

Beyond Isolated Investor: Predicting Startup Success via Roleplay-Based Collective Agents

Zhongyang Liu¹, Haoyu Pei¹, Xiangyi Xiao¹, Xiaocong Du¹, Yihui Li¹, Suting Hong², Kunpeng Zhang³, Haipeng Zhang^{1*}

¹ShanghaiTech University

²Xi'an Jiaotong-Liverpool University

³University of Maryland

{liuzhy12024, peihy2024, xiaoxy, duxc2023, liyh, zhanghp}@shanghaitech.edu.cn,
 Suting.Hong@xjtlu.edu.cn, kpzhang@umd.edu

Abstract

Due to the high value and high failure rates of startups, predicting their success is a critical challenge. Existing approaches typically model startup success from a single decision-maker’s perspective, overlooking the collective dynamics that dominate real-world venture capital (VC) decision-making. We propose **SimVC-CAS**, a collective agent system that simulates VC decisions as a multi-agent interaction process. By designing roleplaying agents and a GNN-based supervised interaction module, we reformulate startup financing prediction as a group decision-making task, capturing both enterprise fundamentals and investor network dynamics. Each agent represents an investor with distinct traits and preferences, enabling heterogeneous evaluations and realistic information exchange over a graph-structured co-investment network. Using both proprietary and public VC data with strict anti-leakage controls, we show that SimVC-CAS significantly improves predictive performance, achieving approximately 25% relative improvement in average precision@10, while exhibiting consistency with real investor decisions. The interaction mechanism is particularly effective for network-central startups, confirming the importance of network in VC decision-making. Analysis of agents’ reasoning for decision changes further reveals how network environments influence decision quality, demonstrating the system’s interpretability. Our approach may generalize to broader group decision-making scenarios. Our code is available at <https://github.com/ZhangDataLab/SimVC-CAS>.

1 Introduction

Startups contributed 60% of global new jobs (2010-2020) [Union, 2019] yet face a 90% failure rate [Kerr *et al.*, 2014]. This significant contrast has made the prediction of early-stage startup success (e.g., financing outcomes

*Corresponding author

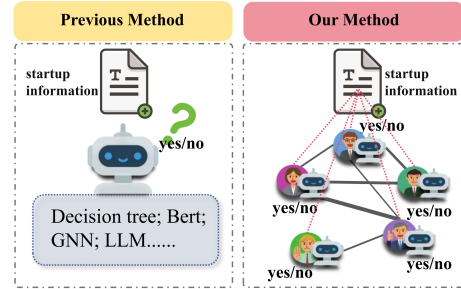


Figure 1: Unlike prior single perspective approaches, our method jointly models the collective behaviors of startups and investors by incorporating multiple decision-makers.

and IPO potential) a central problem in interdisciplinary research [Zhang *et al.*, 2021; Lyu *et al.*, 2025]. Among various success indicators, early-stage financing plays a critical role, serving as the “lifeline” for startup survival and growth [Casar, 2004]. Access to capital not only provides essential financial resources but also shapes long-term strategic direction. Consequently, predicting whether a startup can successfully secure financing has emerged as a core task in startup success prediction [Zhang *et al.*, 2021; Dellermann *et al.*, 2021].

In venture capital (VC), investors typically base their decisions on comprehensive evaluations that consider factors such as team composition, business model, and market potential [Miloud *et al.*, 2012; Kim and Lee, 2022]. Following this logic, prior automated methods have primarily modeled financing outcomes from the perspective of individual decision-making by extracting firm-level features using technical paradigms. Lines of work include: (i) traditional machine learning on engineered features [Krishna *et al.*, 2016; Ünal, 2019]; (ii) graph neural network (GNN)-based modeling of investment networks [Zhang *et al.*, 2021; Lyu *et al.*, 2025]; and (iii) pre-trained language models for textual analysis [Maarouf *et al.*, 2025]. **Despite methodological diversity, these approaches share a common assumption: they aim to simulate a single, idealized investor.**

In reality, investor decisions directly shape startup financing outcomes and subsequent development trajectories. The absence of investment can trigger immediate survival crises,

while investors’ heterogeneous resources, expertise, and networks can profoundly influence long-term growth paths [Diamond and Rajan, 2001; Miloud *et al.*, 2012]. **Importantly, real-world VC evaluations are rarely conducted in isolation. Instead, startups are assessed by multiple potential investors embedded within co-investment networks, where information exchange and peer influence play a central role [Han and Yang, 2013; Liu, 2017; Goldstein *et al.*, 2025].** Our data confirm this dynamic: the average graph distance among investors who eventually co-invest in the same startup is only 2.16, compared to 4.23 across the entire VC network. As a result, financing outcomes often emerge from the collective judgment of an investor group, and startups of similar quality may experience markedly different outcomes depending on the composition of their investor collective. This “investor collective effect” is not captured by existing single-decision-maker models. Given the inherent multi-agent nature of real-world VC decision-making, a multi-agent framework is a natural and necessary modeling choice.

However, existing multi-agent systems fail to capture two defining characteristics of VC forecasting: (i) the **collective nature** of syndicated decision-making and (ii) the **complex relational dynamics** among co-investors. **First**, most frameworks rely on agent specialization to simulate compartmentalized reasoning across information dimensions—assigning individual agents to analyze different dimensions of information [Yu *et al.*, 2024; Zhang *et al.*, 2024b; Wang *et al.*, 2025b; Pei *et al.*, 2026]. While effective for simulating the internal reasoning process of a single investor, this paradigm fails to capture the diversity and interaction of multiple independent investors. **Second**, existing agent interaction mechanisms are largely heuristic-based [Wang *et al.*, 2025a; Qian *et al.*, 2025; Li *et al.*, 2025] and cannot model the nuanced, asymmetric relationships observed in real co-investment behavior—relationships shaped by network topology, investor heterogeneity, and startup-specific context.

To address these challenges, we propose **SimVC-CAS**: a collective agent system for simulating real-world venture capital decision-making. The core idea is to reframe financing prediction as the outcome of a simulated collective decision process carried out by a network of investors, each modeled as an autonomous agent. SimVC-CAS jointly captures enterprise fundamentals and multi-investor decision dynamics. Given a target startup, the system simulates (i) heterogeneous individual evaluations by distinct investor agents and (ii) collective decision updates shaped by a co-investment network. The resulting group-level verdict forms the basis for financing prediction. As shown in Figure 1, this framework enables joint modeling of both company-level attributes and investor-collective behaviors, offering a more realistic and interpretable approach to VC forecasting.

Specifically, SimVC-CAS consists of three modules: **1. Startup Panoramic Portrait.** This module integrates diverse startup data - including basic information, founding team backgrounds, and financing history - to construct a comprehensive profile that serves as shared context for all investor agents. **2. Heterogeneous Investor Portraits.** We construct the potential investor pool from the startup’s his-

torical investors and their past co-investment partners. Empirically, 77.35% of investors in new financing rounds originate from this group in our dataset. Each investor agent is assigned a distinct profile based on background information, experience, and historical investment behavior. Leveraging the role-playing capabilities of large language models (LLMs) [Wang *et al.*, 2025a], agents interpret startup information through personalized lenses, capturing heterogeneous preferences, cognitive biases, and decision patterns. **3. Collective Agent Interaction Modeling.** Investor interactions depend not only on their own attributes and network structure, but also vary with the characteristics of the evaluated startup, forming startup-centric interaction modes [Elfring and Hulsink, 2007; Lee *et al.*, 2015]. To model these dynamics, we design a graph attention network with virtual nodes (VGAT), where a virtual node represents the target startup and serves as an information hub connected to all investor agents. Investor nodes are linked via edges derived from historical co-investment relationships. Through VGAT, SimVC-CAS captures multi-dimensional interaction patterns that balance investor heterogeneity, startup characteristics, and graph topology, thereby approximating realistic group deliberation processes.

Our contributions are summarized as follows:

- We introduce a simulation-based paradigm that reframes startup financing prediction as a collective decision-making process by integrating LLM role-playing with multi-agent modeling.
- We propose **SimVC-CAS**, which jointly models company fundamentals and investor dynamics through scalable profiles and graph-based interactions, achieving significant performance improvements on proprietary and public datasets, with behavioral alignment with real investors.
- We demonstrate that the interaction mechanism is particularly effective for network-central startups, confirming the importance of network structure in VC decision-making. Analysis of agents’ reasoning for decision changes further reveals how network environments influence decision quality, demonstrating the system’s interpretability. Beyond VC, our design may generalize to other group decision-making problems.

2 Related Work

2.1 Startup Success Prediction

Predicting early-stage startup success is a critical task. Early research identified key predictive factors [Song *et al.*, 2008; Sevilla-Bernardo *et al.*, 2022] which informed manual feature engineering efforts. Initial methods used commercial datasets such as PitchBook, Crunchbase to design features based on domain knowledge and heuristics for traditional machine learning models [Ünal, 2019; Bargagli-Stoffi *et al.*, 2021]. As large-scale datasets become available, the field evolved toward three approaches: (1) graph neural networks (GNNs) [Zhang *et al.*, 2021; Lyu *et al.*, 2025], which mine relational structures for investment patterns; (2) pre-trained

language models (LMs) [Maarouf *et al.*, 2025], which extract semantic features from company descriptions and filings; and (3) multi-agent systems [Wang *et al.*, 2025b; Pei *et al.*, 2026], which assign agents to analyze complementary aspects of a company’s information, aiming to address the interpretability and scalability limitations.

Despite these advances, existing models all focus on single-investor decision frameworks, failing to capture the multi-perspective evaluation and group dynamics typical of the real world. Even when employing multi-agent approaches, they merely use multiple agents to analyze different dimensional information [Wang *et al.*, 2025b], remaining within a single decision-maker perspective.

2.2 LLM Agents in Decision Simulation

Large Language Models have shown significant potential in simulating human intelligence and are increasingly used as a foundational component in autonomous agent systems [Luo *et al.*, 2025; He *et al.*, 2025; Zhang *et al.*, 2025]. In decision simulation, LLM agents have been deployed to model user search behavior and predict click-through rates [Zhang *et al.*, 2024a], forecast legislative bill outcomes based on policymaker behavior [Li *et al.*, 2025], and replicate group voting patterns from demographic data [Argyle *et al.*, 2023].

However, the application of LLMs to simulate investor decision-making in VC remains largely underexplored. Current approaches are insufficient for modeling the heterogeneous, relational dynamics inherent in VC investment networks. To address this gap, we propose a graph neural network-based multi-agent interaction framework for simulating VC investment decisions. Beyond VC, this framework offers a generalizable foundation for decision modeling in other complex, heterogeneous network environments.

3 Task Definition

This study focuses on predicting the success of early-stage startups. Following prior work [Zhang *et al.*, 2021], we define early-stage startups as those that have completed their first formal financing round (seed or angel), but have not yet raised subsequent rounds (e.g., Series A). These companies enter the venture capital ecosystem for the first time through initial investors. While success is often measured by whether a startup secures Series A funding [Zhang *et al.*, 2021; Dellermann *et al.*, 2021], prior studies have used different observation windows, potentially introducing time biases. To alleviate this, we adopt a consistent one-year observation window. Accordingly, the core task is defined as: **predicting whether a startup secures subsequent funding within one year of its initial financing.**

4 Method

This section details the core framework of our proposed method, shown in Figure 2. We begin by constructing a comprehensive profile for each target startup, integrating basic information, team composition, and historical financing records. This profile serves as the fundamental input for investment evaluation. Meanwhile, we generate heterogeneous

profiles for potential investors, encompassing personal information, investment experience, and professional background. To simulate investor decision-making, we employ an LLM-based agent to model the investment preferences of individual investors through role-playing. Building on this, we introduce a collective agent interaction system: for each startup, we construct a co-investment network linking its potential investors based on historical co-investment relationships. In addition, we designed a GAT with virtual nodes (VGAT) to capture the investor interaction patterns centered on startups in this network, enabling us to approximate the collective decision-making dynamics in the real world. This collective modeling enhances the predictive accuracy of startup success.

4.1 Startup Panoramic Portrait

To predict success, we construct a panoramic profile capturing the target startup’s core characteristics as input for subsequent agents.

Basic Information: This includes founding date, industry, description, location, products, keywords, and other core attributes. These characteristics provide fundamental signals for screening investment opportunities and reducing information asymmetry between entrepreneurs and investors [Kaplan and Strömberg, 2004; Spence, 1973].

Team Composition: We collect key members’ backgrounds, including gender, education, prior employment, investment experience, and roles. This reflects the “bet on the jockey” hypothesis in VC, which emphasizes human capital as a key predictor of venture success [Kaplan *et al.*, 2009; Gompers *et al.*, 2020].

Historical Financing Records: We summarize past funding rounds, including amounts raised and investor backgrounds. These records reflect capital staging mechanisms for monitoring progress and mitigating agency costs, while also providing cues on valuation trajectory and external validation [Gompers, 1995; Sahlman, 1990].

Affiliated Companies: We incorporate information on startups sharing key individuals with the target. Such links enable information spillovers and social endorsements, allowing affiliated firms’ attributes to serve as signals of the target’s potential [Stuart *et al.*, 1999; Hochberg *et al.*, 2007].

4.2 Heterogeneous Investor Portraits

For each target startup, we construct heterogeneous profiles of its potential investors. The candidate investor set follows this rule: **We randomly sample k individuals from the union of the startup’s historical investors and their past co-investment partners.** Empirically, 77.35% of investors in new financing rounds originate from this group in our dataset.

Based on this set, we generate investor profiles and simulate their behavior through a role-playing agent architecture. Each agent represents a distinct investor persona, offering differentiated analysis and decision-making support. The profiling system is extensible and integrates demographic information, professional experience, and past investment records.

Personal Information This includes the investor’s name, gender, and educational background.

Past Investment Experience This section summarizes the investor’s funding history, including: (i) early-stage invest-

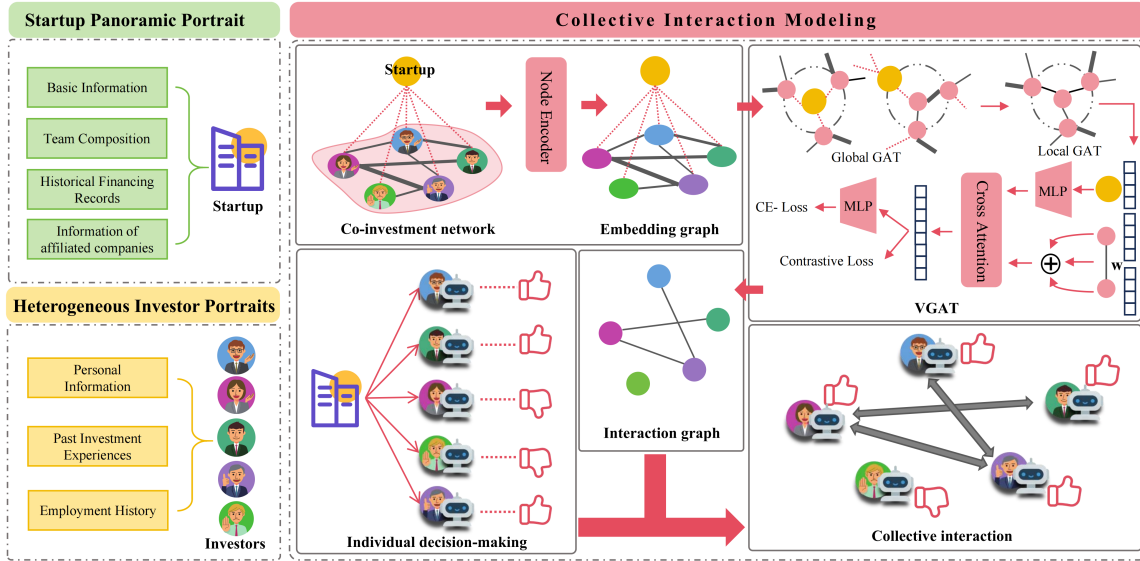


Figure 2: The overall framework of our proposed method.

ments, mainly seed and Series A rounds, as primary references; and (ii) other mid-to-late-stage or non-startup investments as supplementary context. Each investment record is paired with structured startup summaries to characterize the investor’s track record.

Employment History This details the investor’s professional background, including previous roles and affiliations.

4.3 Collective Agent Interaction Modeling

To better approximate real-world investor decision-making dynamics, we design a collective agent interaction framework supervised by a graph neural network. In this system, each investor agent first performs an individual evaluation of the target startup based on their profile and outputs an initial investment decision. Next, the historical co-investment network is used (along with VGAT) to determine the interaction structure among these investor agents. Finally, the agents interact according to the learned topology, updating their beliefs and assessments based on peer influence. This dynamic group reasoning process allows agents to revise their initial decisions, resulting in a final, aggregated investment decision that reflects both individual judgment and collective insight.

VGAT

To model investor interaction patterns shaped jointly by the characteristics of the target startup, the attributes of individual investors, and the topology of the co-investment network, we design the Virtual Node Graph Attention Network (VGAT). In this architecture, the target company is introduced as a virtual node, serving as a global query to attend to relevant investor interactions.

Formally, let $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \omega)$ be an undirected weighted graph. The node set is defined as $\mathcal{V} = \mathcal{V}_R \cup \{v_{target}\}$, where $\mathcal{V}_R = \{v_1, \dots, v_n\}$ represents real investor nodes and $v_{target} \notin \mathcal{V}_R$ denotes the virtual node (i.e., the target company). The edge set consists of two parts: $\mathcal{E} = \mathcal{E}_R \cup \mathcal{E}_V$, where $\mathcal{E}_R \subseteq \mathcal{V}_R \times \mathcal{V}_R$ represents real investor connections,

and $\mathcal{E}_V = \{(v_{target}, v_i) \mid \forall v_i \in \mathcal{V}_R\}$ connects the virtual node to all real nodes. A weight function $\omega : \mathcal{E} \rightarrow \mathbb{R}^+$ assigns positive weights to all edges.

The VGAT model processes $\mathcal{G}$ through three sequential components. First, a global Graph Attention (GAT) layer updates embeddings for all nodes:

$$\{h_v^{(1)}\}_{v \in \mathcal{V}} = \text{GAT}(\mathcal{G}). \quad (1)$$

Second, a local GAT layer refines the embeddings specifically for real investor nodes:

$$\{h_v^{(2)}\}_{v \in \mathcal{V}_R} = \text{GAT}(\mathcal{G}[\mathcal{V}_R]). \quad (2)$$

Concurrently, the virtual node embedding is transformed via an MLP to serve as the global query vector:

$$h_{v_{target}}^* = \text{MLP}(h_{v_{target}}^{(1)}). \quad (3)$$

To determine the importance of each co-investment relationship, we compute an attention score for each real edge $e_{ij} = (v_i, v_j) \in \mathcal{E}_R$. The cross-attention mechanism uses the virtual node as the query and the edges as keys. The normalized attention coefficient α_{ij} is computed via a Softmax operation over all edges in the batch:

$$\alpha_{ij} = \text{Softmax}_{e_{ij} \in \mathcal{E}_R} \left(\frac{(h_{v_{target}}^* \mathbf{W}_Q)(\mathbf{m}_{ij} \mathbf{W}_K)^\top}{\sqrt{d_k}} \right), \quad (4)$$

where d_k is the scaling dimension. The final edge embedding e_{ij} is then obtained by weighting the value projection:

$$e_{ij} = \alpha_{ij}(\mathbf{m}_{ij} \mathbf{W}_V), \quad (5)$$

where the base edge representation $\mathbf{m}_{ij}$ concatenates the node features and edge weight:

$$\mathbf{m}_{ij} = [h_{v_i}^{(2)} \parallel h_{v_j}^{(2)} \parallel \omega_{ij}]. \quad (6)$$

Here, $\mathbf{W}_Q$, $\mathbf{W}_K$, and $\mathbf{W}_V$ are learnable projection matrices.

Method	Precision	Recall	F1
GAT	69.62	77.76	73.47
VGAT	76.18	82.70	79.31

Table 1: Performance comparison on the link prediction task.

We formulate the interaction prediction as a **binary classification task**. Since direct data on investor communications is private and unobservable, we employ **future co-investment relationships** as a proxy for effective interaction. The rationale is that if two investors co-invest in the same startup within a specific future window, they likely engaged in meaningful interaction or shared similar decision logic to mitigate selection uncertainty [Lerner, 1994; Hochberg *et al.*, 2007]. Thus, the goal is to predict the existence of a valid future co-investment link (Class 1) versus a non-existent one (Class 0).

The model is optimized jointly using cross-entropy and contrastive losses. The Cross-Entropy Loss is defined as:

$$\mathcal{L}_{CE} = -\frac{1}{|\mathcal{E}_R|} \sum_{e_{ij} \in \mathcal{E}_R} \sum_{c=1}^{N_C} y_{ij}^c \log(\hat{y}_{ij}^c), \quad (7)$$

where $N_C = 2$ is the number of classes. y_{ij}^c indicates the ground truth label derived from future co-investment data, and $\hat{y}_{ij}^c$ is the predicted probability derived from e_{ij} .

The Contrastive Loss enhances the distinctiveness of edge embeddings:

$$\mathcal{L}_{Contrast} = -\frac{1}{|\mathcal{P}|} \sum_{(e_{ij}, e_{kl}) \in \mathcal{P}} \log \frac{\exp\left(\frac{e_{ij}^\top e_{kl}}{\tau \|e_{ij}\| \|e_{kl}\|}\right)}{\sum_{e_{mn} \in \mathcal{N}_{ij}} \exp\left(\frac{e_{ij}^\top e_{mn}}{\tau \|e_{ij}\| \|e_{mn}\|}\right)}, \quad (8)$$

where $\mathcal{P}$ is the set of positive pairs, $\mathcal{N}_{ij}$ represents negative samples, and τ is the temperature parameter. The final loss is:

$$\mathcal{L} = \mathcal{L}_{CE} + \lambda \mathcal{L}_{Contrast}, \quad (9)$$

where λ is a hyperparameter balancing the two terms.

Training and Validation: We strictly adhered to the temporal split strategy of the main experiment (see Sec.5.1) to train and evaluate VGAT: data prior to September 2021 was utilized for training and validation. As shown in Table 1, VGAT significantly outperforms the GAT baseline (F1: 79.31% vs. 73.47%). This confirms that explicitly modeling the startup as a global query node enables the network to more accurately discern context-specific investor interaction patterns. (See Appendix D for training data construction.)

Investor Agent Decision-Making

Given a set of k investor agents $\{A_1, A_2, \dots, A_k\}$, we construct a heterogeneous profile P_i for each investor A_i . Through role-playing, conditioned on the startup profile $\mathcal{S}$, each agent generates an initial independent decision $D_i^{(0)}$ using a frozen LLM:

$$D_i^{(0)} = \text{LLM}(\mathcal{S}|P_i) \quad (10)$$

We construct a co-investment network $\mathcal{G}$ where nodes $\mathcal{V} = \{A_1, \dots, A_k\} \cup \{v_{target}\}$ represent investors and the target

startup. Edges encode historical co-investments and connections to the target. Using a pre-trained¹ encoder, we embed investor profiles and the startup profile:

- Investor embedding: $\mathbf{h}_i = \text{NodeEncoder}(P_i)$.
- Startup embedding: $\mathbf{h}_{v_{target}} = \text{NodeEncoder}(\mathcal{S})$.

These embeddings form the input graph $\mathcal{G}_{\text{embed}}$ for VGAT to infer interaction edges:

$$\mathcal{E}_{\text{interact}} = \text{VGAT}(\mathcal{G}_{\text{embed}}). \quad (11)$$

These edges define the peer influence topology for the second decision round.

Agents revise their decisions by incorporating insights from interaction neighbors. Specifically, agent A_i updates $D_i^{(0)}$ to $D_i^{(1)}$:

$$D_i^{(1)} = \text{LLM}\left(\left(D_i^{(0)}, \left\{D_j^{(0)}, P_j \mid (A_i, A_j) \in \mathcal{E}_{\text{interact}}\right\}\right) \mid (\mathcal{S}, P_i)\right). \quad (12)$$

The overall investment confidence score C_{invest} is defined as the proportion of agents opting to invest:

$$C_{\text{invest}} = \frac{1}{k} \sum_{i=1}^k \delta_i, \quad (13)$$

where $\delta_i = 1$ if $D_i^{(1)}$ is an investment decision, and 0 otherwise. This score aggregates individual assessments and collective peer influence.

5 Experiments

5.1 Datasets

PitchBook. We compile a comprehensive global venture capital dataset from PitchBook² covering the period from 2005 to November 2023. The dataset includes 263,729 startups and 1,014,157 individuals, with rich information on investment histories, demographic backgrounds, and detailed company profiles. For evaluation, we focus on 2,507 startups that received their initial funding between September 2021 and November 2022. The task is to determine whether a startup secures follow-on financing within one year of its initial funding round, resulting in 533 positive samples and 1,974 negative samples. To avoid look-ahead bias, all startup profiles are constructed exclusively using information available prior to the initial funding event.

Crunchbase. Besides the private dataset, we used an open-source 2013 Crunchbase snapshot³ as a supplementary experiment. We report results in Appendix E.

5.2 Evaluation Metrics

While standard classification metrics (e.g., F1 score) are reported, we prioritize Average Precision at K (AP@K) to align with the practical VC objective of identifying top-tier opportunities from a large candidate pool. In our proposed SimVC-CAS, to reflect real-world co-investment dynamics, a startup

¹<https://huggingface.co/jinaai/jina-colbert-v2>. We selected this model for its superior long-context processing capabilities.

²<https://pitchbook.com/>

³<https://github.com/datahoarder/crunchbase-october-2013>

Method	AP@10	AP@30	Precision	Recall	F1
Random	—	—	19.98	19.02	19.49
BERT Fusion	24.67	25.33	23.63	24.95	24.27
SHGMNN	25.41	26.22	20.51	82.37	32.97
GST	25.71	27.86	21.75	83.54	34.51
GNN-RAG	27.53	26.17	22.81	71.10	34.54
SSFF	30.02	27.26	23.23	69.41	34.81
SimVC-CAS	37.52	30.83	24.19	74.07	36.47

Table 2: Performance comparison on the test set.

is considered positively predicted if at least one investor agent chooses to invest; rankings are then determined by the confidence score C_{invest} , mirroring how investment decisions aggregate in practice. Following prior work [Sharchilev *et al.*, 2018; Zhang *et al.*, 2021; Lyu *et al.*, 2025], we use Precision at K (P@K) to measure the success rate among the top-K ranked startups. To ensure temporal robustness, AP@K is computed by averaging P@K across monthly rolling windows. This metric directly quantifies the model’s effectiveness in shortlisting high-potential investments over time.

5.3 Implementation Details

We employ GPT-3.5⁴ (knowledge cutoff: Sep 2021) on a test dataset collected September 2021 to **ensure strict temporal isolation and prevent data leakage**. For reproducibility, the temperature is fixed at 0. **The number of candidate investors k is set to 10, which we empirically found to balance performance gains and computational efficiency (detailed analysis in Appendix A)**. All reported results represent the average of five independent runs with random investor sampling.

5.4 Baselines

To evaluate the effectiveness of our proposed method, we benchmark it against six diverse baselines: **Random**, which predicts based on historical success rates; **BERT Fusion** [Maarouf *et al.*, 2025], which combines structured variables with BERT-encoded unstructured text for classification; **SHGMNN** [Zhang *et al.*, 2021], a heterogeneous graph neural network that learns representations end-to-end; **GST** [Lyu *et al.*, 2025], a spatiotemporal graph model that employs incremental graph self-attention to capture dynamic node representations; **GNN-RAG** [Mavromatis and Karypis, 2024], which synergizes GNN embeddings with RAG-retrieved shortest paths as context for LLMs; and **SSFF** [Wang *et al.*, 2025b], a multi-agent framework that, unlike our interactive approach, analyzes dimensions in isolation before aggregating insights via a single centralized decision-maker.

5.5 Experimental Results

The main experimental results on the PitchBook dataset are presented in Table 2. The proposed model significantly outperforms baselines across all evaluation metrics except recall, demonstrating its strong effectiveness in predicting the early-stage startup success. In particular, our SimVC-CAS

⁴<https://platform.openai.com/docs/models/gpt-3.5-turbo>

Method	AP@10	AP@30	F1	Decision Consistency
<i>w/o roleplay (single LLM)</i>	26.58	23.44	33.14	-
<i>w/o interaction</i>	35.43	28.04	35.62	0.68
<i>FullInteraction</i>	32.14	27.85	33.64	0.62
<i>NetworkInteraction</i>	36.43	29.28	35.87	0.67
<i>GATInteraction</i>	36.89	29.74	36.07	0.69
SimVC-CAS	37.52	30.83	36.47	0.71
SimVC-CAS (4o-mini)	38.11	31.24	37.02	0.73
SimVC-CAS (Claude 3.5)	38.27	31.79	37.39	0.74

Table 3: Ablation study and comparison of interaction methods.

achieves substantial gains on ranking-based metrics, with AP@10 improving by 25.0%, and AP@30 by 13.1% relative to the strongest baseline. **These results highlight SimVC-CAS’s superior capability to prioritize and rank startups with high success potential.**

Further comparative analysis shows that graph-based models, including both traditional GNN approaches (e.g., SHGMNN and GST) and LLM RAG-enhanced variants like GNN-RAG, consistently underperform relative to SimVC-CAS. This underscores the importance of explicitly modeling potential investor relationships surrounding startups. Additionally, the notably weaker performance of the traditional multi-agent method SSFF further **validates the advantage of incorporating multi-decision-maker interaction dynamics over single-decision mechanisms.**

Furthermore, we find that while GNN methods (such as GST) achieve high recall (83.54%) through aggressive prediction, their precision drops significantly (21.75%). SimVC-CAS demonstrates overwhelming superiority in the most decision-critical AP@10 metric (37.52% vs 25.71%, a relative improvement of 46.1%), with precision and overall F1 scores also significantly outperforming GNN methods.

Results on the public Crunchbase dataset align with our main findings, showing improvements of +6.82% in AP@10, +2.66% in AP@30, and +1.66% in F1 over the strongest baseline (see Appendix E for details).

5.6 Ablation Study

Table 3 shows ablation results. *w/o roleplay*, where a single LLM jointly processes the startup profile and all k investor resumes without simulating individual personas, suffers the most significant performance drop. **This further validates that centralized reasoning cannot substitute for the collective intelligence of diverse agents.** Similarly, *w/o interaction* removes the interaction module and underperforms the full model, demonstrating the effectiveness of our interaction design.

We further compare different interaction strategies. *FullInteraction*, where all agents communicate indiscriminately, performs worse than the non-interaction baseline (*w/o interaction*) (F1: 33.64 vs 35.62), **suggesting that unstructured communication introduces detrimental noise.** In contrast, *NetworkInteraction*, which restricts interactions to historical co-investors, outperforms both (F1: 35.87), **validating the importance of leveraging established co-investment topologies.** Finally, while *GATInteraction* im-

proves upon *NetworkInteraction*, it still lags behind our full model (F1: 36.07 vs 36.47). **This confirms that our VGAT mechanism, by explicitly modeling the startup context, captures interaction patterns more effectively than standard GNNs.**

Finally, benchmarking with upgraded backbones (GPT-4o-mini⁵, Claude 3.5 Sonnet⁶) yields consistent improvements, confirming the robustness of our framework across different LLMs. However, we acknowledge that **data leakage** in these newer models likely contributes to the gains.

5.7 Consistency with Real Investors

To evaluate whether our agents simulate real investor behavior, we introduce the **Decision Consistency** metric. For all investor-startup pairs, this metric measures the alignment between simulated and actual decisions, defined as the percentage of simulated decisions matching ground truth, for which the random baseline is 0.50.⁷ Despite potential noise where investors simply missed unseen opportunities rather than rejecting them, this metric validates that our agents capture specific preferences beyond chance.

As shown in Table 3, SimVC-CAS achieves the **highest Decision Consistency (0.71)**, outperforming *w/o interaction* (0.68) and the random baseline (0.50). The failure of *FullInteraction* (0.62) confirms that **our selective, graph-based interaction design is essential** for modeling real VC dynamics.

5.8 Decision Benefit from Modeling Interactions

To quantify the marginal contribution of the collective interaction mechanism, we introduce **Confidence Alignment Gain (ΔC)**, which captures whether interaction steers predictions closer to ground truth:

$$\Delta C = (C_{\text{invest}}^{(1)} - C_{\text{invest}}^{(0)}) \cdot (2y - 1), \quad y \in \{0, 1\}, \quad (14)$$

where $C_{\text{invest}}^{(1)}$ and $C_{\text{invest}}^{(0)}$ are confidence scores after and before interaction, and y is the ground-truth label. The term $(2y - 1)$ ensures that $\Delta C > 0$ indicates improved alignment with ground truth—higher confidence for successful startups or lower confidence for failed ones.

We investigate the correlation between this gain and startup network centrality (Figure 3). We calculate the PageRank score for each startup in the test set based on the investor-startup bipartite graph and compute the mean ΔC across PageRank bins. A linear regression with 95% confidence interval reveals a clear trend: confidence alignment gain increases with PageRank. A higher PageRank indicates that the startup is embedded in a denser VC network, making it more critical to capture complex investor interaction dynamics. The trend validates that **for startups embedded in active investment networks, our interaction design is effective.**

⁵<https://platform.openai.com/docs/models/gpt-4o-mini>

⁶<https://docs.anthropic.com/en/docs/about-claude/models>

⁷We use a balanced dataset (all positive samples + equal random negatives) to calculate Decision Consistency. In raw VC data, “Invest” decisions are rare, so consistency is dominated by “Pass” decisions—allowing trivial “always Pass” strategies to score high. Balancing forces the metric to equally weight the agent’s ability to identify “Invest” opportunities and correctly “Pass”.

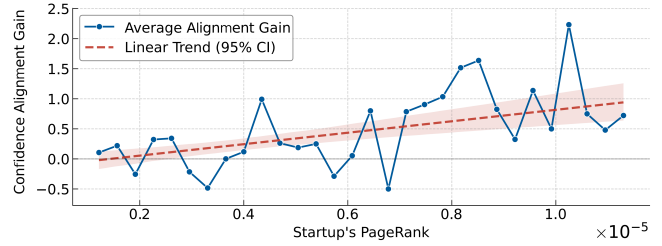


Figure 3: Impact of startup network centrality on collective interaction gain.

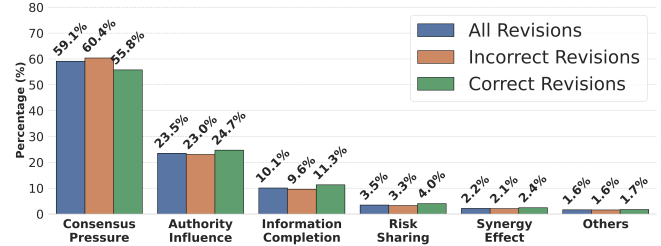


Figure 4: Reasons for decision revisions across groups.

5.9 Why Change Decisions After Interaction?

To understand what drives decision changes, we prompt agents who revised their decisions after interaction to select from predefined reasons. As shown in Figure 4 (bars in blue), **Consensus Pressure (59.1%)**, **Authority Influence (23.5%)**, and **Information Completion (10.1%)** emerge as the most frequently cited factors. We further distinguish between correct revisions (changes aligning with ground truth, in green) and incorrect revisions (orange). Notably, **correct revisions rely less on conforming to majority opinions** from surrounding agents (55.8% vs 60.4%), but **place more weight on deferring to established investors** (24.7% vs 23.0%) and **incorporating new factual details** through the network (11.3% vs 9.6%), suggesting that **effective decisions prioritize expert opinions and facts from the network over surrounding consensus** [Sørensen, 2007; Banerjee, 1992]. See Appendix F for classification details.

6 Conclusion

From the perspective of investor groups, this study proposes SimVC-CAS, a collective agent system designed to simulate the decision-making dynamics of venture capital. Unlike traditional approaches that overlook the investor collective effect, SimVC-CAS explicitly models the interactions among investors during the investment decision-making process. Our experiments demonstrate that this collective modeling approach achieves significant performance improvements and reveals how network environments shape decision quality, particularly for network-central startups. The SimVC-CAS framework is designed to be general and may extend to other complex group decision-making settings.

Appendices

A Effect of Candidate Investor Scale (k)

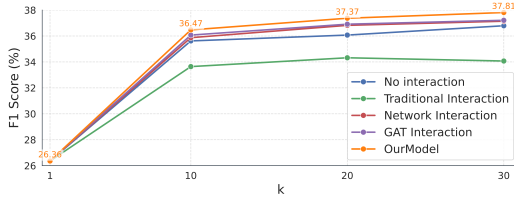


Figure 5: F1 scores under different candidate investor scales.

We evaluate the sensitivity of our model to the number of candidate investors $k \in \{1, 10, 20, 30\}$. As Figure 5 shows, performance improves with increasing k , but the marginal gains diminish beyond $k = 10$. Compared with $k = 1$, using $k = 10$ improves F1 by about 8%, demonstrating the benefit of multiple perspectives while maintaining computational efficiency. Therefore, $k = 10$ is used in all experiments as a cost-effective trade-off.

B Prompt Display

Figure 6 shows the prompts used for both initial investment and interactive revision decisions.

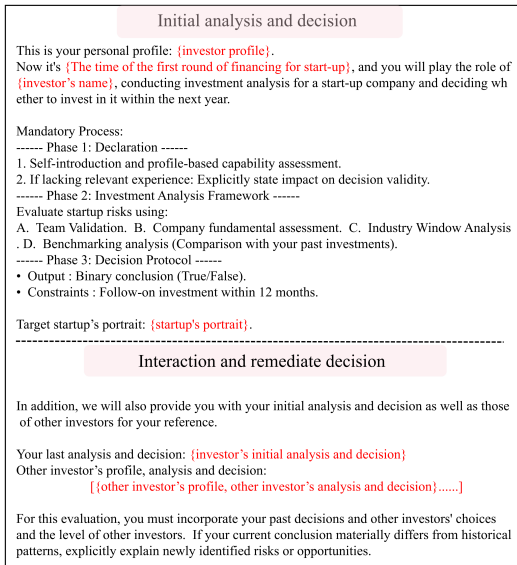


Figure 6: Prompt display.

C Consistency Analysis

Since large language models may produce varying results across runs, we evaluated the consistency of SimVC-CAS by repeating experiments 20 times on 20 randomly selected startups. As shown in Figure 7, only a negligible number of samples yielded different results across runs, demonstrating the strong consistency of our model.

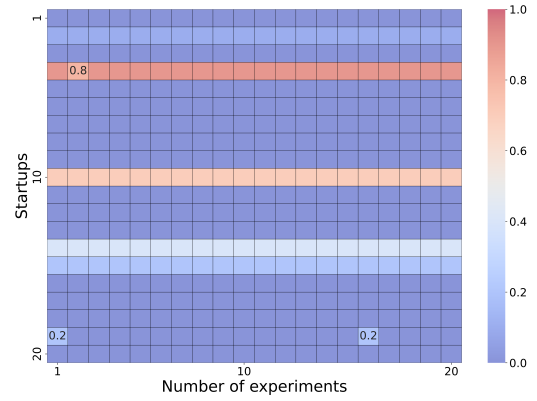


Figure 7: The consistency experiment results.

D VGAT Training Data Construction

The VGAT models investor interaction patterns jointly influenced by startup characteristics, investor attributes, and co-investment network topology. Since direct investor communication data is unobservable, we use future co-investment relationships as proxy supervision signals.

To avoid overlap with the main evaluation period, VGAT is trained and evaluated using startups whose first financing dates precede the main test period. The data are chronologically split by startups' first financing dates: training set from Oct. 2016 to Sep. 2019, validation set from Oct. 2019 to Mar. 2020, and test set from Apr. 2020 to Sep. 2020. For each eligible startup, candidate investors are sampled with $k \in \{10, 20, 30\}$, and an edge is labeled positive if the corresponding two investors co-invest within the following year; otherwise, it is labeled negative. Thus, the latest future co-investment supervision window ends on September 1, 2021, before the main evaluation period, whose initial funding events occur after September 1, 2021 and span September 2021 to November 2022. Networks with extremely low positive-to-negative ratios (< 0.05) are discarded to reduce uninformative sparsity. The final dataset includes 1,992/247/222 train/validation/test networks, with an average positive-to-negative ratio of 0.2062.

E Crunchbase Public-Dataset Evaluation

We further evaluate on the public Crunchbase 2013 snapshot, which covers investment records from 2005 to 2013. For evaluation, we focus on startups that received seed/angel funding between 2012 and 2013. The task is to determine whether these startups raise Series A funding within one year of their initial funding round, yielding 4,375 labeled samples. To avoid look-ahead bias, for each test case, company profiles, investor profiles, and investment relationship contexts are constructed exclusively using information available prior to its seed/angel funding event.

SimVC-CAS outperforms all baselines by +6.82% AP@10, +2.66% AP@30, and +1.66% F1, consistent with the main PitchBook results. Performance is lower due to the smaller scale and less comprehensive information of the public snapshot.

Start-up: ScienceIO
Investor: Austin Rief
Self-introduction: I'm Austin Rief, co-founder and Chief Operating Officer of Morning Brew. My experience mainly revolves around establishing and expanding enterprises, especially in the fields of news briefs and digital media. I have also been involved
Independent analysis and decision-making: The founding team of ScienceIO includes co-founders Lara Yuan, Will Manidis and Gaurav Kaushik. Lara Yuan graduated from Princeton University, which indicates that she has a strong academic background. The experiences and educational backgrounds of Will Manidis and Gaurav Kaushik were not elaborated, but having a co-founder with a good academic background is a positive indicator. However, due to the lack of detailed information about their previous entrepreneurial or industry-specific experiences, I am concerned about whether they can handle the complexity of the medical technology field..... However, in the fierce competition, as established enterprises consolidate their positions, the window for new entrants is narrowing..... In my investment experience, I have collaborated with companies in the medical and technology sectors, such as Foundation Medicine, which focuses on cancer diagnosis. Basic medicine has become a leader in this field. When there is a suitable product market and implementation, it proves the potential for the success of medical technology..... Based on team analysis, industry window analysis and benchmark comparison, I think investing in ScienceIO is a mistake. The team lacks sufficient validation in terms of experience. With the increasingly fierce market competition, enterprises must clearly define their differentiation strategies if they want to succeed. Final investment decision: **False**
Final decision after interaction: After making the initial decision, I had the opportunity to learn the views of several other investors, including Ankur Nagpal and Scott Belsky, who all decided to invest in ScienceIO. Ankur pointed out that although the team's experience might be limited, their academic background is very strong, and innovations in the field of medical technology often come from teams with a solid academic foundation. Scott emphasized the huge potential of the biomedical data analysis market, believing that even in a highly competitive environment, companies with a clear value proposition can still gain a considerable market share. Joshua Buckley also sharedThrough the viewpoints of other investors. I realized that I might have focused too much on the team's lack of experience and failed to fully consider their academic background and innovative potential. Although experience is important, in the field of medical technology, a profound academic background and a deep understanding of problems are equally crucial. I have also rethought the essence of market competition. Although the medical technology field is indeed highly competitive, it is also a huge and constantly growing market with ample room to accommodate multiple successful players. ScienceIO focuses on the specific field of biomedical data simplification and analysis, which gives them a clear positioning. Furthermore, I re-examined my investment experience in the field of basic medicine. Looking back at the time when I invested in basic medicine, I also faced similar challenges. However, with its powerful products and clear market strategy, it has successfully established a leading position in the market. Final investment decision: **True**

Figure 8: ScienceIO: Investor decision-making and revision process.

Start-up: First Automation
Investor: Yoshihiko Kinoshita
Self-introduction: I am Yoshihiko Kinoshita, an executive at Skyland Ventures with a background in investment, particularly in the technology and software sectors. My investment experience includes various companies, such as Aidemy and Audiostock.....
Independent analysis and decision-making: The team of First Automation includes Yuya Murakami, Masao Ito and Shun Sakurai. Although I didn't work directly with Yuya or Shun, Masao Ito has a reliable track record in investing in various technology companies, including Aidemy. The presence of experienced investors can be seen as a positive signal, but the core team appears to lack substantial professional backgrounds in the fields of automation and hardware, which is concerning. "First Automation" was established in 2020 and is in a relatively early stage in the automation industry. In the post-pandemic era, as enterprises pursue efficiency improvement, the market for automated services is growing. However, the company's classification as "other hardware" may indicate that its market positioning is not clear enough, which could limit its growth potential. When comparing "First Automation" with the companies I have invested in (such as Aidemy and Cover Corporation), I find that my successful investments are mainly concentrated in the fields of software and educational technology rather than hardware. For instance, Aidemy has achieved success in the field of educational software, which is quite different from the hardware automation services that "First Automation" focuses on. This difference in industry focus might limit the applicability of my previous investment insights to this startup. Based on the limited experience in hardware investment, the core team's insufficient expertise in automation and hardware, and the ambiguity of the market positioning of "First Automation", I initially believe that the investment decision should be negative. Final investment decision: **False**
Final decision after interaction: By comparing with the viewpoints of other investors, I have gained a more comprehensive perspective: Masao Ito came to the conclusion that investing in First Automation is definite (True), citing the existence of credible investors and the growth potential of the automation industry. However, he also admitted that further due diligence was needed on the core team. Shun Sakurai reached a negative (False) conclusion, emphasizing the core team's insufficient expertise in automation and hardware fields and the specific market conditions of the automated services. The divergence in investors' conclusions highlights a key risk: the uncertainty surrounding the core team's capabilities and the specific market dynamics of the automation services industry. Although Masao Ito saw potential based on overall market growth, Shun Sakurai and I were both concerned about the team's limited professional background in automation hardware and unclear market positioning. This indicates that although there may be opportunities in the field of automation, the execution risk remains high due to the team's inadequate industry-specific experience. Taking everything into consideration, my decision is consistent with Shun Sakurai's assessment, emphasizing caution due to the team's insufficient automation hardware expertise and my limited experience in the hardware field. Therefore, regarding the investment in First Automation, I maintain the final conclusion of negative (False). Final investment decision: **False**

Figure 9: First Automation: Investor decision-making and revision process.

Method	AP@10	AP@30	Prec.	Rec.	F1
BERT Fusion	25.57	25.26	23.54	48.10	31.61
SHGMNN	26.38	26.13	18.85	64.82	29.12
GST	26.54	26.32	19.64	68.45	30.48
GNN-RAG	27.12	25.26	20.88	58.34	30.70
SSFF	27.36	25.41	22.15	51.92	31.05
SimVC-CAS	34.18	28.98	24.05	54.12	33.27

Table 4: Performance on public Crunchbase data.

F Decision Revision Reason Classification

Agents whose decisions change after peer interaction are classified into six categories: *Information Completion*, *Authority Influence*, *Consensus Pressure*, *Synergy Effect*, *Risk Sharing*, *Others*.

For clarity, the six categories are defined as follows: **Information Completion** refers to revising a decision after ac-

quiring previously missing market or team information; **Authority Influence** refers to being swayed by the judgment of a senior or lead investor; **Consensus Pressure** refers to adjusting a decision upon observing that the majority holds a contrary opinion; **Synergy Effect** refers to investing due to expected complementary resources from co-investors; **Risk Sharing** refers to investing after others join, thereby reducing perceived individual risk; and **Others** covers reasons not captured above.

We used GPT-4o-mini to assign each agent's self-reported revision rationale to one of the six predefined categories.

G Explainability Display

As shown in Figures 8 and 9, we present two examples (ScienceIO and First Automation) to illustrate the interpretability of SimVC-CAS prediction results. In each example, an investor first introduces themselves, presents their independent analysis and initial decision, then revises it after interacting with peers.

Acknowledgements

This project is supported by the Shanghai Soft Science Research Program from the Science and Technology Commission of Shanghai Municipality (No.25692108200) and National Natural Science Foundation of China (No.72302148).

References

- [Argyle *et al.*, 2023] Lisa P Argyle, Ethan C Busby, Nancy Fulda, Joshua R Gubler, Christopher Rytting, and David Wingate. Out of one, many: Using language models to simulate human samples. *Political Analysis*, 31(3):337–351, 2023.
- [Banerjee, 1992] Abhijit V Banerjee. A simple model of herd behavior. *The Quarterly Journal of Economics*, 107(3):797–817, 1992.
- [Bargagli-Stoffi *et al.*, 2021] Falco J Bargagli-Stoffi, Jan Niederreiter, and Massimo Riccaboni. Supervised learning for the prediction of firm dynamics. In *Data Science for Economics and Finance: Methodologies and Applications*, pages 19–41. Springer International Publishing Cham, 2021.
- [Cassar, 2004] Gavin Cassar. The financing of business startups. *Journal of business venturing*, 19(2):261–283, 2004.
- [Dellermann *et al.*, 2021] Dominik Dellermann, Nikolaus Lipusch, Philipp Ebel, Karl Michael Popp, and Jan Marco Leimeister. Finding the unicorn: Predicting early stage startup success through a hybrid intelligence method. *arXiv preprint arXiv:2105.03360*, 2021.
- [Diamond and Rajan, 2001] Douglas W Diamond and Raghuram G Rajan. Banks, short-term debt and financial crises: theory, policy implications and applications. In *Carnegie-Rochester conference series on public policy*, volume 54, pages 37–71. Elsevier, 2001.
- [Elfring and Hulsink, 2007] Tom Elfring and Willem Hulsink. Networking by entrepreneurs: Patterns of tie—formation in emerging organizations. *Organization studies*, 28(12):1849–1872, 2007.
- [Goldstein *et al.*, 2025] Itay Goldstein, Yan Xiong, and Liyan Yang. Information sharing in financial markets. *Journal of Financial Economics*, 163:103967, 2025.
- [Gompers *et al.*, 2020] Paul A Gompers, Will Gornall, Steven N Kaplan, and Ilya A Strebulaev. How do venture capitalists make decisions? *Journal of Financial Economics*, 135(1):169–190, 2020.
- [Gompers, 1995] Paul A Gompers. Optimal investment, monitoring, and the staging of venture capital. *The Journal of Finance*, 50(5):1461–1489, 1995.
- [Han and Yang, 2013] Bing Han and Liyan Yang. Social networks, information acquisition, and asset prices. *Management Science*, 59(6):1444–1457, 2013.
- [He *et al.*, 2025] Junda He, Christoph Treude, and David Lo. Llm-based multi-agent systems for software engineering: Literature review, vision, and the road ahead. *ACM Transactions on Software Engineering and Methodology*, 34(5):1–30, 2025.
- [Hochberg *et al.*, 2007] Yael V Hochberg, Alexander Ljungqvist, and Yang Lu. Whom you know matters: Venture capital networks and investment performance. *The journal of finance*, 62(1):251–301, 2007.
- [Kaplan and Strömberg, 2004] Steven N Kaplan and Per Strömberg. Characteristics, contracts, and actions: Evidence from venture capitalist analyses. *The Journal of Finance*, 59(5):2177–2210, 2004.
- [Kaplan *et al.*, 2009] Steven N Kaplan, Berk A Sensoy, and Per Strömberg. Should investors bet on the jockey or the horse? evidence from the evolution of firms from early business plans to public companies. *The Journal of Finance*, 64(1):75–115, 2009.
- [Kerr *et al.*, 2014] William R Kerr, Ramana Nanda, and Matthew Rhodes-Kropf. Entrepreneurship as experimentation. *Journal of Economic Perspectives*, 28(3):25–48, 2014.
- [Kim and Lee, 2022] Dohyeon Kim and Su Yong Lee. When venture capitalists are attracted by the experienced. *Journal of Innovation and Entrepreneurship*, 11(1):31, 2022.
- [Krishna *et al.*, 2016] Amar Krishna, Ankit Agrawal, and Alok Choudhary. Predicting the outcome of startups: less failure, more success. In *2016 IEEE 16th international conference on data mining workshops (ICDMW)*, pages 798–805. IEEE, 2016.
- [Lee *et al.*, 2015] Insuk Lee, Eiru Kim, and Edward M Marcotte. Modes of interaction between individuals dominate the topologies of real world networks. *PloS one*, 10(3):e0121248, 2015.
- [Lerner, 1994] Joshua Lerner. The syndication of venture capital investments. *Financial Management*, 23(3):16–27, 1994.
- [Li *et al.*, 2025] Hao Li, Ruoyuan Gong, and Hao Jiang. Political actor agent: Simulating legislative system for roll call votes prediction with large language models. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 388–396, 2025.
- [Liu, 2017] Ying Liu. Why do large investors disclose their information? *Swiss Finance Institute Research Paper*, (18-17), 2017.
- [Luo *et al.*, 2025] Junyu Luo, Weizhi Zhang, Ye Yuan, Yusheng Zhao, Junwei Yang, Yiyang Gu, Bohan Wu, Binqi Chen, Ziyue Qiao, Qingqing Long, et al. Large language model agent: A survey on methodology, applications and challenges. *arXiv preprint arXiv:2503.21460*, 2025.
- [Lyu *et al.*, 2025] Shiwei Lyu, Xiaofeng Li, Suting Hong, Qing Ke, Jinjie Gu, Kunpeng Zhang, and Haipeng Zhang. Help me screen: Analyzing and predicting the success of start-ups in dynamic venture capital networks. *ACM Trans. Intell. Syst. Technol.*, 16(6), October 2025.
- [Maarouf *et al.*, 2025] Abdurahman Maarouf, Stefan Feuerriegel, and Nicolas Pröllochs. A fused large language model for predicting startup success. *European Journal of Operational Research*, 322(1):198–214, 2025.

- [Mavromatis and Karypis, 2024] Costas Mavromatis and George Karypis. Gnn-rag: Graph neural retrieval for large language model reasoning. *arXiv preprint arXiv:2405.20139*, 2024.
- [Miloud *et al.*, 2012] Tarek Miloud, Arild Aspelund, and Mathieu Cabrol. Startup valuation by venture capitalists: an empirical study. *Venture Capital*, 14(2-3):151–174, 2012.
- [Pei *et al.*, 2026] Haoyu Pei, Zhongyang Liu, Xiangyi Xiao, Xiaocong Du, Suting Hong, Kunpeng Zhang, and Haipeng Zhang. The gaining paths to investment success: Information-driven llm graph reasoning for venture capital prediction, 2026.
- [Qian *et al.*, 2025] Chen Qian, Zihao Xie, YiFei Wang, Wei Liu, Kunlun Zhu, Hanchen Xia, Yufan Dang, Zhuoyun Du, Weize Chen, Cheng Yang, Zhiyuan Liu, and Maosong Sun. Scaling large language model-based multi-agent collaboration. In Y. Yue, A. Garg, N. Peng, F. Sha, and R. Yu, editors, *International Conference on Learning Representations*, volume 2025, pages 41488–41505, 2025.
- [Sahlman, 1990] William A Sahlman. The structure and governance of venture-capital organizations. *Journal of Financial Economics*, 27(2):473–521, 1990.
- [Sevilla-Bernardo *et al.*, 2022] Javier Sevilla-Bernardo, Blanca Sanchez-Robles, and Teresa C. Herrador-Alcaide. Success factors of startups in research literature within the entrepreneurial ecosystem. *Administrative Sciences*, 12(3), 2022.
- [Sharchilev *et al.*, 2018] Boris Sharchilev, Michael Roizner, Andrey Rumyantsev, Denis Ozornin, Pavel Serdyukov, and Maarten De Rijke. Web-based startup success prediction. In *Proceedings of the 27th ACM international conference on information and knowledge management*, pages 2283–2291, 2018.
- [Song *et al.*, 2008] Michael Song, Ksenia Podoyntsyna, Hans Van Der Bij, and Johannes IM Halman. Success factors in new ventures: A meta-analysis. *Journal of product innovation management*, 25(1):7–27, 2008.
- [Sørensen, 2007] Morten Sørensen. How smart is smart money? a two-sided matching model of venture capital. *The Journal of Finance*, 62(6):2725–2762, 2007.
- [Spence, 1973] Michael Spence. Job market signaling. *The Quarterly Journal of Economics*, 87(3):355–374, 1973.
- [Stuart *et al.*, 1999] Toby E Stuart, Ha Hoang, and Ralph C Hybels. Interorganizational endorsements and the performance of entrepreneurial ventures. *Administrative science quarterly*, 44(2):315–349, 1999.
- [Ünal, 2019] Cemre Ünal. Searching for a unicorn: A machine learning approach towards startup success prediction. 2019.
- [Union, 2019] OE Union. The missing entrepreneurs 2019: Policies for inclusive entrepreneurship, 2019.
- [Wang *et al.*, 2025a] Haochun Wang, Sendong Zhao, Jingbo Wang, Zewen Qiang, Bing Qin, and Ting Liu. Beyond frameworks: Unpacking collaboration strategies in multi-agent systems. In Wanxiang Che, Joyce Nabende, Ekaterina Shutova, and Mohammad Taher Pilehvar, editors, *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 21361–21375, Vienna, Austria, July 2025. Association for Computational Linguistics.
- [Wang *et al.*, 2025b] Xisen Wang, Yigit Ihlamur, and Fuat Alican. Ssf: Investigating llm predictive capabilities for startup success through a multi-agent framework with enhanced explainability and performance, 2025.
- [Yu *et al.*, 2024] Yangyang Yu, Zhiyuan Yao, Haohang Li, Zhiyang Deng, Yuechen Jiang, Yupeng Cao, Zhi Chen, Jordan W Suchow, Zhenyu Cui, Rong Liu, et al. Fincon: A synthesized llm multi-agent system with conceptual verbal reinforcement for enhanced financial decision making. *Advances in Neural Information Processing Systems*, 37:137010–137045, 2024.
- [Zhang *et al.*, 2021] Shengming Zhang, Hao Zhong, Zixuan Yuan, and Hui Xiong. Scalable heterogeneous graph neural networks for predicting high-potential early-stage startups. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, pages 2202–2211, 2021.
- [Zhang *et al.*, 2024a] Erhan Zhang, Xingzhu Wang, Peiyuan Gong, Yankai Lin, and Jiaxin Mao. Usimagent: Large language models for simulating search users. In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 2687–2692, 2024.
- [Zhang *et al.*, 2024b] Wentao Zhang, Lingxuan Zhao, Hao-chong Xia, Shuo Sun, Jiaze Sun, Molei Qin, Xinyi Li, Yuqing Zhao, Yilei Zhao, Xinyu Cai, et al. A multimodal foundation agent for financial trading: Tool-augmented, diversified, and generalist. In *Proceedings of the 30th acm sigkdd conference on knowledge discovery and data mining*, pages 4314–4325, 2024.
- [Zhang *et al.*, 2025] Qiang Zhang, Keyan Ding, Tianwen Lv, Xinda Wang, Qingyu Yin, Yiwen Zhang, Jing Yu, Yuhao Wang, Xiaotong Li, Zhuoyi Xiang, et al. Scientific large language models: A survey on biological & chemical domains. *ACM Computing Surveys*, 57(6):1–38, 2025.