

Evolutionary role of startups and its relevance to the success in the blockchain field based on temporal information networks

Ying Wang(王颖) and Qing Guan(管青)

Citation: Chin. Phys. B, 2025, 34 (**8**): 088901. DOI: 10.1088/1674-1056/add247
http://cpb.iphy.ac.cn; http://iopscience.iop.org/cpb

What follows is a list of articles you may be interested in

Node ranking based on graph curvature and PageRank

Hongbo Qu(曲鸿博), Yu-Rong Song(宋玉蓉), Ruqi Li(李汝琦), Min Li(李敏), and Guo-Ping Jiang(蒋国平)
Chin. Phys. B, 2025, 34 (**2**): 028901. DOI: 10.1088/1674-1056/ad9a9b

Information flow between stock markets: A Koopman decomposition approach

Semba Sherehe, Huiyun Wan(万慧云), Changgui Gu(顾长贵), and Huijie Yang(杨会杰)
Chin. Phys. B, 2022, 31 (**1**): 018902. DOI: 10.1088/1674-1056/ac16c9

Shortest path of temporal networks: An information spreading-based approach

Yixin Ma(马一心), Xiaoyu Xue(薛潇雨), Meng Cai(蔡萌), and Wei Wang(王伟)
Chin. Phys. B, 2020, 29 (**12**): 128902. DOI: 10.1088/1674-1056/abb302

Feedback arcs and node hierarchy in directed networks

Jin-Hua Zhao(赵金华), Hai-Jun Zhou(周海军)
Chin. Phys. B, 2017, 26 (**7**): 078901. DOI: 10.1088/1674-1056/26/7/078901

MDSLB：A new static load balancing method for parallel molecular dynamics simulations

Wu Yun-Long (武云龙), Xu Xin-Hai (徐新海), Yang Xue-Jun (杨学军), Zou Shun (邹顺), Ren Xiao-Guang (任小广)
Chin. Phys. B, 2014, 23 (**2**): 028903. DOI: 10.1088/1674-1056/23/2/028903

SPECIAL TOPIC — Computational programs in complex systems

Evolutionary role of startups and its relevance to the success in the blockchain field based on temporal information networks

Ying Wang(王颖)[1] and Qing Guan(管青)[2,†]

[1] *School of Economics and Management, China University of Geosciences, Beijing* 100083, *China*
[2] *School of Information Engineering, China University of Geosciences, Beijing* 100083, *China*

(Received 31 December 2024; revised manuscript received 9 April 2025; accepted manuscript online 30 April 2025)

Startups form an information network that reflects their growth trajectories through information flow channels established by shared investors. However, traditional static network metrics overlook temporal dynamics and rely on single indicators to assess startups' roles in predicting future success, failing to comprehensively capture topological variations and structural diversity. To address these limitations, we construct a temporal information network using 14547 investment records from 1013 global blockchain startups between 2004 and 2020, sourced from Crunchbase. We propose two dynamic methods to characterize the information flow: temporal random walk (sTRW) for modeling information flow trajectories and temporal betweenness centrality (tTBET) for identifying key information hubs. These methods enhance walk coverage while ensuring random stability, allowing for more effective identification of influential startups. By integrating sTRW and tTBET, we develop a comprehensive metric to evaluate a startup's influence within the network. In experiments assessing startups' potential for future success — where successful startups are defined as those that have undergone M&A or IPO — incorporating this metric improves accuracy, recall, and F1 score by 0.035, 0.035, and 0.042, respectively. Our findings indicate that information flow from key startups to others diminishes as the network distance increases. Additionally, successful startups generally exhibit higher information inflows than outflows, suggesting that actively seeking investment-related information contributes to startup growth. Our research provides valuable insights for formulating startup development strategies and offers practical guidance for market regulators.

Keywords: startup, temporal networks, information flow, network analysis, startup success prediction

PACS: 89.65.Gh, 07.05.Tp

DOI: 10.1088/1674-1056/add247 **CSTR:** 32038.14.CPB.add247

1. Introduction

The blockchain field is gaining significant attention from various industries, where blockchain innovation is led by startups.[1,2] Compared to traditional large-scale enterprises, startups exhibit characteristics such as smaller scale, flexibility in strategic planning, swift decision-making, and streamlined services.[3,4] These traits make them more adept at meeting market demands and creating new employment opportunities, thus playing a crucial role in the economic development of various nations worldwide.[5–9]

As financing is one of the vital avenues for startups to address funding shortages, investments play a pivotal role in the growth of these startups.[10,11] The investment relationship between investors and investee startups is widely described in complexity science as an investment network.[12] Investment networks enable mutual economic benefits between investors and startups, from the perspectives of both the individuals associated with startup ventures and the composition of investor portfolios.[12,13] These networks also facilitate optimized resource allocation and enhanced capital utilization efficiency among startups.[14–16] From the perspective of business to business, startups as nodes in the network can be interconnected by the way in which investors invested in different companies at different times to transfer information, thus constituting an information network between startups. Information networks characterize the position,[17] density,[18] and diffusion level of startup ventures within the network,[19] uncovering valuable insights pertaining to startup value, location selection, and sustainable development. However, the majority of research on entrepreneurial information networks has predominantly focused on static local network information, with limited observation of dynamic information flow between startups.[12,20,21] Temporal networks are a network modeling approach that can be used to dynamically observe interactions between startups in a complex network domain by considering information about the dynamic evolution of behavior between subjects.[22,23] In particular, node importance measures that take into account temporal information can better identify the evolutionary characteristics of key startups over time and identify startups with potential for success in an evolving information network environment. In terms of node importance methods, random walks and betweenness centrality provide more accurate predictions for prioritizing connections to ac-

[†]Corresponding author. E-mail: guanqing35@126.com

© 2025 Chinese Physical Society and IOP Publishing Ltd. All rights, including for text and data mining, AI training, and similar technologies, are reserved.

cess points in network information dissemination.[24]

In order to be able to take into account the variability of topological morphology and the diversity of topological patterns in startup networks, we choose to integrate both random walk and betweenness centrality approaches. Specifically, random walk methods can identify the trajectories of information flow from a global perspective, while betweenness centrality methods can discover the intermediaries in information flow from an individual startup's perspective. These two methods complement each other in considering startups that are significantly important in information dissemination at both the global and individual levels. However, the current temporal random walk approach neglects the role of initial nodes, which can lead to the walking falling into a dead loop,[25,26] and existing studies on temporal betweenness centrality tend to overlook the influence of real factors on networks,[27] failing to capture the characteristics of network information flow. Therefore, in order to better characterize the dynamic evolution of startup ventures in information networks, it is necessary to improve temporal random walk and temporal betweenness centrality from the perspectives of initial nodes and temporal boundaries, thus enabling research on the important roles of startup ventures in information flow.

The objective of this paper is to explore the growth trajectory characteristics of startups through the lens of investment information flows and to examine the correlation between startup characteristics and success. Specifically, we address two key questions: how to better characterize the dynamic growth of startups and what distinguishes the information flow patterns of successful startups from those of other firms. To this end, we construct an unweighted startup investment information network, where nodes represent startups and edges indicate shared investors between them. First, we propose sTRW, a temporal random walk method that enhances walk coverage by emphasizing the starting node. Next, we introduce tTBET, which leverages temporal betweenness centrality to ensure the fastest path to the next node throughout the entire time cycle, thereby identifying key players within startup industries. Finally, we develop a comprehensive index that integrates both information flow trajectories and intermediary roles to quantify startups' influence within the information network. The key contributions of this paper are as follows: (i) We observe the overall distribution of information flow among startups to understand the general situation of the market. (ii) We identify the startups that play an important role in information flow and analyze the reasons for the popularity and relevance of the key startups. (iii) We explore the topological characteristics of successful startups and analyze the correlation between other startups and successful startups in terms of information flow. Understanding the characteristics of the information flow of established startups can provide new insights for policymakers and founders about the growth of startups.

The remainder of this paper is organized as follows. Section 2 presents the relevant literature. Section 3 showcases the data sources, network modeling, sTRW, tTBET, and indicators for evaluating startup success. Section 4 displays the results of the analysis of startup networks, the identification of key startups, and the analysis of successful startups. Section 5 provides the discussion and conclusion.

2. Literature review

Exploring various aspects of startups, such as industry characteristics, geographic location, and skill levels, is beneficial for identifying new entrepreneurial opportunities and promoting socioeconomic development.[28] The influence of entrepreneurs and stakeholders on startups' networks should not be overlooked,[29] thus it is important to explore from the perspective of relevant stakeholders in the realm of startups. Entrepreneurs play a significant role in both their learning activities and social networks, contributing to the new venture performance of startups and their benefits in adaptive strategies.[30–32] Moreover, investors in startups also hold crucial significance. Research suggests that investors act as channels for knowledge flow, assuming intermediary roles by forming a topological structure of knowledge sharing among startups.[12]

In addition to examining the perspective of stakeholders associated with startups, we can also analyze startups from the standpoint of startups' networks. Regarding firm performance, an analysis of the heterogeneity, homogeneity, social network theory, and resource-based view of startups' networks suggests that social networks serve as the primary avenue for startups to acquire external resources, thereby facilitating an enhancement of the firms' innovation performance.[33] In terms of collaboration and competition among firms, the utilization of actor networks, social networks, and inter-firm networks explores cooperation and overall competitiveness among firms under different strategies, revealing key issues such as the excessive dependence of startups on large-scale technology corporations.[34,35] Moreover, startups play a critical role in sustaining the development and success of new ventures by engaging in collaborations with different participants at various stages.[36] Although there are several perspectives to deeply study the startups, they are always viewed as static physical objects, whereas the startup is essentially growing dynamically due to the existence of timeless information communication between the startup and the investment behavior of the investors, all of which will have an impact on the next step of the startup. In order to achieve a dynamic description of startups, we subsequently utilize a temporal network.

However, choosing which metrics to use to identify and measure impactful startups is a challenging problem. A random walk is a fundamental model for studying network diffusion, which can be used to characterize network topology, such as node centrality or the presence of communities.[37–39] When the network transitions from a static to a dynamic temporal state, the walking behavior is affected by the temporal duration between events. The primary determinants of walking behavior are associated with two distinct drivers:[40] firstly, centered around the node, which enables traversal of all adjacent edges of the current node in preparation for the next walk; secondly, focused on the edge, where activation of edges surrounding the node may occur at different times, thereby altering the trajectory of the walk. The ancillary impacts of walking are related to whether it is active or passive. An instance of walking is deemed active if it reaches a node or edge, resulting in the resetting of the waiting time of the node or edge, irrespective of the heterogeneous distribution of the activation probability of the node and the equable distribution of the random walk throughout the network.[41] Conversely, a walk is passive when it quietly follows the activation of a node or edge. In this case, in a multilayer network, the waiting time of edges varies across different layers, and the precedence relationship among layers and edges dictates the course of the walk.[42] Due to the coexistence of different time scales in the temporal network, the random walk model can exhibit a rich spectrum of possibilities.[40] However, these ignore the importance of initial nodes and only have access to diffusion paths for information flow in the network.

Betweenness centrality is a measure of a node's centrality property in a network. It is typically calculated by determining the percentage of a node that lies on the shortest path between any pair of nodes in the network. This metric can depict the influence of a node on the dissemination of information in the network, and it provides a more accurate prediction of the preferential connection of access points compared to other centrality measures.[24,43] Betweenness centrality is widely used across various fields such as spatial propagation, supply chain management, and communication.[44,45] In recent years, researchers have undertaken a deep exploration of its characteristics and extensions.[46,47] In recent years, betweenness centrality has received attention from research in the field of investment. Certain researchers have argued that betweenness centrality has a positive impact on creative team success.[48] The aim of this study is to further explore the correlation between important and successful startups in a group of startups.

3. Data and methods

3.1. Data

This study retrieved information on investors and startups in the blockchain sector from the Crunchbase database (https://www.crunchbase.com/) for the period from 2004 to 2023. Crunchbase provides data on startup names, funding rounds, team sizes, and other relevant details, serving as a data source for research on industry connections, corporate online activity, and investment behavior prediction.[49–52]

Emerging startups are more likely to attract investors specializing in the same sector. Therefore, this study focuses exclusively on investors and startups within the blockchain domain, disregarding cross-sector investment activities. The selected time span is used to assess the growth trajectories of individual startups and the evolution of their interconnections. Between 2004 and July 2020, the global economic landscape witnessed several major events, such as the 2008 financial crisis and the COVID-19 pandemic in 2019. These events are of critical importance for understanding the growth dynamics of startups.

To establish investment relationships between startups, we constructed a startup–startup investment dataset based on shared investors. This dataset retains key attributes such as startup name, investor name, investment date, announcement date, funding status, industry classification, and location. Startups lacking investor information or without shared investors were excluded. Additionally, duplicate startup investment records were removed, resulting in a final dataset comprising 1013 startups and 14547 investment transactions. A simplified entity-relationship diagram (ERD) of the table is presented in Fig. 1.

Specifically, we obtain both startup and investor tables from CrunchBase and construct the investor–startup table and the startup–startup table. The startups table describes the information of 6004 startups with the blockchain field as the industry keyword, and the specific information includes the startup ID, headquarters location, founded date, industries, total funding amount in currency (in USD), acquisition status, IPO status, and other related information. The investor table contains 3730 investment data points with investment behaviors between 2004 and July 2020, and each investment data point includes relevant information such as investor ID, announced date, startup ID, startup industries, money raised in USD, and transaction name. We alphabetically sort the startup names to generate unique IDs for each startup.

Since a certain investment may include multiple investors, we decompose the multiple investors in each data point in the investors table into multiple investment data points to form the one-to-one investment information between the investor and the startup, which is saved into the investor–startup table, and each investor–startup data point includes the investor ID, startup ID, announced date, acquisition status, and IPO status. According to the investor's investment behavior of different startups in the investor–startup table, we look for different startups that have been invested by the same investor and save them to the startup–startup table, and each startup–startup data includes two startup IDs, the common investor ID, and the investment time for the two startups.

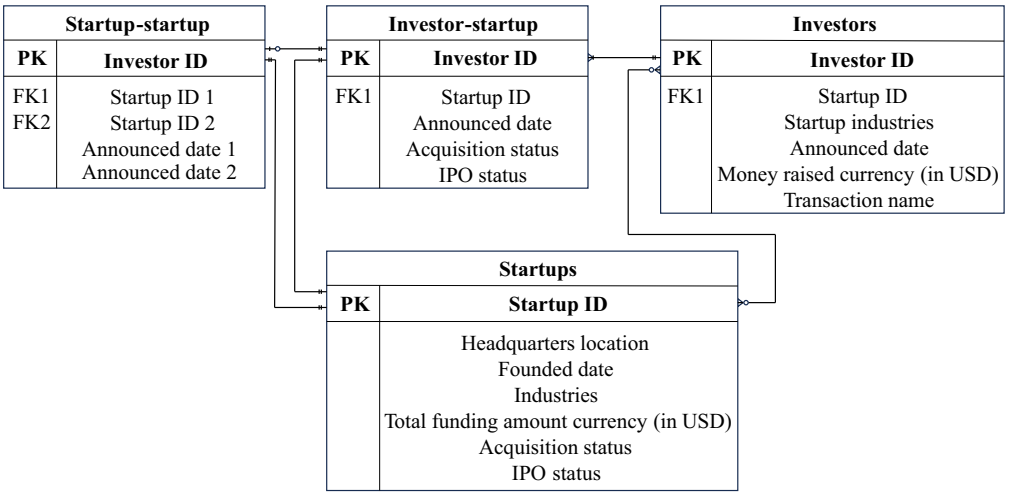


Fig. 1. Simplified ERD diagram of Crunchbase data.

We further processed the investor and business information form data from 2004 to July 2020 to form sample 1 data for the subsequent establishment of the information network. The data processing mainly includes the following parts: 1) removes the startups that do not have common investors with other startups; 2) the startups with acquisition status of made acquisitions or was acquired, or IPO status of public in the startups table are defined as successful startups; and 3) the time of the first M&A or IPO of the successful startups is recorded as the time of success. The sample 1 data processing includes 1013 startups in the blockchain field (of which 24 have been successful) and 14547 startup information with common investors. We processed the startup information form data from 2020 through July 2023 to form sample 2 data, which only serves as a true value for subsequent predictions of successful startups. 1) screening the startups for successful startups; and 2) comparing the status of the startups in sample 1, retaining only those that have changed their startup status from unsuccessful to successful. The processed sample 2 includes information on the 29 most recent successful startups. In this study, all experiments used sample 1 data to ensure coverage of all blockchain startups. Sample 2 data was only used as a test set in the experiment predicting successful startups. We mark successful startups in the dataset as 1 and unsuccessful startups as 0. The dataset is unbalanced in zeros and ones, corresponding to the real-world situation where successful startups are only a tiny minority.

3.2. The model of information networks for startups

For startups, shared investors can be considered a form of information. The interaction of investment behavior among startups constitutes a flow of information, while the timing of investment behavior serves as a crucial reference point for the growth and development of startups. We introduce complex network methods to simulate the flow of investment information among startups. Complex networks have been widely applied in various fields, such as the economy, where they reveal non-linear relationships between entities through multiple nodes and edges.[12,53] This model can characterize both the relationships between startup companies and the individual evolutionary features of each company. We propose to construct an undirected temporal network in which startup companies serve as nodes, and edges are created between them if they share common investors.

We construct a network $G(V,E,T)$, where $V = \{v_1, v_2, \ldots, v_n\}$ represents the set of network nodes, with v_n representing different startups. E denotes the set of edges, which can be represented using a matrix

$$E = \begin{bmatrix} e_{11} & \cdots & e_{1n} \\ \vdots & \ddots & \vdots \\ e_{m1} & \cdots & e_{mn} \end{bmatrix}. \tag{1}$$

In this context, e_{mn} represents the set of edges that connect the startup companies v_m and v_n through common investors. $T = \{t_1, t_2, \ldots, t_k\}$ represents the set of timestamps that record the first instance of common investment between startups v_m and v_n by investors. The method used to construct the network structure is illustrated in Fig. 2.

In Fig. 2, nodes A, B, C, and D represent four distinct startups, while a, b, c, and d represent four different investors. The time stamps are $t_1 < t_2 < t_3 < t_4 < t_5$. In Fig. 2(a), three different colored lines represent the information of co-investors among startups. For example, the purple line indicates that both startups A and B have investment information from investor a at time t_1, in which case a is the co-investor of startup AB at time t_1. There is a common investor between startups A and B, beginning at time t_1. There are a common investor d between startups B and C at time t_3, a common investor c at time t_4, and common investors a and b between startups C and D at time t_1. In Fig. 2(b), the horizontal axis represents

time, while the vertical axis represents investors. The timeline depicts the startups invested in by each investor during each time period. For investor a, investing in both startups A and B lasts from time t_1 to t_3, while investing in both startups C and D lasts from time t_1 to t_2. In the time period from t_1 to t_5, the connection between BC appears later, indicating that both startups B and C have some advantages in attracting investment in the later stages of investment. Conversely, the CD connection only appears in the early stages, indicating that startups lack investment appeal in the later stages, and this information cannot be reflected in a static network.

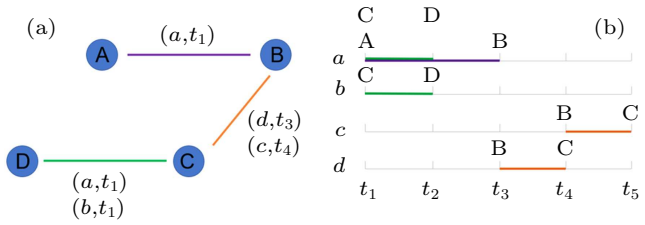


Fig. 2. Schematic representation of co-investor information among startups in a temporal network.

3.3. Improved random walk method considering temporal information

A random walk is a frequently employed technique in complex networks to identify crucial network nodes by exploring the propagation routes of information within the network, thus unraveling the evolutionary mechanisms of the network. In this study, random walk has unveiled pivotal startups involved in both the reception and dissemination of information within the information network.

We randomly select a node within the network as the starting point for the random walk and proceed to search for the next node among its neighbors. If the initial node is an isolated node, from the perspective of the random walk, it may result in a situation where the walk remains at the same node indefinitely. However, from a topological standpoint, a key startup involved in information propagation requires multiple investors to disseminate the information. Consequently, key startups tend to have a substantial number of investors, and the likelihood of isolation is extremely low. Therefore, selecting an isolated node as the initial point would lose its significance in identifying key nodes. Thus, we redefine the starting point until it has neighbors. We set the investment time of the initial node as the initial time and mark it as such. We then search for the neighbors of the current node within the entire time period. Among the neighbors of the current node, we randomly choose an unmarked node as the next node for the walk, and we set the investment time of the next node as the current time. The purpose of marking nodes is to maximize the coverage of the walk and prevent it from getting stuck in a loop. We repeat the above steps until the current node has no unmarked neighbors or all the investment times of the neighboring nodes are earlier than the current time.

To avoid the deviation caused by random walks and conduct multiple iterations of experiments, we express the temporal walk of nodes as

$$\mathrm{RW}^{\mathrm{temp}}(i) = \frac{1}{S}\sum_{j\in U} T_j, \tag{2}$$

where j denotes the current iteration, U represents the total number of iterations, T signifies the frequency at which nodes are traversed within each iteration, and S denotes the predetermined number of steps taken per walk. The necessity of temporal random walks lies in their ability to reflect the changes in investment markets over time and to conform to the evolutionary patterns of investment networks. Through these walks, turning points can be identified and analyzed for their centrality, as well as the frequency and count of investors. The innovation of temporal random walks lies in their ability to not only consider investment time but also reflect changes in investment markets, as compared to static walks. Moreover, they prevent walks from getting trapped in local subgraphs and being unable to escape. Additionally, the more neighbors a node has, the greater the probability of a walk on that node.

3.4. Improved betweenness centrality method considering temporal information

Betweenness centrality, which assesses the proportion of node involvement in paths between other nodes within a network, is one of the key measurement methods for characterizing important structural properties of networks. Starting from the perspective of shortest paths, we compute the sum of the ratio between the number of shortest paths passing through a node u and the total number of shortest paths between any two points as a measure of the node's influence. In the context of all possible shortest paths for information dissemination, a higher number of shortest paths passing through a particular node indicates that the node controls a greater flow of information. Therefore, the node possesses higher importance and influence within the entire network. The computation formula is as follows:

$$\mathrm{BET}(u) = \sum_{i\in V}\sum_{j\neq i\in V}\frac{\delta_{ij}(u)}{\delta_{ij}}, \tag{3}$$

where V represents the set of nodes, δ_{ij} denotes the total number of shortest paths between nodes i and j, and $\delta_{ij}(u)$ signifies the shortest path from node i to node j that passes through node u.

The static betweenness centrality overlooks the network changes caused by information flow over time. Therefore, it is necessary to incorporate time into the static betweenness centrality. This can be achieved by introducing a dependency on time t and calculating scores for detecting paths that are either the shortest or fastest in terms of time traversal through vertices. There are two choices available: one emphasizes

paths with the minimum number of contacts, while the other focuses on traversing the fastest paths. Regarding time boundaries, two options exist: utilizing the entire observation period or considering only the fastest time-related paths starting at each vertex j closest to time t. Considering only the shortest or fastest paths may deviate significantly from the actual network dynamics since path durations are continuous and there may exist paths that are only slightly slower. Therefore, this paper considers paths that reach the next node in the fastest time across the entire period. The following formula is presented:

$$\mathrm{dist}_{ij}^{\mathrm{temp}} = \min_{t_0} \mathrm{dist}_{ij}^{\mathrm{temp}}, \tag{4}$$

$$\delta_{ij}^{\mathrm{temp}} = \sum_{t_0} \mathrm{dist}_{ij}^{\mathrm{temp}}, \tag{5}$$

$$\mathrm{BET}^{\mathrm{temp}}(u) = \sum_{i \in V} \sum_{j \neq i \in V} \frac{\delta_{ij}^{\mathrm{temp}}(u)}{\delta_{ij}^{\mathrm{temp}}}, \tag{6}$$

where $\mathrm{dist}_{ij}^{\mathrm{temp}}$ represents the shortest path between nodes i and j, regardless of any starting time. $\delta_{ij}^{\mathrm{temp}}$ denotes the number of shortest paths from node i to node j for all time periods. For any two points i and j in the set V, the ratio of the number of $\delta_{ij}^{\mathrm{temp}}$ through point u to the total number of $\delta_{ij}^{\mathrm{temp}}$ is the tTBET value of node u, denoted $\mathrm{BET}^{\mathrm{temp}}(u)$.

3.5. Startup success evaluation indicator considering temporal information

The success of startups has always been a popular topic of research, as it facilitates the exploration of developmental and relational patterns among individual startups. We define successful startups as those that have undergone mergers or become publicly listed. By utilizing existing studies on sTRW and tTBET, we examine the influence of successful startups on information dissemination within networks. In order to quantify the impact of successful startups in the network, we introduce the betweenness centrality-random walk (BR) metric

$$\mathrm{Ave}(i) = \frac{1}{V} \sum_{i \in V} \mathrm{BET}^{\mathrm{temp}}(i) + \mathrm{RW}^{\mathrm{temp}}(i), \tag{7}$$

$$\mathrm{BR}(i) = \frac{\mathrm{BET}^{\mathrm{temp}}(i) + \mathrm{RW}^{\mathrm{temp}}(i)}{\mathrm{Ave}(i)}. \tag{8}$$

The process can be mainly divided into two steps. Firstly, the average values $\mathrm{Ave}(i)$ of tTBET $\mathrm{BET}(i)$ and sTRW $\mathrm{RW}(i)$ values for all nodes are calculated. Secondly, the BR indicator value is determined by utilizing the tTBET and sTRW results of node i. The value of BR ranges from 0 to 10. If the startup's role in the flow of information is more important, the closer the BR value of that startup is to 10, otherwise, the closer it is to 0. The BR indicator value can be used to measure the influence of the company in the entire network, with a higher BR value indicating greater influence.

4. Results and analysis

4.1. Static network analysis of startups

By observing the overall distribution pattern of information dissemination among startups in a static network, we can understand the static individual characteristics of startups. Figure 3 depicts the overall layout of the static network composed of startups and measures the degree of distribution. In Fig. 3(a), the nodes represent startups, the edges represent the existence of common investors between companies, and the nodes of different colors represent different modular classes. The size of the node is related to the number of neighbors it has, with larger nodes having more neighbors. Nodes of the same color represent startups that share a large number of common investors. Startups within the same community are likely influenced by the same group of investors, forming a tightly connected network structure. Obviously, nodes of the same color are closer, indicating that investors are more inclined to invest in companies of the same class. Due to some companies acting as bridges, a cluster of startups related to them seems to have formed in the network, indicating the existence of associations between startups from the same investment source.

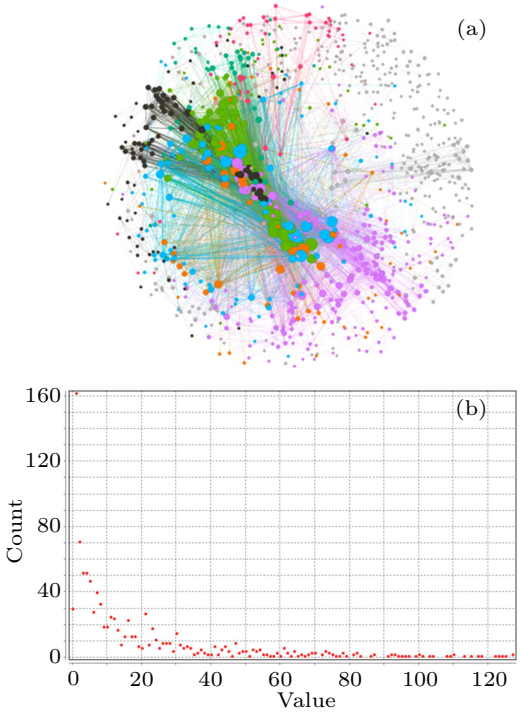


Fig. 3. (a) Static networks of startups and (b) degree distribution.

In Fig. 3(b), the number of neighbors of a node is referred to as the degree of the node. The distribution of node degrees reveals that the majority of nodes have a maximum of 1–10 neighbors, indicating that most startups acquire investments from diverse sources rather than relying on a single investor. The majority of investors tend to focus exclusively on specific domains rather than employing a broad investment strat-

egy. Nonetheless, the average degree of the network stands at 18.521, primarily due to a small fraction of startups receiving substantial attention from numerous investors.

4.2. Key startups identification

4.2.1. key startups identified by sTRW

Regarding the static network of startups, in order to balance the walking coverage and accurately identify crucial nodes, we compared three scenarios: short-range, mid-range, and long-range walks. We conducted experiments with step values of 10, 500, and 1000 in order to avoid the interference caused by the instability of randomness and the varying computational complexity of the experiments. To mitigate experimental contingencies, each experiment was repeated multiple times. Through three sets of experiments, we ensured an equal number of nodes were traversed each time to ensure the identification of key nodes in the static network through random walks. The frequencies of occurrence for startups during the three sets of experiments were summed and sorted to obtain the top 20 startups for each experiment. To facilitate a clear and equitable comparison of startup importance, we normalized the experiments by using the number of nodes traversed in each experiment as the denominator and the actual occurrence frequency of each node in each experiment as the numerator. The rankings and proportions shown in Fig. 4 were obtained. The horizontal axis represents the ID of the key startups, while the vertical axis represents the proportion of steps taken by the startup in the current walk. The three colors correspond to experiments involving three different step lengths.

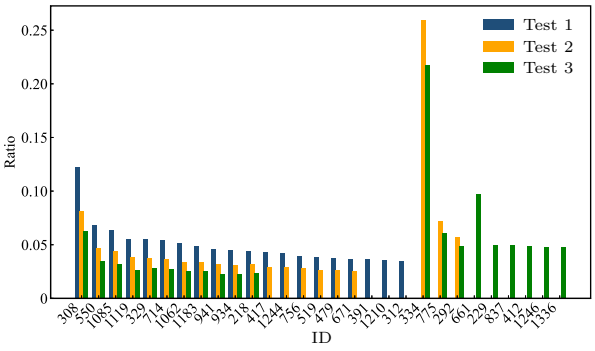


Fig. 4. Static random walk identified key startups proportion chart.

The key startup data trends in the three experiments are similar, and the similarity of the experimental results indicates the successful implementation of the functionality to identify key startup nodes through static network random walks. However, the proportions of random walks for the three startups (IDs correspond to 292, 334, and 775), which differ noticeably from the other experimental data in two instances. This discrepancy is attributed to randomness, resulting in a situation where random walks are predominantly confined to these three companies. Consequently, the coverage of the walks is reduced, leading to significant data bias. Hence, excessive randomness causes inevitable deviations in the results of static random walks.

Startups can only observe the final results of network evolution in static networks, lacking dynamic observations of the growth process. Hence, we analyze companies in the context of investment information flow using a temporal network. In the case of startups' temporal networks, starting from each node, we record the number of steps taken in random walks as initial nodes. The statistics reveal an average walk length of 2.55 steps, with the longest walk reaching 7 steps. Among these, walks consisting of 2 steps account for a frequency of 45.62%, while walks consisting of 3 steps account for a frequency of 22.44%. These two step lengths cover more than 60% of the frequencies. The significant difference between temporal random walk steps and those of static networks can be attributed to two main factors. First, in static network experiments, nodes may be visited multiple times to achieve higher coverage. In contrast, temporal networks visit nodes in chronological order without repetition. Secondly, upon analyzing the data on startups' investments as depicted in Fig. 5, it is observed that the red bar represents startups receiving multiple investments simultaneously, the blue bar indicates that the maximum time interval between startups receiving multiple investments is less than 4 years, and the green bar signifies that the maximum time interval between startups receiving multiple investments exceeds 4 years. The vertical coordinates denote the proportion of startups falling into each category. It is evident from this analysis that there is a time delay between investing in startups. This manifests in the network as a large time span between pairwise connected nodes, leading to fewer walk steps between nodes within a certain period.

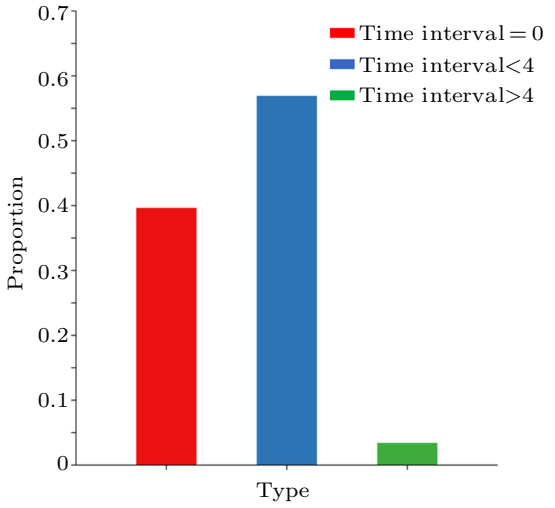


Fig. 5. Time interval of investment.

The coverage rate of random walks reflects, to some extent, whether the walks consider global nodes. A low coverage rate indicates that some crucial nodes may be overlooked. To

increase coverage, we employ a neighbor-marking approach, where already visited nodes are not reselected for the next node. Although this method enhances coverage to a certain extent, it interferes with randomness. To find the optimal node coverage while preserving randomness, we vary the number of iterations, thereby avoiding bias in subsequent experiments. Considering the temporal aspect of the network, we set the number of walk steps. To ensure the accuracy of identifying important nodes through the walks, we conduct iterative experiments. To ensure fairness in the walks, we take into account the coverage rate, represented by the formula $\mathrm{cov} = \frac{1}{I}\sum i$, where i represents the walked nodes and I represents the total number of nodes.

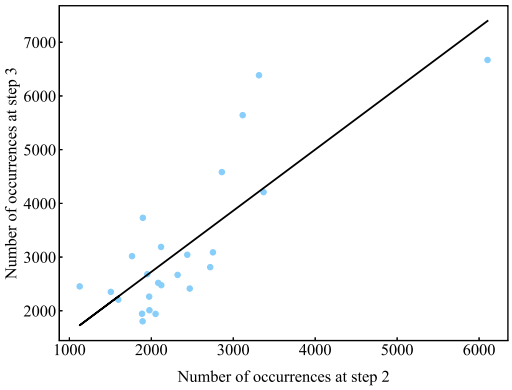


Fig. 6. Relevance of different steps of sTRW identification for startups.

In order to ensure the highest coverage rate, a comparative experiment was conducted on two scenarios involving random walk steps of 2 and 3. Since the frequencies of startup appearances may not be exactly the same, they were summed and subjected to correlation analysis in both experiments. The correlation among the top 20 key startups in the two experiments is illustrated in Fig. 6, with the x and y axes representing the number of times key startups were visited in random walks with steps 2 and 3, respectively. The trend line in Fig. 6 is close to the diagonal line, and the correlation coefficient (R-value) was determined to be 0.8, indicating that although the number of steps differed, the key startups obtained through the two walks showed minimal differences and exhibited similar overall trends, thereby confirming the stability of sTRW.

4.2.2. Key startups identified by tTBET

Although the distribution of key startups in static and temporal random walks appears similar, it is difficult to ascertain whether this similarity persists in other indicators. Therefore, we consider analyzing key startups using betweenness centrality. The key startups identified through the betweenness centrality of the static network are presented in Table 1. It is evident that the majority of these startups have their headquarters established in the United States, and their core operations are closely associated with the information industry. Furthermore, their establishment dates predominantly fall within the period from 2012 to 2014. One possible reason for this trend could be the thriving growth of the US information industry around 2012, which attracted widespread investor attention. Establishing headquarters in the United States would have been more appealing for these industry-related startups. Another reason is that startups concentrated in the same region facilitate the exchange of information and communication, as well as staying updated on the latest developments in the industry.

Table 1. Static betweenness centrality identified key startup information.

ID	Headquarters location	Founded date	Industry groups
829	San Mateo, California, United States	2013/9/1	Information Technology, Internet Services, Messaging and Telecommunications, Privacy and Security, Software
550	Reno, Nevada, United States	2012/4/1	Financial Services, Hardware, Internet Services, Payments, Software
646	San Francisco, California, United States	2014/4/26	Financial Services, Payments, Software
736	Palo Alto, California, United States	2010/1/1	Data and Analytics, Software
660	Lexington, Kentucky, United States	2014/4/1	Financial Services, Payments, Software
1377	Milan, Lombardia, Italy	2018/1/1	Apps, Mobile, Software, Sustainability, Transportation
95	Reykjavk, Gullbringusysla, Iceland	2012/1/1	Financial Services, Information Technology, Payments, Privacy and Security, Professional Services
329	San Francisco, California, United States	2012/6/1	Commerce and Shopping, Financial Services, Payments, Software
1085	San Francisco, California, United States	2012/1/1	Financial Services, Internet Services, Payments, Software
308	Boston, Massachusetts, United States	2013/1/1	Financial Services, Lending and Investments, Payments, Software

Similar to random walks, tTBET also allows for the dynamic observation of information flow among startups. The key startups identified through tTBET are presented in Table 2. Apart from Ripple Inc., none of the other startups are listed in Table 1, indicating a significant disparity between centrality in the static and temporal networks, unlike random walks. Firstly, although the headquarters are still predominantly located in the United States, the majority of startups in Table 1 were established prior to 2014. In contrast, the establishment dates of key startups in Table 2 exhibit a more evenly distributed pattern, indicating

that the key startups in the network change over time with the flow of information. Secondly, the industries represented in Table 2 are more concentrated, with a strong focus on financial services, software, and payment systems as their core operations. This suggests that these types of businesses possess high information flow dynamics and attract significant investor attention.

Table 2. The tTBET identified key startup information.

ID	Headquarters location	Founded date	Industry groups
1423	Denver, Colorado, United States	2015/1/1	Financial Services, Payments, Privacy and Security, Software
1244	Sunnyvale, California, United States	2017/1/1	Blockchain, Cryptocurrency, Software
1116	Los Angeles, California, United States	2014/1/1	Financial Services, Lending and Investments, Payments, Software
1085	San Francisco, California, United States	2012/1/1	Financial Services, Internet Services, Payments, Software
1210	Chicago, Illinois, United States	2018/1/1	Financial Services, Payments, Software
1437	Singapore, Central Region, Singapore	2016/1/1	Financial Services
1306	San Francisco, California, United States	2017/1/1	Commerce and Shopping, Financial Services, Payments, Software
1352	Gibraltar, Gibraltar, Gibraltar	2018/1/1	Financial Services, Lending and Investments
1107	San Francisco, California, United States	2017/1/1	Financial Services, Information Technology, Payments, Software
1119	Denver, Colorado, United States	2014/7/1	Financial Services, Lending and Investments, Payments, Software

Upon comparing Tables 1 and 2, it is evident that the majority of key startups are geographically located in the United States (80%). This trend aligns with the implementation of various US policies in the blockchain space, encompassing legal regulations, public services, procurement, and education aimed at fostering the growth of blockchain technology within the industry.[54,55] Additionally, we compare the distribution of the centrality values of the key startups identified under the two methods of betweenness centrality for static networks and tTBET for temporal networks. In Fig. 7, horizontal coordinates represent the top 100 startups in both methods, and vertical coordinates represent the centrality values, where blue triangles represent the tTBET method and red circles represent the betweenness centrality method. We find that, among the two methods, tTBET can highlight the information transit effect of key startups more. Figure 8 illustrates the distribution differences of key startups identified by two methods in terms of their founding time and industry groups. Specifically, figure 8(a) shows the key startups identified by the betweenness centrality method, while figure 8(b) displays those identified by the tTBET method. The results indicate that the key startups identified by the tTBET method exhibit a more dispersed distribution over time, enabling it to more effectively capture the dynamic characteristics of information flow among startups. Consequently, tTBET provides a more accurate reflection of the evolutionary process of the information network.

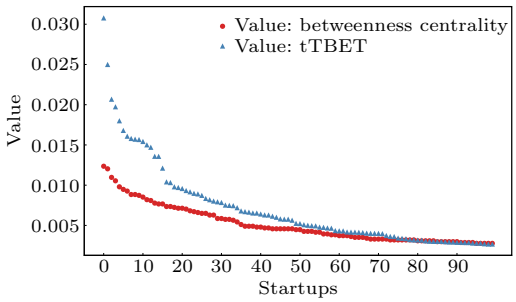


Fig. 7. Distribution of key startups' values.

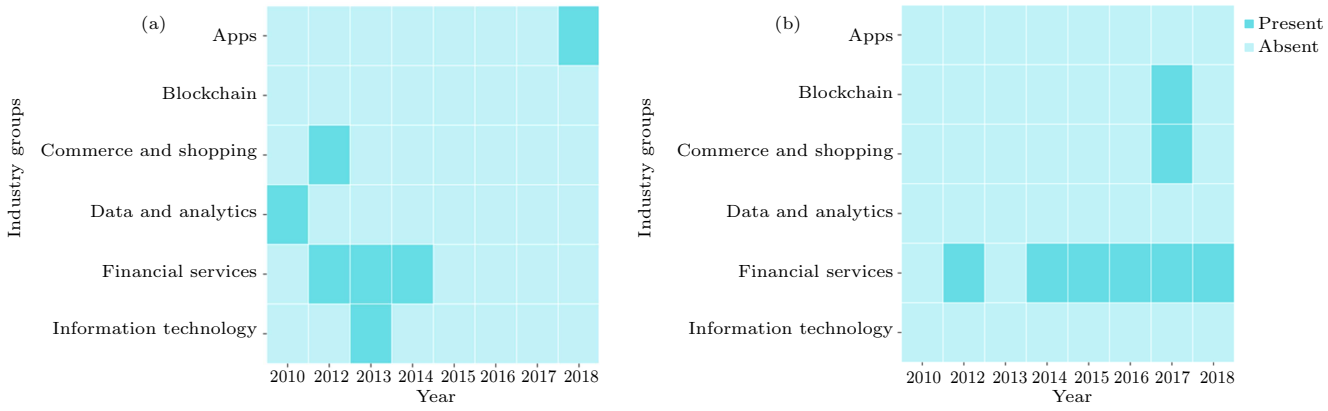


Fig. 8. Distribution of key startups founded date and industry distribution.

4.2.3. Comparing tTBET and sTRW identified startups

We use the classical learning algorithm random forest (RF) in the Python machine learning package Scikitlearn to predict the probability of success of startups in two network models, static and temporal networks, in order to test the performance of the two methods sTRW and tTBET in these networks. Random forest was used in many previous works on Crunchbase data to predict business success.[51] The random forest algorithm is fast to train and has high accuracy. Each decision tree in the algorithm finds the best feature among the randomly selected features, maximizing the gain obtained from the model.

In this experiment, the number of trees is 10, and the maximum eigenvalue is automatic for the best results. The sample set was divided into a 90% training set and a 10% test set. The random walks and betweenness centrality values of the startups are used as eigenvalues, and the success of the target variable startups is defined as two possible values: 1 (successful) or 0 (unsuccessful), which reduces the problem of predicting success to a binary classification, which is solved by supervised learning. In supervised learning, the dataset is partitioned into a subset of features (matrix X) and target variables (vector y). During the training process, the model tries to predict the value of y when given a set of features X. The predicted value is compared with the true value y.

Through Table 3, we can see that the startup states can be identified more accurately under the temporal network using the sTRW and tTBET methods. Although the static network has a higher success rate of identifying successful startups using random walk and betweenness centrality, the success rate of identifying unsuccessful startups is lower than that of the temporal network. When the static network utilizes the sTRW and tTBET methods to improve the recognition success rate for unsuccessful startups, it directly leads to a correct recognition rate of 0 for all successful startups. For the temporal network, the sTRW and tTBET methods outperform the unimproved method in all aspects.

Table 3. RF classification results for static and temporal models.

Model	Methods	Train set accuracy	Test set accuracy	Precision	Recall	F1	Success startups accuracy percentage	Unsuccess startups accuracy percentage
Static	RW & BC	0.980	0.403	0.781	0.403	0.505	40%	59.26%
Static	sTRW & tTBET	0.983	0.639	0.857	0.639	0.709	0%	100%
Temporal	RW & BC	0.987	0.915	0.837	0.915	0.874	33.33%	66.67%
Temporal	sTRW & tTBET	0.992	0.926	0.831	0.926	0.890	41.67%	66.67%

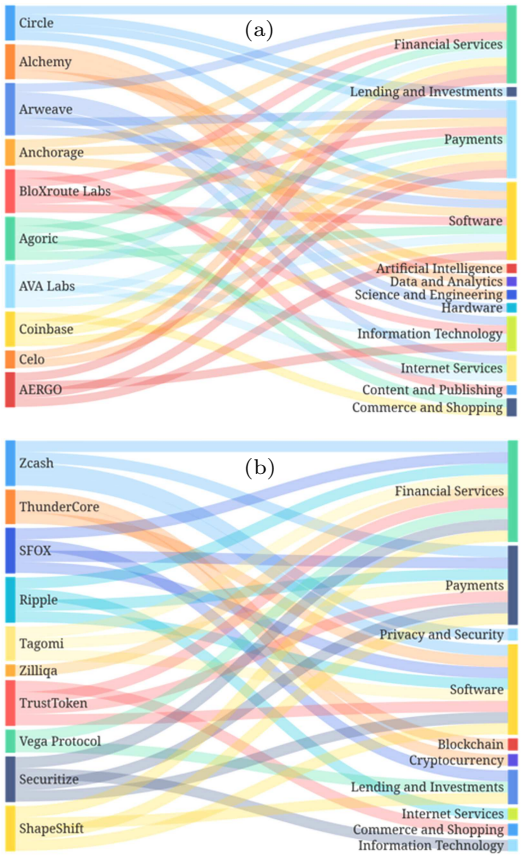


Fig. 9. The (a) sTRW and (b) tTBET identified key startups.

We compare the startups identified using the methods of sTRW and tTBET to determine whether the recognized startups can serve as key players in the transmission of information within the network. Figure 9 displays the names of the key startups identified by both methods, along with their primary business activities.

We have observed that key startups in the field of blockchain, identified through either random walks or betweenness centrality in the information network, primarily engage in three categories of business: financial services, payments, and software. These three types of businesses have garnered widespread attention and appeal from investors, resulting in a significant influx of information. Furthermore, these business domains lack substantial industry barriers, allowing investors to diversify their reliance and leading to a certain level of information outflow. In the context of information circulation, these startups play a pivotal role in network interactions.

4.3. Correlations of network roles measured by random walk and betweenness centrality

There are distinctions among the key startups identified through different methods in the information network. It remains unclear whether these distinctions stem from the time-

liness of the network or the varying merits of the node importance methods. Therefore, we conduct a comparative analysis of the results obtained from random walks and betweenness centrality in both static and temporal networks. In Fig. 10, A represents the key startups identified through betweenness centrality in the static network, B represents the key startups identified through tTBET, C represents the key startups identified through random walks in the static network, and D represents the key startups identified through sTRW. "Both" indicates that the startup ranks within the top 20 in both scenarios, while "none" suggests that the startup does not rank within the top 20 in either scenario.

Figure 10(a) illustrates a remarkably high correlation between temporal networks and static networks in the context of random walks. The nodes are relatively concentrated around the regression line. Notably, six key nodes are present in both the temporal and static networks, and these nodes possess elevated weights within the network. The high weights associated with these nodes signify the greater significance of these startups within the information network, indicating their pivotal role in information circulation, both during the growth process and in the final state.

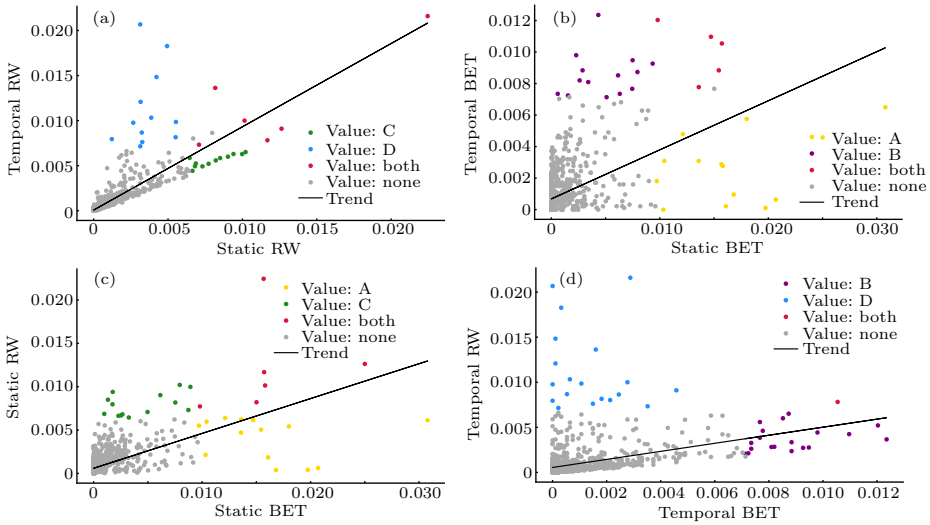


Fig. 10. Relevance of key startups.

Figure 10(b) reveals that there is not a pronounced correlation between temporal networks and static networks in terms of betweenness centrality. Although the betweenness centrality values in the temporal network are generally lower than those in the static network, nodes that perform well in both networks exhibit minimal differences in their betweenness centrality values. This suggests that, in the process of information dissemination, most startups have fewer actual connections with other entities compared to an ideal state. Only a small number of startups serve as primary hubs for transmitting information. These pivotal information hubs maintain their critical role in the information network, regardless of the stage of information flow or the final state.

A high level of betweenness centrality indicates that the startups attract multiple investors over the entire observed time period and play a significant bridging role. A high frequency of random walks signifies that the startup is connected to multiple other entities, demonstrating its essential role in connectivity. In Fig. 10(c), it can be observed that in the static network, both betweenness centrality and random walks exhibit relatively uniform distributions around the regression line. Therefore, in a static network, the differences between these two methods cannot be adequately reflected, and a specific analysis of the characteristics of startup information flow is not feasible.

From Fig. 10(d), it can be observed that, for the temporal network, only one node is identified as a critical node under both random walk and betweenness centrality measures. This node exhibits similar performance to other highly central nodes (represented by purple in Fig. 10(d)) in terms of betweenness centrality but performs worse than other highly random walk nodes (represented by blue in Fig. 10(d)). These findings suggest that although there are fewer direct associations between startups and key startups in the information flow process, there is still very close information exchange between them. The farther the startup is associated with the key startup network, the less information it receives.

4.4. Dynamic topological characterization of successful startups in information networks

The dominant position of key startups in the information flow enables them to acquire more informational resources. Most startups strive for further development through mergers and acquisitions or by going public. Therefore, it is worth investigating whether acting as an information hub to obtain more informational resources contributes to the success of

startups.

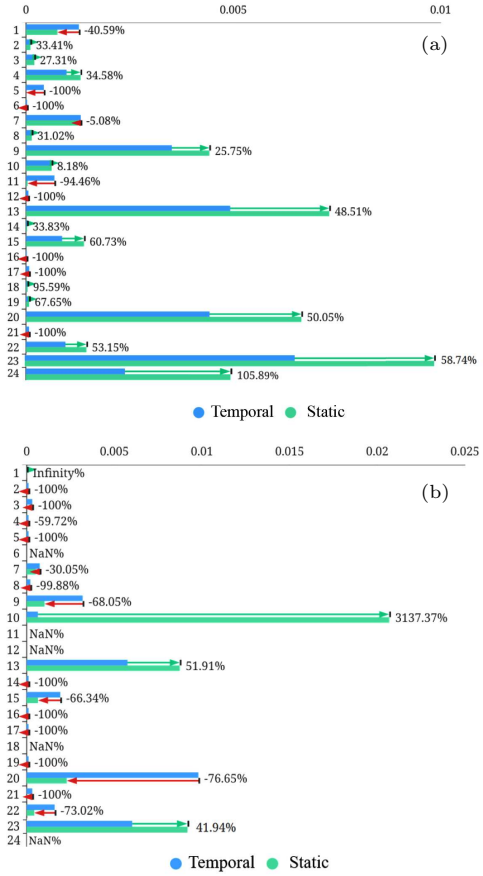


Fig. 11. Percentage of successful startups identified by (a) sTRW and (b) tTBET.

Figure 11 illustrates the proportions of successful startups identified by different networks using two methods: random walk and betweenness centrality. Since isolated companies do not engage in information flow with other startups, we do not consider isolated successful startups. In Fig. 11(a), the 24 non-isolated successful startups are sequentially arranged on the x-axis, while the y-axis represents the proportions of random walks in the static and temporal networks. The red arrow indicates a higher proportion of temporal networks in random walks, whereas the green arrow represents the opposite. The numerical value of the arrow indicates the difference in proportions between the two networks. Some startups have a temporal network proportion of 100%, indicating that the static network fails to recognize the significance of these successful startups. This suggests that the temporal network is better at capturing successful startups compared to the static network. Figure 11(b) shares the same x-axis as Fig. 11(a), while the y-axis represents the proportions of betweenness centrality in the static and temporal networks. The red arrow indicates a higher proportion of temporal network in betweenness centrality, while the green arrow represents the opposite. The numerical value of the arrow indicates the difference in proportions between the two networks. For each successful startup, the temporal network shows a significantly higher proportion of tTBET compared to the static network. This indicates that the temporal network is more sensitive and precise in capturing successful startups. Some of the startups fail to recognize the importance of betweenness centrality in the static network, suggesting that the probability of capturing successful startups through tTBET is higher and covers a broader range.

We use sample 2 as the test set. Excluding the startups in sample 2, all the remaining startups in sample 1 are used as the training set. The success of the target variable startup is defined as two possible values: 1 (successful) or 0 (unsuccessful). Using the RF method, which performs better in Table 3, the sTRW value,tTBET value, and BR value of the startups are used as the eigenvalues, and the comparison is made to verify whether the BR value can be used as one of the features to judge the future success of the startups.

Table 4. RF method for predicting successful startups under different eigenvalues.

	(sTRW, Kim[56])	(tTBET, Kim[56])	(sTRW, tTBET)	(sTRW, tTBET, BR)
Train set accuracy	0.978	0.975	0.983	0.983
Test set accuracy	0.292	0.278	0.286	0.321
Precision	1.0	1.0	1.0	1.0
Recall	0.281	0.275	0.286	0.321
F1	0.385	0.401	0.444	0.486

In Table 4, precision denotes the probability of correctly predicting a positive sample. We compared the existing temporal method[56] and found that, since the test set consists entirely of positive samples (successful startups), the precision is 1, indicating that the model achieves perfect accuracy in predicting positive cases within this set. However, this does not necessarily mean the model has strong discriminative ability in distinguishing between successful and unsuccessful startups. Therefore, accuracy and F1 score provide a more comprehensive evaluation of the model's overall performance. When the BR value is incorporated as a feature, the accuracy improves to 0.321, an increase of 0.035, while the F1 score rises from the original 0.444 to 0.486. This suggests that the BR value can serve as a useful feature in studying the future success of startups.

The importance of the role of successful startups in information flow is unknown. Since the BR value integrates sTRW and tTBET, the higher BR value represents the more impor-

tant role of startups in information flow, so we analyze the role of successful startups in information flow by calculating the BR value of successful startups and other startups in the temporal network. We found that the BR average value for successful startups was 1.315, while the BR average value for non-successful startups was 0.989. This indicates that successful startups have a more abundant flow of information compared to other startups, obtaining more resources through the exchange of information. However, the BR value for successful startups was lower than that of key startups, which suggests that a large amount of information flow for startups does not necessarily bring good resources. Although incoming information represents investors' favor, when information is output, it also means investors' loss or instability. These losses or instability may prevent the accumulation of capital for startups to achieve mergers or listing conditions. Therefore, startups need both the flow of information to obtain resources from investors and the improvement of their own industry skills to control the outflow of investors.

5. Discussion and conclusion

This study explores the characteristics of information flow between key startups and successful startups through dynamic observations of the information network. Our focus is on addressing the following key issues: 1) Examining the overall distribution patterns and individual static features of information flow in startups. 2) Identifying key startups within the information flow network. 3) Analyzing the information flow characteristics of key startups. 4) Comparing the relevance of successful startups to key startups in the context of information flow. To answer these questions, we analyze the temporal startup network from 2004 to 2020, leveraging both the tTBET and sTRW methods to track the dynamic evolution of startups, identify information flow hubs, and measure correlations across different networks. Finally, we introduce the BR index to evaluate the role of successful startups in information flow. The main findings and discussions are as follows:

1) With the support of policies, startups in the blockchain field have been developed rapidly with the medium of investment information.

Through policy enactment, countries led by the United States have fostered the development and application of technology in the blockchain field. A collaborative relationship exists among startups within the same investment network, leading to the formation of clusters in the network via information bridges. From the perspective of individual startups, most have a small number of investors, while only a few startups attract a large number of investors.

2) Startups that play a significant role as information hubs tend to have similar core business operations.

Within the information flow network, key startups primarily engage in three categories: financial services, payments, and software. These three sectors not only attract extensive attention from investors but also lack substantial industry barriers. Investors do not rely exclusively on a single startup. Therefore, due to the inflow and outflow of information, these startups emerge as pivotal players in network interactions.

3) The flow of information between startups gradually diminishes.

We have observed that during the process of information transmission among startups, although the number of sub-key startups directly associated with key startups, which have a significant amount of information input and dissemination, is not substantial, the sub-key startups experience a higher volume of information flow. The amount of information among other startups connected to sub-key startups is relatively reduced. Following this pattern, as the network distance from key startups increases, the volume of information flow among startups decreases.

4) Successful startups experience a greater inflow of information and a lesser outflow.

Our research demonstrates that a high volume of information flow for startups does not necessarily guarantee favorable resources for them. Incoming information represents investor investments in the startups, generating profits for the startups. However, when information is output, it signifies investors diverting their investments towards other startups. The potential loss or instability of these invested resources may impede the accumulation of capital for the startups. Consequently, startups require both information flow to access investor resources and the enhancement of their industry-specific skills to control the outflow of investors. We hope that this study sparks the interest of other researchers to delve further into the growth characteristics of startups in information flow and the social influence of startups in the financial industry. Such exploration can provide valuable recommendations for political decision-making regarding the establishment of market rules.

Our study has several limitations. First, the model's performance may be influenced by feature selection and weight allocation settings, as different feature combinations can affect its effectiveness. Future studies could explore adjusting feature weights or incorporating additional dynamic features to enhance the model's accuracy. Second, while this study reveals a correlation between the structural features of startups in the investment information network and their success, it does not imply causality. Future research could further explore the causal mechanisms between network structure and startup success. Finally, this study primarily examines successful startups, with limited analysis of failed ones. Future studies could incorporate failed startups to compare the network structure differences between successful and failed enterprises, offer-

ing more comprehensive insights into the factors influencing startup success.

Program availability

The relevant code of this article has been published on GitHub at https://github.com/forever-wy/01 and is also openly available in Science Data Bank at https://doi.org/10.57760/sciencedb.23263. The example dataset used in this paper is available upon request.

Acknowledgements

The authors are grateful for the funding from the National Natural Science Foundation of China (Grant Nos. 42001236, 71991481, and 71991480) and Young Elite Scientist Sponsorship Program by Bast (Grant No. BYESS2023413). The authors would like to express their gratitude to Jianhe Guan who helped a lot during this work with precious analysis suggestions.

References

[1] Fanning K and Centers D P 2016 *Journal of Corporate Accounting & Finance* **27** 53
[2] Park G, Shin S R and Choy M 2020 *Journal of Business Research* **109** 64
[3] Katila R, Chen E L and Piezunka H 2012 *Strategic Entrepreneurship Journal* **6** 116
[4] Hellmann T and Puri M 2002 *The Journal of Finance* **57** 169
[5] Song L Z, Song M and Parry M E 2010 *Journal of Product Innovation Management* **27** 130
[6] Jung Y and Pawlowski S 2015 *Telematics and Informatics* **32** 193
[7] Chung W Y, Jo Y and Lee D 2021 *Telematics and Informatics* **56** 101482
[8] Hasani T, Bojei J and Dehghantanha A 2017 *Telematics and Informatics* **34** 655
[9] Xu Y, Li X, bin Mustakim F, Alotaibi F M and Abdullah N N 2022 *Information Processing & Management* **59** 103055
[10] Koster S and van Stel A 2014 *Papers in Regional Science* **93** 203
[11] Lin W S, Chen H R and Yang M L 2020 *Telematics and Informatics* **52** 101428
[12] Lim S, Kwon O and Lee D H 2018 *Telematics and Informatics* **35** 1887
[13] Wu S, Chen R, Wei G, Gao X and Huo L 2021 *Complexity* **2021** 1
[14] Wadhwa A, Phelps C and Kotha S 2016 *Journal of Business Venturing* **31** 95
[15] Wassmer U 2010 *Journal of Management* **36** 141
[16] Yang Y, Narayanan V K and De Carolis D M 2014 *Strategic Management Journal* **35** 1993
[17] Liu T, Qu S and Scherpereel C M 2020 *Sustainability* **12** 491
[18] Butler J S, Garg R and Stephens B 2020 *Information Systems Research* **31** 198
[19] Xu S, Zhang Q, Lü L and Mariani M S 2020 *Information Sciences* **515** 103
[20] Baum J A, Calabrese T and Silverman B S 2000 *Strategic Management Journal* **21** 267
[21] Bonaventura M, Ciotti V, Panzarasa P, Liverani S, Lacasa L and Latora V 2020 *Scientific Reports* **10** 345
[22] Sherehe S, Wan H and Gu C 2022 *Chin. Phys. B* **31** 018902
[23] Chai Y, Wang Y G and Yan J 2023 *Chin. Phys. B* **32** 090202
[24] Abbasi A, Hossain L and Leydesdorff L 2012 *Journal of Informetrics* **6** 403
[25] Starnini M, Baronchelli A, Barrat A and Pastor-Satorras R 2012 *Phys. Rev. E* **85** 056115
[26] Zhining L, Dawei Z, Yada Z, Jinjie G, and Jingrui H 34*th AAAI Conference on Artificial Intelligence*, February 7–12, 2020, New York, USA, p. 4973
[27] Holme P and Saramäki J 2012 *Physics Reports* **519** 97
[28] Stephens H M, Mack E A and Mann J 2022 *Telematics and Informatics* **74** 101873
[29] Keidel J, Bican P M and Riar F J 2021 *Sustainability* **13** 6184
[30] Sullivan D M, Marvel M R and Wolfe M T 2021 *Technovation* **101** 102209
[31] Balachandran S 2024 *Strategic Management Journal* **45** 1117
[32] Liu W, Xu Y, Wu C H and Luo Y 2023 *Information Processing & Management* **60** 103406
[33] Gao J, Cheng Y, He H and Gu F 2021 *Entrepreneurship Research Journal* **13** 31
[34] Yang J, Zhu M, Zhang M and Yao K 2021 *Journal of Management & Organization* **30** 2275
[35] Ferratti G M, Neto M S and Candido S E 2021 *Technology in Society* **66** 101623
[36] Passaro R, Quinto I, Rippa P and Thomas A 2020 *Sustainability* **12** 9437
[37] Masuda N, Porter M A and Lambiotte R 2017 *Physics Reports* **716** 1
[38] Lambiotte R and Rosvall M 2012 *Phys. Rev. E* **85** 056107
[39] Delvenne J C, Yaliraki S N and Barahona M 2010 *Proc. Natl. Acad. Sci. USA* **107** 12755
[40] Petit J, Lambiotte R and Carletti T 2019 *Applied Network Science* **4** 1
[41] Moinet A, Starnini M and Pastor-Satorras R 2019 *New J. Phys.* **21** 093032
[42] Gueuning M, Cheng S, Lambiotte R and Delvenne J C 2020 *Journal of Complex Networks* **8** cnz027
[43] Newman M E 2005 *Social Networks* **27** 39
[44] Tokito S, Kagawa S and Hanaka T 2022 *Economic Systems Research* **34** 111
[45] Bloodgood J M, Hornsby J S, Rutherford M and McFarland R G 2017 *International Entrepreneurship and Management Journal* **13** 525
[46] Brandes U, Borgatti S P and Freeman L C 2016 *Social Networks* **44** 153
[47] Keidel K, Ettinger U, Murawski C and Polner B 2022 *Journal of Research in Personality* **96** 104166
[48] Lutter M and Weidner L 2021 *Poetics* **87** 101535
[49] Arroyo J, Corea F, Jimenez-Diaz G and Recio-Garcia J A 2019 *Ieee Access* **7** 124233
[50] Gangwani D and Zhu X 2024 *Artificial Intelligence Review* **57** 44
[51] Kim J, Kim H and Geum Y 2023 *Technological Forecasting and Social Change* **193** 122614
[52] bikowski K and Antosiuk P 2021 *Information Processing & Management* **58** 102555
[53] Kwon O, Yun S G, Han S H, Chung Y H and Lee D H 2018 *Computational Economics* **51** 847
[54] Nicola F 2017 *IEEE International Conference on Internet of Things*, June 21–23, 2017, Exeter, United Kingdom, p. 578
[55] Kuo C C and Shyu J Z 2021 *Sustainability* **13** 6893
[56] Kim H and Anderson R 2012 *Phys. Rev. E* **85** 026107