Round 2
	1	2	3	4
Community	1.130***	1.184***	1.065*	0.961
	(4.46)	(3.76)	(1.87)	(−1.08)
No. of obs.	67,081	67,081	19,104	14,267
Controls	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
Industry fixed effects	Yes	Yes	Yes	Yes

https://doi.org/10.1017/S002210901900005X Published online by Cambridge University Press

estimates in the form of an exponentiated hazards ratio where a coefficient greater than 1 for a variable indicates that it speeds exit. The hazard ratio for community VCs is 1.13 and is significant at the 1% level. Thus, community VC financing is associated with faster investment exit.

Specifications 2–4 of Table 13 report estimates of competing hazards models. Specification 2 defines IPOs as a success but recognizes that IPOs are observed only conditional on not having a prior M&A.[27] Here too, community VC financing is associated with a quicker exit. Finally, in specifications 3 and 4, we consider rounds 1 and 2 separately. Although this test reduces the sample size, it serves as a conservative robustness check that uses each round only once.[28] We find that community coefficients are greater for round 1 financings. Sourcing financing from a community VC appears to have a greater effect in earlier rounds, which are likely to involve more intense scrutiny and greater ambiguity about business prospects.

VIII. Conclusion

Syndication is a pervasive feature of venture capital financing. Over a period of time, the typical VC syndicates often and with multiple partners. We show that many VCs exhibit a preference for a small set of other VCs as syndicate partners. These preferences lead to the formation of groups of VCs, or preferred-partner clusters, that we call "VC communities." These groups are economic agglomerates with the property that their members are likely to pick partners within the group, although not exclusively so. We make precise the mathematical construct underlying communities, discuss the computational techniques to identify communities, and implement community detection on a large sample of VC syndications to understand the economics and economic activities associated with communities.

We find that the number of communities and VCs in communities varies over time. On average, 11% of VCs belong to communities. The existence of VC communities is consistent with the view of strong and weak ties in the sociology literature (Granovetter (1973)). Strong ties to familiar partners are useful because they lower transaction costs and information asymmetries. However, ties to less familiar partners are also beneficial because they generate new learning and avoid groupthink. Communities reflect precisely this phenomenon, the joint existence of a core of strong ties for a VC along with many other weaker ones.

We find a pecking order among VCs. Community VCs are more active, make larger investments, are more likely to syndicate, have greater amounts of assets under management, are better networked, and invest greater amounts in portfolio firms with more diverse portfolios. At the other end are small, less networked VCs that do not belong to communities or partner with community VCs. Those that partner with community VCs but are not themselves members of communities display intermediate characteristics. Community members tend to be homogeneous in key attributes. Machine-learning tools classify realized communities

[27] We get similar results if we consider IPOs alone, without accounting for the merger hazard.

[28] We also obtain positive community coefficients if we consider community formation based on first-round financings alone.

https://doi.org/10.1017/S002210901900005X Published online by Cambridge University Press

into 3 clusters roughly ordered according to the age and the reach of VCs. We use supervised learning to better understand the drivers of these types of communities. Statistics based on the reduction in Gini impurity show that VC age, portfolio firm size, and centrality have the highest discriminatory power relative to the other variables.

We find a significant association between sourcing financing from community VCs and economic outcomes. Community VC financing is associated with innovation, as measured by patenting data, and the effects are particularly pronounced for early-stage firms and firms without prior patenting. Community VC financing is also associated with a greater likelihood of maturation of portfolio firms, as reflected in investment exit. These effects could reflect the value-addition by community VCs or the selection effects of better firms being financed by community VCs.

Our approach to analyzing the existence and nature of partner preferences has many applications outside VCs. The potential applications include loan syndications, underwriting, and even interfirm alliances. As Robinson (2008) highlights, simultaneous, nonoverlapping interfirm collaborations are quite common even between firms that compete in some product markets. An interesting question is whether these types of collaborations also have preferred-partner clustering. Our study provides a coherent framework and a set of practical tools to analyze such questions. One can then better understand the *entire set* of partnerships of firms rather than the one-off ties that firms form.

Finally, communities appear to be interesting organizational forms that lie in between formal conglomerates and firms demarcated by legal organizational boundaries. Spot contracting between legally separate entities helps avoid the inflexibility and complexity of running large conglomerates. However, it also compromises the benefits of soft information flows and relationships from an integrated conglomerate. Communities can be regarded as organizational intermediates that provide some benefits of both forms of organization that lie somewhere in between hard-boundary conglomerates that internalize all transactions and arm's-length spot contracting with outside partners.

Appendix. Variable Definitions

This Appendix defines the variables used in our study. A row in the VentureXpert data set we download is at the level of a portfolio company × financing round × VC firm × VC fund; these aspects are identified by the variable names company_id, deal_no, firm_id, and fund_id, respectively. The standard geographic indicators include state and MSA, and in addition, the data set includes a region identifier for the VC firm and the portfolio company in the round. The industry classification of the portfolio company is the variable company_veic6c. The financing round is identified with a stage by the variable stage1 in the database. Other important primary fields for the VC firm include the firm's ownership type, assets under management for the fund, and the fund's investment in the round, which are firm_type, fund_size, and est_fund_total, respectively. We use these primary fields in combination with other primary fields in the VentureXpert database to derive a number of variables employed in the econometric specifications, as we describe next. The HHI for an attribute across multiple buckets is the squared sum of the shares of the attribute for each bucket.

https://doi.org/10.1017/S002210901900005X Published online by Cambridge University Press

AGE: Number of years between a VC's last investment in year t and the VC firm's founding year (firm_founded_year).

AUM_ROUND: Natural logarithm of 1 plus the average AUM (in $millions) of the participating VCs' funds that invested until the year prior to the financing round.

AVERAGE_VC_AUM: the average AUM of all funds of the VC firm in the 5-year period

CENTRALITY: For each VC, we compute the eigenvector centrality based on syndicated rounds in a 5-year rolling window.

COMMUNITY: Equals 1 if there is at least one community VC in the financing round, and 0 otherwise.

COMMUNITY_SIZE: The number of VC firms in a community.

COMPANY_GEOGRAPHICAL_CLUSTER: A dummy variable for whether the state of the portfolio company headquarters is California or Massachusetts.

COMPANY_REGION_HHI: HHI based on shares of deals in the portfolio company geographic region (the variable company_ region_code, as noted previously) in the relevant 5-year rolling window.

COMPANY_REGION_RANK: The rank of the geographic region according to the amount invested by all VCs in a 5-year rolling window.

COMPANY_REGION_VARIATION: The proportion of a VC's deals in a region minus the average proportion across all other VCs in the region in a 5-year rolling window. The community-level variable is similarly defined at the community level.

CORPORATE_VC: A dummy variable equal to 1 if at least one VC in a round has firm_type equal to "Corporate PE, Venture."

EARLY_STAGE: A dummy variable equal to 1 if the variable stage1 for the round is either "Early Stage" or "Startup-Seed."

EARLY_STAGE_FOCUS: The natural logarithm of 1 plus the proportion of early-stage companies that the VCs participating in a round invested in until prior to the financing round.

EXPERIENCE: The natural logarithm of 1 plus the average age of the participating VCs 1 year prior to the round winsorized at the 1% level.[29]

FI_VC: A dummy variable equal to 1 if at least one VC in a round has firm_type containing the terms "Financial," "Bank," "Insurance," or "Pension."

INDUSTRY_FOCUS: The natural logarithm of 1 plus the proportion of companies funded by the participating VCs that have the same industry code company_veic6c as the portfolio company until the year prior to the financing round.

INDUSTRY_HHI: HHI based on shares of deals in the portfolio company industry (the variable company_veic6c) in the relevant 5-year rolling window.

INDUSTRY_RANK: The rank of the industry according to the amount invested by all VCs in a 5-year rolling window.

INDUSTRY_VARIATION: The proportion of a VC's deals in an industry (company_veic6c) minus the average proportion across all other VCs in the industry in a 5-year rolling window. The community-level variable is similarly defined at the community level.

[29] Our definition modifies the Lindsey (2008) definition on two fronts. First, we consider age based on the VC firm's founding year rather than its entry into Venture Economics. Second, we consider a VC's experience based on time periods prior to the financing round in question.

https://doi.org/10.1017/S002210901900005X Published online by Cambridge University Press

IPO_RATE: The natural logarithm of 1 plus the average of each participating VC's ratio of IPOs to number of portfolio companies invested in over the last 3 years prior to the financing round (Krishnan and Masulis (2011)).

PAST_INNOVATION: The natural logarithm of 1 plus the average number of 3-year forward citations per patent received by the portfolio company in the 3 years prior to the financing-round date.

STAGE_HHI: HHI based on shares of deals in the portfolio company stage (the variable stage1) in the relevant 5-year rolling window.

STAGE_RANK: The rank of the stage according to the amount invested by all VCs in a 5-year rolling window.

STAGE_VARIATION: The proportion of a VC's deals in a stage (stage1) minus the average proportion across all other VCs in the stage in a 5-year rolling window. The community-level variable is similarly defined at the community level.

SYNDICATED: A dummy variable equal to 1 if the count of unique VCs in a round exceeds 1, and 0 otherwise.

VC_AUM: Aggregate of fund_size across all funds of the VC firm (in $millions).

VC_GEOGRAPHICAL_CLUSTER: A dummy variable for whether the state of the VC firm headquarters is California or Massachusetts.

VC_MSA_HHI: HHI based on the proportion of VCs in a community from each MSA in the relevant 5-year rolling window.

VC_OWNERSHIP_HHI: HHI based on the proportion of VCs in a community from each ownership type in the relevant 5-year rolling window.

VC_REGION_HHI: HHI based on the proportion of VCs in a community from each geographical region in the relevant 5-year rolling window.

VC_STATE_HHI: HHI based on the proportion of VCs in a community from each U.S. state in the relevant 5-year rolling window.

References

Barabasi, A. L., and R. Albert. "Emergence of Scaling in Random Networks." *Science*, 286 (1999), 509–512.

Bengtsson, O., and D. Hsu. "How Do Venture Capital Partners Match with Startup Founders?" Working Paper, University of Pennsylvania (2010).

Bernstein, S.; X. Giroud; and R. Townsend. "The Impact of Venture Capital Monitoring." *Journal of Finance*, 71 (2016), 1591–1622.

Bhagwat, V. "Manager Networks and Coordination of Effort: Evidence from Venture Capital Syndication." Working Paper, University of Oregon (2013). Available at http://dx.doi.org/10.2139/ssrn.1962146.

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

Brander, J. A.; R. Amit; and W. Antweiler. "Venture-Capital Syndication: Improved Venture Selection vs. the Value-Added Hypothesis." *Journal of Economics and Management Strategy*, 11 (2002), 423–452.

Breiman, L. *Classification and Regression Trees*. New York, NY: Routledge (1984).

Burdick, D.; S. R. Das; M. Hernandez; H. Ho; G. Koutrika; R. Krishnamurthy; L. Popa; I. R. Stanoi; and S. Vaithyanathan. "Extracting, Linking and Integrating Data from Public Sources: A Financial Case Study." *IEEE Data Engineering Bulletin*, 34 (2011), 60–67.

Cestone, G.; J. Lerner; and L. White. "The Design of Syndicates in Venture Capital." Working Paper, Harvard Business School. Published as Fundacion BBVA White Paper (2006).

Chen, H.; P. Gompers; A. Kovner; and J. Lerner. "Buy Local? The Geography of Successful Venture Capital Expansion." *Journal of Urban Economics*, 67 (2010), 90–102.

https://doi.org/10.1017/S002210901900005X Published online by Cambridge University Press

Chidambaran, N. K.; S. Kedia; and N. R. Prabhala. "CEO-Director Connections and Fraud." Working Paper, University of Maryland (2010).
Cohen, L.; A. Frazzini; and C. Malloy. "Sell-Side School Ties." *Journal of Finance*, 65 (2010), 1409–1437.
Cohen, L.; A. Frazzini; and C. Malloy. "Hiring Cheerleaders: Board Appointments of Independent Directors." *Management Science*, 58 (2012), 1039–1058.
Cook, E. F., and L. Goldman. "Empiric Comparison of Multivariate Analytic Techniques: Advantages and Disadvantages of Recursive Partitioning Analysis." *Journal of Clinical Epidemiology*, 37 (1984), 721–731.
Cornelli, F., and O. Yosha. "Stage Financing and the Role of Convertible Securities." *Review of Economic Studies*, 70 (2003), 1–32.
Currarini, S.; M. Jackson; and P. Pin. "An Economic Model of Friendship: Homophily, Minorities, and Segregation." *Econometrica*, 77 (2012), 1003–1045.
Da Rin, M.; T. Hellmann; and M. Puri. "A Survey of Venture Capital Research." In *Handbook of the Economics of Finance*, G. M. Constantinides, M. Harris, and R. M. Stulz, eds. Amsterdam, Netherlands: Elsevier, North Holland (2013).
Du, Q. "Birds of a Feather or Celebrating Differences? The Formation and Impact of Venture Capital Syndication." Working Paper, University of British Columbia (2011). Available at http://dx.doi.org/10.2139/ssrn.1360545.
Engelberg, J.; P. Gao; and C. Parsons. "The Value of a Rolodex: CEO Pay and Personal Networks." *Review of Financial Studies*, 26 (2013), 79–114.
Fortunato, S. "Community Detection in Graphs." *Physics Reports*, 486 (2009), 75–174.
Fracassi, C., and G. Tate. "External Networking and Internal Firm Governance." *Journal of Finance*, 67 (2012), 153–194.
Gertler, M. S. "Being There: Proximity, Organization and Culture in the Development and Adoption of Advanced Manufacturing Technologies." *Economic Geography*, 7 (1995), 1–26.
Goldfarb, B.; D. Kirsch; and D. Miller. "Was There Too Little Entry in the Dot Com Era?" *Journal of Financial Economics*, 86 (2007), 100–144.
Gompers, P., and J. Lerner. "The Venture Capital Revolution." *Journal of Economic Perspectives*, 15 (2001), 45–62.
Gompers, P.; V. Mukharlyamov; and Y. Xuan. "The Cost of Friendship." *Journal of Financial Economics*, 119 (2012), 626–644.
Gonzalez-Uribe, J. "Venture Capital and the Diffusion of Knowledge." Working Paper, London School of Economics (2014). Available at http://ssrn.com/abstract=2405362.
Gorman, M., and W. Sahlman. "What Do Venture Capitalists Do?" *Journal of Business Venturing*, 4 (1989), 231–248.
Granovetter, M. "The Strength of Weak Ties." *American Journal of Sociology*, 78 (1973), 1360–1380.
Granovetter, M. "Economic Action and Social Structure: The Problem of Embeddedness." *American Journal of Sociology*, 91 (1985), 481–510.
Grossman, S., and O. Hart. "The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration." *Journal of Political Economy*, 94 (1986), 691–719.
Guiso, L.; P. Sapienza; and L. Zingales. "The Role of Social Capital in Financial Development." *American Economic Review*, 94 (2004), 526–556.
Gulati, R. "Does Familiarity Breed Trust? The Implications of Repeated Ties for Contractual Choice in Alliances." *Academy of Management Journal*, 38 (1995), 85–112.
Hall, B.; A. Jaffe; and M. Trajtenberg. "The NBER Patent Citations Data File: Lessons, Insights and Methodological Tools." Working Paper 8498, National Bureau of Economic Research (2001).
Hall, B.; A. Jaffe; and M. Trajtenberg. "Market Value and Patent Citations." *RAND Journal of Economics*, 36 (2005), 16–38.
Hart, O., and J. Moore. "Property Rights and the Nature of the Firm." *Journal of Political Economy*, 98 (1990), 1119–1158.
Hegde, D., and J. Tumlinson. "Does Social Proximity Enhance Business Relationships? Theory and Evidence from Ethnicity's Role in US Venture Capital." *Management Science*, 60 (2014), 2355–2380.
Hellmann, T. J., and M. Puri. "Venture Capital and the Professionalization of Start-Up Firms: Empirical Evidence." *Journal of Finance*, 57 (2002), 169–197.
Ho, T. K. "Random Decision Forests." In *Proceedings of the 3rd International Conference on Document Analysis and Recognition*, Piscataway, NJ: Institute of Electrical and Electronics Engineers (1995).
Ho, T. K. "The Random Subspace Method for Constructing Decision Forests." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20 (1998), 832–844.

https://doi.org/10.1017/S002210901900005X Published online by Cambridge University Press

Hoberg, G., and G. Phillips. "Product Market Synergies and Competition in Mergers and Acquisitions: A Text-Based Analysis." *Review of Financial Studies*, 23 (2010), 3773–3811.
Hochberg, Y.; L. Lindsey; and M. Westerfield. "Resource Accumulation through Economic Ties: Evidence from Venture Capital." *Journal of Financial Economics*, 118 (2015), 245–267.
Hochberg, Y.; A. Ljungqvist; and Y. Lu. "Whom You Know Matters: Venture Capital Networks and Investment Performance." *Journal of Finance*, 62 (2007), 251–301.
Hochberg, Y.; A. Ljungqvist; and Y. Lu. "Networking as a Barrier to Entry and the Competitive Supply of Venture Capital." *Journal of Finance*, 65 (2010), 829–859.
Hsu, D. "What Do Entrepreneurs Pay for Venture Capital Affiliation?" *Journal of Finance*, 59 (2004), 1805–1844.
Hwang, B., and S. Kim. "It Pays to Have Friends." *Journal of Financial Economics*, 93 (2009), 138–158.
Kaplan, S. N.; B. Sensoy; and P. Stromberg. "How Well Do Venture Capital Databases Reflect Actual Investments?" Working Paper, University of Chicago (2002).
Kaplan, S. N., and P. Stromberg. "Financial Contracting Theory Meets the Real World: Evidence from Venture Capital Contracts." *Review of Economic Studies*, 70 (2003), 281–316.
Kaplan, S. N., and P. Stromberg. "Characteristics, Contracts and Actions: Evidence from Venture Capital Analyses." *Journal of Finance*, 59 (2004), 2177–2210.
Krishnan, C. N. V., and R. W. Masulis. "Venture Capital Reputation." In *The Oxford Handbook of Entrepreneurial Finance*, Douglas J. Cummings, ed. Oxford, England: Oxford University Press (2011).
Lerner, J. "The Syndication of Venture Capital Investments." *Financial Management*, 23 (1994), 16–27.
Lerner, J.; F. Hardymon; and A. Leamon. *Venture Capitalist and Private Equity: A Casebook*, Hoboken, NJ: John Wiley and Sons (2007).
Lindsey, L. "Blurring Firm Boundaries: The Role of Venture Capital in Strategic Alliances." *Journal of Finance*, 63 (2008), 1137–1168.
Maats, F.; A. Metrick; B. Hinkes; A. Yasuda; and S. Vershovski. "On the Completeness and Interchangeability of Venture Capital Databases." Working Paper, University of California, Davis (2008).
McPherson, M.; L. Smith-Lovin; and J. Cook. "Birds of a Feather: Homophily in Social Networks." *Annual Review of Sociology*, 27 (2001), 415–444.
Neher, D. V. "Staged Financing: An Agency Perspective." *Review of Economic Studies*, 66 (1999), 255–274.
Newman, M. "Scientific Collaboration Networks: II. Shortest Paths, Weighted Networks, and Centrality." *Physical Review* E, 64 (2001), 016132.
Newman, M. *Networks: An Introduction*. New York, NY: Oxford University Press (2010).
Pons, P., and M. Latapy. "Computing Communities in Large Networks Using Random Walks." *Journal of Graph Algorithms and Applications*, 10 (2005), 191–218.
Porter, M. A.; P. Mucha; M. Newman; and A. J. Friend. "Community Structure in the United States House of Representatives." *Physica A: Statistical Mechanics and Its Applications*, 386 (2007), 413–438.
Porter, M. E. "Location, Competition and Economic Development: Local Clusters in a Global Economy." *Economic Development Quarterly*, 14 (2000), 15–34.
Robinson, D. "Strategic Alliances and the Boundaries of the Firm." *Review of Financial Studies*, 21 (2008), 649–681.
Robinson, D., and B. Sensoy. "Private Equity in the 21st Century: Cash Flows, Performance, and Contract Terms from 1984–2010." Working Paper, Ohio State University (2011).
Robinson, D., and T. Stuart. "Financial Contracting in Biotech Strategic Alliances." *Journal of Law and Economics*, 50 (2007), 559–596.
Seru, A. "Firm Boundaries Matter: Evidence from Conglomerates and R& D Activity." *Journal of Financial Economics*, 111 (2014), 381–405.
Sorensen, M. "How Smart Is Smart Money? A Two-Sided Matching Model of Venture Capital." *Journal of Finance*, 62 (2007), 2725–2762.
Sorensen, M. "Learning by Investing: Evidence from Venture Capital." Working Paper, Columbia University 2008.
Sorenson, O., and T. Stuart. "Syndication Networks and Spatial Distribution of Venture Capital Investment." *American Journal of Sociology*, 106 (2001), 1546–1588.
Tian, X. "The Causes and Consequences of Venture Capital Stage Financing." *Journal of Financial Economics*, 101 (2011), 132–159.
Uzzi, B. "Social Structure and Competition in Interfirm Networks: The Paradox of Embeddedness." *Administrative Science Quarterly*, 42 (1997), 35–67.

https://doi.org/10.1017/S002210901900005X Published online by Cambridge University Press