How to Better Identify Venture Capital Network Communities: Exploration of A Semi-Supervised Community Detection Method

Hong Xiong  
School of Systems Science, Beijing Normal University, Beijing 100875, China

Ying Fan  
School of Systems Science, Beijing Normal University, Beijing 100875, China

Follow this and additional works at: https://dc.tsinghuajournals.com/journal-of-social-computing

Part of the Computer Engineering Commons, Computer Sciences Commons, and the Science and Technology Studies Commons

Recommended Citation
Hong Xiong, Ying Fan. How to Better Identify Venture Capital Network Communities: Exploration of A Semi-Supervised Community Detection Method. Journal of Social Computing 2021, 2(1): 27-42.
How to Better Identify Venture Capital Network Communities: Exploration of A Semi-Supervised Community Detection Method

Hong Xiong and Ying Fan

Abstract: In the field of Venture Capital (VC), researchers have found that VC companies are more likely to jointly invest with other VC companies. This paper attempts to realize a semi-supervised community detection of the VC network based on the data of VC networking and the list of industry leaders. The main research method is to design the initial label of community detection according to the evolution of components of the VC industry leaders. The results show that the community structure of the VC network has obvious distinguishing characteristics, and the aggregation of these communities is affected by the type of institution, the source of capital, the background of personnel, and the field of investment and the geographical position. Meanwhile, by comparing the results of the semi-supervised community detection algorithm with the results of community detection using extremal optimization, it can be shown to some extent that the semi-supervised community detection results in the VC network are more accurate and reasonable.

Key words: Venture Capital (VC); complex network; community detection

1 Introduction

In existing studies of community detection, quality functions, such as modularity, were used as the main evaluation indicators for the quality of community detection and to guide the process of community detection[1]. However, in the actual large-scale network, it is difficult to find the so-called real community, and the results obtained from community detection only by using the optimization of quality functions, such as modularity, cannot meet the needs of specific practical research[2, 3]. Therefore, this paper hopes to add a priori information into the network to increase the accuracy of community detection. Furthermore, the traditional community detection method is an unsupervised method, which only finds the community structure in the network according to the topological structure information of the network, which is the connection relationship between nodes[4]. Real networks can be imprecise or incomplete, and the accuracy of these traditional community detection methods depends on the accuracy of the network itself. Thus, when there is noise in the network, the accuracy of the community detection is reduced. In practice, we can often obtain data other than structural information through which we can directly or indirectly obtain some a priori information about the community structure, including node labels and node pair constraints. How to use a priori knowledge to improve the ability of community detection is a very key issue[5].

The industrial network with various cooperative relationships can be described as a relationship network, one representative of which is the joint investment network of the Venture Capital (VC) industry (hereinafter referred to as VC network). The VC industry is one with extremely asymmetric information, which leads to a high imbalance in resource allocation. Under this extremely unstable environment, the formation of inter-enterprise network is beneficial to the survival rate of enterprises. Therefore, Chinese VC companies that cannot obtain institutional advantages in the formal system have to reduce risks by establishing relationship networks[6]. VC firms tend to co-invest with others to obtain judgements on investment projects, lower
investment risks and information asymmetry, and form reciprocal relationship to sugar up their performance\[^7\]. VC companies that jointly invest form links to constitute a joint investment network for the VC industry\[^8\]. VC has the characteristics of high risk and high return, and it is virtually impossible for a VC company to have investment experience in various industries. Therefore, it is an inevitable trend for VC companies to cooperate with others to give full play to their respective advantages in the form of win-win cooperation. This network helps VC companies to share information resources, improve their competitive advantages, and increase the benefits of invested start-up enterprises.

Based on a priori information implied in the VC network of the VC industry, this paper designs a semi-supervised community detection method to be applied to the VC network, and explains the effectiveness of this method, as well as how this method is superior to Weighted Extremal Optimization (WEO) community detection method in terms of accuracy and rationality of community detection. Further, in order to explore the community structure and characters of the VC network, this paper analyzes the semi-supervised community detection results of the VC network, compares them with the unsupervised results applied to the same VC network, and tries to contribute to key knowledge in the field of VC industry based on these discoveries. Two pieces of a priori information were used to conduct the semi-supervised community detection of VC network for this paper. The first is that industry leaders in Chinese VC industry will form small clusters, and some VC industry circles will be formed around VC industry leaders, which will further gather into communities\[^9\]. The second is the list of VC industry leaders in the Chinese VC industry which were found through the methods of expert opinion research and qualitative data mining\[^10\]. We decided to use information about how industry leaders form clusters and how their circles realize a semi-supervised community detection for the VC network.

2 Semi-Supervised Community Detection Method

2.1 Structural characteristics of the Chinese VC industry

Since the 1980s, the VC industry has been growing rapidly, and a large number of VC companies have sprung up. Joint investment activities between them have become increasingly common, and the cooperative relationship between VC companies has shown significant network characteristics. However, in the research of VC network, few scholars have studied the community structure of the VC network. Decisions on Reform of Science and Technology System, issued by Chinese government in March 1985, marked the official start of Chinese VC industry, soon after which, a number of domestic VC firms were established and a number of foreign VC firms entered the Chinese mainland market. However, most domestic VC companies lacked the proper methods to exit and thus ended in failure. It was not until the No.1 Proposal, Proposal on Developing Venture Capital in China as soon as Possible, proposed by Cheng Siwei passed at the Chinese People’s Political Consultative Conference (CPPCC) in 2001 that Chinese domestic VC companies really developed. In this initial stage of development, the Chinese VC industry presented a distinct characteristic whereby foreign VC companies and domestic VC companies would form clusters separately and did not co-invest with each other. However, with the establishment of domestic VC’s dollar fund and the exit channel’s expansion and development of the RMB fund, the joint investment behavior between foreign and domestic VC companies gradually became apparent\[^11\], and the community structure of VC industry presents a dynamic characteristic of blending.

Since VC industry began to develop in China, it gradually came to display unique investment trends and business logic that differed from the Western VC industry. Comparatively speaking, Chinese VC companies are more inclined to complete joint investment among VC companies by establishing relationships and circles. In addition, although influential investors have their own circles, the VC industry itself is a small-world network, that is to say, there are often bridges between circles to connect them, and the role of bridge is usually the leader of important circle networks\[^12\]. VC network has the characteristics of small world, and VC companies are easy to huddle up to form joint investment. The VC network has an obvious community structure, in which big differences exist between the VCs of some core communities in terms of location and type of capital. Furthermore, the start-up enterprises which these VCs invest in also differ in regards to the location, development stage, and investment industry\[^13\].

Previous methods of semi-supervised community detection proposed by previous studies are no longer
applicable to the VC network. The majority of existing researches adopt traditional unsupervised community detection methods. Applying analysis via traditional community detection algorithms on the VC network cannot ensure accuracy and rationality, and can only provide weak conclusion. In addition, due to the lack of multi-dimensional comparative analysis in previous studies, the implicit information and dynamic behavior of VC communities were not observed. These remaining problems need to be solved. Therefore, this paper proposes a new semi-supervised community detection method for VC network—the concept that the VC network should be divided into communities with the idea of semi-supervision. By comparing and analyzing the results of different types of community detection, this paper hopes to improve the accuracy and rationality of detection results via the semi-supervised community detection of VC network, and to reveal the phenomenon of VC network community and other key principles which previous studies failed to reveal.

2.2 Semi-supervised community structure detection algorithm

Machine learning can be divided into three categories: supervised learning, unsupervised learning, and semi-supervised learning. The applied conditions of semi-supervised learning were made between supervised learning and unsupervised learning, with only a small amount of labeled data. Researchers can try to gather useful information from small amount of labeled data and a large amount of unlabeled data. Currently, a priori information forms which are used in semi-supervised community detection can be divided into two categories: individual labels and pairwise constraints. The guidance form of individual labels considers that nodes with the same label belong to the same community. Pairwise constraints can help determine whether a pair of individuals belong to the same community (must-link constraints) or different communities (cannot-link constraints).

Label propagation algorithm is a semi-supervised learning method based on graph. The basic idea behind it is to use the label information of marked nodes to predict the label information of unmarked nodes. The relationship between samples is used to establish a complete graph model of the relationship. In the complete graph, nodes include labeled and unlabeled data, and edges represent the similarity between two nodes. The labels of nodes are passed between nodes according to the similarity. Label data are like a source, which can mark unlabeled data. The greater similarity between nodes means that it is easier for the label to spread between them.

Louvain algorithm is an algorithm based on multi-level round by round heuristic iterative modularity optimization. The modularity function was originally used to measure the quality of the results of community detection algorithm. It can describe the tightness of the community found by the algorithm, so the modularity function can be used as an optimization function. This means that while the node is added to the community where one of its neighbors is, if the modularity value of the current community structure can be improved, the iterative optimization is acceptable. The definition of modularity of community detection is as follows:

$$Q = \frac{1}{2m} \sum_{ij} (A_{ij} - P_{ij}) \delta(C_i, C_j)$$ (1)

where $m$ is the total number of edges in the network, $A_{ij}$ represents the edge weight between nodes $i$ and $j$, $P_{ij}$ is the expectation of the weight of the edges of nodes $i$ and $j$ in the empty model, and $\delta$ function means that when nodes $i$ and $j$ belong to the same community ($C_i = C_j$), the value of $\delta$ is 1, otherwise it is 0.

The idea of Extremal Optimization (EO) algorithm is derived from the idea of self-organized criticality, which is a heuristic search method inspired by the Bak-Snappen model of biological co-evolution. According to the evolution process of natural species, the least adaptable species in a community are either eliminated or go through mutation in order to survive. Suppose there are $n$ species in the community and the adaptation value of species is $\lambda_i \in [0, 1]$. At each step of evolution, the species with the poorest adaptation value are selected, and their adaptation value is updated to a random number between 0 and 1. In EO community detection, the global variable to be optimized is modularity $Q$, and the definition of the local variable used in the extremal optimization problem should be related to the contribution of a single node $i$ to the sum of $Q$ values when a particular community detection of network is given in Eq. (1). At each time step of the heuristic search evolution, the system self-organizes by moving nodes with low adaptation value (extremal) from one community to another. This process is repeated until the optimal state corresponding to the maximum $Q$ value is reached.

Label propagation algorithm is a classical semi-supervised learning method. According to the principle
of community detection algorithm, the detection result of Louvain algorithm can better reflect the community-core effect than that of label propagation algorithm. Furthermore, considering that the label propagation algorithm itself ignores the topological structure of network, there is a certain degree of error between the detection result and reality. The Louvain algorithm based on the topological structure of network makes up for this unreliability. Conversely, if only the Louvain algorithm is used, there would be too many kinds of communities, making detection unreasonable. The initial label information provided by the label propagation algorithm can improve the detection rationality of the Louvain algorithm. To sum up, combining the two algorithms, we can initially design a semi-supervised community detection algorithm (Label Propagation Algorithm (LPA) + Louvain community detection algorithm, hereinafter referred to as LL algorithm) based on label propagation and modularity optimization algorithm, which is added with the a priori information of community-core’s initial label. Meanwhile, this paper attempts to show that the results of LL algorithm are more reasonable by comparing the detection results of LL algorithm and EO algorithm in the VC network between the years of 2000 and 2013.

2.3 Data description

The original adjacency list data of the VC network were obtained from Qingke SiMuTong Database, which includes a list of investment activities in the form of “a VC company invests in a start-up enterprise one day”. From this list, we can establish the adjacency list data of VC joint investment in each year from 2000 to 2017, enabling us to build adjacency matrix for calculation and analysis. The research object of the proposed semi-supervised community detection is the largest component, with 1950 VC nodes in the VC network during the period of 2000–2013. The VC network in the period of 2000–2013 has 2218 nodes, and the sum of its edge weights is 8690. Its largest component has 1950 nodes (accounting for 87.9%), and the sum of edges’ weight is 8502 (accounting for 97.8%).

This paper is based on a cooperative project. According to research results of Ref. [12], this research had obtained the list of 42 (first-class) VC industry leaders during the period of 2000–2013 through data mining of the largest component of VC network and the selection of VC industry’s own indicators, combined with the rooted truth of interviews with VC industry professionals[12].

3 Semi-Supervised Community Detection of VC Network

3.1 Design of semi-supervised community detection algorithm

The initial labels were designed based on the cluster information of VC industry leaders. The VC nodes with initial labels are used as part of the input of the LPA, and the output of LPA algorithm is used as the initial community relationship vector of the modular optimization algorithm. On this basis, the output of the modularity optimization algorithm is an optimal community detection result under current conditions. This process is the concrete embodiment of the semi-supervised community detection method (LL algorithm) proposed in this paper.

The algorithm consists of two stages:

The first stage (label propagation):

1. Initialization: build edge weight matrix $w_{ij}$, and get the similarity between nodes.
2. According to $w_{ij}$ obtained by Step 1, use Eq. (2) to calculate the propagation probability from node $j$ to $i$.
3. A standard matrix $Y$ of $l \times u \times D$ dimension is defined ($l$ is the number of labeled nodes; $u$ is the number of unlabeled nodes; and $D$ is the number of label categories).
4. Each node adds the labeled values of its surrounding nodes’ propagation by edge weight according to the propagation probability and updates its own probability distribution.
5. The probability distribution of labeled nodes is reassigned to the initial value. Repeat Step 4 until the convergence emerges, and output the labels of all nodes which represent their communities.

The second stage (modular optimization algorithm):

6. Initialization: take the label information of all nodes obtained in Step 5 as the initial basis of community detection, and divide each node into corresponding communities.
7. For each node $i$, try to allocate node $i$ to the community of each neighbor node in turn, and calculate the change of modularity $\Delta Q$ after allocation according to Eq. (3). If $\Delta Q$ is positive, accept the allocation; otherwise, abandon.
8. Repeat Step 7 until the modularity cannot be increased.
(9) Recombine all nodes in the same community into a new community, and continue with Steps 7 and 8 until the community structure no longer changes.

\[ T_{ij} = P(j \rightarrow i) = \frac{w_{ij}}{\sum_{k=1}^{n} w_{kj}} \]  

(2)

\[ Q = \frac{1}{2M} \sum_{i,j} \left( A_{ij} - \frac{k_{i}k_{j}}{2M} \right) \delta(c_i, c_j) \]  

(3)

where both \( T_{ij} \) and \( P(j \rightarrow i) \) represent the propagation probability of node \( j \) to \( i \). \( M = \frac{1}{2} \sum_{i,j} A_{ij} \) is the sum of weights of all edges in the network. \( k_{i} = \sum_{j} A_{ij} \) is the sum of the weights of all the edges connected to node \( i \). \( k_{j} = \sum_{i} A_{ij} \) is the sum of the weights of all edges connected to node \( j \). \( c_{i} \) represents the community to which node \( i \) is assigned and \( c_{j} \) represents the community to which node \( j \) is assigned.

\[ \delta(c_i, c_j) = \begin{cases} 
1, & i \text{ and } j \text{ are divided into the same community;} \\
0, & i \text{ and } j \text{ are divided into different communities.} 
\end{cases} \]

According to the above algorithm steps, the semi-supervised community detection algorithm flow is designed (see Fig. 1).

### 3.2 Design initial labels

From the data of VC network in the period of 2000–2013, we extracted data of 42 VC industry leaders, and then established the network of 42 VC industry leaders.

---

Fig. 1  Flow chart of semi-supervised community detection algorithm based on label propagation and modularity optimization algorithm combining a priori information of community-core’s initial label.
The knowledge of how VC industry leaders cluster was calculated and analyzed, served as a priori information for the semi-supervised community detection algorithm.

We built a weighted network of 42 VC industry leaders, for those edges with a certain weight of $k_{ij}$, set the edge breaking condition $k_{ij} < b (b = 1, 2, \ldots, 21)$. For $b = 1, 2, \ldots, 21$ (the biggest weight of these edges is 20) while $k_{ij} < b$, cut the connection between nodes $i$ and $j$, and obtain the corresponding network of 42 VC industry leaders. The parameter $b$ is the edge breaking condition threshold, which represents a measure of the degree of testing the relationship between VC nodes. When $b$ is larger, it indicates that the greater the degree of testing the relationship between VC nodes is, the more difficult it is to establish connections between VC nodes. When the value of $b$ changes from small to large, it means that the relationship between VC nodes is facing more and more severe tests, and more and more relationship edges are broken because they can not withstand the test. Different components could be divided from the network. By arranging these 21 situations, we can see the evolution trend of components in the network of these 42 VC industry leaders. Among them, for the VC nodes that “go all the way to the end hand in hand”, constraint pairs were added to them as a priori information of semi-supervised community detection. Figure 2 shows the evolution of the components of the VC network of 42 VC industry leaders, which actually reflects how they cluster. From this, we can extract constraint pairs of VC industry leaders as shown in Table 1. All VC companies appearing in this paper are referred to in abbreviations consisting of English capital letters, such as SCG for Shenzhen Capital Group and IT for iD TechVentures. A detailed list of abbreviations and English names of VC companies can be found in Table A1 in the Appendix.

In Fig. 2, each column represents a situation, and each row represents the community affiliation of a VC company and its changes. Each grid in Fig. 2 corresponds to the state of VC company in the current situation, and different colors represent different components. The blank grids in Fig. 2 represent isolated nodes in the network. In addition to a visual figure, there is also a trend chart of the number of components of the 42 VC industry leaders, as shown in Fig. 3.

From Fig. 3, we can see two points with relatively
Table 1 Constraint pairs among 42 VC industry leaders.

| Constraint pair | Constraint pair  |
|-----------------|-----------------|
| SCG-GPCP        | ICIC-SC         |
| DFJ-DCMCL       | LCM-DCM         |
| ITC-JI          | NLVC-NEA        |
| SOYH-IG         | JVCIMC-QWVCM    |

Fig. 3 Variation tendency of number of components of 42 VC industry leaders.

large change values, corresponding to $b = 4$ and $b = 9$, that is, from $b = 3$ to $b = 4$ and from $b = 8$ to $b = 9$, the number of components suddenly increases. At the same time, we can see in Fig. 2 that the disappearance of the color of the VC node’s grid mostly occurs in $b = 4$ and $b = 9$. That is to say, in these two “moments”, the relationship between VC industry leaders faces its most severe test throughout the whole relationship development period.

The initial labels were designed as follows:

Firstly, we found the corresponding situation when the number of components changes most obviously in Fig. 2 (this corresponds to column 9 in Fig. 2, which has the most different types of the color). We then translated other colors in other columns to column 9, and finally deleted other columns and blank points in column 9. The remaining 19 VC nodes with color (except blank) can be initially labeled according to their colors, as shown in Table 2.

Chinese VC industry leaders play an important role in the industry, and when multiple industry leaders cooperate frequently, their followers will naturally have cooperative relationships and participate in the formation of communities with these industry leaders as the core. Therefore, the cooperation among industry leaders is of more significance in community detection. However, it is difficult to determine which companies are industry leaders by objective data. Although the company’s status is usually represented by the feature centrality of the network, it is different from investment experience and performance in that it reflects people’s cognition of the company’s status in the social hierarchy system\(^1\). Compared with the algorithm simulation of joint investment, the views of industry insiders can better capture the company’s position. Therefore, by harnessing expert opinion investigation and research, VC industry experts were invited to choose whether they agreed that a VC company on the list should be deemed as a leading company in the industry. The final list of leading companies obtained from this method better identifies VC industry leaders than by relying on simple network data. After focusing on the more solid cooperation relationships among leading companies selected by industry insiders in the network model, we can significantly improve the accuracy of community detection\(^2\).

Table 2 Initial labels designed according to the cluster situation of 42 VC industry leaders.

| VC company | Initial label | VC company | Initial label |
|------------|---------------|------------|---------------|
| SCG        | 1             | NLVC       | 4             |
| SFVC       | 1             | KPCB       | 4             |
| GPCP       | 1             | NEA        | 4             |
| ICIC       | 2             | DCMCL      | 5             |
| SC         | 2             | DFJ        | 5             |
| QWVCM      | 2             | ITC        | 6             |
| JVCIMC     | 2             | JI         | 6             |
| LCM        | 3             | SOYH       | 7             |
| IT         | 3             | IG         | 7             |
| DCM        | 3             |            |               |
3.3 Semi-supervised community detection in VC network

Using the seven initial labels shown in Table 2 as the input of LL algorithm, we apply them to the largest component of VC network in the period of 2000–2013 to acquire a semi-supervised community detection result. The largest component has a total of 1950 nodes, of which 194 VC nodes’ labels change after completion of the second stage of LL community detection algorithm, compared with the export of the first stage (the results of LPA algorithm, as shown in Table 3).

The community detection result shown in Table 4 represents a classic semi-supervised community detection result according to 19 VC companies with initial labels obtained from the evolution of components of 42 VC industry leaders and LL semi-supervised community detection algorithm. In this result, apart from the small size of the LL6 and LL7 communities, the other five communities have a certain scale and are gathered together according to certain characteristics.

As can be seen from Fig. 3, when \( b = 4 \) and \( b = 9 \), the number of components in the VC network of 42 VC industry leaders suddenly increased, which means that in these two “moments”, the relationship between VC industry leaders faces its most severe test throughout their whole relationship development period. However, \( b = 4 \) occurs in the early stage during the period when the number of VC nodes that are separated from the industrial leader cluster is small, which is excluded by the industrial leader cluster earlier. Based on this, we suggest that the VC companies with \( b \geq 4 \) (VC companies of which number of joint investments is four or higher) have a relatively close relationship. In order to more clearly show the results of LL community detection, we consider 42 VC industry leaders and 57 VC companies that have invested with them on four or more times, with the corresponding results shown in a visual network diagram (Fig. 4). Different colors in Fig. 4 show different communities. The connection edge is established according to amount of times that joint investment was conducted of these 99 VC companies during the period of 2000–2013. The community is determined according to the LL detection result of the largest component of VC network in the period of 2000–2013.

Specifically, Community LL1 is a large-scale community led by SCG and dominated by domestic VC companies. Community LL2 is led by ICIC and SC and dominated by foreign VC companies, which gathers many VC elites. The two leaders of Community LL3 are backed by the world’s top Personal Computer (PC) manufacturers in China, and they have combined a number of domestic and foreign VC companies (the predecessor of LCM is “Lenovo investment”, and the original name of IT is “Acer venture capital”).

| VC industry leader | Community number | Number of VC nodes | Proportion (%) |
|-------------------|------------------|--------------------|---------------|
| SCG, SIDVCH       | LPA1             | 984                | 50.46         |
| ICIC, SC          | LPA2             | 573                | 29.38         |
| LCM, IT           | LPA3             | 113                | 5.79          |
| NLVC, KPCB        | LPA4             | 133                | 6.82          |
| DCMCL, DFJ        | LPA5             | 53                 | 2.72          |
| ITC, JI           | LPA6             | 47                 | 2.41          |
| SOYH              | LPA7             | 47                 | 2.41          |

| VC industry leader | Community number | Number of VC nodes | Proportion (%) |
|-------------------|------------------|--------------------|---------------|
| SCG, SFVC         | LL1              | 943                | 48.4          |
| ICIC, SC          | LL2              | 514                | 26.4          |
| LCM, IT           | LL3              | 158                | 8.1           |
| NLVC, KPCB        | LL4              | 142                | 7.3           |
| DCMCL, DFJ        | LL5              | 84                 | 4.3           |
| SUVC               | LL6              | 71                 | 3.6           |
| SOYH, SIDVCH      | LL7              | 38                 | 1.9           |

| VC industry leader | Community number | Number of VC nodes | Proportion (%) |
|-------------------|------------------|--------------------|---------------|
| SCG, SFVC         | LL1              | 943                | 48.4          |
| ICIC, SC          | LL2              | 514                | 26.4          |
| LCM, IT           | LL3              | 158                | 8.1           |
| NLVC, KPCB        | LL4              | 142                | 7.3           |
| DCMCL, DFJ        | LL5              | 84                 | 4.3           |
| SUVC               | LL6              | 71                 | 3.6           |
| SOYH, SIDVCH      | LL7              | 38                 | 1.9           |
Community LL4 is also dominated by foreign VC companies and has a strong American background in corporate management. Community LL5 can be seen as the leader of foreign VC. DCMCL and DFJ are building their own lineage and their investment fields greater overlap. The top three leaders of Community LL7 are all headquartered in Suzhou, Jiangsu Province, and PCIM, which follows closely, pays more attention to the development of Suzhou Industrial Park. In a word, among other factors, we can see that the aggregation of these communities is affected by the type of institution, capital sources, personnel background, investment fields, and geographical location.

On the whole, the Chinese VC industry shows a typical clustering culture dominated by homogeneity\cite{20}. According to Fig. 4, Table 3, and related attributes of VC companies, two central and representative communities exist, among which Community LL1 is the largest community in the Chinese VC industry, and its head nodes are mainly composed of domestic VC companies located in Shenzhen. It represents a community based on security and stability that fully embodies the characteristics of the domestic VC circles. Community LL2, although mainly composed of VC companies from USA, has also absorbed domestic VC companies with international investment as the main investment focus, strategic investment with companies with IT backgrounds, and even industrial funds led by government. Their institutional headquarters are mostly located in large cities, such as Beijing and Shanghai, which fully reflect the integration of VC companies with different backgrounds, representing a more diverse community driven by the need for complementary resources.

But meanwhile, it should be noted that the number of communities detected by the classic initial labels is low and the detection is not sufficiently detailed. Therefore, in the subsequent research, we can consider whether we can detect more communities by increasing the number and types of initial labels, such as in some hidden small communities and remote communities.

### 4 Comparison of Community Detection Results

#### 4.1 A comparison represented by industry leaders and their close followers

This paper compares the unsupervised result with the result of LL algorithm in VC network during the period of 2000–2013 to show that the semi-supervised community detection result combined with a priori information (initial label) is better than the traditional unsupervised community detection result in terms of accuracy and rationality.

According to the EO community detection of VC network in the period of 2000–2013, we obtained 14 communities, as shown in Table 5.

In order to specifically illustrate the results of the LL community detection, we considered the capital types and institutional headquarters location attributes of 42 VC industry leaders and 57 VC companies that have jointly invested with them on four or more times.

For the community detection results (hereinafter referred to as LL detection and EO detection) of LL and EO algorithms of these 99 VC companies, the associations between the community to which the VC company belongs and the capital type are shown in Tables 6 and 7.

Next, we introduce the Aggregation Degree

### Table 5 EO community detection results.

| VC industry leader | Community number | Number of VC nodes | Proportion (%) |
|--------------------|------------------|--------------------|----------------|
| ICIC, LCM          | EO1              | 371                | 19.0           |
| SCG, NV            | EO2              | 267                | 13.7           |
| SFVC, KJCH         | EO3              | 275                | 14.1           |
| GPCP, JGCG         | EO4              | 312                | 16.0           |
| DCMCL, SOYH        | EO5              | 227                | 11.6           |
| CVS, DC            | EO6              | 147                | 7.5            |
| SC, ZV             | EO7              | 124                | 6.4            |
| KPCB, WI           | EO8              | 100                | 5.1            |
| SLC, THPEMP        | EO9              | 64                 | 3.3            |
| STVC, BCEG         | EO10             | 35                 | 1.8            |
| PCIM, SVG          | EO11             | 17                 | 0.9            |
| SSMEVCH, XXII      | EO12             | 6                  | 0.3            |
| WEIM               | EO13             | 2                  | 0.1            |
| GPE, PMC           | EO14             | 3                  | 0.2            |

### Table 6 Association of communities and capital types in LL detection of 42 VC industry leaders and 57 VC companies that have jointly invested with them on four or more times.

| Community number | Number of VC nodes | Including domestic VC | Including joint VC | Including foreign VC |
|------------------|--------------------|-----------------------|--------------------|---------------------|
| LL1              | 24                 | 21                    | 1                  | 2                   |
| LL2              | 48                 | 8                     | 0                  | 40                  |
| LL3              | 9                  | 1                     | 1                  | 7                   |
| LL4              | 8                  | 1                     | 0                  | 7                   |
| LL5              | 2                  | 0                     | 0                  | 2                   |
| LL6              | 1                  | 0                     | 1                  | 0                   |
| LL7              | 7                  | 6                     | 0                  | 1                   |
Table 7  Association of communities and capital types in EO detection of 42 VC industry leaders and 57 VC companies that have invested with them on four or more times.

| Community number | Number of VC nodes | Number of capital types Including domestic VC | Including joint VC | Including foreign VC |
|------------------|-------------------|-----------------------------------------------|-------------------|---------------------|
| EO1              | 35                | 5                                             | 1                 | 29                  |
| EO2              | 8                 | 6                                             | 1                 | 1                   |
| EO3              | 5                 | 5                                             | 0                 | 0                   |
| EO4              | 8                 | 7                                             | 0                 | 1                   |
| EO5              | 20                | 4                                             | 1                 | 15                  |
| EO6              | 3                 | 2                                             | 0                 | 1                   |
| EO7              | 8                 | 4                                             | 0                 | 4                   |
| EO8              | 10                | 2                                             | 0                 | 8                   |
| EO10             | 1                 | 1                                             | 0                 | 0                   |
| EO11             | 1                 | 1                                             | 0                 | 0                   |

Coefficient (ADC) of related attributes:

\[
ADC = \sum_{v=1}^{V} \frac{N_v}{N}
\]  

(4)

where \(N\) represents the total number of VC nodes considered, \(N_v\) represents the number of VC companies contained in the attribute with the largest proportion in the community \(v\), and \(V\) represents the number of communities considered.

It is calculated that the capital type ADC of these 99 VC companies in LL detection is 0.848; the capital type ADC of these 99 VC companies in EO detection is 0.788. The aggregation degree coefficient of LL detection is higher than that of EO detection, which reflects the higher aggregation degree of capital types in LL detection results, and to a certain extent, indicates that LL detection results are more reasonable.

At the same time, according to the number of VCs included in the relevant attributes, we mainly made an analysis on the aggregation degree of the 99 VC companies which are headquartered in China (Shenzhen), China (Shanghai), and America, as shown in Fig. 5, in the form of “peak map”. In Fig. 5, the vertical coordinate is the number of VC companies in the corresponding community, and the horizontal coordinate of its maximum value is set to be 0, and other values are arranged in descending order on both sides according to the number of vertical coordinates.

We can create the “peak map” as shown in Fig. 5. It can be seen from Fig. 5 that the aggregation degree of headquarters of these 99 VC companies in the LL detection is higher than that in the EO detection. This is intuitively reflected by the more concentrated curve corresponding to the LL detection in Fig. 5, which is similar to the peak shape with sharper angle.

4.2 A comparison represented by head VC nodes

Then we took another method to compare the results
of community detection. Initially, in order to rank the VC nodes in the community, we designed a new sorting index. Firstly, three kinds of node centrality indexes (degree, kcoreness, and closeness) were calculated and normalized. Then the proportion coefficient was distributed according to \(0.4:0.3:0.3\). Three centrality values were multiplied by the proportion coefficient to get a new centrality index value NCI (New Centrality Indicators). We found the nodes of which NCI is greater than 0.6 in LL detection and EO detection results, and listed them in Tables 8 and 9. NCI is set to better sort VC nodes according to their importance, and consider degree, kcoreness, and closeness, and give the degree index a relatively higher weight, because the impact of degree on node importance is relatively greater in VC network. Before the NCI is calculated, three indicators have been normalized, so the corresponding proportion coefficient of the three indicators should be equal to 1.

How can we compare the results of the above two kinds of community detections to show their comparative advantages and disadvantages? To do so, we look at EO detection based on LL detection. Assuming that LL detection is accurate, there are at least three errors in EO detection:

(1) ICIC should not be linked with LCM;
(2) DCMCL should not be linked with SOYH;
(3) NLVC and KPCB should not be separated.

Then we calculate the distance between these VC nodes.

From the calculation results in Table 10, it can be seen that the distance between ICIC and LCM is significantly greater than that between ICIC and SC. The distance between DCMCL and SOYH is significantly greater than that between DCMCL and DFJ. The distance between NLVC and ICIC is significantly greater than that between NLVC and KPCB. Table 9 shows that ICIC and LCM are joining, but ICIC and SC are not; NLVC and ICIC are joining, but NLVC and KPCB are not. This implies that the three errors of the above EO detection have been confirmed to some extent.

The average distance of 42 VC industry leaders is 0.6418, and the average distance of 19 VC industry leaders with initial labels is 0.4217. Meanwhile, we calculate the two-point distance of constraint pairs in VC industry leaders assumed in Section 3.2, as shown in Table 11.

It is calculated that the average value of the distance between two nodes of constraint pairs in these VC industry leaders is \(d_c = 0.1334\). To some extent, this value can be used as a measure of whether a pair of VC nodes constitute a constraint relationship. In other words, when the distance between a pair of VC industry leaders is less than \(d_c\), it is more reasonable to form a constraint relationship and be divided into the same community. When the distance between a pair of VC industry leaders is greater than \(d_c\), it does not constitute

### Table 8 VC companies with NCI greater than 0.6 in LL detection results.

| Community number | VC company                  |
|------------------|-----------------------------|
| LL1              | SCG, SFVC, NV, GPCP, SCVCI, and JGCG |
| LL2              | ICIC, SC, WI, ITC, QWVCM, TGSG, JJ, SBVVC, and ZV(NCI>0.7) |
| LL3              | IT, DCM, VVC, SI, HI, KV, and CMFAM |
| LL4              | NLVC, KPCB, and NEA          |
| LL5              | DCMCL and DFJ                |
| LL6              | SOYH                        |

### Table 9 VC companies with NCI greater than 0.6 in EO detection results.

| Community number | VC company                  |
|------------------|-----------------------------|
| EO1              | ICIC, LCM, NLVC, ITC, QWVCM, TGSG, JJ, SBVVC, and IT(NCI>0.7) |
| EO2              | SCG and NV                  |
| EO3              | SFVC                        |
| EO4              | GPCP and JGCG               |
| EO5              | DCMCL, SOYH, DFJ, GC, GVM, MVC, ERVC, MGI, CIM, and CIC |
| EO6              | SC, ZV, JVCIMC, and LIC     |
| EO7              | KPCB, WI, RCV, and QV       |
| EO8              | SCVCI                       |

### Table 10 Distance between various VC industry leaders.

| VC node pair | Distance | VC node pair | Distance |
|--------------|----------|--------------|----------|
| ICIC-SC      | 0.1      | DCMCL-SOH    | 0.7917   |
| ICIC-LCM     | 0.2909   | NLVC-KPCB    | 0.0769   |
| DCMCL-DFJ    | 0.125    | NLVC-ICIC    | 0.1667   |

### Table 11 Distance between two nodes of each constraint pair of VC industry leaders.

| Constraint pair on VC industry leaders | Distance |
|----------------------------------------|----------|
| SCG-GPCP                               | 0.0909   |
| DFJ-DCMCL                              | 0.125    |
| ITC-JI                                 | 0.1111   |
| SOYH-IG                                | 0.25     |
| ICIC-SC                                | 0.1      |
| LCM-DCM                                | 0.05     |
| NLVC-NEA                               | 0.0625   |
| JVCIMC-QWVCM                           | 0.2778   |
a constraint relationship, and it is more reasonable to be divided into different communities.

Based on this, looking at the two-point distance results of some VC node pairs in the LL and EO communities calculated previously, we can see that not only the two-point distance of VC node pairs numbered 1, 3, and 5 is less than the two-point distance of VC node pairs numbered 2, 4, and 6, respectively, but also the two-point distance of the former is less than \( d_c = 0.1334 \), and the distance between the two nodes of the latter is greater than 0.1334. This means that the constraint relationship of VC node pairs numbered 1, 3, and 5 is established, while that of VC node pairs numbered 2, 4, and 6 is not. However, the community detected by EO algorithm destroys this constraint relationship. That is to say, in the VC industry leaders of the EO community detection results, there are errors in that the two nodes of a VC pair that should be constrained are divided into different communities and the two nodes that should not be constrained are divided into the same community. However, this kind of error is effectively avoided in the industry leader community detected by LL algorithm. VC circles will be formed around industry leaders, and the vast majority of joint investment behaviors occur between VC leaders. In fact, most VC companies follow their leaders. This means that the behavior of VC leaders is enough to represent a circle or community. Therefore, this paper only uses a few VC industry leaders’ node pairs to illustrate the detection quality as much as possible, while it is difficult to compare all nodes or the whole network.

Meanwhile, the number and its proportion of joint investment of VC companies were calculated for the constraint pair used for comparison in LL detection and EO detection results, as shown in Table 12.

It can be seen that for VC node pairs ICIC-SC and ICIC-LCM, the number of joint investments of the former is 5 times as big as that of the latter, and the number of joint investments of the former exceeds the threshold 9, while the number of joint investments of the latter is less than the threshold 4 (the thresholds 4 and 9 were explained in Section 3.2). Meanwhile, the proportion of joint investment times of the former is 5.6% higher than that of the latter.

For VC node pairs DCMCL-DFJ and DCMCL-SOYH, the number of joint investments of the former is 8 times as big as that of the latter, and the number of joint investments of the former exceeds the threshold 4, while the number of joint investments of the latter is less than the threshold 4. Meanwhile, the proportion of joint investment times of the former is 7.8% higher than that of the latter.

For VC node pairs NLVC-KPCB and NLVC-ICIC, the number of joint investments of the former is 2.2 times as big as that of the latter, and the number of joint investments of the former exceeds the threshold 9, while the number of joint investments of the latter is less than the threshold 9. Meanwhile, the proportion of joint investment times of the former is 7.7% higher than that of the latter.

To sum up, it can be preliminarily considered that VC node pairs ICIC-SC, DCMCL-DFJ, and NLVC-KPCB constitute the constraint relationship and should be divided into the same community; VC node pairs ICIC-LCM, DCMCL-SOYH, and NLVC-ICIC do not constitute the constraint relationship and should not be divided into the same community.

In other words, the constraints of ICIC-SC, DCMCL-DFJ, and NLVC-KPCB are far stronger than those of ICIC-LCM, DCMCL-SOYH, and NLVC-ICIC. In the EO detection, the two nodes of VC node pairs ICIC-SC and NLVC-KPCB are separated, while those of ICIC-LCM and NLVC-ICIC are hand in hand. At the same time, SOYH is inserted between DCMCL and DFJ, which is incorrect. However, this kind of error is better avoided in LL detection.

Although the VC community nowadays has become more diversified, it is still rare to see industry leaders,

| VC node pair | Number of joint investments | Proportion (ICIC) | Proportion (SC) | Proportion (LCM) |
|--------------|-----------------------------|-------------------|----------------|-----------------|
| ICIC-SC      | 10                          | 10/143 (7.0%)     | 10/134 (7.5%)  | –               |
| ICIC-LCM     | 2                           | 2/143 (1.4%)      | –              | 2/114 (1.8%)    |
| DCMCL-DFJ    | 8                           | 8/90 (8.9%)       | 8/67 (11.9%)   | –               |
| DCMCL-SOYH   | 1                           | 1/90 (1.1%)       | –              | 1/74 (1.4%)     |
| NLVC-KPCB    | 13                          | 13/91 (14.3%)     | 13/84 (15.5%)  | –               |
| NLVC-ICIC    | 6                           | 6/91 (6.6%)       | –              | 6/143 (4.2%)    |

Table 12 Number of joint investments and its proportion of VC companies in relevant VC node constraint pairs.
the core of a community, coming from different backgrounds. VC firms with different backgrounds might cooperate to acquire complementary resources, mostly in status asymmetry conditions, so that the high status VC dominates the cooperation. However, frequent cooperation among industry leaders must be based on high mutual trust, consistent investment logic, and common organizational routines. Therefore, the EO algorithm divides ICIC and LCM, DCMCL and SOYH, the two pairs of foreign-dominate (heterogeneous) VC company combinations, into the same community, respectively, which is not as reasonable as the result of LL detection. Both NLVC and ICIC belong to foreign-funded communities and have American backgrounds and they represent two types of foreign-funded communities. One is a foreign-funded VC company with ICIC as the representative, which is gradually localized operating under the support of foreign VC companies. The other company has a relatively pure American background in personnel and management represented by NLVC. Therefore, LL algorithm that separates these two types of foreign VC communities is also more reasonable by comparison.

It can be concluded from the above that the result of LL detection is more accurate and reasonable than that of EO detection. This paper suggests that the LL semi-supervised community detection algorithm based on the initial label information which is obtained from VC industry leaders’ behavior of forming clusters is superior to the traditional EO community detection algorithm that is only based on network topology in terms of accuracy and rationality of detection results.

5 Discussion and Conclusion

Based on the data of VC joint investment and the list of VC industry leaders obtained from previous studies, this paper realizes a semi-supervised community detection of VC network. The results show that the community structure of the VC network has obvious distinct characteristics. The gathering of VC companies and the formation of VC communities will be affected by factors including the type of organization, capital source, personnel background, investment field, and geographical location. At the same time, by calculating the distance between the nodes, the results of the semi-supervised community detection algorithm designed in this paper are compared with the results of the WEO community detection, which show that the semi-supervised community detection results of VC network are more accurate and reasonable to some extent. Based on this point of view, through the empirical analysis of the joint investment network of Chinese VC industry, and combined with the industry characteristics of Chinese VC, this paper explores and obtains a semi-supervised community detection method. At the core of this method is designing the initial label of semi-supervised community detection according to the information of VC industry leaders’ behavior of forming clusters which is learned from the evolution of their components. This leads to a discussion on comparing semi-supervised and unsupervised community detection results by quantitative method.

It has been well recognized that the network structure of an enterprise will influence its innovation performance[21, 22]. Most existing studies analyze the enterprises’ ego network or global network, but in recent years, the community network between the ego network and the global network has attracted increasing attention. Community network provides network information different from that of the ego network and global network, and enables us to study the distribution of heterogeneous information and resources in a social system[23]. For example, it has been found that members’ innovation benefits most when the turnover frequency of community members is at a medium level[23]. The concentration level of a business group positively influences the performance within companies, but this positive impact is weakened in environments with high uncertainty[24]. As these streams of research rely on the accurate recognition of community, the importance of community detection methods increase. The more accurate results of community detection obtained in this paper can help pave the way for further study on the impact of community structure on investment performance, of which the semi-supervised community detection method adopted in this paper can also help provide a good research starting point.

Apart from the consequence of network structures, the dynamic change and evolution mechanism of inter-enterprise network structure is also a classic topic within the fields of sociology and management[20]. The dimension of network community level network dynamic helps us better understand the evolution mechanism of network[25]. In particular, the community of Chinese VC companies is often formed around one or two core VC companies. In the VC network, the center of each community is often a bridge connecting different communities. The VC network as a whole presents the
characteristics of an elite small world\cite{6,9}. The core nodes in the community may be in a key position both inside and outside the community, which may affect the dynamic performance of the entire industrial network. Trends of the whole industrial network may first appear in some specific community networks, and thus the network characteristics of the elite small world mean that the network characteristics and dynamic changes at the community level are crucial to the evolution of the overall network structure and the basis of the community network’s evolution is community detection. Based on the results of this paper, we can further investigate the evolution of community structure and behavior patterns of VC network in a more sophisticated and dynamic manner.

Yet, various defects in the community detection method adopted in this paper are present, mainly due to the difficulty faced in the identification of some VC companies that have not participated in the forming of clusters group but are still likely to become community leaders when designing the initial labels, by which the lack of some initial labels will lead to the results of semi-supervised community detection to be inaccurate, and some hidden marginal communities may be absorbed into core communities. To address this, a further in-depth analysis of the VC industry leaders’ clustering-forming situation can be undertaken and various ways can be used to increase the number of initial labels. The various results can then be compared and properly summarized so the rules can form a more detailed and convincing theoretical scheme about the initial label design in the semi-supervised community detection method. In addition, we have to admit that the most commonly used index to measure the quality of community detection based on network topology is modularity function, but it is suitable for unsupervised community detection, but not for semi-supervised community detection. When comparing the results of semi-supervised and unsupervised community detection algorithms, there is no corresponding and more intuitive comprehensive index in this paper, which will be an important research direction based on this paper. Meanwhile, the premise of the proposed algorithm in this paper is that the network data have relatively accurate initial labels. In theory, this algorithm can be used to detect the community in any relational network with similar initial labels. There is no doubt that the generality of the algorithm proposed in this paper needs to be tested on more kinds of more general networks to verify its universality, which is also a research direction based on this paper.

### Acknowledgment

We are grateful for the financial support from the Chinese Natural Science Foundation Project “Social Network in Big Data Analysis: A Case in Investment Network” (Nos. 71372053 and 71731002), as well as the support from the Tencent Research Institute Project “Research on Identification of Opinion Leaders Based on QQ Big Data” (No. 20182001706).

### Appendix

**Table A1** Detailed list of abbreviations and English names of VC companies.

| VC company                                  | Abbreviation |
|---------------------------------------------|--------------|
| Tuspark Venture Co., Ltd.                   | TV           |
| Shenzhen Capital Group Co., Ltd.            | SCG          |
| NewMargin Ventures                          | NV           |
| Zhengzhou Britic Innovation Fund Investment Co., Ltd. | ZBIFI  |
| Shenzhen Carecall Capital Investment Co., Ltd. | SCCI        |
| Shenzhen Futian Investment Development Co.  | SFID         |
| Zhongyi Venture Investment Fund             | ZVIF         |
| Xi’an Innovation Investment Management Co., Ltd. | XIIM   |
| Shenzhen Luohu Laterite VC Co., Ltd.        | SLLV         |
| Shenzhen HTI Venture Capital Co., Ltd.      | SHVC         |
| Shenzhen Fortune Venture Capital Co., Ltd.  | SFVC         |
| China Venture Capital Co., Ltd.             | CVC          |
| Jiangsu Addor Equity Investment Fund Management Co., Ltd. | JAEIFM |
| Beijing Jiahua Venture Capital Co., Ltd.    | BJVC         |
| Green Pine Capital Partners Co., Ltd.       | GPCP         |
| Shenzhen CDF-Capital Co., Ltd.              | SCC          |
| Jiansu Govtor Capital Group Co., Ltd.       | JGCG         |
| Shenzhen Dongfang Fuhai Venture Capital Enterprise (L.P.) | SDFVCE    |
| Mingxin China Growth Fund                   | MCGF         |
| Shenzhen Co-power Venture Capital Co., Ltd. | SCVC         |
| Wuhan Huagong Venture Capital Co., Ltd.     | WHVC         |
| Delta Capital                               | DC           |
| Hubei High-tech Industry Investment Group Co., Ltd. | HHIIG |
| Shenzhen Co-win Venture Capital Investment Ltd. | SCVCI      |
| SB China Venture Capital                    | SBCVC        |
| Investor Growth Capital Asia Ltd.           | IGCA         |
Table A1 Detailed list of abbreviations and English names of VC companies. (continued)

| VC company                                         | Abbreviation | VC company                                         | Abbreviation |
|----------------------------------------------------|--------------|----------------------------------------------------|--------------|
| The Goldman Sachs Group, Inc.                      | TGSG         | Redpoint China Ventures                            | RCV          |
| IDG Capital Investment Consultant (Beijing) Co. Ltd. | ICIC         | Qualcomm Ventures                                 | QV           |
| Intel Capital                                      | ITC          | iD TechVentures, Inc.                              | IT           |
| WI Harper Group                                    | WHG          | New Enterprise Associates                          | NEA          |
| Headland Capital Partners Limited (HSBC Private Equity (Asia) Ltd.) | HCPL | Shanghai Bio Veda Investment Management Consulting Co., Ltd. | SBVIMC |
| JAFCO Investment (Hong Kong) Ltd.                  | JI           | Keytone Ventures                                  | KTV          |
| SAIF Partners                                      | SAIF         | Lilly Asia Ventures                               | LAV          |
| SIG Asia Investment Fund                           | SP           | Kleiner Perkins Caufield & Byers                  | KPCB         |
| The CID Group                                      | TCG          | Lightspeed Management Company, LLC                 | LMC          |
| Harbinger Venture Management                       | HVM          | Draper Fisher Jurvetson                           | DFJ          |
| CDH Venture                                        | CV           | DT Capital Management Company Limited              | DCMCL        |
| Mustang Ventures                                   | MV           | Taishan Angel Venture Fund                         | TAVF         |
| Qiming Weichuang Venture Capital Management (Shanghai) Co., Ltd. | QWVC | Suzhou National Development Capital Investment, Co., Ltd. | SNDCCI |
| Shanghai Trust Bridge Partners Management Co., Ltd. | STBPM        | Suzhou Oriza Yuanhe Holdings Co., Ltd.             | SOYH         |
| Infotech Ventures Co., Ltd.                        | IV           | Infinity Group                                    | IG           |
| GGV Capital                                        | GC           | Suzhou Industrial Park Rongfeng Investment Management Co., Ltd. | SIPRIM  |
| TDF Capital                                        | TC           | Suzhou Industrial Park Venture Capital Guide Fund Management Centre | SIPVCGF |
| Ceyuan Investment Consulting (Beijing) Co., Ltd.    | CIC          | Tianjin PrelPO Venture Capital Management Co., Ltd. | TPVCM |
| Venrock Associates                                 | VA           | Pingan Caizhi Investment Management Co., Ltd.      | PCIM         |
| Mayfield Fund                                      | MF           | China Renaissance Capital Investment Inc.          | CRCI         |
| GSR Ventures Management Co., Ltd.                  | GVM          | Mitsui Global Investment Ltd.                      | MGI          |
| Steamboat Ventures                                 | SV           | CDH Investment Management Co., Ltd.                | CIM          |
| Tenaya Capital (previously Lehman Brothers Asia Ltd.) | TNC          | Shanghai Investment (Shanghai) Co., Ltd.           | SI           |
| Oak Investment Partners                            | OIP          | Singapore UOB Venture Capital Management Co., Ltd. | SUVC         |
| Tiger Global Management                            | TGM          | Suzhou International Development Venture Capital Holding Co., Ltd. | SIVCH |
| Chengwei Ventures Shanghai LLC.                    | CVS          | Kunwu Jiuding Capital Holdings Co., Ltd.           | KJCH         |
| SEQUOIA CHINA                                      | SC           | Shanghai Leading Capital Co., Ltd.                 | SLC          |
| Zero2IPO Venture                                   | ZV           | Tianjin Heguang Private Equity Management Partnership (L.P.) | THPEMP |
| Jingwei Venture Capital (Beijing) Investment Management Consulting Co., Ltd. | JVCIMC | Shenzhen Tongchuang Venture Capital Co., Ltd.       | STVC         |
| Lightspeed China Partners                          | LCP          | Beijing China Equity Group Inc.                    | BCEG         |
| Lanchi Investment Consulting (Shanghai) Co., Ltd.   | LIC          | SND Ventures Group Co., Ltd.                       | SVG          |
| K2 Ventures Capital Ltd.                           | KVC          | Sinkiang Small and Medium-sized Enterprise         | SSMEVCH      |
| Beijing Zhen Fund Management Co., Ltd.             | BZFM         | Venture Capital Holding Co., Ltd.                  |              |
| Shanghai Pudong Science And Technology Investment Co., Ltd. | SPSAT | Xinjiang Xintou Industry Investment Co., Ltd.      | XXII         |
| Walden International                               | WI           | Wuhan East-lake Investment Management Co., Ltd.    | WEIM         |
| GOBI Partners Inc.                                 | GP           | Guoyuan Private Equity Co., Ltd.                   | GPE          |
| Shanghai Venture Capital Co., Ltd.                 | SVC          | Prime Mont Capital Co., Ltd.                       | PMC          |
| Bay Partners                                       | BP           |                                                  |              |
| InterWest Partners                                 | IP           |                                                  |              |

References

[1] M. E. J. Newman and M. Girvan, Finding and evaluating community structure in networks, *Phys. Rev. E*, vol. 69, no. 2, p. 026113, 2004.

[2] S. Fortunato and M. Barthélemy, Resolution limit in community detection, *Proc. Natl. Acad. Sci. USA*, vol. 104, no. 1, pp. 36–41, 2007.
[3] X. F. Wang and Y. B. Liu, Overview of algorithms for detecting community structure in complex networks, (in Chinese), Journal of University of Electronic Science and Technology of China, vol. 38, no. 5, pp. 537–543, 2009.

[4] X. J. Li, P. Zhang, Z. R. Di, and Y. Fan, Community structure in complex networks, (in Chinese), Complex Systems and Complexity Science, vol. 5, no. 3, pp. 19–42, 2008.

[5] M. Du, Semi-supervised algorithm research for community structure detection, (in Chinese), Master dissertation, Hefei University of Technology, Hefei, China, 2014.

[6] J. D. Luo, L. Qin, and L. Zhou, Circle in Chinese venture capital industry, (in Chinese), Chinese Journal of Management, vol. 11, no. 4, pp. 469–477, 2014.

[7] J. Lerner, The syndication of venture capital investments, Financial Management, vol. 23, no. 3, pp. 16–27, 1994.

[8] Y. H. Jin and Q. Zhang, Network characteristics and community structure of Chinese venture capital networks, (in Chinese), Journal of Systems Engineering, vol. 31, no. 2, pp. 166–177, 2016.

[9] W. W. Gu, J. D. Luo, and J. F. Liu, Exploring small-world network with an elite-clique: Bringing embeddedness theory into the dynamic evolution of a venture capital network, Social Networks, vol. 57, pp. 70–81, 2019.

[10] H. Yang, J. D. Luo, Y. Fan, and L. Zhu, Using weighted k-means to identify Chinese leading venture capital firms incorporating with centrality measures, Information Processing & Management, vol. 57, no. 2, p. 102083, 2020.

[11] M. Sytch and A. Tatarynowicz, Exploring the locus of invention: The dynamics of network communities and firms’ invention productivity, Academy of Management Journal, vol. 57, no. 1, pp. 94–102, 2018.

[12] Pediaiy Website. Brief History of China’s Venture Capital. Beijing, China: People’s Post and Telecommunications Press, 2017.

[13] J. D. Luo and Y. Fan, How does computational social science take root in the truth: How to find the “industry leader” of venture capital in China, (in Chinese), Exploration and Contention, no. 7, pp. 94–102, 2018.

[14] J. Yang, Q. Y. Wang, and W. J. Guo, Venture capital network community detection and community differences analysis, (in Chinese), Soft Science, vol. 33, no. 3, pp. 20–24, 2019.

[15] J. Kong, Research for community detection algorithm based on semi-supervised learning, (in Chinese), Master dissertation, Liaoning Normal University, Dalian, China, 2010.

[16] V. D. Blondel, J. L. Guillaume, R. Lambiotte, and E. Lefebvre, Fast unfolding of communities in large networks, Journal of Statistical Mechanics: Theory & Experiment, vol. 2008, no. 10, p. P10008, 2008.

[17] M. E. J. Newman, Fast algorithm for detecting community structure in networks, Phys. Rev. E Stat. Nonlin. Soft Matter Phys., vol. 69, p. 066133, 2004.

[18] J. Duch and A. Arenas, Community detection in complex networks using extremal optimization, Phys. Rev. E, vol. 74, no. 2, p. 027104, 2006.

[19] M. Sauder, F. Lynn, and J. M. Podolny, Status: Insights from organizational sociology, Annual Review of Sociology, vol. 38, no. 1, pp. 267–283, 2012.

[20] W. W. Powell, D. R. White, K. W. Koput, and J. Owen-Smith, Network dynamics and field evolution: The growth of interorganizational collaboration in the life sciences, American Journal of Sociology, vol. 110, no. 4, pp. 1132–1205, 2005.

[21] R. Gulati, N. Nohria, and A. Zaheer, Strategic networks, Strategic Management Journal, vol. 21, no. 3, pp. 203–215, 2000.

[22] J. M. Podolny, Networks as the pipes and prisms of the market, American Journal of Sociology, vol. 107, no. 1, pp. 33–60, 2001.

[23] M. Sytch and A. Tatarynowicz, Exploring the locus of invention: The dynamics of network communities and firms’ invention productivity, Academy of Management Journal, vol. 57, no. 1, pp. 249–279, 2014.

[24] I. P. Mahmood, H. J. Zhu, and A. Zaheer, Centralization of intragroup equity ties and performance of business group affiliates, Strategic Management Journal, vol. 38, no. 5, pp. 1082–1100, 2017.

[25] G. Ahuja, G. Soda, and A. Zaheer, The genesis and dynamics of organizational networks, Organization Science, vol. 23, no. 2, pp. 299–306, 2012.

Ying Fan received the BEng and PhD degrees in physics from Beijing Normal University in 1994 and 1999, respectively. She is a professor at School of Systems Science, Beijing Normal University, China. Her current research interests include complex systems theory, complex network, and analysis on social and economic systems.

Hong Xiong is pursuing the MS degree at School of Systems Science, Beijing Normal University, China. He received the BEng degree from Beijing Institute of Technology in 2018. His current research interests include complex network science and social network analysis.