

# Venture Capital Network Structure and Nestedness Analysis

Disclaimer: this is a intermediary report over the results gotten until now. Colored rectangles represent "to do" or "ongoing" activities to be done in subsequent work until the final version of this article is complete.

## 1 Introduction

Two or more venture capital (VC) firms co-investing on the same enterprise is known in economics as syndication. In innovation networks, investors tend to behave like this so they can reduce the risk of investing in something not yet completely validated or functional. In that case, reputation and centrality play an important role, once VCs will not only try to measure the potential return on investment (ROI), but also use its co-investors characteristics as a signal that interferes on their decisions.

The rising of syndicated investments in the last decades is evidence that innovation networks are by far a socialized network, where agents are not acting isolated, randomly, but in communities, being influenced by its peers. This phenomena is known as embeddedness.

Vast literature show how heavy tailed degree distributions emerge from such kind of interactions in social networks. In this same manner, innovation networks are not an exception, specially when it comes to number of connections of a certain player, as concentrated hubs of strongly connected

agents can normally be seen, while most part of the sample have only few ties, leading to a heterogeneous distribution of connectance - also known as power-law degree distribution.

The widespread presence of power law degree distributions has incentivised numerous studies focused on uncovering plausible mechanisms behind their emergence, as well as exploring their impact on processes such as spreading dynamics [5] and network robustness [1]

When it comes to spreading dynamics, social and economic scientists have already explored how novel ideas are spread through networks, and how formation of bridge edges (with high betweenness) impact the chances of, for instance, novelty to spread in innovation networks. Well established ideas, like the Strength of Weak Ties theory, are normally used as theoretical bases for this kind of assumption.

In the other hand, measure the robustness to social and, being more specific, innovation networks is still a theoretical and practical challenge. Impressively, ecology came to play an important role to face it, and metrics like nestedness and the ecological consequences of its presence started to be transposed to social networks [7].

On that paper, network theory is used to represent syndicated investments as edges of a network where investors are nodes with broad set of characteristics (geographic, financial, sectorial, etc.) This mathematical representation opens horizons to better visualize and interpret characteristics of this syndication network structure (or sub-networks inside of it) through ecology and economics lenses.

Special attention is given for the fact that nestedness was observed among a certain group of early and late state investors.

## 2 Methodology

### 2.1 Data Source and Preprocessing

This study uses data from Crunchbase, a broad database containing information about startups, venture capital firms, and investment rounds. The dataset includes information about companies, investors, investments, and funding rounds in the United States market. International venture capital firms from other countries also appear in the dataset when they participate in US startup investments.

The data preprocessing follows established methodologies from entrepreneurship literature [4]. The cleaning process implemented includes several steps: (1) removal of companies with incomplete information, (2) exclusion of companies founded after 2017 to allow sufficient time for investment patterns to emerge, (3) removal of companies with exit status (bankruptcy, acquisition, or IPO), and (4) application of a minimum funding threshold of \$150,000 to focus on substantive investment relationships.

### 2.2 Investment Network Construction

The analysis focuses on venture capital co-investment patterns across different funding stages. Investment stages are categorized into two main groups:

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

A bipartite network is constructed where nodes represent venture capital firms and edges represent co-investment relationships in the same company. The network is bipartite because it connects two distinct sets of investors: those participating in early-stage rounds (right nodes) and those participating in late-stage rounds (left nodes).

This approach allows us to study how early-stage and late-stage investors interact in the investment ecosystem.

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

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

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

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

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. Further more, investors that participated in both early and late stages receive a suffix so they can be treated as distinct agents for each phase.

Clearly show the overlap or number of connections made between the same investors but in distinct phases ex. VC1\_serieA-VC1\_serieC

## 2.3 Community Detection

Community structure in the bipartite network is identified using the greedy modularity optimization algorithm [3]. This method 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:

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

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.

The algorithm identifies communities of venture capital firms that frequently co-invest together, revealing 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 [2] 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 [2]. For a bipartite adjacency matrix  $M$  with rows and columns sorted by decreasing degree, NODF is calculated as:

$$NODF = \frac{NODF_{rows} + NODF_{columns}}{2} \quad (5)$$

where:

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

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

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  $i$ .

With this method, NODF vary between 0 and 1 (perfect nestedness).

## 2.5 Statistical Significance Testing

To determine whether observed nestedness values are significantly higher than expected by chance, we employ a null model approach using the Curve-

ball algorithm [6]. This algorithm generates randomized matrices that preserve the degree sequence of both node sets while randomizing the connection patterns.

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

$$Z = \frac{NODF_{observed} - \mu_{null}}{\sigma_{null}} \quad (8)$$

where  $\mu_{null}$  and  $\sigma_{null}$  are the mean and standard deviation of the null distribution. Communities with  $p < 0.05$  (where  $p$  is the proportion of null models with  $NODF \geq \text{observed NODF}$ ) are considered to have significantly high nestedness.

Better explain Z-core and P-values interpretation and relationships

## 3 Results

### 3.1 Network Characteristics

The Crunchbase dataset, following the cleaning processes described in the "Methodology" section, yields 147,832 investment registers, representing transactions among 22,527 companies and 38,843 investors.

Exclusion of non-venture capital investors reduces the dataset to 104,618 investment records and 16,932 unique companies with venture capital funding.

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.

Add network visualization showing bipartite structure

### 3.2 Community Structure and Size Distribution

Community detection using greedy modularity optimization identifies 175 distinct communities, 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.

Analysis focuses on communities with at least 150 nodes to ensure statistical power for nestedness analysis. Such a threshold yields 5 communities.

Table 1 shows the size distribution of the largest communities identified by the modularity optimization algorithm.

Rationale of threshold

| Community ID | Number of Pairs |
|--------------|-----------------|
| 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.

Mention literature, as this phenomena is somehow well-known

Add figure of community size distribution

### 3.3 Nestedness Findings

Nestedness analysis across investor communities reveals heterogeneous structural patterns. Among the 5 communities examined, one exhibits statistically significant nestedness ( $p < 0.01$ ) relative to degree-preserving null models generated through the Curveball algorithm.

Figure 1 presents the comparison between observed and null model nestedness scores, where each data point represents a distinct community positioned according to its observed NODF value against the corresponding null model mean.

Community 2 demonstrates the most pronounced nestedness, exhibiting an NODF score of 0.088 with statistical significance of  $p = 0.00001$ . This hierarchical structure indicates that 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 contributes to the nested structure by creating hierarchical dependencies between investor types.

The following sections provide detailed characterization of this nested community through comparison with Communities 0 and 1, which serve as contrasting examples of similar-sized but differently structured investor networks.

### 3.4 Communities Characterization

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



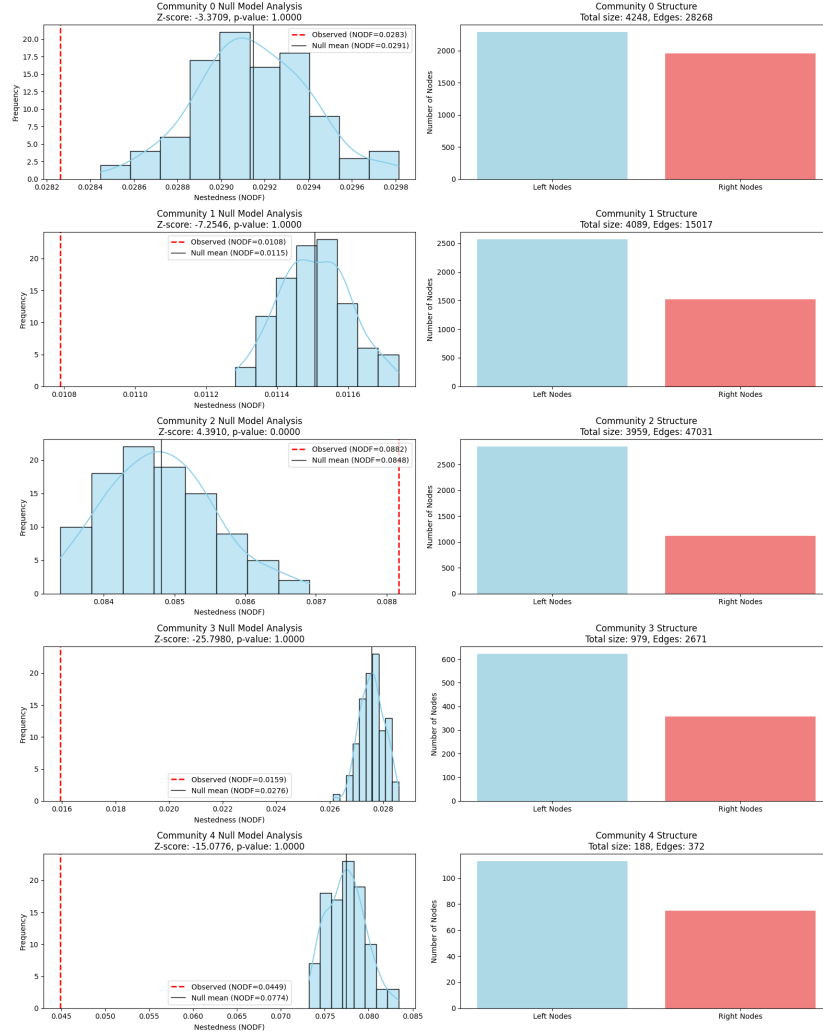


Figure 1: Comparison of observed versus null model nestedness scores for the five largest investor communities. The diagonal line represents equal observed and expected values, with points above the line indicating higher-than-random nestedness. "Left" stands for late stage investors, and "Right" for early stage ones.

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       | Number of Investments | Relative Proportion |
|-----------------|-----------------------|---------------------|
| Community 0     | 32,164                | 19.4%               |
| Community 1     | 17,301                | 10.4%               |
| Community 2     | 55,863                | 33.6%               |
| Community 3     | 2,904                 | 1.7%                |
| Cross-community | 58,329                | 35.1%               |
| <b>Total</b>    | <b>166,561</b>        | <b>100%</b>         |

Table 2: Distribution of syndicated investments across investor communities

The analysis reveals important patterns in investment activity distribution. Community 2, which exhibits significant nestedness, 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 within the nested community suggests that hierarchical investor structures may facilitate higher transaction volumes. Notably, cross-community investments represent over one-third of all transactions, indicating substantial interconnectedness across community boundaries.

### 3.4.1 Funding Characteristics

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

The funding characteristics analysis indicates that nested communities exhibit concentrated capital deployment patterns, with higher-degree investors participating in larger funding rounds while maintaining broader portfolio diversification. This suggests that hierarchical investor organization may

Add  
at-  
tach-  
ment  
with  
statis-  
tical  
proofs

plot a  
graph  
of dis-

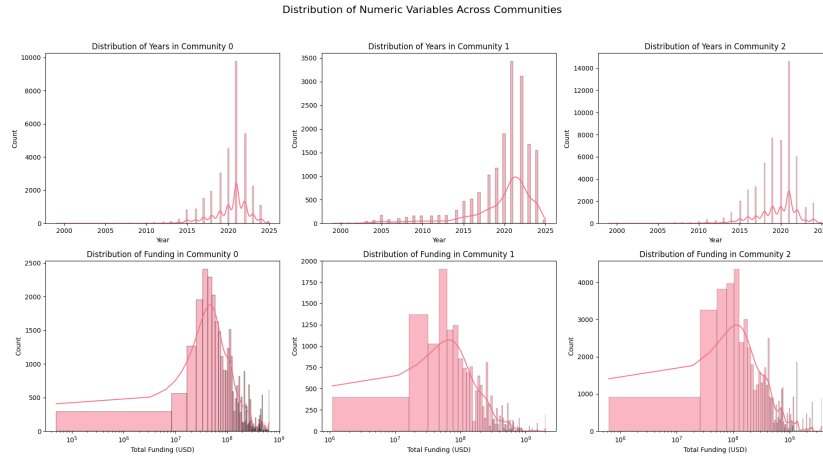


Figure 2: Funding characteristics across the three largest investor communities. The analysis reveals systematic differences in investment amounts, round frequency, and funding patterns between communities, with nested structures exhibiting distinct capital deployment strategies compared to randomly organized investor groups.

create more efficient capital allocation mechanisms compared to randomly structured networks.

### 3.4.2 Geographic Distribution

Geographic analysis reveals distinct spatial clustering patterns across the three largest communities. Figure 3 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 contain only american startups' investments, which include international investors).

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 compo-

add  
liter-  
ature  
base  
for  
this  
strong  
as-  
sump-  
tion

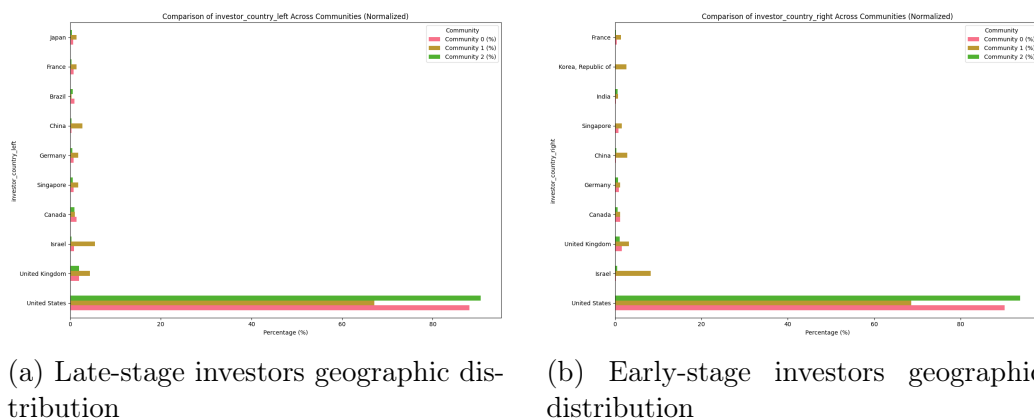


Figure 3: Geographic distribution of venture capital investors across the largest communities. The bipartite structure reveals differential geographic clustering between late-stage (left) and early-stage (right) investor networks, with Community 1 exhibiting greater international diversification compared to the U.S.-concentrated Communities 0 and 2.

sition 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.

### 3.4.3 Investment Stage Preferences

The investment stage distributions reveal systematic specialization patterns across communities that align with their geographic profiles and transaction volumes. Figure 4 demonstrates the distribution patterns of investment types within the bipartite network structure.

**Comment investment stages distribution**

**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 par-

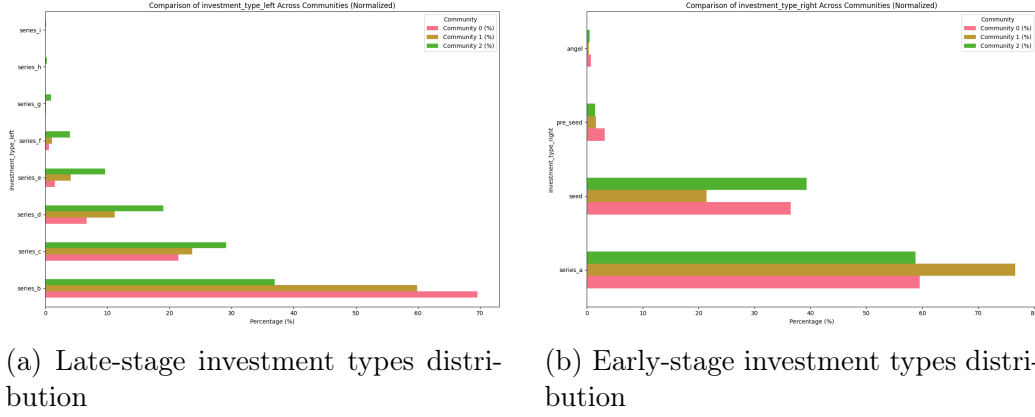


Figure 4: Investment stage distribution across the three largest communities. The distribution patterns reveal stage-specific specialization within investor communities.

ticular 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 the nested structure in Community 2 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.

### 3.4.4 Sectoral Focus

The sectoral analysis reveals distinct specialization patterns that align with each community’s structural and geographic characteristics. Figure 5 illus-

trates the distribution of investment focus across technology sectors, demonstrating how different communities exhibit varying degrees of sectoral concentration.

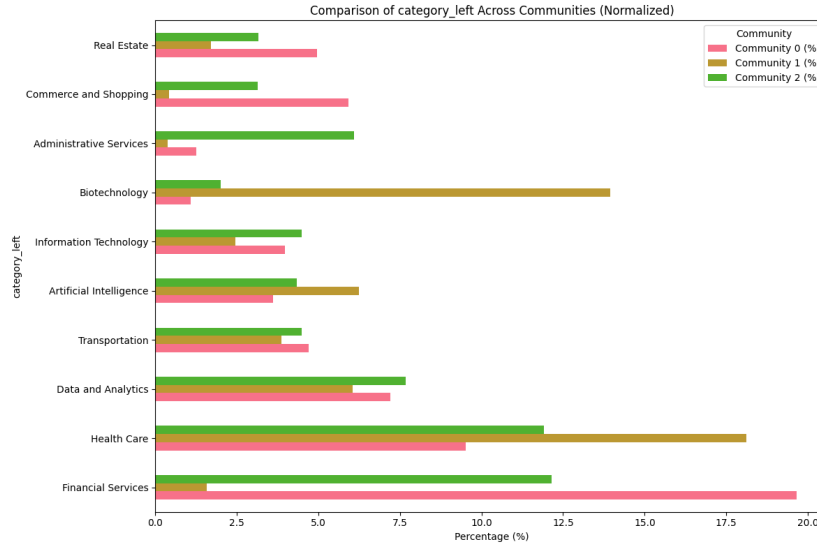


Figure 5: Sectoral distribution across the three largest investor communities. The analysis reveals differential industry focus patterns, with certain communities demonstrating concentrated investment strategies in specific technology sectors while others maintain broader sectoral diversification.

#### Comment sectorial distribution

**Community 0:** Concentrates in real estate, commerce and shopping, and financial services, while maintaining standard representation in information technology and artificial 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:** The nested 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 nested structure facilitates broader sectoral participation rather than concentrated specialization.

Community 0 serves as an effective structural baseline for comparison with nested Community 2, given their similar geographic profiles and certain sectoral overlaps, while differing significantly in network organization and transaction volumes.

Add overall comparison

Comment more on funding characteristics

## 3.5 Implications and Future Directions

The discovery of significantly nested communities within the 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.

### 3.5.1 Network Robustness and Resilience

The 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 [3].

Following insights from nestedness research in complex networks [?], the hierarchical organization observed in Community 2 may confer distinct robustness properties to the venture capital ecosystem. In mutualistic networks, 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 while 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 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 nestedness creates hierarchical fragility patterns.

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

### 3.5.3 Dynamic Network Evolution

The temporal evolution of community nestedness represents another critical research direction. Investigating how nested structures emerge, persist, and potentially dissolve over time could illuminate the underlying mecha-

Add  
the  
anal-  
ysis  
per-  
formed  
to this  
part  
and  
re-



nisms driving venture capital ecosystem organization. This analysis should examine whether nestedness develops gradually through preferential attachment processes or emerges rapidly through strategic alliance formation.

Future research should also explore the predictability of link formation and dissolution within nested communities. The hierarchical structure may create predictable patterns of investor syndication, with new partnerships more likely to form between investors already connected to common highly-connected nodes. Conversely, link dissolution might follow predictable patterns based on position within the nested hierarchy.

#### **3.5.4 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. Chapter 5 of "Nestedness in complex networks" [?] 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 nested communities provide superior or inferior outcomes for portfolio companies compared to randomly organized investor groups represents a crucial research priority. 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.

#### **3.5.5 Policy and Ecosystem Development Implications**

Understanding these patterns may inform policy discussions about startup ecosystem development and investor network formation. If 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. Conversely, if 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 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 nested organization.

This analysis provides the foundation for deeper investigation into how nested investor communities influence entrepreneurial ecosystems and capital allocation efficiency, which will be the focus of subsequent research phases.

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

Analyze the temporal evolution of community nestedness to understand how these structures emerge and persist over time, including prediction of link formation and dissolution patterns.

Examine whether nested communities provide better or worse outcomes for portfolio companies compared to random investment patterns, considering both efficiency gains and systemic risks.

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.

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.

## References

- [1] Réka Albert, Hawoong Jeong, and Albert-László Barabási. Error and attack tolerance of complex networks. *Nature*, 406(6794):378–382, 2000.
- [2] 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.
- [3] Stephen P. Borgatti and Daniel S. Halgin. On network theory. *Organization Science*, 22(5):1168–1181, 2011.
- [4] Jean-Michel Dalle et al. Accelerator-mediated access to investors among early-stage start-ups. *Research Policy*, 2025. In press.
- [5] Romualdo Pastor-Satorras and Alessandro Vespignani. Epidemic spreading in scale-free networks. *Physical Review Letters*, 86(14):3200–3203, 2001.

- [6] 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.
- [7] Author Theophile. Complex networks in entrepreneurial ecosystems. 2024. Working paper.