國立臺灣大學管理學院資訊管理學研究所

碩士論文

Department of Information Management

College of Management

National Taiwan University

Master's Thesis

利用融資訊息和圖神經網路進行新創創投鏈接推薦

Finding the Perfect Match: Leveraging Funding Information for Startup-Venture Capital Link Recommendation Using Graph Neural Networks

劉品君

Pin-Jun Liu

指導教授：魏志平 博士

Advisor: Chih-Ping Wei, Ph.D.

中華民國 113 年 7 月

July 2024

doi:10.6342/NTU202404115

致謝

兩年碩士生涯倏忽即逝，過程中時常體認到自己的不足，又激發想要讓研究更臻完美的意志，或許這就是學術研究的迷人之處吧！

首先，誠摯感謝我的指導教授魏志平博士，魏老師從大學專題到碩士論文一路作為我的學術明燈，以溫潤如玉的態度指引大家在嚴謹的研究中更上層樓。魏老師的學術熱忱與圓融處世是我一路以來學習的標竿。老師以熱忱與嚴謹投身學術，以寬容和關懷提攜後進，在我遇見困境茫然之時慷慨伸出援手，成為汪洋中的巍峨燈塔。期許未來自己有能力協助後進時，能將魏老師的恩情傳承下去。接著，感謝方思涵學姊在新創相關資料集的統整與分析，讓我能奠基於此，進一步發展建模。過去一年與學姊日夜相繼討論資料與發展研究的情景歷歷在目，成為我的學術養分，未來也將持續推進這些研究。最後，感謝兩位口試委員—楊錦生老師及胡雅涵老師撥空參與論文口試，提供寶貴的建議，使我的研究更加完整。

感謝虹鈞姊姊安排實驗室的大小事務，讓我們能沒有後顧之憂地專注在研究之上。感謝在台大資管一起奮鬥的同學們，互相激勵扶持，在來回討論中獲取靈感與啟發，很高興能跟大家一起抵達求學路上的里程碑。感謝實驗室的學長姐、學弟妹，在學術討論及實驗室運作等諸多面向，給予不可或缺的協助。

最後，感謝我的家人們，給予我無條件的支持與關愛，包容我時常忙碌，讓我在學術之路能更有底氣勇敢前行。深深感謝您們的栽培與愛護，希望今日的我能令你們引以為榮。

劉品君 謹識

于台灣大學資訊管理學研究所

中華民國一百一十三年八月

doi:10.6342/NTU202404115

摘要

新創企業是社會中重要的創新來源，卻受困於極高的失敗率。超過三成早期倒閉的新創將失敗的主因歸咎於募資困難。創業投資家（VC）作為新創企業最重要的資金來源之一，從合適的 VC 取得融資對於新創的成功有很大幫助。然而，新創企業與創投之間存在著高度資訊不對稱，導致配對困難。本研究旨在開發一種基於深度學習的推薦系統，協助新創企業找到適合的投資者。

本研究的核心貢獻在於將融資活動的時間序列資料引入新創與 VC 的關係建模中，並以新創視角將過往一對一的鏈結預測問題視為 VC 排序問題，致力於推薦合適的投資者給新創公司。我們取用 SDC VentureXpert 資料庫 2009 年至 2020 年期間的新創企業融資資料。我們提出了一個深度學習模型，結合了靜態公司特徵、動態融資資訊以及動態 VC 共投關係，以預測新創企業與 VC 之間的投資配對可能性。

為了評估模型效能，我們採用了全類平均準確率（mAP）、平均倒數排名（MRR）和前 K 準確率（Precision@K）等指標。實驗結果顯示，我們的模型在所有指標上均優於傳統模型，證明了深度學習與時間序列資料在解決新創與 VC 匹配問題上的有效性。此外，消融實驗結果顯示，公司特徵以及過往投資者相關特徵對於提升模型效能至關重要。

本研究的發現對於新創企業和 VC 具有重要的實務意涵，通過精準的推薦系統，新創企業可以更有效地尋找到潛在投資者，VC 也能接觸更多具有潛力的投資機會。然而，本研究仍存在一些限制，例如未提供第一輪預測、資料集的完整性等，未來研究可以進一步擴大研究範圍，並整合不同資料源的資訊。

關鍵字：新創公司分析、人工智慧、機器學習、圖表示學習、鏈結預測、鏈結推薦、創投-新創公司投資鏈結預測

doi:10.6342/NTU202404115

Abstract

Startups are a critical engine of innovation but face high failure rates. Over 30% of early-stage startups attribute their failure to difficulties in securing funding. Venture capitalists (VCs) are a primary source of funding for startups, and securing investment from suitable VCs is crucial for their success. However, the information asymmetry between startups and VCs hinders effective matching. This study aims to develop a deep learning-based recommendation system to assist startups in finding suitable investors.

The core contribution of this study lies in incorporating sequential data of funding activities into the modeling of relationships between startups and VCs. By adopting a startup-centric perspective, we transform the traditional one-to-one link prediction problem into a VC ranking problem, focusing on recommending the most suitable investors for each startup. Using funding data from the SDC VentureXpert database spanning 2009 to 2020, we propose a deep learning model that combines startup profile features, funding-related features, historical investor behaviors, round information, and dynamic VC co-investment relationships to predict the likelihood of investment between startups and VCs.

To evaluate the model's performance, we employ metrics such as mean average precision (mAP), mean reciprocal rank (MRR), and precision@K. Experimental results

doi:10.6342/NTU202404115

show that our model outperforms traditional models on all metrics, demonstrating the effectiveness of deep learning and sequential data in solving the startup-VC matching problem. Moreover, ablation studies reveal that company features and historical investor-related features are crucial for improving model performance.

The findings of this study have significant practical implications for both startups and VCs. By providing an accurate recommendation system, startups can more efficiently identify potential investors, while VCs can discover more promising investment opportunities. However, the study has some limitations, such as the lack of first-round predictions and data completeness. Future research can further expand the scope and integrate information from various data sources.

Keywords: Startup Analytics, Artificial Intelligence, Machine Learning, Graph Representation Learning, Link Prediction, Link Recommendation, Venture-capital-Startup Investment Link Prediction

doi:10.6342/NTU202404115

Table of Contents

doi:10.6342/NTU202404115

doi:10.6342/NTU202404115

doi:10.6342/NTU202404115

List of Figures

doi:10.6342/NTU202404115

List of Tables

doi:10.6342/NTU202404115

Chapter 1 Introduction

1.1 Background

Startups, as nascent, often technology-driven entities, have emerged as pivotal drivers of innovation across diverse sectors. Their agility, coupled with a relentless pursuit of disruptive ideas, has catalyzed transformative changes in industries ranging from healthcare and transportation to entertainment and finance. The role of startups in fostering economic growth and creating new job opportunities cannot be overstated. They serve as incubators for groundbreaking concepts, challenging established norms, and ultimately enhancing the quality of life for individuals and societies worldwide.

However, the startup landscape is fraught with challenges that contribute to a high failure rate. Statistics underscore the precarious nature of this entrepreneurial journey. A staggering two out of three startups fail to achieve sustained success (Salamzadeh & Kawamorita Kesim, 2015), with more than half closing their doors within the first five years of operation (Gonzalez, 2017). A significant contributor to these failures is a chronic lack of funding, cited by over one-third of unsuccessful ventures (CB Insights, 2024). This financial constraint often hampers startups' ability to develop their products, attract talent, and compete effectively in the marketplace.

doi:10.6342/NTU202404115

The complexities inherent in startup financing exacerbate the challenges faced by these nascent enterprises. High levels of uncertainty surround the potential returns on investment, as startups operate in dynamic and unpredictable environments. Moreover, information asymmetry between entrepreneurs and investors creates an imbalance of power, hindering the ability of startups to communicate their value proposition effectively. The intangible nature of many startup assets, such as intellectual property and human capital, further complicates the valuation process. To compound matters, startups are particularly vulnerable to economic fluctuations, as their growth trajectories are closely tied to broader market conditions. These factors contribute to the formidable obstacles that startups encounter in securing the capital necessary for their survival and growth (Gompers & Lerner, 2004).

Venture capital (VC) stands as a cornerstone of financing for startups. The infusion of capital from VC firms is instrumental to the survival and growth of these nascent enterprises. The absence of institutional financing poses significant challenges to entrepreneurial endeavors (Fernandez, 2021). Relying solely on informal sources of capital can be a precarious proposition, often limiting a startup's ability to scale and compete effectively.

Recognizing the critical role of VC funding, researchers have explored innovative

doi:10.6342/NTU202404115

approach to optimize the matching process between startups and investors. Link recommendation in venture capital networks offers a promising avenue for identifying potential synergies between these two parties. By leveraging data-driven insights, this methodology can assist startups in identifying suitable financial partners while providing VC firms with an objective framework for investment decisions. In essence, link recommendation serves as a valuable tool for enhancing the efficiency and effectiveness of the startup funding ecosystem. For VC firms, it can serve as a counterbalance to the often subjective nature of investment decisions, offering a data-driven approach to supplement intuition and experience (Ghassemi et al., 2020).

While seeking venture capital is paramount for startup success, the process of identifying and securing the right investors is fraught with challenges. Several factors contribute to the difficulty of aligning startups with suitable VC firms.

A primary obstacle lies in the inherent information asymmetry between startups and investors. Startups often possess limited financial and operational histories, making it challenging to convey their potential value and mitigate investor concerns about risk. The intangible nature of many startup assets, such as intellectual property and human capital, further complicates the valuation process and increases uncertainty for potential investors (Gompers & Lerner, 2004). Moreover, the startup landscape is characterized by high

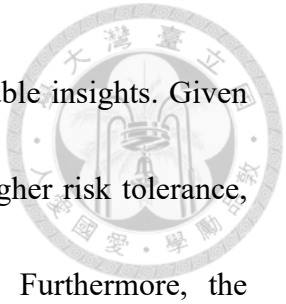
doi:10.6342/NTU202404115

volatility, with market conditions and technological advancements rapidly evolving. This dynamic environment introduces additional uncertainties, making it difficult for investors to assess the long-term prospects of these young ventures (Kerr & Nanda, 2011).

1.2 Research Motivation

Existing research has demonstrated the potential of link recommendation in bipartite startup-venture capital networks. These studies have primarily employed two main approach: similarity-based methods, which assess latent similarities between network nodes, and learning-based methods, which utilize maximum likelihood estimation to predict unobserved connections. However, a critical gap exists in the current literature. Extant studies formulate the startup-venture capital link recommendation task as either a pairwise link prediction task or a startup recommendation to VCs. Previous studies have primarily treated the startup-venture capital (VC) linking recommendation problem as a pairwise link prediction task or as a ranking problem where startups are recommended to VCs. These approach have overlooked the perspective of recommending VCs to startups, thereby missing an opportunity to better support the growth of social innovation. Pairwise link prediction, which requires comparing every startup with every VC, is computationally intensive and lacks a clear focus, making it difficult to provide startups with structured guidance on seeking investment opportunities. In contrast, ranking-based

doi:10.6342/NTU202404115



recommendations tailored to a specific party can offer more actionable insights. Given the critical role of startups in driving social innovation and their higher risk tolerance, startups are more motivated to actively seek suitable funding. Furthermore, the information asymmetry in the startup-VC market often arises from the lack of a startup's historical data. Therefore, enabling startups to proactively seek out suitable investors can significantly enhance their chances of securing the most favorable funding. As highlighted by the Harvard Business Review (Fayard et al., 2024), matching startups with suitable VCs is essential for sustaining and fostering a thriving innovation ecosystem. Beyond the task formulation, there is also room for improvement in the data used in previous models. Previous studies have relied heavily on topological features extracted from startup social networks, focusing primarily on relational information, they have overlooked the crucial role of funding activities, which unfold as sequential data. This oversight presents a significant opportunity to enhance link recommendation accuracy and provide more valuable insights into the dynamics of startup-venture capital relationships.

1.3 Research Objectives

This study proposes a novel deep learning approach integrating company profile features, funding-related features, historical investor information, round information, and

doi:10.6342/NTU202404115

topological features extracted from the venture capital co-investment networks to enhance startup-venture capital link recommendation. From the perspective of recommending VCs to startups, our proposed method aims to preserve the scalability and relational information of existing deep learning models while capitalizing on the valuable insights embedded within sequential information retrieved from funding activities. Both topological features extracted from the venture capital co-investment networks and non-topological features retrieved from funding records are utilized. Since the funding activities are sequential data in nature, the accumulated information should improve the effectiveness of recommendations in later rounds.

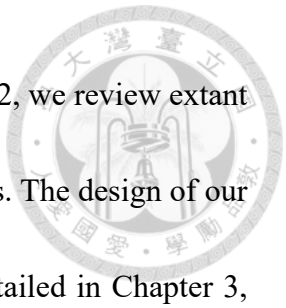
To evaluate the effectiveness of our proposed method and proposed feature sets on startup-venture capital link recommendation during different time periods, the following research questions will be addressed:

RQ1: What is the effect of our proposed method on the effectiveness of startup-venture capital link recommendation?

RQ2: What are the effects of our proposed funding-related features on the effectiveness of startup-venture capital link recommendation?

RQ3: What are the performance differentials of our proposed method over different funding rounds?

doi:10.6342/NTU202404115



The remainder of this paper is organized as follows: In Chapter 2, we review extant literature on link recommendation in startup-venture capital networks. The design of our proposed startup-venture capital link recommendation method is detailed in Chapter 3, followed by our empirical evaluation design and important evaluation results in Chapter 4. Chapter 5 provides a summary and discussion of our contributions, together with limitations and future research directions.

doi:10.6342/NTU202404115

Chapter 2 Literature Review

In this section, we review the literature relevant to this study. We review prior predictive studies on startup-venture capital link recommendation and analyze their limitations to justify the research motivation for the current study.

Link recommendation is a classic problem in social media analytics, which aims to predict missing or future relationship between entities. Approach to the problem fall into two categories: similarity-based, and learning-based (Yuliansyah et al., 2020). Similarity-based methods evaluate latent similarity between pair of nodes. While they achieve good performance on small graph, they suffer from poor scalability. Learning-based methods estimate the possibility of unobserved links by maximum likelihood method. They require more computational resources but achieve greater scalability and prediction accuracy.

2.1 Existing Startup-Venture Capital Link Prediction Studies

Within the domain of bipartite startup-venture capital networks, existing research has applied both similarity-based and learning-based approach. We will review them along four dimensions: 1) data source, 2) formulation type, 3) prediction approach, and 4) features used.

doi:10.6342/NTU202404115

2.1.1 Similarity-based approach

The similarity-based approach for link prediction in bipartite networks can be categorized into local and latent similarity methods (Kunegis, 2010). Local methods focus on immediate network neighbors, while latent methods consider the underlying structure of the entire network. Given the dynamic nature of the startup-venture capital network, local similarity methods without the assumption of triangular closure are more suitable for link prediction in this context. For instance, Zhang et al. (2015) employ a preferential attachment approach on a directed bipartite network constructed from link history collected from Crunchbase. They formulated the task as a pairwise link recommendation and adopted the preferential attachment method with the following equation

$$PA(s, v) = \frac{\mathrm{d}(s)\mathrm{d}(v)}{2(|\mathrm{E}|)},$$

where $\mathrm{d}(s)$ and $\mathrm{d}(v)$ denotes the number of neighbors of startup s and venture capital v in the startup-venture capital investment network respectively, and $|E|$ denotes the total number of edges in the startup-venture capital investment network.

2.1.2 Learning-based approach

In the learning-based approach for startup-venture capital link prediction, machine learning methods leveraging SNA metrics were brought out first. Liang and Yuan (2016) predict investor funding behavior using Crunchbase social network features. They

doi:10.6342/NTU202404115

formulated the problem as pairwise link recommendation. Machine learning methods including CART Decision Tree, SVM, Naïve Bayes were applied with pairwise scores for shortest paths, Adamic/Adar, Jaccard Coefficient, common neighbors, preferential attachment and number of shortest paths as input data.

Then, prior studies also proposed deep learning methods on startup-venture capital investment network representation. Gu et al. (2019) propose an architecture named DeepLinker, which is an extension of Graph attention network (GAT). They collect investment relationship between startups and VCs from Zero2IPO. The task is formulated as pairwise link recommendation. For each pair consists of a startup and a venture capital, the representation of both entities are retrieved. A possibility of link existence was generated based on the similarity between both representations.

After knowing the effectiveness of deep learning methods on representing investment network. Xu et al. (2023) proposed SocioLink, a framework leveraging relational information in heterogeneous startup social media knowledge graph. They constructed a knowledge graph based on social network connectivity data collected from Crunchbase. The target of the study is to recommend suitable startups for each VC. Hence, the problem was formulated as ranking among startups. The prediction model is Gradient Boosting Decision Tree (GBDT) with input data including knowledge graph

doi:10.6342/NTU202404115

representation from learning-based model, TransD, and self-defined meta-path counts.

Existing research on link recommendation within bipartite startup-venture capital networks primarily relies on data extracted from startup social networks, with a predominant focus on relational information. Consequently, the critical role of funding activities as a sequential dataset has been overlooked. Moreover, previous studies have predominantly framed the problem as either pairwise link recommendation or startup ranking for venture capitalists. While these approach offer valuable insights, formulating the problem from the perspective of ranking venture capitals for startups could provide more actionable information for entrepreneurial ventures.

doi:10.6342/NTU202404115

Table 1: Existing Studies on Startup-Venture Capital Link Recommendation

Source	Approach	Formulation	Methods	Input data	Data
(Zhang et al., 2015)	Similarity-based	Pairwise	Preferential attachment	Startup-VC investment network	Crunchbase
(Liang & Yuan, 2016)	Learning-based	Pairwise	CART Decision Tree, SVM (rbf kernel), Naïve Bayes	shortest paths, Adamic/Adar, Jaccard Coefficient, common neighbors, preferential attachment and number of shortest paths	Crunchbase
(Gu et al., 2019)	Learning-based	Pairwise	Graph attention network (GAT)	Startup-VC investment network	Zero2IPO
(Xu et al., 2023)	Learning-based	Startup ranking	TransD + Gradient Boosting Decision Tree (GBDT)	knowledge graph from startup social media, meta-path count, non-relational	Crunchbase

Chapter 3 Methodology

3.1 Design of Our Proposed Startup-Venture Capital Link Recommendation Method

This study introduces a novel deep learning architecture designed to predict and rank the likelihood of venture capital investment in startups. By leveraging fund-related and investor-related information embedded within funding activities, our proposed method aims to enhance the accuracy of link recommendations. Given a specific startup and a pool of potential investors, our model predicts the probability of each venture capital firm participating in the startup's next funding round. The resulting probability distribution is then utilized to generate a ranked list of recommended investors for the startup.

The architecture incorporates both non-topological and topological features extracted from the startup-venture capital network, treating these features as sequential data. This approach builds upon the strengths of previous deep learning models while introducing a novel perspective that emphasizes the dynamic nature of funding activities. By considering the temporal evolution of relationships within the network, our proposed method seeks to improve the predictive power of link recommendation systems.

doi:10.6342/NTU202404115

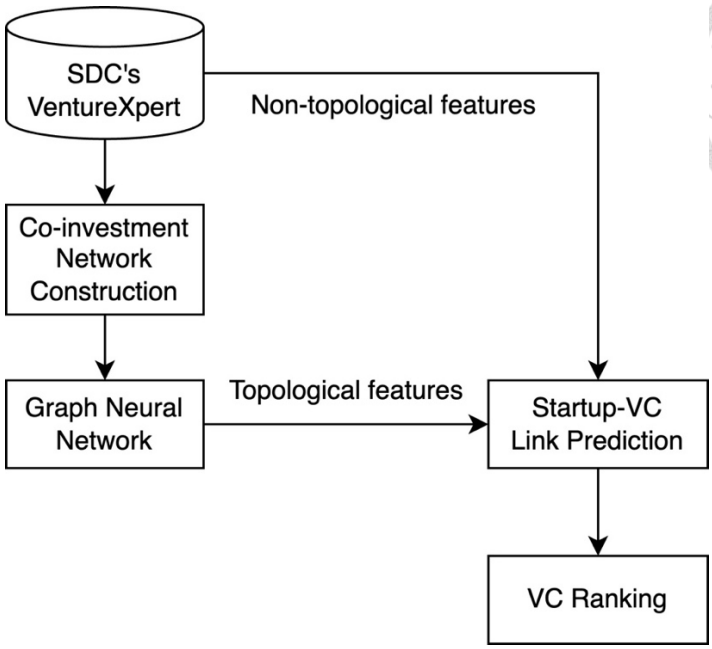


Figure 1: Process of Our Proposed Startup-VC Link Recommendation Method

3.2 Non-topological Features for Our Proposed Method

To effectively capture the characteristics of startups, investors, and their interactions, a comprehensive set of non-topological features is incorporated into our model. These features encompass various aspects of the startup and investment landscape, providing valuable insights into the dynamics of the startup-venture capital ecosystem. Features under four categories are designed, including startup profile, round information, funding-related, historical investor features.

3.2.1 Startup profile features

This category encompasses a startup's fundamental information. We consider four features that are commonly adopted in past studies: location, industry, customer type, and

doi:10.6342/NTU202404115

age (Alexy et al., 2012; Freiberg & Matz, 2023; Hoenen et al., 2014; Ko & McKelvie, 2018; Shetty & Sundaram, 2019; Talaia et al., 2016; Wise et al., 2022; Zhou et al., 2016). We slightly adjust definition of the three features to fit in our dataset.

- *World Region*: This feature indicates the continent where the focal startup is located. The values include Americas, Europe, Asia, Pacific, and Africa. By classifying startups according to their *World Region*, we can analyze regional patterns and trends in link prediction between startups and VCs, capturing how geographical factors influence investment relationships.
- *State*: The *state* feature identifies the specific US state in which the startup operates. The options are limited to the top five states with the highest number of startups in our dataset: California (CA), New York (NY), Massachusetts (MA), Texas (TX), and Pennsylvania (PA). Startups located in other states within the US are categorized as 'US-Other', while startups located outside the US are labeled as 'Other'. This feature allows for a more granular analysis within the US, where certain states like California and New York may have more robust VC networks and startup ecosystems compared to others. It helps in assessing how location within the US and internationally impacts the startup's connection with VCs.
- *SIC*: The Standard Industrial Classification (SIC) code is a numerical system used

doi:10.6342/NTU202404115

to classify industries by a specific number code. For this research, the first three digits of the SIC code are used to denote the industry of the startup. This feature allows for an understanding of the industry in which the startup operates, which can influence the types of venture capital that might be interested in the startup and how industry-specific factors affect investment decisions.

- *Customer Type*: This feature represents the primary type of customers the startup targets. The customer types can include categories such as business-to-business (B2B), business-to-consumer (B2C), business-to-government (B2G), and mixed type (All). Understanding the customer type helps in evaluating the market demand and potential revenue streams for the startup. It can also influence how VCs perceive the startup's market potential and scalability. For instance, targeting enterprise clients versus individual consumers can affect the startup's growth prospects and funding appeal.
- *Age*: The *age* feature measures the number of years since the startup's inception. It provides insights into the maturity of the startup, which can influence its likelihood of attracting VCs. Older startups may be perceived as more established and less risky, while younger startups might be seen as high-risk but with high growth potential. This feature helps in assessing how the startup's age correlates with its ability to

doi:10.6342/NTU202404115

attract VC and the typical funding stages across different ages.

3.2.2 Round information

This category captures the essential details of each funding round. These features are available at the start of each round, enabling the prediction of links for round *n* using round *n*'s information.

- *Round number*: This feature represents the ordinal number of the funding round. It is an important indicator of the stage and maturity of the startup in its funding lifecycle. As the round number increases, it often signifies that the startup has successfully navigated through previous funding stages, thereby reducing the information asymmetries between VCs and the startup itself (Hoenen et al., 2014). Higher round numbers are associated with greater transparency and trust, as the startup's track record becomes clearer and more verifiable through subsequent rounds of investment.
- *Months have passed since last round*: This feature measures the number of months that have passed since the previous funding round. It captures the temporal aspect of funding activities and can indicate the startup's funding rhythm and urgency. A longer duration since the last round may suggest either a period of steady development or potential difficulties in securing additional funding. Conversely,

doi:10.6342/NTU202404115

shorter intervals between funding rounds could imply rapid growth and aggressive expansion strategies. This temporal measure helps in understanding the dynamics and pace at which a startup is progressing and seeking additional investment.

3.2.3 Funding-related features

This category encompasses the fundraising information for each round. Larger funding amount, higher investor amounts, and more evenly distributed funding sources are indicators of successful fundraising (Hoenen et al., 2014). These features become available after the round concludes (lagged indicators). To predict links for round N, funding features from rounds up to N-1 are utilized. The funding amounts are recorded in units of a thousand USD with logarithm applied to reduce value variation.

- *Cumulative funding amount*: The *cumulative funding amount* represents the total amount of funding raised by the focal startup across all rounds. This feature provides a comprehensive view of the startup's financial strength and its ability to attract investment over time. Higher cumulative funding often indicates strong investor confidence and a robust business model. It also reflects the startup's growth trajectory and potential for scalability, making it a crucial factor in predicting future linkages with VCs.
- *Amount raised in the last round*: The *amount of capital raised in the last round* refers

doi:10.6342/NTU202404115

to the capital secured by the startup during its most recent funding round. It provides insight into the startup's current financial momentum and investor confidence. A substantial amount raised in the last round can signal to potential investors and VCs that the startup is on an upward trajectory, making it a more attractive candidate for future funding.

- *Average VCs*: The *average VCs* denotes the average number of disclosed VCs participating across all funding rounds of the startup. This feature reflects the level of interest and support the startup has received from the VC community. A higher average suggests broad-based confidence and backing from multiple investors, which can enhance the startup's credibility and stability. It also indicates the startup's ability to engage with and maintain relationships with a diverse set of VCs, which is beneficial for long-term growth and access to resources.
- *Capital Concentration Rate (CCR)*: *CCR* is a measure inspired by the Herfindahl-Hirschman Index, used to estimate the concentration of capital from VCs within a given startup. This metric provides insight into how investment capital is distributed among the VCs backing the startup. Specifically, we estimate capital concentration rate (CCR) as the sum of the squares of the capital shares of all VCs within the focal startup, as shown in the equation:

doi:10.6342/NTU202404115

$$CCR = \sum_{i=1}^{v} S_i^2,$$

where *v* is the number of VCs invested in the focal startup, and S_i is the percentage of funding amount invested by VC *i*.

The value of CCR ranges from 0 to 1, where a higher CCR indicates a higher concentration of capital among fewer VCs, suggesting that the startup relies heavily on a small number of VCs for funding. If these VCs are reputable and well-established within the industry, the high CCR could indicate a strong endorsement of the startup's potential. Conversely, a lower CCR indicates a more even distribution of capital among multiple VCs, reflecting a more diversified investment base. This feature is crucial in understanding the influence and potential control exerted by VCs within the startup, as well as the startup's reliance on specific investors.

3.2.4 Historical investor features

This category comprises aggregated features of venture capitalists who have previously invested in the target startup. Similar to funding-related features, these features become available after the round concludes. To predict links for round N, historical investor features from rounds up to N-1 are employed.

- *Average Distance from Startup to VCs*: This feature represents the average

doi:10.6342/NTU202404115

geographical distance between the startup and its investors. It measures the spatial separation between the startup's location and the locations of the VCs that have invested in it. By assessing this average distance, one can evaluate how proximity or remoteness might influence investment decisions and relationships. A shorter distance may indicate easier communication and collaboration, potentially enhancing the likelihood of investment, while a greater distance might reflect challenges in managing and supporting the startup from afar.

- *Average Age of VCs*: This feature calculates the average age of the ~~venture capitals~~ VCs who have invested in the startup. Older VCs may bring more experience and established networks to the table, which could influence their investment strategies and preferences. Conversely, younger VCs might be more inclined towards innovative and emerging sectors. Understanding the average age of VCs can help in analyzing how the experience level and generational perspectives of investors impact their investment decisions and the potential success of the startups they support.
- *Average Success Rate:* The *average success rate* is a metric that evaluates the effectiveness of VCs who have invested in the focal startup. Specifically, it calculates the average success rate of all startups that these VCs have backed over the past

doi:10.6342/NTU202404115

decade. The success rate is determined by the proportion of a VC's portfolio companies that achieve notable milestones, such as reaching the expansion stage, later stage, mergers, acquisitions, leveraged buyouts, or going public within five years of their founding. This measure reflects a VC's track record in selecting and nurturing high-potential startups. A higher success rate is indicative of a VC's ability to identify and support startups that thrive, which in turn enhances the VC's reputation. A VC's proven success in the past not only bolsters its reputation but also attracts further investment, as other investors are keen to be associated with startups backed by proven successful VCs. This increased investor interest can lead to higher capital inflows in subsequent funding rounds.

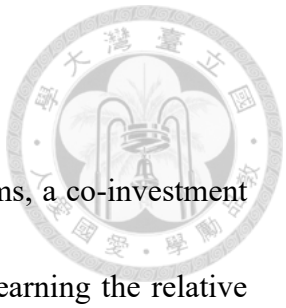
A comprehensive overview of the definitions of all features is presented in Table 2.

doi:10.6342/NTU202404115

Table 2: Non-topological Features Employed in Our Proposed Method

Category	Features	Description
Startup profile	World region	Which of the five continents is the focal startup located on?
	State	In which US state is the focal startup based?
	SIC	Industry of the focal startup (according to the first 3-digit of SIC code)
	Customer type	Customer type of the focal startup
	Age	Age of the focal startup
Round information	Round number	Number of funding rounds received by the focal startup
	Months have passed since last round	Months have passed since the last funding round
Funding-related	Cumulative funding amount	Cumulative amount raised across all funding rounds
	Amount raised in the last round	Amount raised in the last funding round
	Average VCs	Average number of VCs invested in the focal startup across all funding rounds
	Capital concentration rate	Capital concentration rate of VCs investing in the focal startup
Historical investor	Average distance from startup to VCs	Average distance from the focal startup to VCs invested in the focal startup across all funding rounds
	Average age of VCs	Average age of VCs invested in the focal startup across all funding rounds
	Average success rate	Average success rate of the VCs invested in the focal startup across all funding rounds

doi:10.6342/NTU202404115



3.3 Topological Features for Our Proposed Method

To capture the intricate relationships among venture capital firms, a co-investment network is constructed. This network serves as the foundation for learning the relative positions of VCs within a vector space. A link is established between two VCs if they have co-invested in the same startup during the same funding round. To incorporate a historical perspective, the co-investment network for a specific year t is constructed by considering all investment records within a ten-year window preceding year t.

To address the challenge of isolated nodes in the co-investment network, which could result in inadequate representations for certain VCs, we incorporate VC profile features into the model. Specifically, industrial preference, stage preference, and geographical preference are transformed into three-dimensional representations each through learnable fully connected layers. These representations serve as initial node embeddings for VCs that lack connections in the co-investment network.

Subsequently, GCN (Graph Convolutional Network) (Kipf & Welling, 2016), one of the most representative graph embedding method for homogeneous graph, is applied to refine the VC representations by considering their connections within the network. The GCN process iteratively updates the node embeddings, incorporating information from neighboring nodes and refining the learned representations. The final output is an eight-

doi:10.6342/NTU202404115

dimensional node embedding for each VC, capturing both the structural information from the co-investment network and the preference information from the VC profile features.

To ensure temporal consistency in VC embeddings, an incremental update approach is adopted (Barros et al., 2021). Embeddings from the previous year are utilized as initializations for the corresponding nodes in the subsequent year's network. The GCN process is then applied to refine these embeddings based on the new network structure. The final output is an eight-dimensional node embedding for each VC, capturing both the structural information from the co-investment network, the semantic information from the VC profile features, and the temporal dynamics of VC preferences.

3.4 Deep Learning Model for Our Proposed Method

To integrate the diverse features and capture the temporal dynamics of the startup-venture capital relationship, a deep learning architecture is proposed. The model incorporates both static and dynamic features, with the latter processed through a Long Short-Term Memory (LSTM) network to effectively handle sequential data. The LSTM captures the sequential patterns within non-profile features, enabling the model to learn from the historical evolution of startups and investors. The output of the LSTM is subsequently concatenated with startup profile features and fed into a fully connected layer to generate the final prediction.

doi:10.6342/NTU202404115

A three-dimensional learnable embedding is employed for each categorical input to mitigate the limitations of traditional encoding methods such as one-hot encoding and label encoding, which can introduce dimensionality issues or misleading ordinal relationships.

To account for the varying availability of feature information across different time periods, a differentiated input mechanism is implemented. For features that are known prior to a specific round (e.g., round information), the corresponding values for that round are directly input into the model. In contrast, for features that become available only after a round's outcome (e.g., funding-related features, historical investor features), the cumulative information up until the previous round is utilized. This approach ensures that the model effectively leverages available information while avoiding the use of future data.

doi:10.6342/NTU202404115

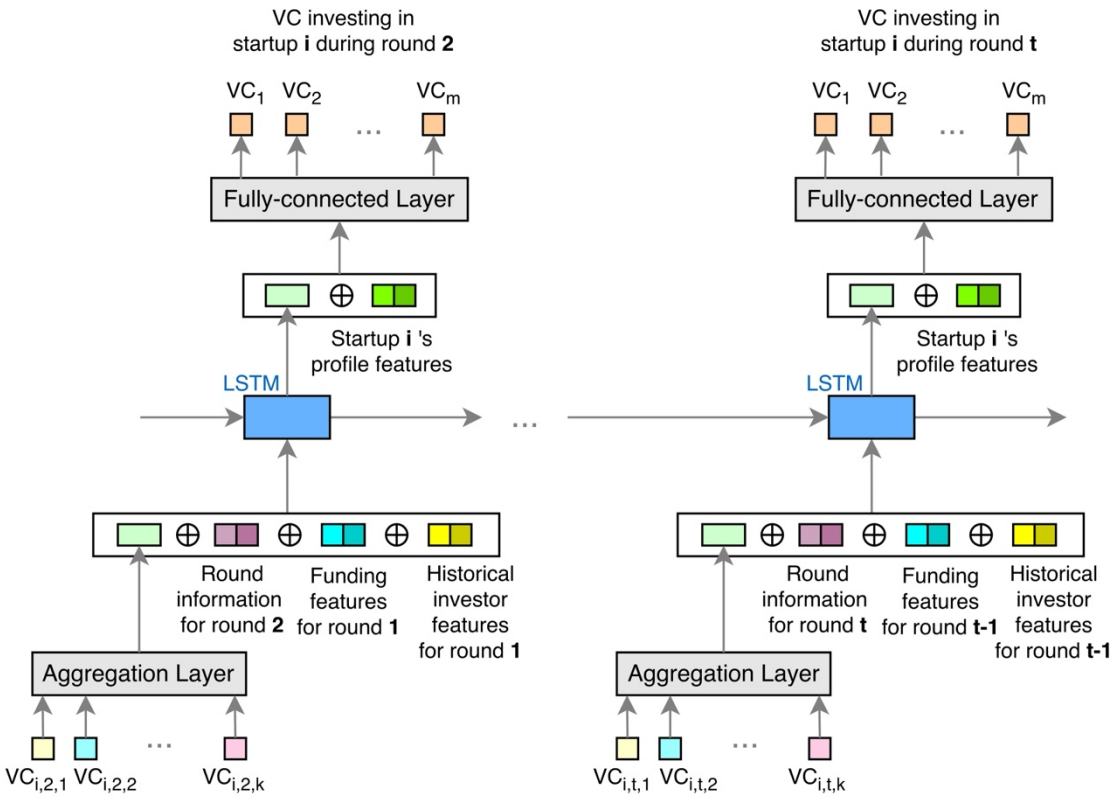


Figure 2: Overall Architecture of Our Proposed Method

doi:10.6342/NTU202404115

Chapter 4 Experiments

4.1 Data Collection

The primary data source for this study is the SDC Platinum Database – VentureXpert. A comprehensive dataset encompassing all startup funding instances between 2009 and 2020 was collected. This timeframe was selected to capture the post-financial crisis landscape and to analyze the evolution of startup-venture capital dynamics over a substantial period.

To refine the dataset for subsequent analysis, three filtering criteria were applied:

1. *Industry focus:* The study concentrates on high-tech sectors with significant innovation potential. Consequently, the dataset was filtered to include startups operating in the biotech (SIC: 283), information and communications technology (ICT, SIC: 356, 367, 366, 369), and software (SIC: 737) industries.
2. *Startup characteristics:* To ensure a homogeneous sample of early-stage startups, the dataset was further restricted to companies founded between 2009 and 2014 that secured their initial funding at the seed or early stage. Additionally, startups with a founding-year investment exceeding $10 million were excluded to minimize the inclusion of spinoffs from established corporations.
3. *Data completeness:* To maintain data integrity and facilitate sequential analysis,

doi:10.6342/NTU202404115

startups with non-continuous funding rounds were removed from the dataset.

The filtering process resulted in a dataset comprising 11,365 companies that underwent a total of 23,201 funding rounds. These rounds involved 7,445 unique venture capital firms operating within the specified timeframe of 2009 to 2020.

Table 3 provides descriptive statistics of our dataset. There are several interesting insights we can gain from the descriptive statistics. The majority of the startups in the dataset are located in the United States (US), with a notable concentration in California, the home of Silicon Valley. This geographical distribution reflects the prevalence of high-tech startups in these regions. Around nine out of ten startups within our sample belong to the software sector (SIC 737). This aligns with the overall focus on high-tech industries in the data collection process. Over half of the startups primarily target business-to-business (B2B) customers. This suggests that the companies are catering to other businesses rather than directly to end consumers. However, the number of investors each new startup can attract in a fundraising round varies greatly. While some companies received investments from as many as 43 registered VCs in a single round, the majority of startups struggle to secure even one investor. Even when considering only successful fundraising rounds, the average number of registered VCs per round is less than three, highlighting the competitive nature of the startup investment landscape.

doi:10.6342/NTU202404115

Table 3: Descriptive Statistics of Our Dataset

Features	Descriptive Statistics
World region	Americas: 3697 (67.72%); Asia: 897 (16.43%); Europe: 798 (14.62%); Pacific: 51 (0.93%); Africa: 16 (0.29%)
State	Other: 2033 (37.24%); CA: 1470 (26.93%); US-Other: 1020 (18.68%); NY: 423 (7.75%); MA: 279 (5.11%); TX: 125 (2.29%); PA: 109 (2.00%)
SIC	737: 4894 (89.65%); 283: 311 (5.70%); 367: 165 (3.02%); 366: 54 (0.99%); 357: 34 (0.62%); 369: 1 (0.02%)
Customer type	Business: 2820 (51.66%); All: 1493 (27.35%); Consumer: 1134 (20.77%); Government: 11 (0.20%); Other: 1 (0.02%)
Age	μ=4.09, σ=2.17, max=10.00, min=0.00; Missing: 0 (0.00%)
Round number	μ=3.20, σ=1.68, max=22, min=2; Missing: 0 (0.00%)
Months have passed since last round	μ=13.65, σ =11.02, max=119.00, min=0.00; Missing: 0 (0.00%)
Cumulative funding amount	μ=8.86, σ=1.69, max=16.95, min=0.26; Missing: 0 (0.00%)
Amount raised in the last round	μ=8.32, σ=1.64, max=15.52, min=0.00; Missing: 0 (0.00%)
Average VCs	μ=2.76, σ=1.67, max=43.00, min=1.00; Missing: 0 (0.00%)
Capital concentration rate	μ=0.610, σ=0.310, max=1.000, min=0.063; Missing: 1156 (9.77%)
Average distance from startup to VCs	μ=1332.04, σ=1690.68, max=11025.31, min=0.00; Missing: 465 (3.93%)
Average age of VCs	μ=17.34, σ=12.29, max=151, min=0; Missing: 314 (2.65%)
Average success rate	μ=0.634, σ=0.261, max=1.000, min=0.000; Missing: 5418 (45.78%)

doi:10.6342/NTU202404115

4.2 Benchmark Design

To evaluate the performance of our proposed model, a set of benchmark models is established for comparison. These benchmarks represent different methodological approach to the link prediction problem.

4.2.1 Similarity-based method

A modified version of the preferential attachment method is employed as a benchmark similarity-based model. Traditionally, preferential attachment is calculated as the degree of a node. However, to accommodate the ranking-based problem formulation, the preferential attachment score is normalized to a range of [0, 1] using the formula:

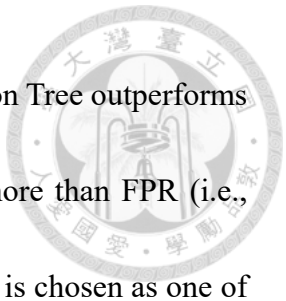
$$\mathrm{PA}_{\mathrm{s,v,t}} = \frac{\deg_{\mathrm{v,t}} - \deg_{\mathrm{min,t}}}{\deg_{max,\mathrm{t}} - \deg_{\mathrm{min,t}}},$$

where $PA_{s,v,t}$ denotes the preferential attachment score of startup s and VC v in year t; $degree_{v,t}$ is degree of VC v in the co-investment network in year t; $deg_{min,t}$ and $deg_{max,t}$ denotes the minimum and maximum degree among all VCs in the co-investment network in year t respectively.

4.2.2 Learning-based method

Previous study (Liang & Yuan, 2016) on startup-venture capital link prediction evaluates Decision Tree and Naïve Bayes with respect to TPR (true positive rate, precision), FPR (false positive rate, recall), and AUC. The evaluation result shows that

doi:10.6342/NTU202404115



Naïve Bayes outperforms Decision Tree in terms of FPR, and Decision Tree outperforms Naïve Bayes in terms of TPR. Since TPR (i.e., precision) means more than FPR (i.e., recall) under the VC ranking context within our study, Decision Tree is chosen as one of our benchmark models.

Given the limitations of applying embeddings to traditional machine learning models, a set of social network analysis (SNA) metrics are utilized as features for the decision tree. These metrics include: degree centrality (DC) of the main investor, betweenness centrality (BC) of the main investor, average BC of the VC's portfolio companies, average DC of the VC's portfolio companies.

By comparing the performance of these benchmark models with the proposed deep learning architecture, the effectiveness of the proposed approach can be assessed.

4.3 Evaluation Design

4.3.1 Evaluation procedure

To assess the performance of the proposed model and the benchmark models, we establish an evaluation framework. Specifically, the dataset is divided into training and testing sets based on the year in which the funding round occurred. The training set comprises data from 2009 to 2019, encompassing 9,166 startups and 18,592 funding rounds. The testing set consists of data from 2020, including 3,846 startups and their

doi:10.6342/NTU202404115

corresponding final funding rounds.

To focus on active and influential investors, a filter is applied to select venture capital firms with a minimum of 30 investment records between 2009 and 2019. This criterion identifies 301 venture capital firms from the initial pool of 7,445. For each startup in the testing set, the model predicts the likelihood of investment from the shortlisted 301 VCs. If a startup does not receive funding from any of these VCs in the actual data, the corresponding prediction is excluded from the evaluation process.

Missing values in a categorical feature are imputed with a new category value, labeled as "Missing." Similarly, for numerical features, an additional category value indicating the presence of missing values is included, and missing values are replaced with zeros.

4.3.2 Evaluation metrics

Since our proposed method formulates the startup-VC link recommendation as a VC ranking task, what we care is the quality of shortlisted recommendation rather than overall accuracy or F1-score. Hence, we adopt the following three metrics: mean average precision (mAP), mean reciprocal rank (MRR), and top-*k* precision (precision@*k*). Mean Average Precision (mAP) measures the average precision of a ranking system over a set of queries. In this context, a query represents a startup seeking investment, and the ranking

doi:10.6342/NTU202404115

is the list of recommended VCs. A higher mAP indicates better overall ranking performance.

$$mAP = \frac{1}{N}\sum_{i=1}^{N} AP_i$$

Mean Reciprocal Rank (MRR) measures the average reciprocal rank of the first relevant item in a ranking list. A higher MRR indicates that relevant VCs are ranked higher in the recommendation list.

$$MRR = \frac{1}{N}\sum_{i=1}^{N} \frac{1}{rank_i}$$

Precision@k calculates the proportion of relevant items among the top-k recommended items. This metric focuses on the quality of the top-k recommendations, providing insights into the model's ability to identify the most promising VCs.

$$Precision@k = \frac{actual\ links\ in\ top\ k\ recommendations}{k}$$

4.3.3 Code implementation and hyper-parameter settings

The code was implemented based on PyTorch 2.3.0 and PyTorch Geometric 2.3.0.

After a thorough fine-tuning process, a hyperparameter settings is adopted. The evaluation results presented in the following section are based on the hyperparameter settings: batch size: 64, dropout rate for embedding layer: 0.2, # of fully-connected layers: 2, dropout rate for the first fully-connected layer: 0.1. No dropout is applied to the second

doi:10.6342/NTU202404115

fully-connected layer.

4.4 Experimental Results

4.4.1 Link recommendation

Table 4 presents a comprehensive comparison of our proposed model's performance against benchmark models across various evaluation metrics. Our method consistently outperforms the benchmark models on all metrics. Notably, our approach achieves an mAP score of 13.79%, significantly surpassing the mAP score of 5.1% revealed in SocioLink (Xu et al., 2023). This is a remarkable achievement given SocioLink elaborate from the perspective of VCs with rich resources. Additionally, our model attains an MRR score of 18.88%, indicating that the correct VC is typically ranked as top five recommendations on average. This provides startups with a more focused and actionable strategy, allowing them to prioritize outreach to a smaller pool of potential investors.

To gain a deeper understanding of the performance across different funding rounds, we analyze the performance of each model for varying numbers of rounds. The results are summarized in Table 5. When employing preferential attachment, the performance remains relatively consistent across rounds. This suggests that the information captured by the degree in the startup-venture capital network does not significantly improve with increasing rounds. In contrast, observing the performance of the decision tree across

doi:10.6342/NTU202404115

rounds reveals a clear upward trend in the metrics. For instance, mAP increases from 5.92% in the second round to 9.76% in the sixth round, and precision@2 rises from 3.59% to 7.58%. This demonstrates that our incorporation of round information, funding-related features, and historical features, which collectively constitute sequential data, effectively captures information from previous rounds and enhances link recommendation performance in subsequent rounds. Furthermore, our proposed method exhibits a consistent upward trend in all metrics as the number of rounds increases. mAP improves from 11.57% in the second round to 27.49% in the sixth round, while MRR increases from 14.80% to 35.24%, representing over twofold growth for both metrics. These results highlight our model's ability to not only leverage the information from past rounds embedded in sequential data but also effectively propagate this information through the LSTM learning mechanism.

To sum up, evaluation results in this section indicates that our proposed method outperforms benchmark models on all evaluation metrics. And, the sequential data we adopted effectively captures information from previous rounds and enhances link recommendation performance. Last but not least, LSTM in our proposed method effectively propagates information from past rounds to improve performance.

4.4.2 Weighting strategy

doi:10.6342/NTU202404115

Motivated by the notion that later funding rounds accumulate more information and insights due to the cumulative nature of investment, we experimented with adjusting the loss function during training to assign higher weights to later rounds. The objective was to investigate whether leveraging the rich information from later rounds could enhance overall prediction performance. The adjusted loss function formula is designed as

$$L_{total} = \sum_{i=1}^{r} L_i \times \min(1, \log_8(i+2)),$$

where L_{total} is the overall loss, L_i is the binary cross entropy of samples in round i, and r is the maximal round number in samples.

Table 6 summarizes the evaluation results. Incorporating this round-weighted strategy not only led to an expected decline in prediction performance for later rounds but also negatively impacted overall prediction performance. However, rounds 3 and 4 did benefit from this weighting strategy, exhibiting slight improvements in mAP and MRR. Therefore, the decision to employ this weighting strategy should be carefully considered based on the specific application scenario and the rounds of interest.

doi:10.6342/NTU202404115

Table 4: Evaluation Results of Benchmark Models and Our Proposed Method

	mAP	MRR	Precision@2	Precision@5	Precision@10
Preferential attachment	4.67%	6.86%	2.54%	1.80%	1.71%
Decision Tree	5.57%	7.80%	3.74%	2.25%	1.35%
Our Proposed method	13.79%	18.88%	7.75%	5.66%	4.63%

Table 5: Round-wise Evaluation Results of Our Proposed Method

Round	mAP	MRR	Precision@2	Precision@5	Precision@10
Preferential Attachment					
Round 2	4.26%	6.32%	2.19%	1.68%	1.68%
Round 3	4.83%	7.39%	3.01%	1.99%	1.81%
Round 4	4.18%	6.67%	2.08%	2.13%	1.71%
Round 5	5.27%	8.16%	3.57%	2.57%	2.43%
Round 6+	4.18%	8.42%	3.93%	2.02%	1.69%
Decision Tree					
Round 2	5.92%	7.79%	3.59%	2.07%	1.30%
Round 3	5.76%	8.34%	3.92%	3.01%	1.84%
Round 4	6.63%	10.02%	5.56%	3.06%	1.67%
Round 5	8.26%	13.44%	7.14%	4.14%	2.50%
Round 6+	9.76%	12.04%	7.58%	4.83%	2.64%
Our Proposed Method					
Round 2	11.57%	14.80%	5.22%	3.96%	3.47%
Round 3	11.46%	17.28%	6.93%	5.19%	4.63%
Round 4	13.31%	19.88%	8.68%	6.53%	5.21%
Round 5	17.33%	24.65%	10.92%	8.51%	6.09%
Round 6+	27.49%	35.24%	17.37%	10.95%	7.89%

doi:10.6342/NTU202404115

Table 6: Evaluation Results of Our Proposed Method under Different Round-Weighting Strategy

	Round	mAP	MRR	Precision@2	Precision@5	Precision@10
Round-weighted	**Overall**	**13.27%**	**18.43%**	**7.65%**	**5.54%**	**4.48%**
	Round 2	10.83%	13.97%	5.11%	4.04%	3.47%
	Round 3	12.61%	18.85%	7.79%	5.11%	4.20%
	Round 4	15.65%	23.35%	10.07%	7.50%	5.90%
	Round 5	14.07%	19.76%	6.90%	7.13%	5.75%
	Round 6+	22.07%	29.83%	16.32%	9.26%	6.63%
Not round-weighted	**Overall**	**13.79%**	**18.88%**	**7.75%**	**5.66%**	**4.63%**
	Round 2	11.57%	14.80%	5.22%	3.96%	3.47%
	Round 3	11.46%	17.28%	6.93%	5.19%	4.63%
	Round 4	13.31%	19.88%	8.68%	6.53%	5.21%
	Round 5	17.33%	24.65%	10.92%	8.51%	6.09%
	Round 6+	27.49%	35.24%	17.37%	10.95%	7.89%

4.4.3 Ablation study

To assess the individual contributions of different feature sets to the overall model performance, an ablation study was conducted. The baseline model consists solely of round information (Round), served as a foundation for comparison. The full model incorporating all four feature sets: startup profile (Profile), funding-related features (Fund), and historical investor features (Past VCs), and VC embeddings (Emb) from GCN, representing the upper performance bound. The study aimed to quantify their individual

doi:10.6342/NTU202404115

impact and potential synergies by sequentially adding and removing feature sets. The results are presented in Table 7, giving us a more granular understanding of the individual contributions of different feature sets to the overall model performance.

The results indicate that company profile features significantly enhance prediction accuracy. Among the sequential features, funding-related features provide slightly more incremental value compared to historical investor features when added to the baseline model. However, the combined effect of profile features and historical investor features surpasses the combination of profile features and funding-related features, suggesting potential overlap in the information captured by these feature sets.

Table 7: Ablation Study on Our Proposed Method

Feature set	mAP	MRR	Precision@2	Precision@5	Precision@10
Baseline (Round)	5.76%	8.65%	2.73%	2.66%	2.30%
Round + Emb	5.37%	7.57%	1.94%	2.32%	2.23%
Round + Fund	7.21%	10.33%	3.43%	2.86%	2.41%
Round + Past VCs	7.61%	10.35%	3.67%	3.06%	2.78%
Round + Profile	11.18%	14.90%	5.36%	4.97%	3.99%
Full − Emb	13.43%	18.14%	7.60%	5.64%	4.66%
Full − Fund	12.98%	17.61%	6.75%	5.26%	4.39%
Full − Past VCs	12.60%	16.81%	6.80%	5.34%	4.58%
Full − Profile	8.62%	12.41%	4.52%	3.56%	3.24%
Full	13.78%	18.34%	7.25%	6.10%	4.76%

doi:10.6342/NTU202404115

Chapter 5 Conclusion

5.1 Conclusion

This study delves into the intricate dynamics of startup-venture capital relationships, aiming to enhance the precision of link recommendation systems. By recognizing the pivotal role of venture capital in the startup ecosystem and the challenges associated with startup financing, this research endeavors to develop a robust model capable of effectively matching startups with suitable investors.

A cornerstone of this research is the incorporation of sequential data, including funding activities and network evolution, into the link prediction framework. By leveraging a deep learning architecture that integrates startup profile features, funding-related features, historical investor behaviors, and round information, the proposed method effectively captures the temporal dependencies and complex interactions within the startup-venture capital landscape. The empirical results demonstrate the superiority of the proposed model over traditional benchmark methods, highlighting the significance of incorporating sequential information and deep learning techniques.

The ablation study further underscores the importance of various feature sets in driving model performance. While company profile features provide foundational information, the inclusion of sequential data, particularly funding-related features,

doi:10.6342/NTU202404115

significantly enhances predictive accuracy. The interplay between these features highlights the complex nature of startup-venture capital relationships and the need for a comprehensive approach to modeling these interactions.

In conclusion, this research contributes to the advancement of link prediction in the startup-venture capital domain by proposing a novel deep learning framework that effectively leverages sequential data and incorporates relevant features. The findings of this study provide valuable insights for both startups and venture capitalists, facilitating more efficient and effective investment decisions. Future research could explore the integration of additional data sources, such as social media and news sentiment, to further enrich the model's predictive capabilities. Additionally, investigating the application of transfer learning techniques to improve model generalization across different industries and time periods would be an interesting avenue for future exploration.

By addressing the challenges inherent in startup financing and leveraging the power of data-driven modeling, this research offers a promising approach to bridging the gap between startups and investors, ultimately fostering innovation and economic growth.

5.2 Limitations and Future Works

This study represents a significant step forward in link recommendation for the startup-venture capital ecosystem. However, several limitations and avenues for future

doi:10.6342/NTU202404115

research emerge.

- **Cold-start problem**: The current model is constrained to providing recommendations after the second funding round due to limitations in model architecture and feature design. To address this, future research could focus on developing additional features or refining the model architecture to enable predictions for the first funding round. Such advancements would be particularly valuable in mitigating the cold-start problem prevalent in the startup-VC market.
- **Graph embedding method**: The current study relies solely on the GCN method for graph representation learning. To explore the potential of alternative approach, future research could experiment with other embedding techniques such as GAT (Veličković et al., 2017) or GraphSAGE (Hamilton et al., 2017). Furthermore, constructing a heterogeneous knowledge graph by integrating data from multiple sources could offer a more comprehensive representation of the startup-VC ecosystem.
- **Multi-task learning**: The incorporation of multi-task learning presents an intriguing avenue for future research. By leveraging related auxiliary tasks, the accuracy of the primary link recommendation task could potentially be enhanced.
- **Interpersonal interaction network**: The availability of data sources imposed

doi:10.6342/NTU202404115

certain limitations on this study. Notably, the interpersonal networks within the startup community were not incorporated. Future research could address this by integrating data from platforms like Crunchbase to capture these valuable relationships.

By addressing these limitations and exploring the proposed research directions, the field of link recommendation for startups and venture capitalists can be further advanced, leading to more accurate and informative matching systems.

doi:10.6342/NTU202404115

References

Alexy, O. T., Block, J. H., Sandner, P., & Ter Wal, A. L. (2012). Social capital of venture capitalists and start-up funding. *Small Business Economics*, *39*, 835-851.

Anne-Laure, F., Jess, M., Martina M., & Rachel, K. (2024). Nurturing Innovation. *Harvard Business Review, 102*(2), 88-97.

CB Insights. (2024). State of venture Q2'24 report. https://www.cbinsights.com/research/report/venture-trends-q2-2024/. CB Insights.

Fernandez, V. (2021). The role of trust and social commitment in start-up financing. *International Review of Financial Analysis*, *75*, 101722.

Freiberg, B., & Matz, S. C. (2023). Founder personality and entrepreneurial outcomes: A large-scale field study of technology startups. *Proceedings of the National Academy of Sciences*, *120*(19), e2215829120.

Ghassemi, M., Song, C., & Alhanai, T. (2020, February). The Automated Venture Capitalist: Data and Methods to Predict the Fate of Startup Ventures. In *AAAI KDF Workshop*.

Gompers, P. A., & Lerner, J. (2004). *The Venture Capital Cycle*. MIT press.

Gonzalez, G. (2017). What factors are causal to survival of a startup. *Muma Business Review*, *1*(9), 97-114.

doi:10.6342/NTU202404115

Gu, W., Gao, F., Lou, X., & Zhang, J. (2019). Link prediction via graph attention network. *arXiv preprint arXiv:1910.04807*.

Hamilton, W., Ying, Z., & Leskovec, J. (2017). Inductive representation learning on large graphs. *Advances in Neural Information Processing Systems*, *30*.

Hoenen, S., Kolympiris, C., Schoenmakers, W., & Kalaitzandonakes, N. (2014). The diminishing signaling value of patents between early rounds of venture capital financing. *Research Policy*, *43*(6), 956-989.

Kerr, W. R., & Nanda, R. (2011). Financing constraints and entrepreneurship. *Handbook of Research on Innovation and Entrepreneurship. Cheltenham: Elgar*, 88-103.

Ko, E.-J., & McKelvie, A. (2018). Signaling for more money: The roles of founders' human capital and investor prominence in resource acquisition across different stages of firm development. *Journal of Business Venturing*, *33*(4), 438-454.

Kunegis, J., De Luca, E. W., & Albayrak, S. (2010, June). The link prediction problem in bipartite networks. In *International Conference on Information Processing and Management of Uncertainty in Knowledge-based Systems* (pp. 380-389). Berlin, Heidelberg: Springer Berlin Heidelberg.

Liang, Y. E., & Yuan, S.-T. D. (2016). Predicting investor funding behavior using crunchbase social network features. *Internet Research*, *26*(1), 74-100.

doi:10.6342/NTU202404115

Salamzadeh, A., & Kawamorita Kesim, H. (2015). Startup companies: Life cycle and challenges. Proceedings of the 4th International conference on employment, education and entrepreneurship (ICEEE), Belgrade, Serbia.

Shetty, S., & Sundaram, R. (2019). Funding acquisition drivers for new venture firms: Diminishing value of human capital signals in early rounds of funding. *Problems and Perspectives in Management* 17(1), 78-94.

Talaia, M., Pisoni, A., & Onetti, A. (2016). Factors influencing the fund raising process for innovative new ventures: an empirical study. *Journal of Small Business and Enterprise Development*, *23*(2), 363-378.

Veličković, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., & Bengio, Y. (2017). Graph attention networks. *arXiv preprint arXiv:1710.10903*.

Wise, S., Yeganegi, S., & Laplume, A. O. (2022). Startup team ethnic diversity and investment capital raised. *Journal of Business Venturing Insights*, *17*, e00314.

Xu, R., Chen, H., & Zhao, J. L. (2023). SocioLink: Leveraging Relational Information in Knowledge Graphs for Startup Recommendations. *Journal of Management Information Systems*, *40*(2), 655-682.

Yuliansyah, H., Othman, Z. A., & Bakar, A. A. (2020). Taxonomy of link prediction for social network analysis: a review. *IEEE Access*, *8*, 183470-183487.

doi:10.6342/NTU202404115

Zhou, X., Zhou, B., Yu, J., Yang, L., & Yang, Y. (2016). Research on initial rotor position estimation and anti-reverse startup methods for DSEM. *IEEE Transactions on Industrial Electronics*, *64*(4), 3297-3307.

doi:10.6342/NTU202404115