INSTITUT POLYTECHNIQUE DE PARIS

Institut Polytechnique de Paris

Master IREN - Network Industries and Digital Economy

Nested Syndication Networks: Community Structure and Hierarchical Organization in Innovation Ecosystems

Master Thesis in Innovation Economy

Authored by João Melga

Advised by Jean-Michel Dalle

August 25, 2025

Abstract

This study analyzes venture capital syndication networks using network theory to understand how investors organize and collaborate in funding startups. We examine 104,618 investment records from Crunchbase involving 38,843 investors and 16,932 companies to identify structural patterns in investor communities.

Using greedy modularity optimization, we discovered 170 distinct investor communities, with three large communities containing over 12,000 investors (75% of the network). These communities exhibit power-law degree distributions typical of scale-free networks but differ significantly in investment activity and organization.

Our key finding is the emergence of significantly high nested structures within Community 2, which is dominated by Silicon Valley investors. This hierarchical organization enables substantially higher transaction volumes (33.6% of all investments) compared to similarly-sized communities. The nested structure suggests that informal investment hierarchies systematically influence funding accessibility for entrepreneurs.

The concentration of nested structures specifically within Silicon Valley indicates that geographic clustering in innovation hubs may facilitate hierarchical investor relationships. High-degree investors in the nested community, particularly early-stage firms like SV Angel, function as network hubs that create efficient capital allocation pathways.

These findings challenge assumptions about random mixing in venture capital markets and suggest that network topology, rather than community size alone, determines investment ecosystem efficiency. The hierarchical organization may enhance stability against random investor departures but creates vulnerability to targeted removal of highly connected nodes.

Our results provide insights for ecosystem development strategies and suggest that understanding network structure is crucial for predicting investment patterns and improving access to capital for entrepreneurs.

Contents

List of Figures

List of Tables

1 Introduction

Two or more investors (e.g., venture capital funds, business angels, banks) co-investing on the same in the same round/startup/venture is known as syndication. This is the standard term in finance, economics, and entrepreneurship research. For example: venture capital syndication, loan syndication, investment syndication.

This collaborative behavior goes beyond simple risk-sharing and reflects information asymmetries and screening challenges inherent in investment markets. Understanding how syndication patterns emerge and evolve requires examining the social, economic, and structural forces that shape investor relationships within innovation ecosystems.

In venture capital markets (innovation financing included), syndication serves as a coordination mechanisms used by Venture Capitalists (VCs) to operate in complex networks where investment decisions are embedded within social structures, geographic clusters, and institutional relationships.

However, traditional approaches to studying venture capital often focus on individual investment decisions or firm-level characteristics, leaving the broader systemic patterns of investor coordination (e.g. syndication) to be further explored.

This study applies network theory to analyze these coordination patterns in the US, revealing how investors organize into distinct network communities with different structural properties and organizational patterns that may influence capital allocation efficiency, deal flow, and market access.

Special attention is given for the fact that a community of proiminent VCs in the Sylicon Valley demonstrate nested hierarchical organization patterns, while others maintain more random, hard-to-define partnership structures.

The following sections examine the theoretical foundations for syndication behavior, drawing from literature in innovation financing, social embeddedness, network theory, and ecological systems to build a comprehensive framework for understanding syndication investments networks organization.

Nevertheless, implications on market efficiency, geographic clustering effects, and the role of innovation hubs in shaping investment ecosystems is discussed, as well as future reasearch opportunities are presented.

1.1 Syndication in Innovation Financing

Theoretical foundations for syndication behavior comes from literature on screening and information sharing among venture capitalists [8]. In summary, in markets characterized by incomplete information (e.g. innovation financing), the decision to syndicate and the choice of syndication partners are aspects that influence both investment outcomes and network position [8].

In innovation financing, when investors face uncertainty about startup quality and potential, syndication serves multiple strategic functions: it enables knowledge pooling from partners with complementary expertise [13], reduces individual exposure to high-risk ventures, and provides valuable signals [16] about investment quality through the revealed preferences of co-investors.

In this context, considering that screening innovative ventures is frequently a multi-stage process (venture capitalists define formal pipeline which startups and often their founders need to pass through), syndication plays an important role to reduce the effort and possibilty the number of screening stages, with the cost of disclousring information, as investors are essencially competitors.

1.2 Social Sciences and Embeddedness

Social science theory provides crucial insights into why syndication patterns emerge through social relationships rather than purely economic calculations. Granovetter's concept of embeddedness demonstrates that economic actions are fundamentally driven by social structures and ongoing relationships, challenging the assumption that markets operate through atomized, independent decision-making [12].

In venture capital markets, embeddedness manifests in several important ways. First, investment decisions are embedded within social networks where venture capitalists develop recurring partnerships based on trust, shared investment philosophies, and successful collaboration histories [12]. These relationships create persistent patterns of co-investment that go beyond individual deal characteristics, forming the basis for community structure within investor networks.

Second, geographic embeddedness plays a critical role in shaping investor relationships. Venture capitalists concentrated in innovation hubs like Silicon Valley develop dense local networks through frequent face-to-face interactions, shared professional communities, and common institutional affiliations [12]. This geographic clustering creates information advantages and relationship-building opportunities that can influence syndication patterns and network organization.

Third, embeddedness creates informal hierarchies of influence and reputation within investor communities. Rather than operating as independent agents, investors form interconnected networks that share information, coordinate investment strategies, and establish social rankings based on past performance and network position [12]. These social hierarchies can evolve into the structural hierarchies measured through nestedness analysis.

The embeddedness perspective helps explain why certain investor communities develop systematic organizational patterns while others remain more randomly structured. Social relationships, trust networks, and geographic proximity may facilitate the emergence of hierarchical organization by creating conditions where, for instance, specialists naturally align with generalists' investment strategies (not purely random). This social foundation provides the context for understanding how network topology emerges from underlying social processes rather than purely market-driven mechanisms.

1.3 Network Theory and Power-law

Network theory provides a mathematical framework for understanding complex systems of interconnected actors through node-link structures [4]. In venture capital ecosystems, this approach enables systematic analysis of syndication by representing investors as nodes and co-investment relationships as edges.

Three network concepts are particularly relevant for understanding venture capital ecosystems. First, network position and centrality metrics quantify an investor's importance within the overall system. Highly central investors often function as gatekeepers who control information flow and access to investment opportunities [4].

In fact, different centrality measures capture distinct aspects of influence: degree centrality identifies investors with many direct partnerships, while betweenness centrality identifies those who bridge otherwise disconnected investor groups.

Second, community structure reveals how investors cluster into distinct groups with denser internal connections than external ones [7]. These communities often reflect shared investment philosophies, geographic proximity, or sector specialization. The modularity optimization methods used in this study identify these natural groupings within the investment network, revealing how social embeddedness manifests in structural patterns.

Last but not least, power-law degree distributions characterize the heterogeneous connectivity patterns in venture capital networks [6]. This mathematical property describes systems where a small number of investors (hubs) maintain extensive partnership networks while most participants have relatively few connections. Power-law distributions arise from preferential attachment mechanisms, where well-connected investors attract disproportionately more new partnerships, reflecting reputation accumulation and resource concentration dynamics.

Empirical evidence shows that venture capital networks consistently exhibit these power-law properties [10]. Hub investors play crucial roles in facilitating deal flow and information diffusion throughout the ecosystem, while also creating potential bottlenecks that may limit access for entrepreneurs outside established networks.

This structural feature creates conditions for the emergence of hierarchical organization patterns measured through, for instance, nestedness analysis.

The methodology section builds directly on these network theory foundations by implementing community detection algorithms that identify investor clusters, while the results section examines how community structure relates to geographic distribution, investment patterns, and hierarchical organization.

1.4 Ecology and Nestedness

The power-law degree distributions discussed above provide important insights into network structure, but they cannot fully capture the hierarchical organization patterns that emerge within investor communities. Ecological network theory offers additional tools for understanding these organizational patterns through the concept of nestedness [14].

Nestedness originates from ecological studies of mutualistic relationships, particularly pollinator-plant networks. In these systems, specialist pollinators (those that visit few plant species) typically interact with subsets of the same plants that generalist pollinators (those that visit many species) choose [14]. This creates a nested hierarchy where specialists operate within the interaction networks established by generalists, rather than forming separate, independent partnerships.

Nestedness describes this specific structural pattern where specialist nodes (those with few connections) tend to connect to subsets of the partners associated with generalist nodes (those with many connections) [14]. In venture capital networks, this means that investors with fewer partnerships typically co-invest with a subset of the same companies that highly connected investors choose. This creates hierarchical organization where less connected investors operate within the partnership networks established by more connected ones.

This hierarchical arrangement has important implications for information flow and market access. In nested networks, information and opportunities tend to flow from generalist nodes (highly connected investors) to specialist nodes (less connected investors), creating asymmetric relationships that can influence deal flow, startup access to capital, and overall market efficiency [14]. Such organization contrasts with random network structures where partnerships form without systematic hierarchical patterns.

Nestedness patterns are particularly relevant for understanding venture capital ecosystems because they can reveal whether certain investors function as gatekeepers who control access to broader investment networks [4]. This has direct implications for startup fundraising success and the overall efficiency of capital allocation within innovation ecosystems [7].

Measuring nestedness requires specialized metrics that can quantify the extent to which observed network patterns deviate from random organization. The NODF (Nestedness based on Overlap and Decreasing Fill) metric provides a standardized approach for measuring these hierarchical patterns and testing their statistical significance against null models of random network formation [14]. This methodological approach enables systematic investigation of whether investor communities exhibit meaningful hierarchical organization or operate through random partnership patterns.

1.5 Research Approach and Contribution

This study addresses the theoretical and methodological challenges outlined above by developing a empirical approach that combines network theory (focusing on bipartite networks) with ecological analysis techniques. We construct a bipartite network representation of venture capital syndication patterns using comprehensive data from the United States market, where co-investment relationships between early-stage and late-stage investors is used as the analytical basis for understanding ecosystem organization.

Our analytical framework integrates community detection algorithms with nestedness analysis. The NODF (Nestedness based on Overlap and Decreasing Fill) metric, combined with statistical significance testing through null model comparisons, enables us to identify investor communities that exhibit systematic hierarchical structures beyond what would be expected from random partnership formation.

The empirical analysis reveals three key contributions to the literature: First, we demonstrate significant heterogeneity in organizational structures across investor communities on US, with some exhibiting pronounced hierarchical patterns while others maintain more distributed partnership networks. Second, we identify a strong relationship between geographic clustering (particularly within Silicon Valley) and the emergence of nested organizational structures that confer measurable advantages in transaction efficiency and capital deployment. Third, we provide quantitative evidence that network topology, rather than community size alone, serves as a critical determinant of investment ecosystem performance.

The remainder of this paper presents our methodological framework, empirical findings, and their implications for understanding venture capital market organization. The Methodology section details our data preprocessing, network construction, and analytical techniques. The Results section provides comprehensive characterization of investor communities, including their geographic distributions, funding patterns, sectoral focus, and quantitative nestedness measurements. Together, these analyses offer new insights into how network structure shapes innovation financing efficiency and market access patterns.

2 Methodology

2.1 Data Cleaning and Preprocessing

This study uses data from Crunchbase, a commercial database that has emerged as a primary source for entrepreneurship and venture capital research [9]. Founded in 2007 as a side project to TechCrunch, Crunchbase has evolved into a crowdsourced platform containing detailed information about startups, venture capital firms, accelerators, and investment rounds globally. The platform relies on multiple data collection methods, including user submissions, automated web crawling, partnerships with data providers, and editorial curation [9].

The dataset was obtained by querying Crunchbase for all investment transactions involving enterprises (startups, scale-ups, and growth companies) domiciled in the United States, which captures the complete investment ecosystem for US-based firms, including both domestic and cross-border capital flows.

Foreign venture capital firms and institutional investors appear in the dataset when they participate in financing rounds of US enterprises. Furthermore, Crunchbase's coverage is particularly good for technology-oriented startups and venture capital transactions, making it well-suited for studies of innovation ecosystems and investment networks.

However, limitations and potential biases must be acknowledged when using Crunchbase data [9], for instance:

- **Geographic bias**: Stronger coverage of US and Western European markets compared to emerging economies.
- **Sectoral bias**: Emphasis on technology and internet companies, potentially underrepresenting traditional industries.

- **Venture capital bias**: Better documentation of VC-backed companies compared to bootstrapped or debt-financed ventures (which may not be problematic for this study).
- **Size bias**: Larger and more successful companies are more likely to be thoroughly documented.
- **Temporal bias**: More recent information tends to be more complete and accurate than historical records.

These biases do not invalidate research using Crunchbase but require careful consideration in study design and interpretation of results. For network analysis of venture capital co-investments, the VC bias may actually enhance data quality by focusing on the target population of interest.

The data preprocessing and cleaning follows established methodologies from entrepreneurship literature [10] and is implemented through the following multi-step procedure:

1. **Company data cleaning**: Removal of companies with incomplete essential information (missing unique identifiers, names, or founding years), exclusion of companies founded after 2017 to allow sufficient time for investment patterns to emerge, and removal of companies with exit status (closed, acquired, or IPO).
2. **Investment data cleaning**: Removal of investment records with missing essential linkage information (company or investor identifiers), elimination of investments with invalid funding amounts (negative or zero values), and validation of funding consistency by excluding rounds where the sum of individual investments does not match the total funding amount reported in the database.
3. **Funding threshold application**: Restriction of the sample to companies that raised

more than $150,000 in total funding to focus on substantive investment relationships and ensure stronger statistical reliability.

4. **Data consistency validation**: Final consistency checks to ensure all investment records reference existing companies in the dataset and all retained companies have at least one valid investment record.

2.2 Network Construction

The network construction process transforms raw investment data into a structured bipartite representation that enables theoretical reasoning about venture capital syndication patterns. This transformation follows the conceptual framework outlined by [4], who distinguish between flow models (focusing on how resources move through networks) and bond models (emphasizing coordination and solidarity among actors).

The construction process involves three sequential transformations, each representing a different conceptualization of network ties, each one described in 1 and in the subsequent paragraphs below.

Stage 1 - Raw event-type investment ties: The initial Crunchbase dataset records discrete transactional relationships between venture capital firms and portfolio companies. These event-type ties represent what [4] characterize as nominalist network constructions, where investment transactions ere defined as the fundamental tie type that constitutes the network in that stage. Each investment event creates a direct dyadic relationship between a venture capital firm and a company, but these isolated ties do not yet reveal the broader structural patterns of co-investment behavior.

Stage 2 - Syndicated investment network construction: The raw event data are aggregated to identify co-investment relationships, where multiple venture capital firms par-

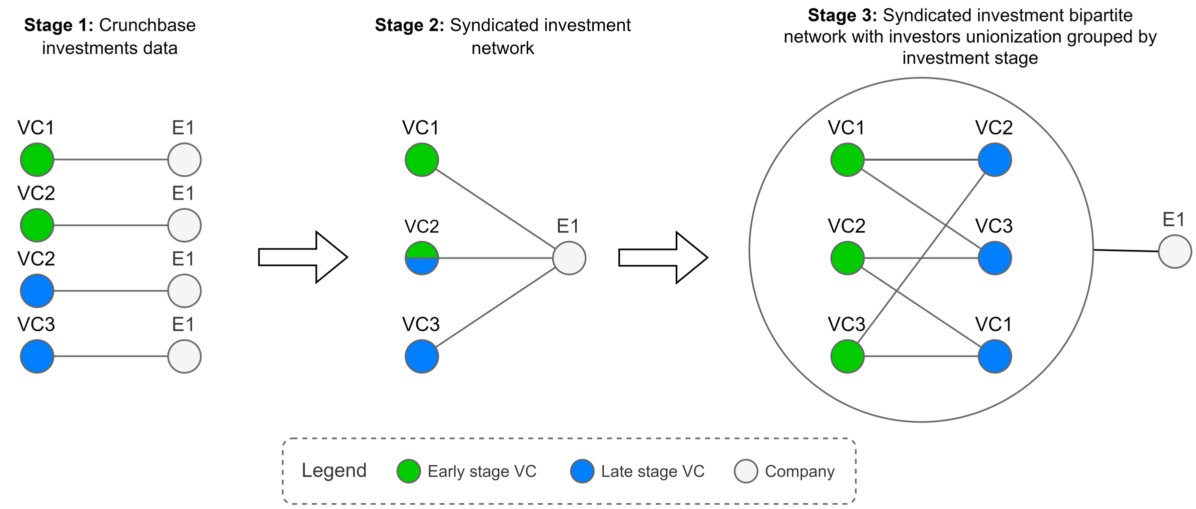


Figure 1: From event-type investment ties to bipartite syndicate networks: structuring vc relations through flow and bond models

ticipate in the same funding round or invest in the same company across different rounds. This aggregation transforms discrete investment events into a network structure where venture capital firms become indirectly connected through their shared ties to portfolio companies. Following [4], this step creates pathways between actors that enable analysis of their relative positions, including measures of centrality and brokerage within the investment ecosystem.

Stage 3 - Bipartite network segmentation by investment stage: The final transformation reorganizes the syndicated network into a bipartite structure that separates investors based on their participation in different funding stages.

First, investment stages are categorized into two main groups:

- Early stages: angel, pre-seed, seed, and Series A
- Late stages: Series B through Series I

This bipartite structure reflects both flow and bond models as described by [4]. The flow model perspective enables analysis of how resources and information traverse across investment stages, capturing the sequential nature of venture capital financing. Simultaneously, the bond model perspective recognizes that co-investors within each stage can be conceptualized as forming coordinated coalitions, reflecting solidarity ties among investors.

An important consideration in this construction involves investors who participate in both early and late stages (for example, a venture capital firm that invests in both Series A and Series C rounds). In such exceptional cases, the same investor appears as distinct agents on both sides of the bipartite network. This separation is achieved by appending the investment stage identifier to the unique investor UUID during the network construction process, effectively creating stage-specific identities for multi-stage investors. This approach ensures that the bipartite structure remains mathematically valid while preserving the analytical ability to study how the same investor behaves differently across investment stages.

The resulting bipartite graph connects two distinct sets of investors: those participating in early-stage rounds and those participating in late-stage rounds. Edges represent co-investment relationships where early-stage and late-stage investors have both invested in the same company, creating a bridge between different phases of the venture capital investment cycle.

Formally, the bipartite graph $G = (U \cup V, E)$ consists of:

$$U = \{u_1, u_2, \ldots, u_m\} \text{ (late-stage VCs)} \tag{1}$$

$$V = \{v_1, v_2, \ldots, v_n\} \text{ (early-stage VCs)} \tag{2}$$

$$E \subseteq U \times V \text{ (co-investment relationships)} \tag{3}$$

To prevent spurious connections from related entities, investor pairs where the first five characters of their names match are filtered out, reducing the likelihood of including different funds from the same parent organization. Furthermore, investors that participated in both early and late stages receive a suffix so they can be treated as distinct agents for each phase.

2.3 Community Detection

Community structure in the bipartite network is identified using modularity-based optimization methods, building on foundational work by [15] who established modularity as a key metric for community detection in complex networks. For bipartite networks specifically, we employ the greedy modularity optimization algorithm following [2], who introduced a modularity definition tailored for bipartite graphs and developed corresponding algorithms that account for the distinct node sets characteristic of bipartite structures.

The choice of greedy optimization is motivated by its computational efficiency and scalability, as demonstrated by [3] in their widely-adopted Louvain algorithm. This approach iteratively merges communities to maximize the modularity score, which measures the density of connections within communities compared to connections between communities.

For a bipartite network, modularity Q is defined as:

$$Q = \frac{1}{2m} \sum_{i,j} \left[A_{ij} - \frac{k_i k_j}{2m} \right] \delta(c_i, c_j) \tag{4}$$

where A_{ij} is the adjacency matrix, k_i is the degree of node i, m is the total number of edges, c_i is the community of node i, and $\delta(c_i, c_j)$ is 1 if nodes i and j are in the same community, 0 otherwise.

This methodology has proven particularly effective in venture capital research, as demonstrated by [5] who used computational methods over three decades of syndication data to identify venture capital communities. Their work establishes empirical precedent for applying community detection algorithms to co-investment networks, showing that such methods can reveal meaningful structural patterns in the investment ecosystem that may not be apparent from individual investment decisions.

2.4 Nestedness Analysis

Nestedness is a structural property commonly observed in ecological networks [1] that describes the tendency for specialists to interact with a subset of the partners of generalists. In the context of venture capital networks, nestedness would indicate that investors with fewer connections tend to co-invest with a subset of the partners of more connected investors.

We measure nestedness using the NODF (Nestedness based on Overlap and Decreasing Fill) metric [1], which has become a standard measure for quantifying nested patterns in bipartite networks. The NODF metric is particularly well-suited for network analysis as it provides a standardized measure that accounts for both the overlap of connections and the degree differences between nodes [11].

For a bipartite adjacency matrix M with rows and columns sorted by decreasing degree, NODF is calculated as:

$$NODF = \frac{NODF_{rows} + NODF_{columns}}{2} \tag{5}$$

where:

$$NODF_{rows} = \frac{100}{R(R-1)/2} \sum_{i=1}^{R-1} \sum_{j=i+1}^{R} \frac{|N_i \cap N_j|}{k_j} \text{ if } k_i > k_j \tag{6}$$

$$NODF_{columns} = \frac{100}{C(C-1)/2} \sum_{i=1}^{C-1} \sum_{j=i+1}^{C} \frac{|N_i \cap N_j|}{k_j} \text{ if } k_i > k_j \tag{7}$$

Here, R and C are the number of rows and columns, N_i represents the set of connections for node i, and k_i is the degree of node i.

Using this method, NODF values range between 0 and 1 (perfect nestedness).

2.5 Statistical Significance Testing

To determine whether observed nestedness values differ significantly from what would be expected by chance, we employ a null model approach using the Curveball algorithm [17]. This algorithm generates randomized matrices that preserve the degree sequence of both node sets while randomizing the connection patterns, representing a "hard" constraint null model that maintains structural properties while randomizing interaction patterns [11].

The choice of degree-preserving null models is critical for nestedness interpretation, as the degree sequence itself can influence apparent nestedness patterns. By preserving degree sequences, we ensure that observed nestedness reflects genuine structural organization rather than mere consequences of heterogeneous node connectivity.

For each community, we generate 100 null matrices using 10,000 Curveball iterations. The statistical significance is assessed by comparing the observed NODF score against the distribution of null model scores:

Generate 1000 null matrices instead

The standardized Z-score is calculated to quantify how many standard deviations the observed nestedness differs from the null expectation:

$$Z = \frac{NODF_{observed} - \mu_{null}}{\sigma_{null}} \tag{8}$$

where μ_{null} and σ_{null} are the mean and standard deviation of the null distribution, respectively.

The p-value is calculated empirically from the null distribution as the proportion of randomized matrices that exhibit nestedness equal to or greater than the observed value:

$$p = \frac{1 + \sum_{i=1}^{N} I(NODF_{null,i} \geq NODF_{observed})}{N + 1} \tag{9}$$

where N is the number of null matrices (100 in our case), $NODF_{null,i}$ is the nestedness score of the i-th null matrix, and $I(\cdot)$ is an indicator function that equals 1 when the condition is true and 0 otherwise. The addition of 1 in both numerator and denominator provides a conservative estimate that avoids p-values of exactly zero.

The p-value represents the probability of observing nestedness as high as or higher than the observed value under the null hypothesis of random co-investment patterns. Communities with $p < 0.05$ are considered to have significantly high nestedness (when observed values exceed null expectations) or significantly low nestedness (when observed values fall below null expectations), indicating that the observed nested structure is unlikely to have arisen by chance alone.

While both Z-scores and p-values assess statistical significance, they provide complementary information: the Z-score quantifies the magnitude of deviation from the null expectation in standardized units, while the p-value provides the probability of observing such deviation under the null hypothesis. In our analysis, we primarily rely on p-values for significance testing as they directly quantify the statistical evidence against the null hypothesis of random network structure.

3 Results

3.1 Dataset Characterization

The dataset scraped from Crunchbase contains data about 22,527 companies (startups, ...), 38,843 investors, and 147,832 invesments.

After the data cleaning process described in the *Methodology* section, the dataset used to further generate the syndication networks yields 104,618 investment records, representing transactions among 16,932 companies and 11,279 investors, totaling $3.19 tri USD invested from 1990 until 2025, even though we are going to concentrate on invesment data among the 20 years window from 2004 to 2024.

3.1.1 Syndication Trends

A significant part of these investments records represent syndicated investments, as showed in figure 2.

3.1.2 Investment Activity

Figure 3 presents the overall investment activity trends, while Figure 4 provides a detailed breakdown of temporal patterns across different investment characteristics.

The investment activity data demonstrates clear temporal variations that might reflect broader economic cycles and venture capital market maturation.

The analysis reveals growth in venture capital activity over the study period, with remarkable acceleration followed by decreasing period in recent years. From 2004 to 2024, the venture capital ecosystem experienced substantial expansion, with total investments

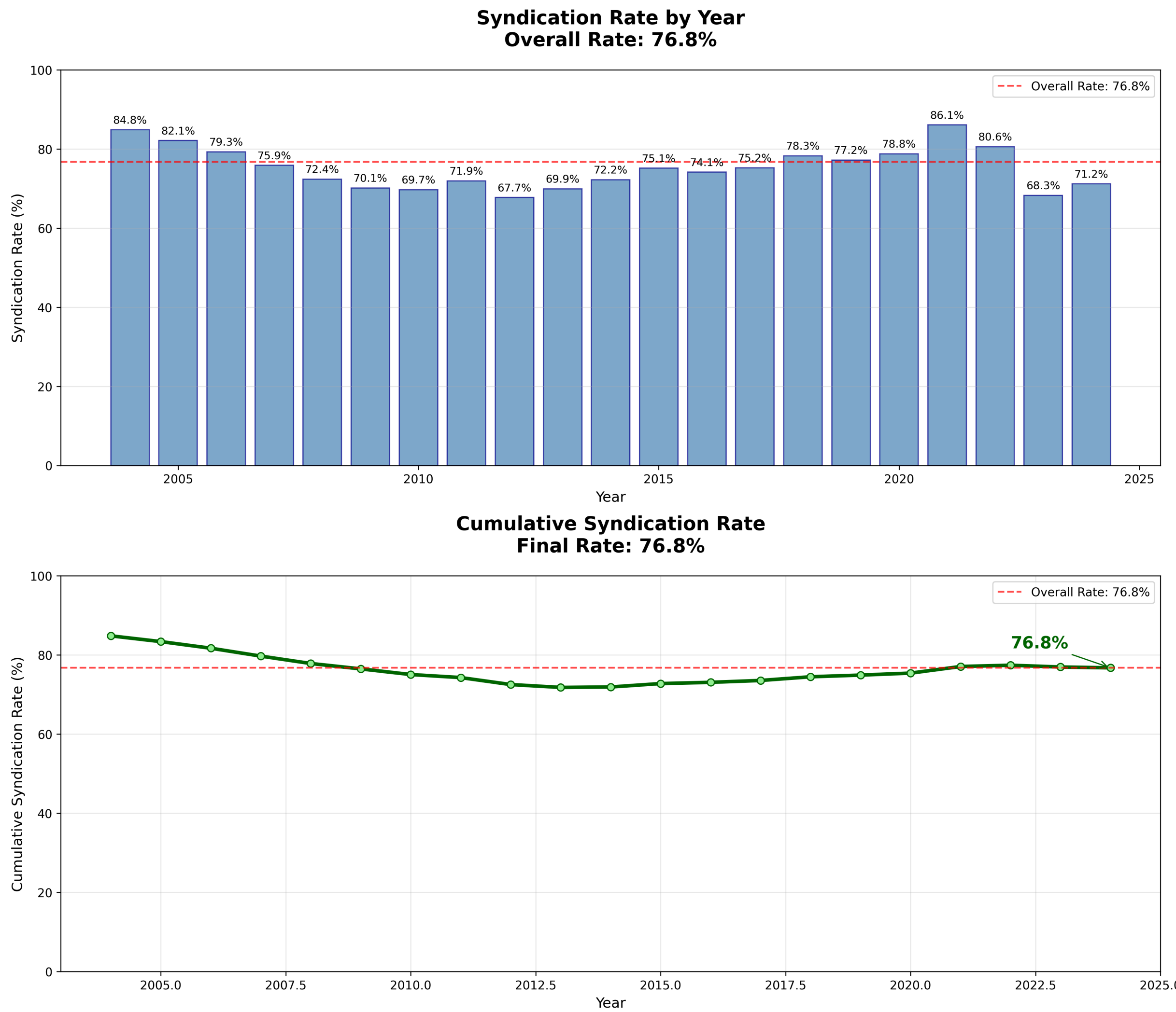


Figure 2: Evolution of syndicated investment trends in the Crunchbase dataset. The figure illustrates changes in the proportion and of syndicated investments over time, highlighting the prevalence of co-investment activity among venture capital investors.

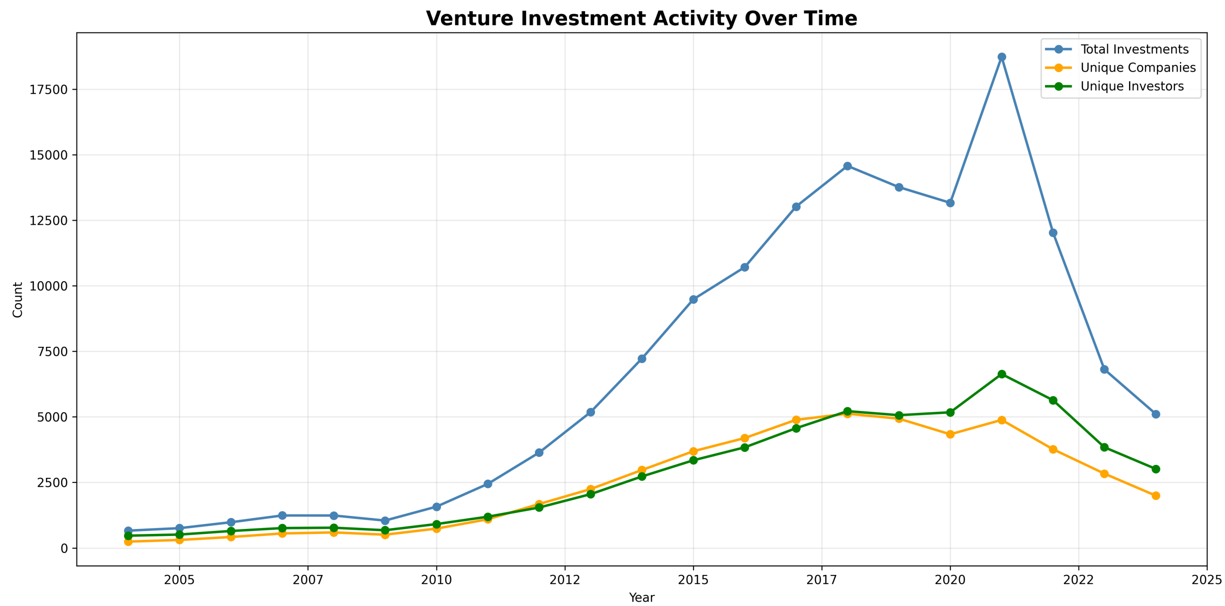


Figure 3: Venture capital investment activity over time. The figure shows the evolution of investment transactions and funding patterns across the study period, highlighting major trends and cyclical patterns in venture capital deployment.

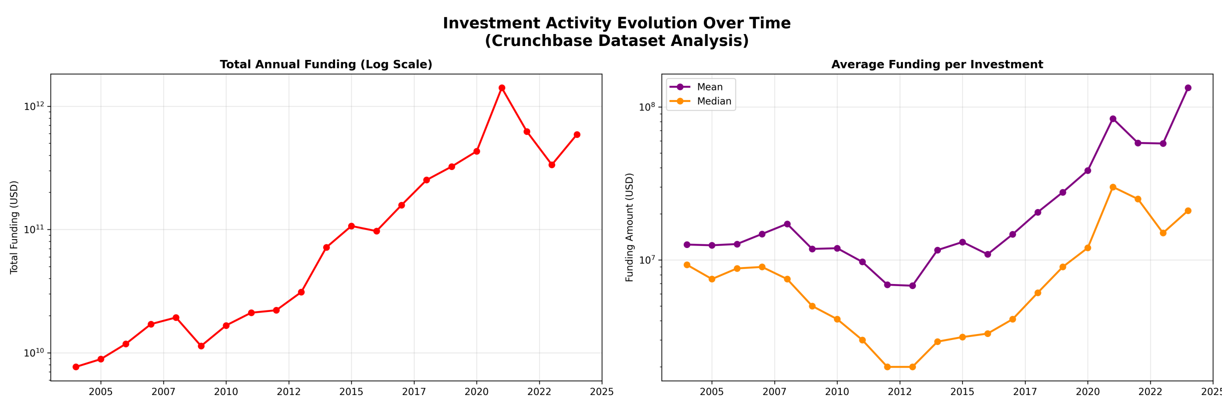


Figure 4: Temporal evolution of investment activity characteristics. This comprehensive view illustrates how different aspects of venture capital investment behavior have changed over time, including variations in funding amounts, transaction frequency, and market participation patterns.

increasing by 674.5% and unique investors growing by 544.4%.

The year 2021 emerged as a unique moment for venture capital activity, representing peak performance across multiple dimensions. This year recorded the highest number of investments (18,741 transactions), the largest total funding amount ($1.41 trillion), and the greatest investor participation (6,636 unique investors). This convergence of peak metrics in 2021 reflects both the maturation of the venture capital industry and the exceptional market conditions during the post-pandemic economic recovery.

The analysis reveals that investment frequency and average funding amounts do not necessarily follow identical temporal patterns. While transaction volumes may increase during certain periods, average investment sizes can exhibit different trends, suggesting varying risk appetites and market conditions across different time windows.

The substantial growth rates observed between 2004 and 2024 indicate not only market expansion but also the increasing institutionalization of venture capital as an asset class.

These temporal patterns provide important context for understanding the network structures observed in the community analysis, as investment behaviors and syndication patterns may be influenced by the broader economic environment during different time periods.

add insights from Crunchbase posts on why investment activities are decreasing

3.1.3 Investment Stages Evolution

Figures 5 and 6 illustrates the analysis of investment stages over time across the venture capital ecosystem

The investment stage distribution demonstrates a clear early-stage focus within the venture capital ecosystem. Seed funding represents the largest single category at 23.3%

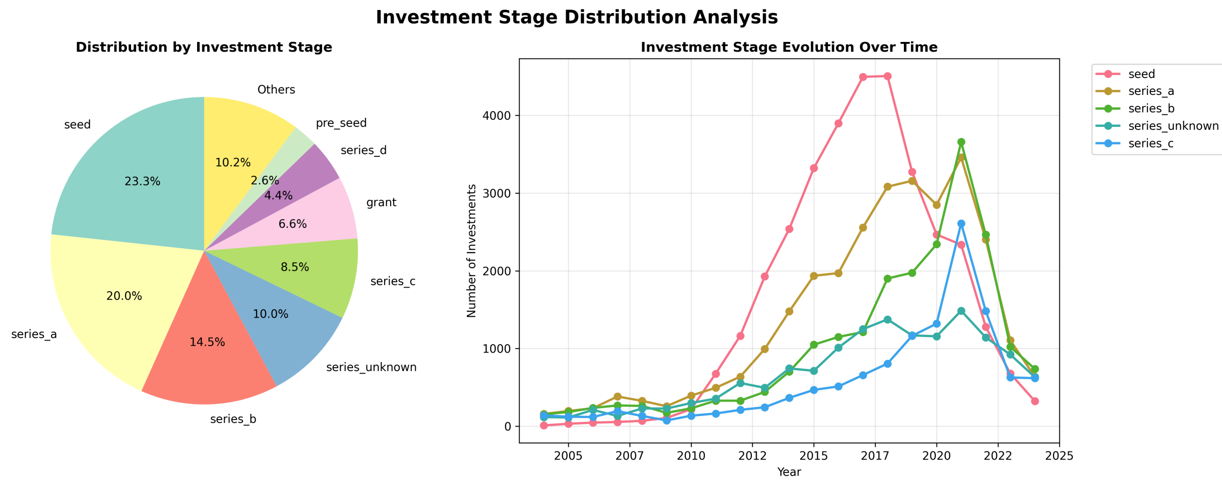


Figure 5: Distribution of investment stages across the venture capital dataset. The figure shows the prevalence of different funding stages, highlighting the concentration of activity in early-stage investments (seed and Series A) which together account for over 43% of all transactions.

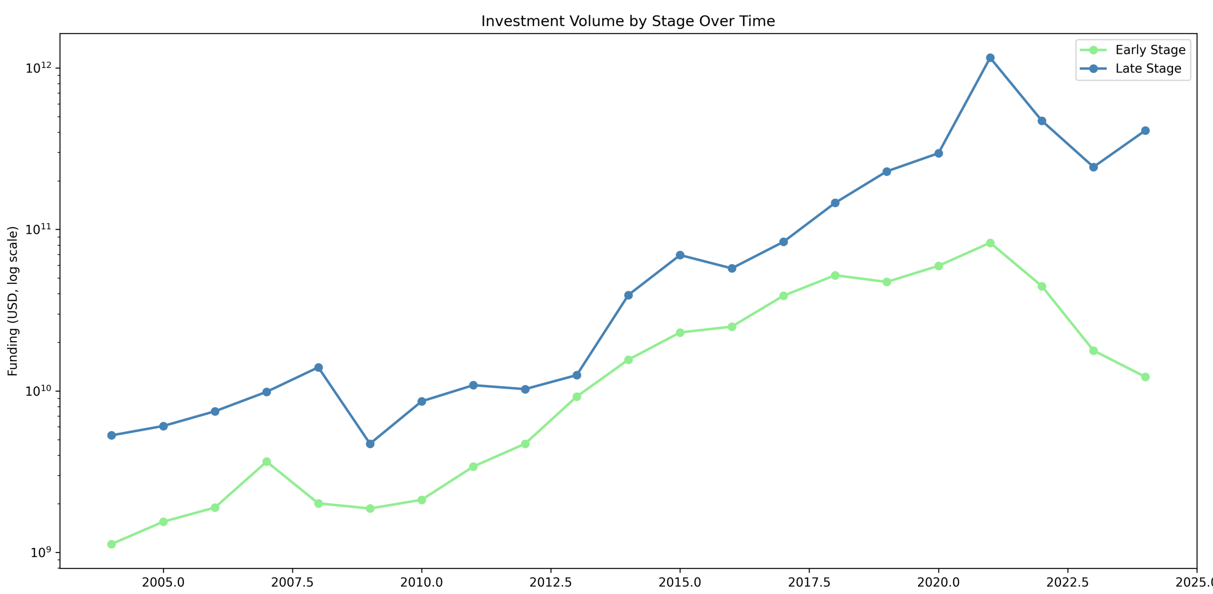


Figure 6: Temporal evolution of investment stage distribution. This chart illustrates how the composition of different funding stages has evolved over time.

of all investments (33,419 transactions), followed closely by Series A at 20.0% (28,703 transactions). This concentration in early-stage funding reflects the fundamental role of venture capital in supporting nascent companies during their most critical development phases.

When aggregated by stage classification, early-stage investments (angel through Series A) comprise 46.7% of all transactions (66,978 investments), substantially exceeding late-stage investments (Series B and beyond) at 30.4% (43,635 investments). The remaining 22.9% consists of alternative funding mechanisms including grants, convertible notes, and specialized instruments.

However, temporal analysis also reveals a shift in capital allocation patterns. While early-stage investments dominate by transaction count, late-stage funding has achieved financial dominance in recent years. From 2004 to 2024, late-stage funding experienced exponential growth, escalating from $5.31 billion in 2004 to $409.52 billion in 2024 (a 77 times increase). In contrast, early-stage funding grew from $1.13 billion to $12.24 billion over the same period, representing an 11 times increase.

This divergence became particularly pronounced after 2013, when late-stage funding began substantially outpacing early-stage investments. By 2021, late-stage funding reached an unprecedented $1.16 trillion, passing the $82.74 billion in early-stage funding by a factor of 14.

Even during the market correction of 2022-2024, late-stage funding maintained its dominance, suggesting a structural trend toward growth-stage capital deployment.

This capital concentration in late-stage investments may indicate the venture capital industry's evolution toward supporting unicorn creation and pre-IPO growth, where individual transactions can exceed hundreds of millions of dollars. The trend suggests that while early-stage investments remain crucial for company formation, the majority of venture capital deployment now focuses on scaling proven business models.

Some strong affirmatives need to have supporting references

3.1.4 Geographic Distribution

As expected, as our dataset query was built to get US companies, the geographic distribution of venture capital investments reveals US-concentration pattern on global scale (Figure 7). At US regional levels, Californian region (where Sylicon Valley is located) concentrates the invesment registers, as illustrated in Figure 8.

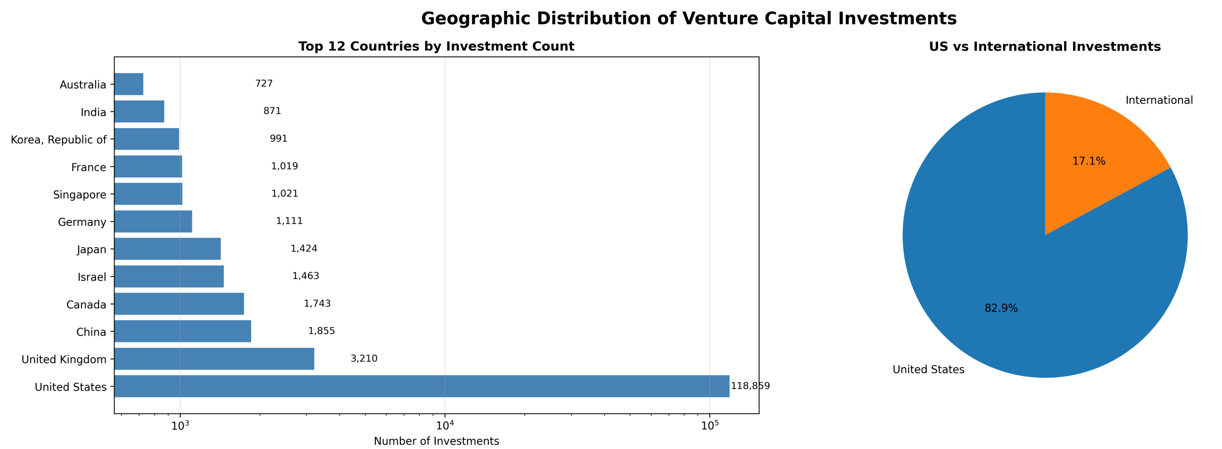


Figure 7: Geographic distribution of venture capital investments by country. The analysis shows the dominance of the United States VCs in venture capital activity on US (naturaly), accounting for 82.9% of all investment transactions, with the top 15 countries representing the vast majority of global venture capital deployment on US.

Globally, the United States accounts for 118,859 investments (82.9% of total activity), while international markets collectively represent 24,579 investments (17.1%). Among international markets, the United Kingdom leads with 3,210 investments (2.2%), followed by China with 1,855 investments (1.3%) and Canada with 1,743 investments (1.2%).

Within the United States, regional concentration patterns mirror the global trend. California dominates with 48,033 investments (40.4% of US total), representing nearly half of all American venture capital activity. New York follows as the second-largest market with 18,755 investments (15.8%), while Massachusetts ranks third with 7,946 investments

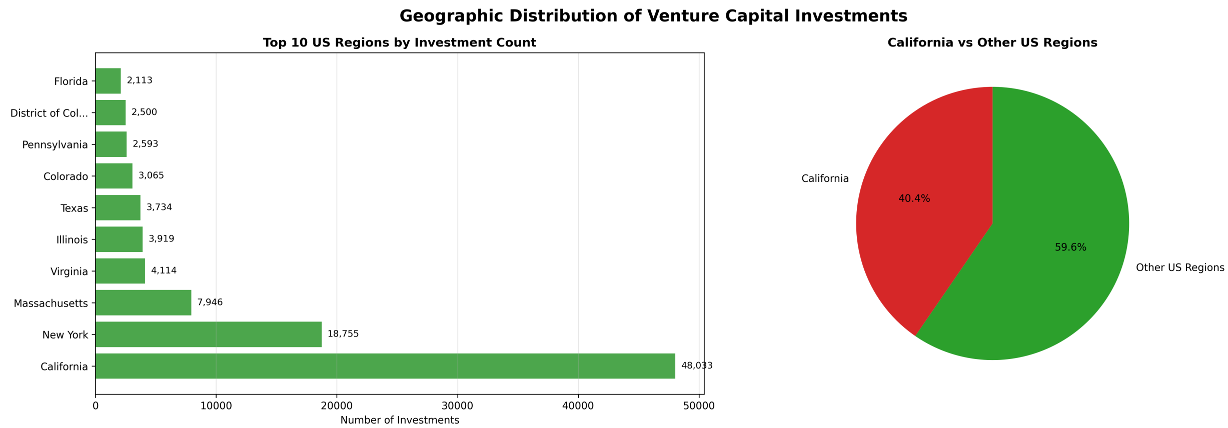


Figure 8: Regional distribution of venture capital investments within the United States. California emerges as the dominant hub with 40.4% of all US investments, followed by New York at 15.8%, demonstrating significant geographic clustering within the American venture capital ecosystem.

(6.7%). The top ten US regions account for approximately 81.5% of all domestic venture capital transactions.

These geographic patterns reflect the clustering effects of venture capital ecosystems around established technology and financial centers. The concentration in California, particularly in Silicon Valley, and New York's financial district demonstrates how proximity to talent, infrastructure, and capital sources influences investment distribution. Similar clustering patterns emerge internationally, with London, Beijing, and Toronto serving as regional venture capital hubs for funds coming to US.

3.1.5 Sectorial Distribution

The sectoral distribution analysis reveals significant concentration patterns in venture capital investment activity. As illustrated in Figure 9, health care dominates the investment landscape with 21,574 investments (15.0% of total activity), followed by financial services at 11,409 investments (8.0%) and data and analytics at 8,626 investments (6.0%).

The investment concentration demonstrates a highly uneven distribution across sectors.

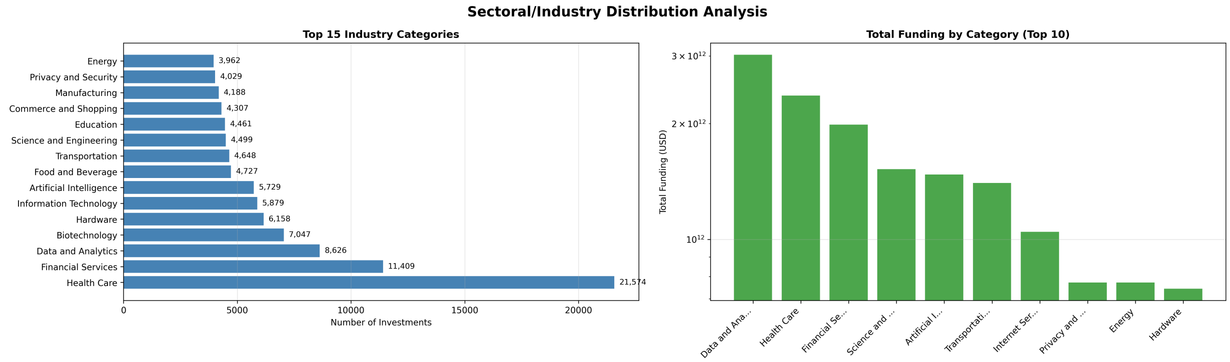


Figure 9: Sectoral distribution of venture capital investments across industry categories. The bar chart shows the concentration of investment activity, with health care as the dominant sector, followed by financial services and data analytics. The top 20 categories account for over 81% of all investment transactions.

The top 5 categories capture 38.2% of total investments, while the top 10 categories account for 56.0% of all activity. This concentration extends to the top 20 categories, which represent 81.2% of total investments across only 42 unique industry categories.

Technology-related sectors represent a notable but minority portion of the ecosystem. The analysis identifies 6 technology categories (artificial intelligence, data and analytics, hardware, information technology, privacy and security, and internet services) that collectively account for 33,998 investments (23.7% of total activity). The remaining 109,440 investments (76.3%) occur in non-technology sectors, indicating that venture capital deployment extends well beyond traditional technology boundaries.

Figure 10 illustrates the temporal evolution of sectoral investment patterns, revealing how industry preferences have shifted over time.

The healthcare sector's dominance reflects the substantial capital requirements and long development cycles characteristic of medical innovation, which align well with venture capital investment strategies. Financial services' strong showing demonstrates the ongoing fintech revolution and the digitization of traditional financial institutions.

The presence of traditional sectors like food and beverage (4,727 investments, 3.3%),

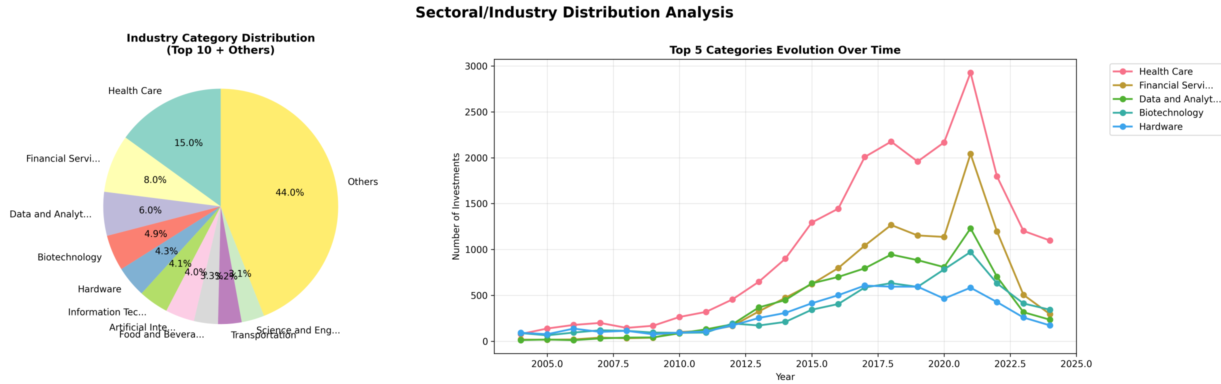


Figure 10: Temporal trends in sectoral investment distribution. The figure shows how investment patterns across different industry categories have evolved over time, highlighting the growth trajectories of major sectors and the emergence of new investment areas.

transportation (4,648 investments, 3.2%), and manufacturing (4,188 investments, 2.9%) among the top categories indicates that venture capital has expanded beyond its historical technology focus to encompass a broader range of innovation-driven industries.

Biotechnology's position as the fourth-largest category (7,047 investments, 4.9%) highlights the significant role of life sciences in the venture capital ecosystem, often requiring specialized knowledge and longer investment horizons compared to other sectors.

3.2 Communities Characterization

The division of venture capital firms into early-stage and late-stage investor groups results in 169,679 investment pairs comprising 3,666 unique startups.

Community detection using greedy modularity optimization identifies approximately 168 distinct communities (the number of communities oscillates between 167 and 175 across different trials), with the largest communities containing over 4000 investors each, followed by 1 community with almost 1000 agents, 4 communities with more than 100 agents, and then several smaller groups, as summarized in Table 1.

Initially, analysis focuses on communities with at least 150 nodes to ensure statistical power for nestedness analysis. This threshold yields 5 communities.

Add rationale for threshold

Community ID	**Number of Investors**
0	4,248
1	4,089
2	3,959
3	979
4	188
5	155
6	137
7	122

Table 1: Size distribution of the largest investor communities identified through greedy modularity optimization

The largest three communities (0, 1, and 2) contain over 12,000 investors combined, representing approximately 75% of all investors in the network. This concentration suggests a highly centralized structure within the venture capital ecosystem, with most investment activity occurring within a small number of large communities, what suggests community size distribution follows a typical power-law pattern observed in many social networks [4].

Mention literature, as this phenomenon is well-documented

Add figure of community size distribution

Community boundaries are defined at the node level, meaning each investor belongs to exactly one community. However, edges (investment relationships) can span community boundaries when investors from different communities co-invest in the same startup.

To analyze investment patterns, we classify each syndicated investment as either: (1)

intra-community if all participating investors belong to the same community, or (2) cross-community if investors from multiple communities participate together. Table 2 presents the resulting investment distribution across communities.

Community	**Co-investments Pairs**	**Relative Proportion**
Community 0	32,164	19.4%
Community 1	17,301	10.4%
Community 2	55,863	33.6%
Cross-community	58,329	35.1%

Table 2: Distribution of syndicated investments across investor communities

The analysis reveals important patterns in investment activity distribution. Community 2 accounts for the largest share of investments (33.6%), containing approximately 50% more investments than Community 0 and over three times more than Community 1. This concentration of investment activity suggests that structural features of Community 2 may facilitate higher transaction volumes. Notably, cross-community investments represent over one-third of all transactions, indicating substantial interconnectedness across community boundaries.

3.2.1 Degree Distribution

Before further examining community-level investment patterns in-depth, we analyze the degree distribution characteristics across the three largest investor communities. This analysis reveals that all communities exhibit power-law patterns typical of scale-free networks, with notable differences in magnitude and scale parameters.

Communities 0 and 2 demonstrate remarkably similar degree distribution magnitudes, while Community 1 exhibits consistently lower magnitude values across all degree ranges. This pattern becomes particularly evident when examining the degree distributions on a logarithmic scale, where Community 2 shows the highest volume of nodes across all degree ranges, suggesting greater overall connectivity within this community.

Compare connectance among communities

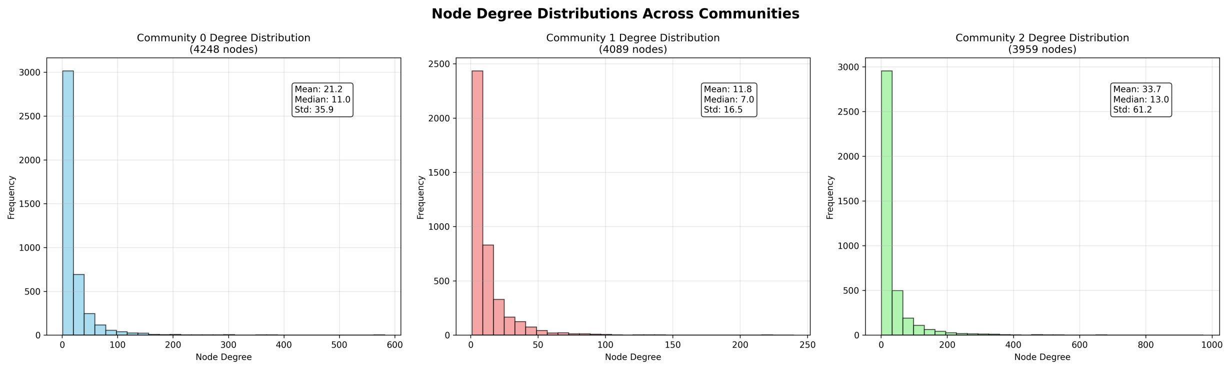


Figure 11: Node degree distributions for the three largest investor communities. The figure shows both the individual degree distributions and their overlap on a logarithmic scale. Communities 0 and 2 display similar heavy-tailed, power-law-like patterns, while Community 1 has consistently lower degree magnitudes. The log-scale overlay highlights that Community 2 maintains the highest node volume across all degree ranges, indicating greater overall connectivity.

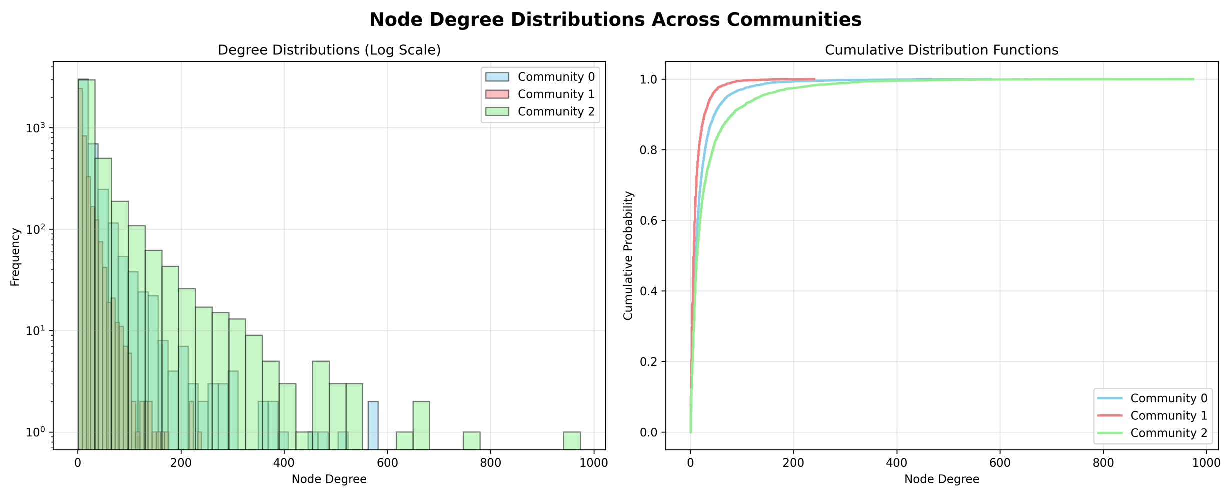


Figure 12: TBD

Analysis of high-degree nodes (95th percentile) reveals distinct patterns across communities. Community 0 demonstrates a mixed composition of hub nodes, with early-stage investors like Techstars-seed (degree: 582) and 500 Global-seed (degree: 564) dominating the highest positions, alongside significant late-stage players such as Gaingels-series_b (degree: 509).

Community 1 exhibits a more balanced distribution with Intel Capital maintaining strong presence across multiple investment stages, while Community 2 shows remarkable concentration among early-stage Silicon Valley investors, with SV Angel-seed achieving the highest connectivity (degree: 974) followed by other prominent early-stage firms including Andreessen Horowitz and Khosla Ventures.

Table 3 presents the top high-degree nodes for each community, highlighting the structural differences in hub organization.

Community	**Investor**	**Degree**	**Type**
0	Techstars-seed	582	Early-stage
	500 Global-seed	564	Early-stage
	Gaingels-series_b	509	Late-stage
	Greycroft-series_a	483	Early-stage
	Bossa Invest-series_b	450	Late-stage
1	Intel Capital-series_b	240	Late-stage
	Norwest Venture Partners-series_a	222	Early-stage
	Canaan Partners-series_a	219	Early-stage
	Intel Capital-series_c	175	Late-stage
	SOSV-series_b	163	Late-stage
2	SV Angel-seed	974	Early-stage
	SV Angel-series_a	754	Early-stage
	Andreessen Horowitz-series_a	664	Early-stage
	Khosla Ventures-series_a	659	Early-stage
	New Enterprise Associates-series_a	619	Early-stage

Table 3: Top 5 high-degree nodes (95th percentile) for the three largest investor communities

As per Figure 13, the relationship between degree centrality and investment activity demonstrates a positive correlation across all communities, with higher-degree nodes exhibiting greater investment frequency. This pattern suggests that network position, as measured by degree centrality, serves as a reliable predictor of investment activity levels within the venture capital ecosystem.

Add more centrality metrics

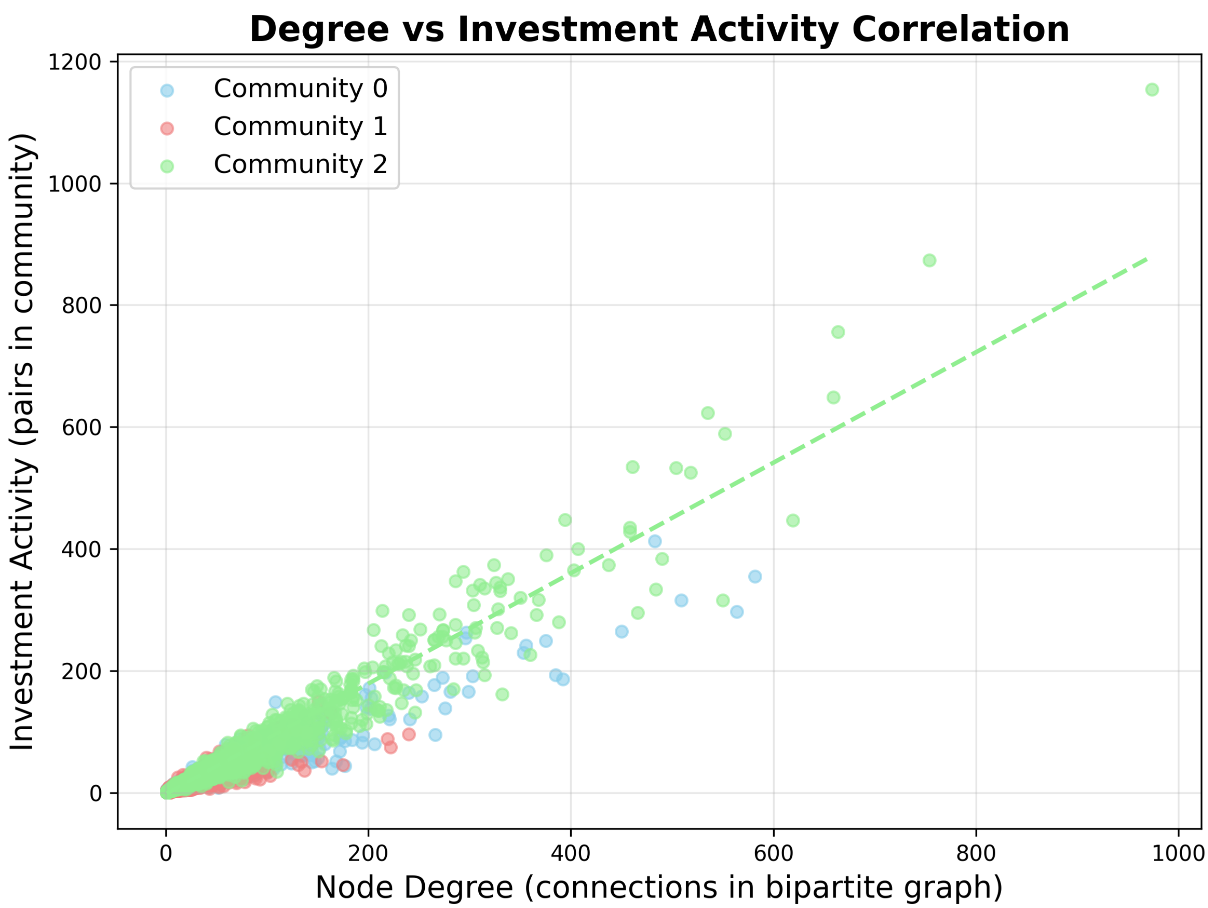


Figure 13: Relationship between node degree and investment activity for the three largest investor communities. The scatter plot demonstrates a positive correlation: higher-degree nodes tend to participate in more investment activities. This pattern is consistent across all communities, supporting the interpretation that network centrality is a strong predictor of investment frequency and influence within the venture capital ecosystem.

3.2.2 Geographic Distribution

Geographic analysis reveals distinct spatial clustering patterns across the three largest communities. Figure 15 illustrates the asymmetric geographic distributions between early-stage and late-stage investment networks.

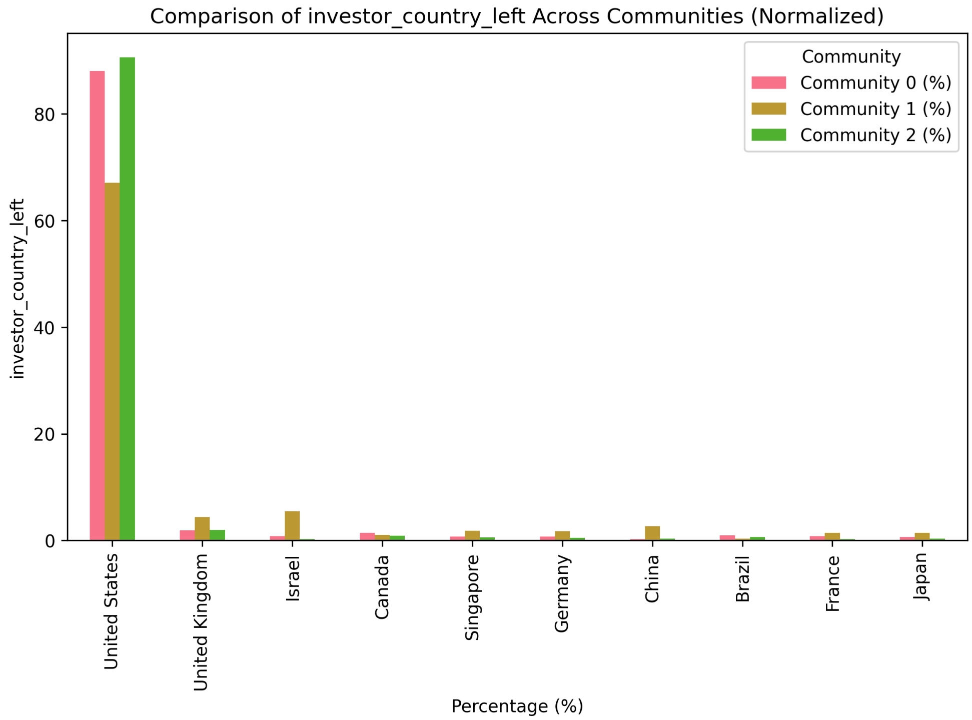
Better format geographic distribution figure

Communities 0 and 2 exhibit similar geographic profiles with predominantly American investors, reflecting the dominance of U.S.-based venture capital (important to remember our dataset contains only American startups' investments, which include international investors).

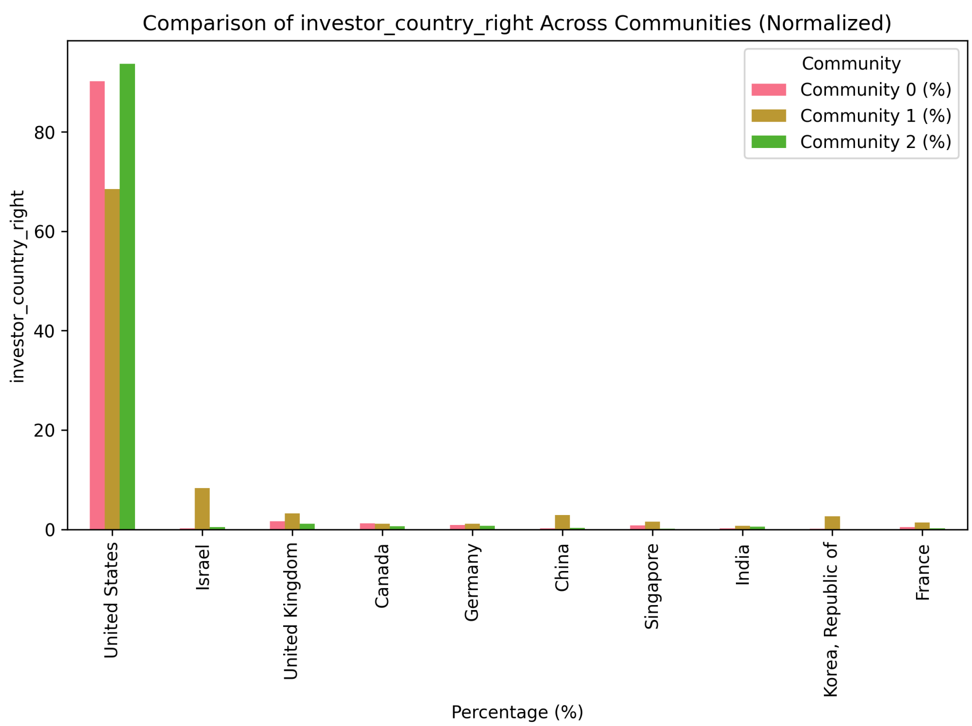
However, regional analysis within the United States reveals a striking pattern: Community 2 demonstrates exceptional concentration in California, particularly Silicon Valley, with approximately 50% more California-based investors than either Community 0 or Community 1 for both late-stage and early-stage investor categories.

This Silicon Valley concentration in Community 2 is particularly notable given the region's status as the world's premier innovation ecosystem. The dominance of California investors in this community, which also exhibits the highest transaction volumes, suggests potential advantages conferred by geographic clustering within innovation hubs.

In contrast, Community 1 demonstrates significantly greater international diversification, with substantial representation from Israel, the United Kingdom, China, South Korea, Singapore, and France. This international composition in Community 1 may reflect different risk tolerance profiles, regulatory environments, or access to cross-border deal flow compared to the more domestically concentrated communities.

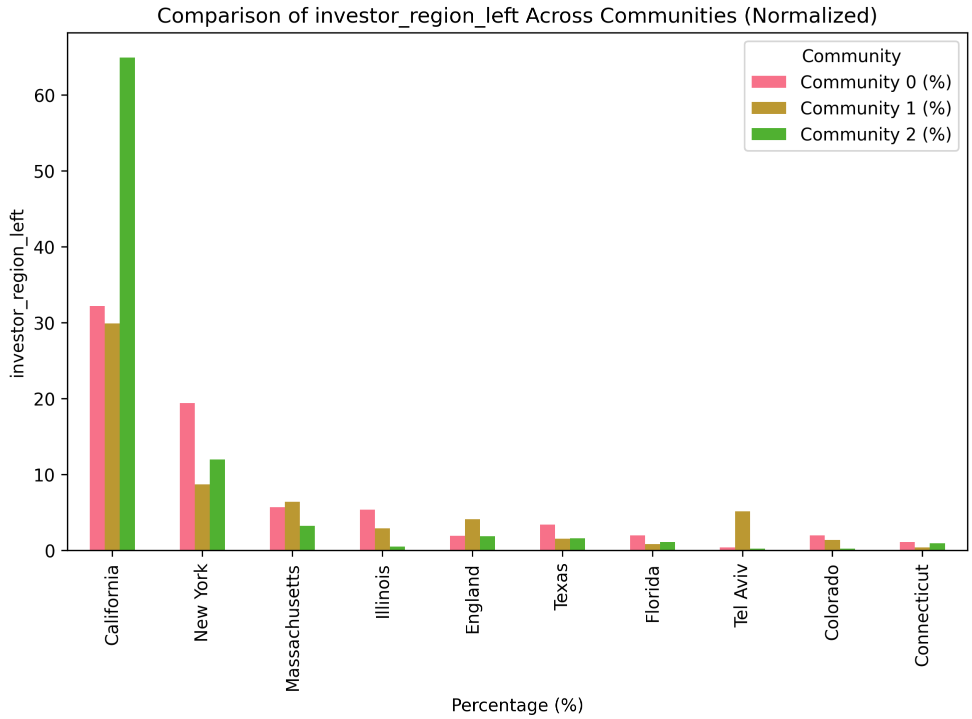


(a) Late-stage investors geographic distribution (countries)

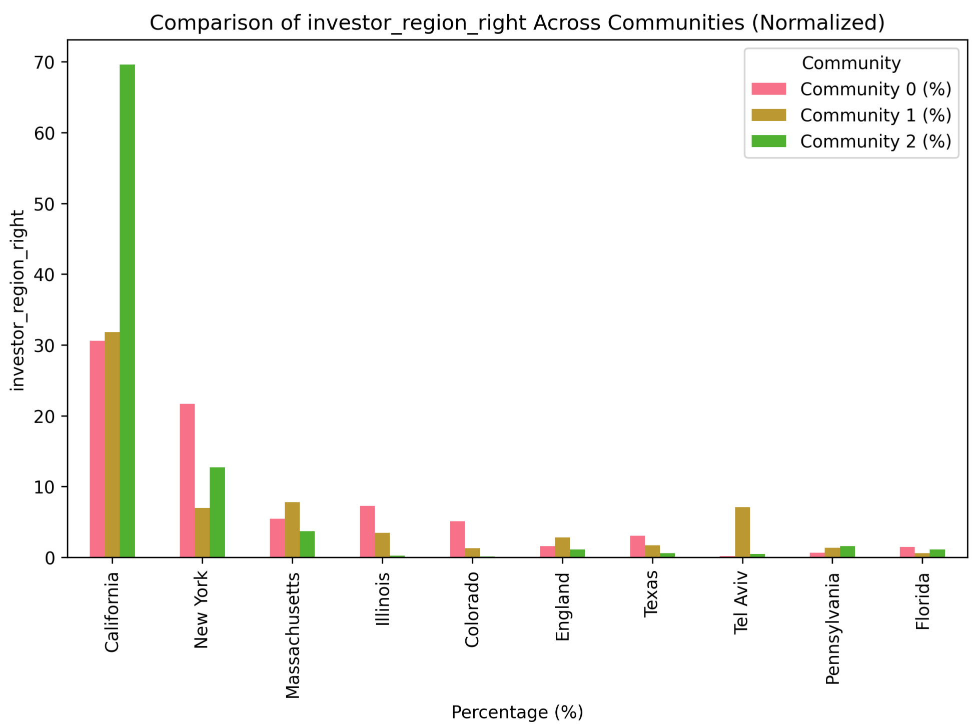


(b) Early-stage investors geographic distribution (countries)

Figure 14: Geographic distribution of venture capital investors across the largest communities. The bipartite structure reveals differential geographic clustering between late-stage (top) and early-stage (bottom) investor networks, with Community 1 exhibiting greater international diversification compared to the U.S.-concentrated Communities 0 and 2.



(a) Late-stage investors geographic distribution (regions)



(b) Early-stage investors geographic distribution (regions)

Figure 15: Geographic distribution of venture capital investors across regions

3.2.3 Funding Characteristics

Analysis of funding patterns reveals that Community 2 exhibits substantially higher funding frequency and larger investment amounts compared to the other communities, as demonstrated in Figure 16.

Add attachment with statistical proofs

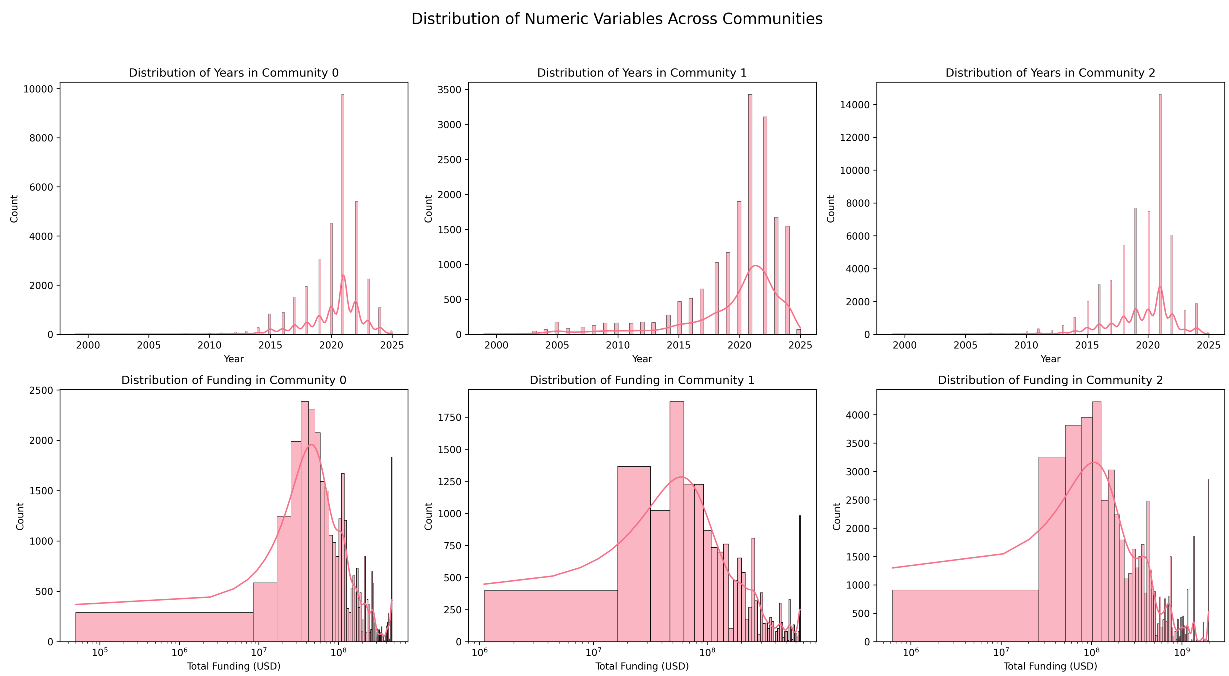


Figure 16: Funding characteristics across the three largest investor communities. The analysis reveals systematic differences in investment amounts, round frequency, and funding patterns between communities, with different organizational structures exhibiting distinct capital deployment strategies.

plot a graph of distribtuion of invesments among degrees of investors

Furthermore, communities exhibit concentrated capital deployment patterns, with higher-degree investors participating in larger funding rounds while maintaining broader portfolio diversification. This suggests that certain organizational structures may create more efficient capital allocation mechanisms compared to other network configurations.

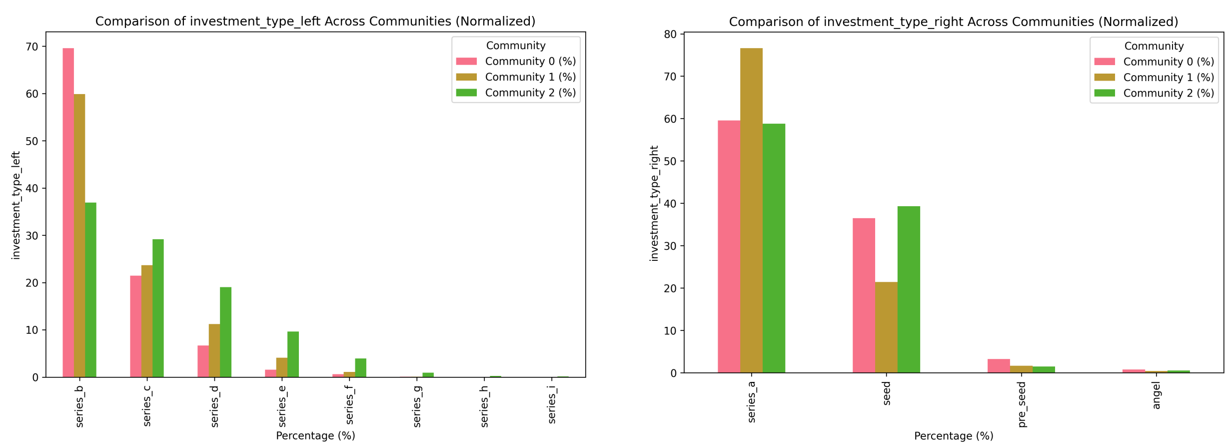
Despite superficial similarities between Communities 0 and 2 in terms of size and geographic concentration within the United States, Community 2 appears to confer distinctive advantages in organizational efficiency. While both communities share comparable scales and American investor bases, Community 2's organizational structure enables it to achieve substantially higher transaction volumes (33.6% vs. 19.4% of total investments) and more comprehensive funding coverage across all investment stages.

This pattern suggests that specific network topologies, rather than community size or geographic distribution alone, may be critical determinants of investment ecosystem efficiency.

add literature base for this strong assumption

3.2.4 Investment Stage Preferences

The investment stage distributions reveal systematic specialization patterns across communities that align with their geographic profiles and transaction volumes. Figure 17 demonstrates the distribution patterns of investment types within the bipartite network structure.



(a) Late-stage investment types distribution

(b) Early-stage investment types distribution

Figure 17: Investment stage distribution across the three largest communities. The distribution patterns reveal stage-specific specialization within investor communities.

Late-stage investment patterns: Community 2 dominates Series C and later funding rounds, demonstrating its role in growth-stage capital deployment. Community 1, despite its smaller transaction volume, shows strong representation in Series B and later stages, with participation rates exceeding Community 0 in Series C and beyond. Community 0 exhibits particular strength in Series B rounds while maintaining lower participation in later stages compared to Community 2.

Early-stage investment patterns: Community 2 shows prominence in seed-stage investments while maintaining comparable levels to Community 0 in Series A funding. Both Communities 0 and 2 participate actively in angel and seed rounds, though Community 0 shows relatively higher pre-seed activity. Community 1 demonstrates concentrated focus on Series A investments, aligning with its international profile and suggesting specialization in cross-border early-growth funding.

These stage-specific patterns suggest that Community 2's organizational structure facilitates participation across the entire funding spectrum, from seed to late-stage rounds, potentially enabling more comprehensive support for portfolio companies throughout their development lifecycle.

3.2.5 Sectoral Focus

The sectoral analysis reveals distinct specialization patterns that align with each community's structural and geographic characteristics. Figure 18 illustrates the distribution of investment focus across technology sectors, demonstrating how different communities exhibit varying degrees of sectoral concentration.

Comment sectorial distribution

Community 0: Concentrates in real estate, commerce and shopping, and financial services, while maintaining standard representation in information technology and artificial

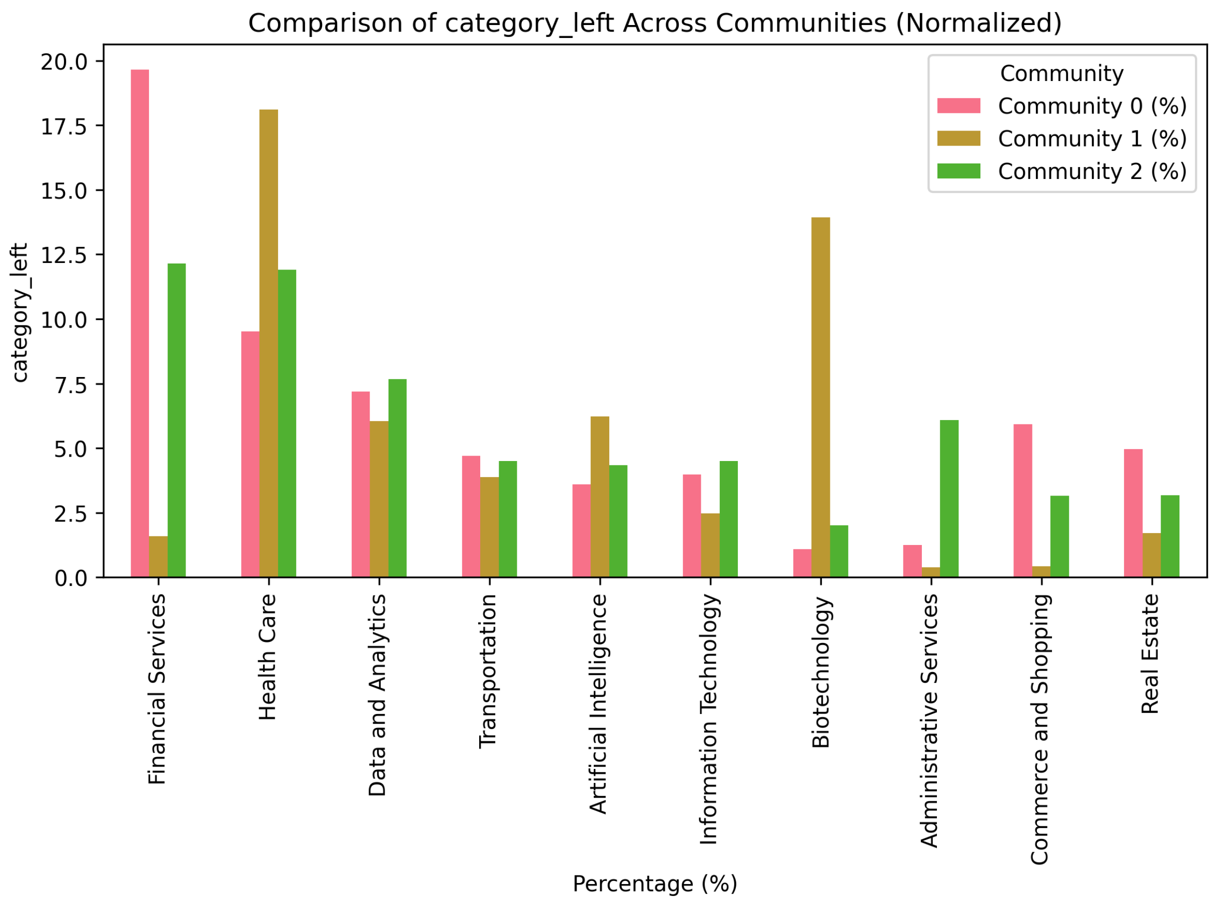


Figure 18: Sectoral distribution across the three largest investor communities. The analysis reveals differential industry focus patterns, with certain communities demonstrating concentrated investment strategies in specific technology sectors while others maintain broader sectoral diversification.

intelligence. Notably absent from administrative services and biotechnology, suggesting specialized expertise in consumer-facing and financial technology sectors.

Community 1: Specializes significantly in biotechnology and healthcare, with enhanced artificial intelligence focus compared to other communities. Shows reduced interest in real estate, commerce and shopping, and administrative services. The biotechnology concentration aligns with its international composition, potentially reflecting access to global biotech innovation hubs.

Community 2: This community exhibits strong representation in administrative services while maintaining comparable levels to Community 0 in information technology and artificial intelligence. Despite its larger transaction volume, it shows lower concentration in real estate and financial services than Community 0, suggesting that its organizational structure facilitates broader sectoral participation rather than concentrated specialization. This sectoral breadth, combined with the community's Silicon Valley concentration, indicates potential advantages of diversified investment strategies within innovation-rich geographic clusters.

Community 0 serves as an effective structural baseline for comparison with Community 2, given their similar geographic profiles and certain sectoral overlaps, while differing significantly in network organization and transaction volumes. The systematic differences observed between these structurally similar communities highlight the potential impact of organizational structures on investment behavior and capital deployment efficiency.

3.3 Overall Nestedness Findings

Nestedness analysis across investor communities reveals heterogeneous structural patterns. Among the 8 communities examined, one exhibits significantly high nestedness ($p < 0.01$) relative to degree-preserving null models generated through the Curveball algorithm.

Figure 19 presents the comparison between observed and null model nestedness scores, where each plot represents a distinct community's null model NODF distribution accompanied by its observed NODF value.

A broader analysis examining nestedness patterns across all analyzed communities reveals a relationship between community size and nestedness significance, as illustrated in Figure 20. This analysis examines 8 communities ranging from 122 to 4,248 investors, providing insights into how community structure varies across different organizational scales.

The analysis reveals that high nestedness significance does not follow a simple relationship with community size. While smaller communities (122-188 investors) tend to exhibit higher absolute nestedness scores (0.045-0.099), these values are statistically significant low when compared to appropriate null models. Medium-sized communities (979-4,248 investors) generally show lower absolute nestedness scores (0.011-0.028), with most falling below low nestedness significance thresholds.

Notably, Community 2 stands out as the only community achieving statistically significant nestedness despite having a moderate absolute score (0.088) and being neither the largest nor smallest community analyzed. This pattern suggests that nestedness emergence depends on specific structural characteristics rather than simple community scale effects.

Community 2 demonstrates the most pronounced nestedness, exhibiting an NODF score of 0.088 with statistical significance of $p = 0.00001$. This indicates a non-random behavior where less-connected investors maintain co-investment relationships with subsets of partners associated with highly-connected investors, creating a hierarchical investment pattern.

Additionally, Community 2 exhibits an asymmetric composition with a pronounced ratio favoring late-stage investors over early-stage investors. This imbalance may contribute to the nested structure by creating hierarchical dependencies between investor types.

Explore how imbalance contribute to nestedness according to literature

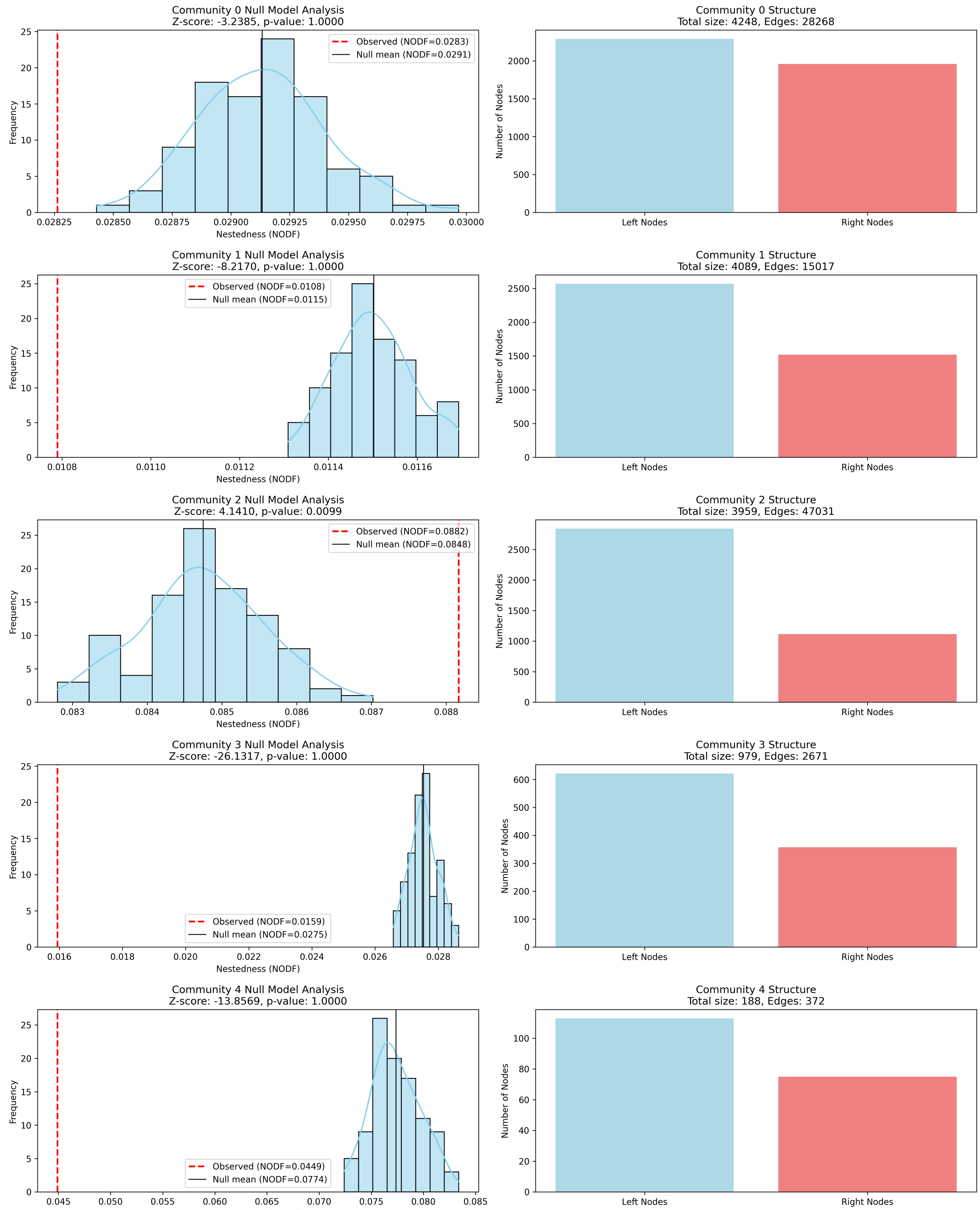


Figure 19: Comparison of observed versus null model nestedness scores for the five largest investor communities. The diagonal line represents equal observed and expected values, with points above the line indicating higher-than-random nestedness. "Left" refers to late-stage investors, and "Right" refers to early-stage investors.

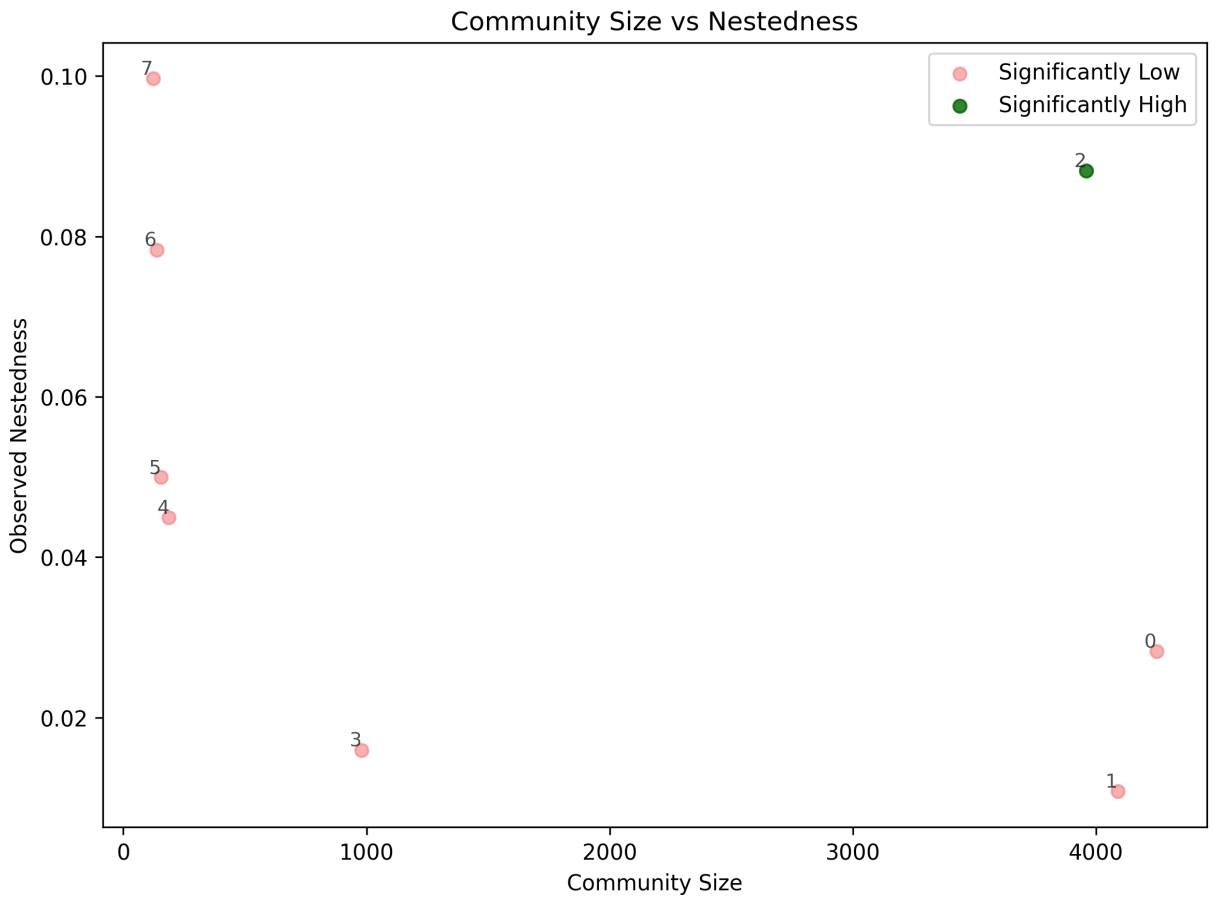


Figure 20: Relationship between community size and nestedness significance across all analyzed investor communities. The figure shows observed nestedness scores (y-axis) against community size (x-axis), with markers indicating statistical significance. Only Community 2 exhibits significantly high nestedness, while all other communities show significant low patterns relative to their respective null models.

The following sections provide detailed characterization of the 3 most relevant communities in terms of number of investors, while analyzing in-depth Community 2 (where nestedness was observed) through comparison with Communities 0 and 1, which serve as contrasting examples of similar-sized but differently structured investor networks.

3.4 Evolution of Nestedness in Silicon Valley Community

To understand the emergence and development of nested structure, we conducted a temporal analysis of Community 2's nestedness evolution using cumulative window. This analysis reveals how the significantly high nested structure observed in the static analysis developed over time and when it became statistically distinguishable from random network configurations.

The temporal analysis spans 18 years (2007-2024) with sufficient data for nestedness calculation. Early periods (2004-2006) contained insufficient investment activity for meaningful analysis. The evolution demonstrates three distinct phases: an early period with significantly low nestedness (2007-2018), a transition period (2019), and a sustained period with significantly high nestedness (2019-2024).

During the period with significantly low nestedness (2007-2018), observed nestedness scores ranged from 0.15 to 0.38, consistently failing to exceed null model expectations. Z-scores remained predominantly negative, indicating that the observed network structure was less nested than expected under random configuration preserving the degree sequence. This suggests that early venture capital network organization in Community 2 followed patterns that were actually less hierarchical than random syndication among investors with equivalent activity levels.

The transition occurred in 2019, marking the first year when Community 2 achieved significantly high nestedness (Z-score: 2.91, p-value < 0.001). This transition coincided with substantial network growth, reaching 2,364 total nodes and 33,805 edges.

Notably, the 2019 transition corresponds to the network achieving a connectance threshold of 0.0264, suggesting that nestedness emergence may be related to specific density conditions within large-scale investment networks.

The sustained significant period (2019-2024) demonstrates consistent statistical significance with progressively strengthening Z-scores, reaching a maximum of 4.76 in 2022.

Interestingly, while absolute nestedness scores decreased from 0.38 in early periods to 0.088 in 2024, the statistical significance increased dramatically. Table 4 presents key statistics from the temporal evolution analysis.

This apparent paradox reflects the fundamental principle that nestedness significance depends on comparison with appropriate null models rather than absolute magnitude [14].

Period	**Years**	**Mean NODF**	**Mean Z-score**	**Years with High Nestedness**
Low Nestedness (2007-2018)	12	0.2547 ± 0.1158	-1.4442 ± 1.1108	0
Transition (2019)	1	0.1177	2.9125	1
High Nestedness (2019-2024)	6	0.1012 ± 0.0121	3.8950 ± 0.7182	6

Table 4: Summary statistics for Community 2 nestedness evolution across temporal periods

The analysis reveals several notable patterns in network structure evolution. Late-stage investor participation increased consistently over time, while early-stage investor numbers showed initial growth followed by stabilization after 2020. This asymmetric growth pattern contributed to the development of the nested structure by creating increasingly hierarchical relationships between investor types.

Connectance exhibited a systematic decline from 0.195 in 2007 to 0.015 in 2024, reflecting the network's evolution toward sparser but more strategically organized connections. Despite this decreased density, the emergence of statistical significance suggests that the

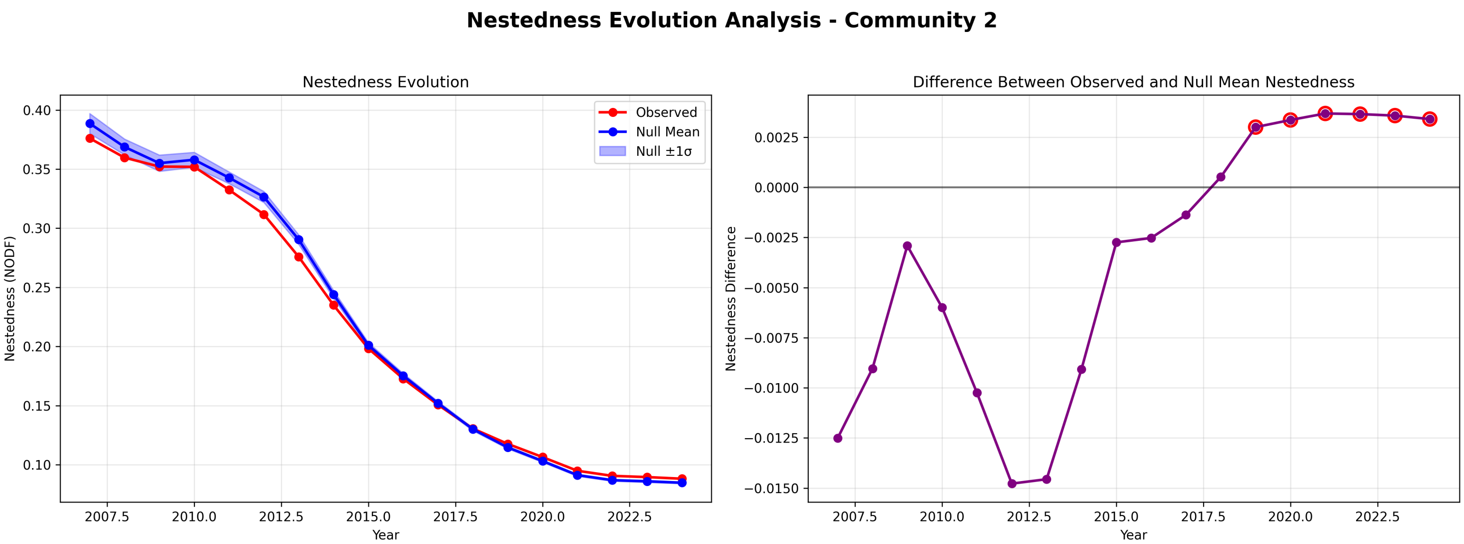


Figure 21: Temporal evolution of nestedness in Community 2. The left subplot shows the observed nestedness (red) and null model mean (blue) over time, with shaded areas representing the null model standard deviation. The right subplot displays the evolution of network size, with total nodes (green) and total edges (purple) over the same period. The figure highlights the phase transition in 2019, where observed nestedness becomes significantly higher than null model expectations and the network grows rapidly.

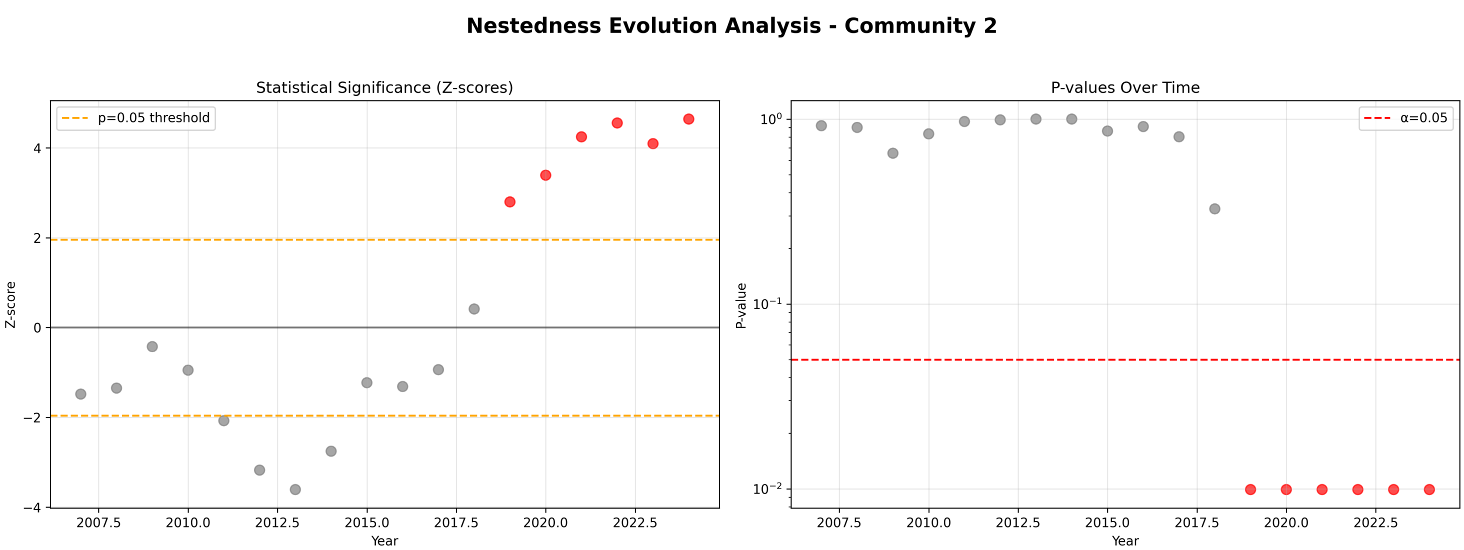


Figure 22: Connectance and statistical significance evolution in Community 2. The left subplot shows the decreasing trend in connectance (network density) over time, while the right subplot presents the evolution of p-values on a logarithmic scale, indicating the emergence of significantly high nestedness after 2019. The figure demonstrates how the network becomes sparser yet more hierarchically organized, with high nestedness emerging as the network reaches critical size and density thresholds.

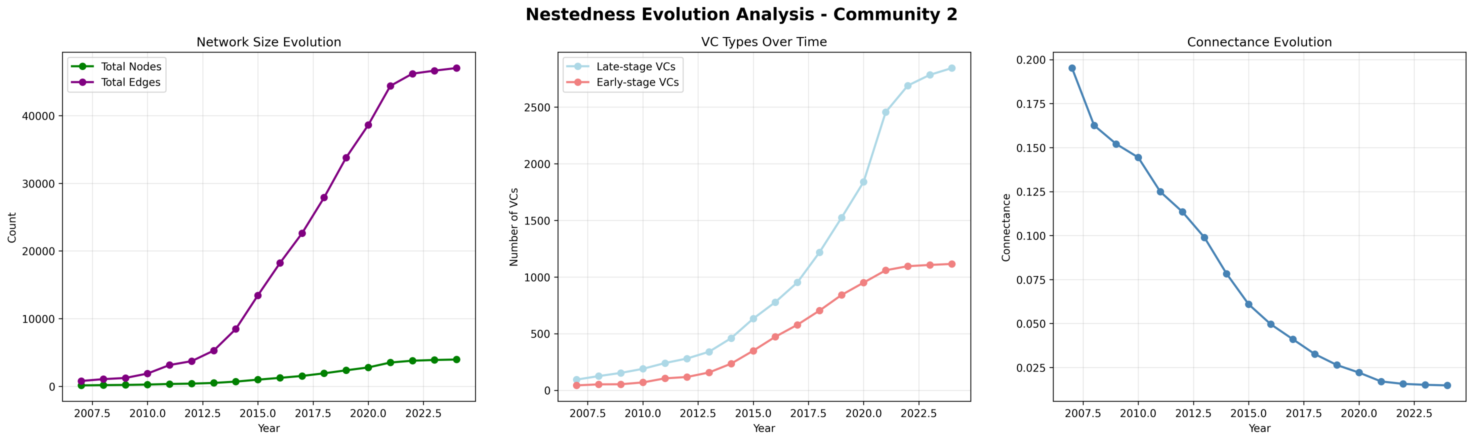


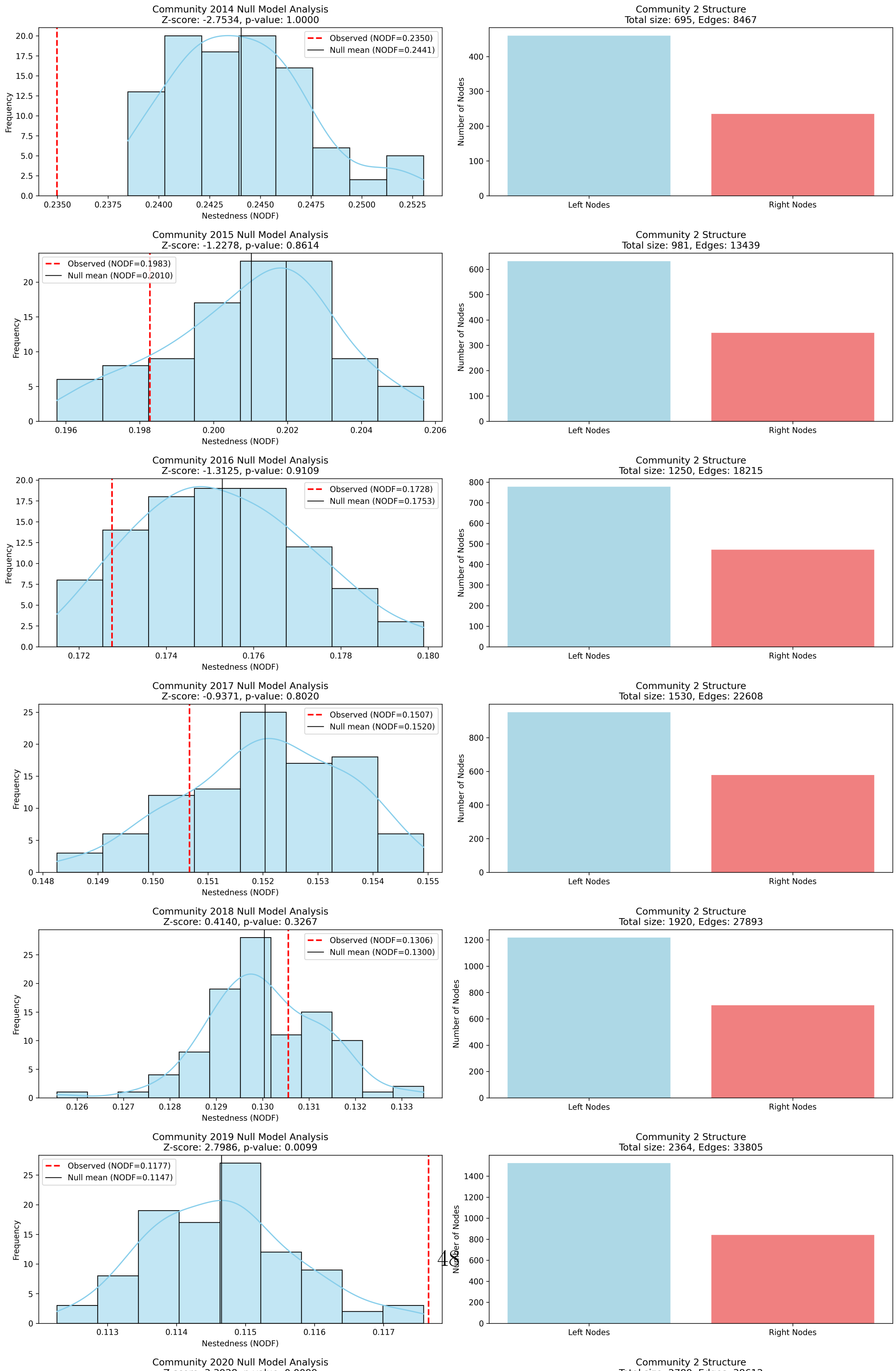
Figure 23: TBD

remaining connections became increasingly hierarchically organized, with less-connected investors maintaining relationships with subsets of the partners of highly-connected investors.

Detailed analysis of the significant periods (2019-2024) reveals consistent patterns in network organization. Each significant year demonstrates similar structural characteristics: large networks (>2,300 nodes), substantial edge counts (>33,000), low connectance (<0.027), and strong statistical significance (Z-scores >2.9). This consistency suggests that the nested structure represents a stable organizational state that persists once established.

The temporal analysis provides evidence that nestedness in venture capital networks emerges through a phase transition process rather than gradual development. The sharp transition from significantly low to significantly high nestedness in 2019, followed by sustained high nestedness, suggests threshold effects in network organization.

This pattern aligns with theoretical frameworks from complex network theory indicating that certain topological properties emerge discontinuously as networks reach critical size or density parameters [14].



Community 2014 Null Model Analysis
Z-score: -2.7534, p-value: 1.0000
Observed (NODF=0.2350)
Null mean (NODF=0.2441)
Frequency
Nestedness (NODF)
Community 2 Structure
Total size: 695, Edges: 8467
Number of Nodes
Left Nodes
Right Nodes
Community 2015 Null Model Analysis
Z-score: -1.2278, p-value: 0.8614
Observed (NODF=0.1983)
Null mean (NODF=0.2010)
Frequency
Nestedness (NODF)
Community 2 Structure
Total size: 981, Edges: 13439
Number of Nodes
Left Nodes
Right Nodes
Community 2016 Null Model Analysis
Z-score: -1.3125, p-value: 0.9109
Observed (NODF=0.1728)
Null mean (NODF=0.1753)
Frequency
Nestedness (NODF)
Community 2 Structure
Total size: 1250, Edges: 18215
Number of Nodes
Left Nodes
Right Nodes
Community 2017 Null Model Analysis
Z-score: -0.9371, p-value: 0.8020
Observed (NODF=0.1507)
Null mean (NODF=0.1520)
Frequency
Nestedness (NODF)
Community 2 Structure
Total size: 1530, Edges: 22608
Number of Nodes
Left Nodes
Right Nodes
Community 2018 Null Model Analysis
Z-score: 0.4140, p-value: 0.3267
Observed (NODF=0.1306)
Null mean (NODF=0.1300)
Frequency
Nestedness (NODF)
Community 2 Structure
Total size: 1920, Edges: 27893
Number of Nodes
Left Nodes
Right Nodes
Community 2019 Null Model Analysis
Z-score: 2.7986, p-value: 0.0099
Observed (NODF=0.1177)
Null mean (NODF=0.1147)
Frequency
Nestedness (NODF)
Community 2 Structure
Total size: 2364, Edges: 33805
Number of Nodes
Left Nodes
Right Nodes
Community 2020 Null Model Analysis
Community 2 Structure

Make reflection about possible link with network effects from economics and industrial organization

The period of nestedness emergence (2019-2024) corresponds with an increase in late-stage investor participation and a relative decrease in early-stage investor numbers within the community. This asymmetric evolution may contribute to the hierarchical structure by creating conditions where early-stage investors increasingly depend on relationships with a subset of the partners associated with highly-connected late-stage investors.

Add reference to figure of nestedness vs null models between 2019 and 2024

However, whether this temporal correlation reflects causal mechanisms or represents coincidental market dynamics requires further investigation with appropriate theoretical frameworks [10].

4 Discussion and Implications

As outlined in the Results section, this article first presents a detailed characterization of investor communities and then examines nestedness within those communities; the discussion below follows that ordering and builds directly on the community-level descriptions provided earlier.

The discovery, documented in the Results section, of communities with significantly high nestedness within the late-early stage venture capital network provides new insights into investor behavior and startup access to capital. The hierarchical structure in Community 2 suggests that informal investment hierarchies may systematically influence funding accessibility for entrepreneurs.

The concentration of significantly high nested structure specifically within Silicon Valley investors adds a significant geographic dimension to these findings. Community 2's exceptional dominance by California-based investors, particularly those in Silicon Valley, suggests that geographic clustering within the world's premier innovation ecosystem may facilitate the emergence of hierarchical investment structures.

This pattern may indicate that dense information networks, frequent face-to-face interactions, and shared risk assessment practices characteristic of innovation hubs naturally give rise to nested investor relationships. The dominance of California investors in the only community with significantly high nestedness suggests a potential relationship between geographic clustering in innovation hubs and the emergence of hierarchical investment structures.

Furthermore, the comparative analysis between Communities 0 and 2 reveals that high nestedness functions as an organizational catalyst that transforms otherwise similar investor communities. Despite comparable sizes and geographic concentration within the United

States, Community 2's highly nested structure enables substantially higher transaction volumes and more comprehensive funding coverage across all investment stages.

This suggests that network topology, rather than community size or geographic distribution alone, may be the critical determinant of investment ecosystem efficiency. The hierarchical organization in Community 2 appears to create more efficient capital allocation mechanisms, with higher-degree investors participating in larger funding rounds while maintaining broader portfolio diversification compared to randomly structured networks.

4.1 Degree Distribution Patterns and Hub Organization

The analysis of degree distributions across communities reveals fundamental organizational differences that complement the nestedness findings. All three communities exhibit power-law degree distributions characteristic of scale-free networks [4], but with distinct magnitude and structural parameters that reflect different organizational strategies.

The similarity in degree distribution magnitudes between Communities 0 and 2, contrasted with Community 1's consistently lower values, suggests that network scale alone does not determine investment efficiency. Community 2's superior performance in investment volumes and nested organization occurs despite degree distribution patterns nearly identical to Community 0.

This finding reinforces that the specific arrangement of connections, rather than their quantity or distribution, drives the observed efficiency advantages.

Analysis of high-degree nodes (network hubs) reveals distinct organizational philosophies across communities. Community 2's hub structure demonstrates exceptional concentration among early-stage Silicon Valley investors, with SV Angel achieving remarkable connectivity across both seed and Series A stages.

This pattern contrasts with Community 0's more diversified hub structure and Community 1's balanced early-late stage distribution.

The concentration of hubs within the early-stage investment category in Community 2 may facilitate the nested structure by creating clear hierarchical pathways from early-stage to late-stage investment relationships.

The positive correlation between degree centrality and investment activity across all communities validates network position as a predictor of investor influence. However, the strength of this relationship appears amplified within the nested Community 2, suggesting that hierarchical organization may enhance the efficiency of high-degree investors in deploying capital and identifying investment opportunities.

4.2 Temporal Dynamics and Phase Transition Emergence

Add observation that literature on other formations were already explored, but VC-VC is rare

The temporal evolution analysis of Community 2 provides unprecedented insight into how highly nested structures emerge within VC-VC investment networks. The identification of a sharp phase transition in 2019, rather than gradual nestedness development, challenges assumptions about evolutionary network organization and suggests threshold-dependent emergence mechanisms.

The three-phase evolution pattern (period with significantly low nestedness between 2007-2018, transition in 2019, and sustained high nestedness during 2019-2024) indicates that nested organization in venture capital networks may represent a distinct organizational state that emerges discontinuously when specific conditions are met.

This finding aligns with theoretical predictions from complex network theory regarding critical transitions in topological properties [14].

The apparent paradox of decreasing absolute nestedness scores (from 0.38 to 0.088) coinciding with increasing statistical significance reflects sophisticated changes in network organization. As the network grew substantially larger, maintaining even modest levels of hierarchical organization became increasingly difficult under random formation processes, making the observed nested patterns more statistically remarkable.

The correspondence between the emergence of high nestedness and specific connectance thresholds (approximately 0.026) provides practical insights for ecosystem development.

This threshold may represent a critical density where hierarchical organization becomes sustainable within large-scale investment networks, offering guidance for policy interventions aimed at fostering similar organizational efficiency in other innovation ecosystems.

The asymmetric evolution of investor types—increasing late-stage participation coupled with stabilizing early-stage numbers—may have facilitated the emergence of highly nested structure by creating conditions favoring hierarchical relationships.

This pattern suggests that the development of highly nested organization may require specific demographic conditions within investor communities, rather than simply network growth or density changes.

4.3 Network Robustness and Resilience

The highly nested structures challenge assumptions of random mixing in venture capital markets, suggesting that certain investors function as "gatekeepers" who control access to broader investment networks. This finding aligns with social network theories about structural holes and brokerage positions [4].

Following insights from nestedness research in complex networks [14], the hierarchical organization observed in Community 2 may confer distinct robustness properties to the venture capital ecosystem.

In mutualistic networks, high nestedness typically enhances stability against random node removal but creates vulnerability to targeted elimination of highly connected nodes. Applied to venture capital, this suggests that highly nested investor communities may be resilient to random investor departures, they could be particularly vulnerable to the exit of key hub investors.

The concept of "mutualistic trade-offs" from ecological network theory provides a framework for understanding these dynamics. In highly nested venture capital communities, less-connected investors maintain relationships with subsets of the partners associated with highly-connected investors, creating dependencies that could influence network stability.

Future research should investigate whether less-connected venture capital firms exhibit higher exit probabilities, which would support the hypothesis that high nestedness creates hierarchical fragility patterns.

5 Conclusion and Future Directions

5.1 Individual Nestedness Contributions

An important avenue for future research involves analyzing individual nestedness contributions within these communities. Rather than treating nestedness as a global network property, examining how specific investors contribute to the overall nested structure could reveal mechanisms driving community formation and persistence.

This approach could help predict which network positions are most vulnerable to disruption and identify critical nodes whose removal would significantly alter community structure.

Understanding individual contributions to high nestedness could also inform strategies for network intervention and ecosystem development. If certain investor positions disproportionately contribute to highly nested stability, targeted support or policy interventions could enhance overall ecosystem resilience.

5.2 Dynamic Network Evolution

The temporal analysis reveals several critical insights about the dynamic processes underlying venture capital network organization. The sharp phase transition observed in Community 2's nestedness evolution suggests that network topology may be subject to discontinuous organizational changes rather than gradual evolution. This finding has important implications for understanding how investment ecosystems develop and potentially collapse.

The three-phase temporal pattern (extended periods with significantly low nestedness, rapid transition, and sustained high nestedness) may represent a general framework for understanding organizational emergence in investment networks. The identification of specific connectance thresholds (approximately 0.026) associated with the emergence of high nestedness provides quantitative targets for ecosystem development strategies.

Future research should investigate whether similar threshold effects exist in other geographic markets and whether policy interventions can facilitate reaching these critical organizational states.

The asymmetric evolution of investor types during the transition period offers insights into the demographic conditions that may facilitate highly nested organization. The pattern of increasing late-stage investor participation coupled with stabilizing early-stage numbers suggests that hierarchical organization may require specific ratios between investor types.

This finding could inform strategies for ecosystem development in emerging markets, where the balance of early-stage and late-stage capital availability is often suboptimal.

The temporal analysis also reveals that high nestedness persistence appears robust once established. The sustained high nestedness observed from 2019-2024, despite substantial network growth and changing market conditions, suggests that highly nested organization may represent a stable attractor state in investment network evolution.

This stability has important implications for long-term ecosystem planning and suggests that successful development of highly nested organization may provide lasting competitive advantages for innovation hubs.

5.3 Causal Mechanisms and Economic Outcomes

The identification of these nested communities opens several avenues for future research into the social and economic mechanisms that drive venture capital ecosystem organization. [14] provides theoretical frameworks for understanding why nestedness emerges in complex systems, including factors such as heterogeneous node fitness, temporal constraints on link formation, and spatial or industry-specific constraints on partnership formation.

Investigating whether highly nested communities provide superior or inferior outcomes for portfolio companies compared to randomly organized investor groups represents a research opportunity. The concentrated capital deployment patterns observed in Community 2 suggest potential efficiency advantages, but these must be weighed against potential risks from reduced diversity and increased systemic vulnerability.

5.4 Policy and Ecosystem Development Implications

The concentration of highly nested structures within Silicon Valley suggests that geographic proximity within innovation hubs may be a prerequisite for the emergence of hierarchical investor relationships. This finding has important implications for ecosystem development strategies in other regions seeking to replicate Silicon Valley's success.

Understanding these patterns may inform policy discussions about startup ecosystem development and investor network formation. If highly nested structures facilitate higher transaction volumes and more comprehensive funding support, as observed in Community 2, policies that encourage the formation of such hierarchical investor relationships might enhance ecosystem efficiency.

However, the geographic specificity of this pattern suggests that simply replicating formal structures may be insufficient—the dense information networks and shared practices of established innovation hubs appear to be necessary conditions for highly nested organization to emerge.

Conversely, if high nestedness creates barriers to entry for new investors or reduces access for certain entrepreneur populations, regulatory interventions might be warranted to promote more equitable network organization.

The dominance of Silicon Valley investors in the highly nested community raises questions about geographic bias in capital allocation and whether hierarchical structures may inadvertently concentrate investment opportunities within established innovation centers.

The possibility of network rewiring , analogous to ecological community adaptation, might confer additional robustness to venture capital ecosystems. Policies that facilitate investor mobility and relationship reformation could enhance system-wide resilience while maintaining the efficiency benefits of highly nested organization.

mention literature about rewiring

Investigate whether the Silicon Valley concentration in nested communities reflects unique geographic advantages, information network density, or institutional factors that could be replicated in other innovation ecosystems.

Examine the relationship between geographic clustering in innovation hubs and the emergence of nested investor structures across different global venture capital markets.

Investigate the economic consequences of nested community structure on startup success rates and funding efficiency.

Analyze individual nestedness contributions to identify critical nodes and understand how specific investor positions contribute to overall community stability and structure.

Investigate the robustness properties of nested venture capital communities, particularly vulnerability to targeted removal of highly connected investors versus resilience to random investor departures.

Examine the relationship between investor position within nested hierarchies and probability of network exit, testing whether less-connected VCs exhibit higher departure rates.

Apply social network theories of structural holes to understand the role of highly connected investors in nested communities and their function as potential gatekeepers.

Investigate whether the nested structure reflects information asymmetries, risk-sharing mechanisms, or industry-specific constraints among investors.

Develop theoretical models to explain the emergence of nested structures in investment networks, incorporating insights from Chapter 5 of "Nestedness in complex networks" regarding heterogeneous node fitness and temporal constraints.

Compare nestedness patterns across different geographic markets and time periods to understand generalizability and cultural influences on network organization.

Investigate the relationship between degree distribution patterns and nestedness emergence, examining whether specific scale-free parameter ranges facilitate hierarchical organization.

Explore the role of hub investor strategies in nested community formation, particularly investigating how early-stage hub concentration may facilitate hierarchical pathway development.

Analyze the predictive power of connectance thresholds for nestedness emergence in other venture capital markets and innovation ecosystems.

Examine the stability mechanisms that maintain nested organization once established, investigating whether demographic balance between investor types is necessary for persistence.

References

[1] Mário Almeida-Neto, Paulo Guimarães, Paulo R. Guimarães Jr, Rafael D. Loyola, and Werner Ulrich. A consistent metric for nestedness analysis in ecological systems: reconciling concept and measurement. *Oikos*, 117(8):1227–1239, 2008.

[2] Michael J. Barber. Modularity and community detection in bipartite networks. *Physical Review E*, 76(6):066102, 2007. arXiv:0707.1616 [physics]. Accessed: 2025-08-24.

[3] Vincent D. Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10):P10008, 2008. arXiv:0803.0476 [physics]. Accessed: 2025-08-24.

[4] Stephen P. Borgatti and Daniel S. Halgin. On network theory. *Organization Science*, 22(5):1168–1181, 2011.

[5] Amit Bubna, Sanjiv R. Das, and Nagpurnanand Prabhala. Venture capital communities. *The Journal of Financial and Quantitative Analysis*, 55(2):621–651, 2020. Accessed: 2025-08-24. JSTOR catalogue.

[6] William D. Bygrave. Syndicated investments by venture capital firms: A networking perspective. *Journal of Business Venturing*, 2(2):139–154, 1987.

[7] Théophile Carniel. *Complex networks in entrepreneurial ecosystems: clustering methodologies and topological structure*. Thèse, Université Paris Cité, 2024. Catalogue: inria.hal.science. Consulted: 2025-07-18 15:30:01. Added: 2025-07-18 15:30:01. Modified: 2025-07-18 15:30:06.

[8] Catherine Casamatta and Carole Haritchabalet. Experience, screening and syndication in venture capital investments. *Journal of Financial Intermediation*, 16(3):368–398, 2007. Accessed: 2024-11-18. 189 citations (Crossref/DOI) [2024-11-22].

[9] Jean-Michel Dalle and Matthijs Den Besten. Using crunchbase for economic and managerial research. OECD Science, Technology and Industry Working Papers 2017/08, OECD, November 2017. Accessed: 2025-08-24. DOI: 10.1787/6c418d60-en. Series: OECD Science, Technology and Industry Working Papers Volume: 2017/08.

[10] Jean-Michel Dalle, Matthijs Den Besten, and Jeremie Morfin. Accelerator-mediated access to investors among early-stage start-ups. *Annals of Operations Research*, 348(3):1925–1952, 2025. In press. Accessed: 2025-05-26. 0 citations (Semantic Scholar/DOI) [2025-07-14]. Added: 2025-05-26 19:10:14. Modified: 2025-07-14 13:29:18.

[11] Carsten Dormann, Jochen Fründ, Nico Blüthgen, and Bernd Gruber. Indices, graphs and null models: Analyzing bipartite ecological networks. *The Open Ecology Journal*, 2:7–24, 2009. Accessed: 2025-08-24. ResearchGate catalogue.

[12] Mark Granovetter. Economic action and social structure: The problem of embeddedness. *American Journal of Sociology*, 91(3):481–510, 1985.

[13] Josh Lerner. The syndication of venture capital investments. *Financial Management*, 23(3):16–27, 1994. Special Issue on Venture Capital. Reprinted in Venture Capital, edited by Michael Wright and Ken Robbie. Aldershot: International Library of Management/Dartmouth Publishing, 1997.

[14] Sebastian Mariani. *Nestedness in complex networks*. Springer, 2019. ISBN: 978-3-030-29984-6.

[15] M. E. J. Newman. Modularity and community structure in networks. *Proceedings of the National Academy of Sciences*, 103(23):8577–8582, 2006. arXiv:physics/0602124. Accessed: 2025-08-24.

[16] Michael Spence. Signaling, screening, and information. In Sherwin Rosen, editor,

Studies in Labor Markets, pages 319–358. University of Chicago Press, 1981. Accessed: 2025-05-17.

[17] Giovanni Strona, Domenico Nappo, Francesco Boccacci, Simone Fattorini, and Jesús San-Miguel-Ayanz. A fast and unbiased procedure to randomize ecological binary matrices with fixed row and column totals. *Nature Communications*, 5(1):4114, 2014.