



---

# A Venture Capital Recommendation Algorithm based on Heterogeneous Information Network

S. Wu, H. Li, L. Liu

## Sen Wu\*

Donlinks School of Economics and Management  
University of Science and Technology Beijing, China  
30 Xueyuan Road, Haidian District, Beijing 100083, China  
\*Corresponding author: wusen@manage.ustb.edu.cn

## Huifei Li

Donlinks School of Economics and Management  
University of Science and Technology Beijing, China  
30 Xueyuan Road, Haidian District, Beijing 100083, China  
HunterLi\_et@qq.com

## Lu Liu

Donlinks School of Economics and Management  
University of Science and Technology Beijing, China  
30 Xueyuan Road, Haidian District, Beijing 100083, China  
liulu1891@163.com

## Abstract

According to its characteristics, venture capital can be described as a typical heterogeneous information network, which includes multiple kinds of nodes and various relations. Getting hints from PathRank algorithm, this paper proposes VC-Recom, a recommendation algorithm based on heterogeneous information network, which helps investment companies find suitable startup projects. Besides, the experimental results show that the proposed algorithm can produce more effective recommendation results for investment firms compared with other methods.

**Keywords:** PathRank, venture capital recommendation, heterogeneous information network, meta path.

## 1 Introduction

Recently, the prosperity of venture capital industry has caught a lot of attention. Venture capital (VC) is a form of private equity, a medium to long-term form of finance provided in return for an equity stake in potentially high growth companies [13]. It can bring high returns to investment companies when they get a promising startup. But it is hard for investment trusts to find suitable startups. While, the recommendation algorithm can recommend suitable items to users based on their

preferences. We try to apply recommendation algorithms to venture capital to help investment companies find the right startups. Currently, there are some commonly used recommendation algorithms, such as the recommendation based on collaborative filtering, content-based recommendation algorithm, model-based recommendation algorithm and hybrid recommendation [3] [17]. However, these recommendation algorithms are based on the traditional homogeneous information network, while venture capital network is a typical heterogeneous information network. A heterogeneous information network is a special type of information network with the underneath data structure as a directed graph [12], which either contains multiple types of objects or multiple types of links [15]. When we use the homogeneous information network to analyze the realistic venture capital behavior, it can only describe the relationship between the same type of objects. And it ignores the attribute information of investment companies and startup projects, such as the location of the venture project, the field and so on, which are included in the heterogeneous information network. Therefore, this paper tries to establish a heterogeneous information network model for venture capital recommendation to get a more accurate result. At present, there are several classic heterogeneous information network recommendation algorithms, such as HeteRecom [11], PathRank [8] and so on. Based on the thought of PathRank algorithm, this paper puts forward a venture capital recommendation algorithm to solve the problem of venture capital recommendation.

## 2 Related work

Venture capital plays an irreplaceable role in the transformation of science and technology into actual productive forces [14] [18]. At present, how to choose a startup for an investment company is still a key issue for venture capital research. Jiang L discussed how to make investment decisions based on the point of view of VPN rules [5]. Dicks D L et al. explained the impact of disgust on investment in innovative projects from the perspective of investor distaste for Knightian's uncertainty [2]. Li M et al. solved the investment decision-making problem by constructing a mathematic optimization model of the integrated gray correlation coefficient that maximizes the decision value, and applied it to choose the startup project for investment companies [9]. Xu et al. adopted the design science approach and developed a prediction model to support corporate venture capital(CVC) investment decisions by identifying a list of potential investees from a large pool of portfolio companies for a CVC investor [16]. Justin Chircop et al. studied whether religiosity influences VC investment decisions. And the result shows that VC located in more religious countries are more risk averse and their investments are of better quality [1]. However, there are few researches on solving the choice of venture project from the perspective of venture capital recommendation. This paper applies the recommendation algorithm to venture capital and helps investors choose the right investment projects. At present, some recommendation algorithms are based on homogeneous information networks. When venture capital behavior is abstracted as homogeneous information networks, some attribute information and relationships between different objects are discarded. In order to express venture capital behavior more completely, this paper sets up a heterogeneous information network model for venture capital, which is the foundation of heterogeneous information network recommendation algorithms.

Current heterogeneous information network recommendation algorithms often ignore the structural information of heterogeneous information networks. In order to solve the problem that heterogeneous networks neglect the structure information of network, many scholars seek help from PageRank algorithm [10], integrating PageRank algorithm and heterogeneous information networks to the recommendation system to achieve the node ranking in the network and improve the effectiveness of the recommendation system. Both Jeh et al. and Kleinberg have made great strides in node ranking algorithm based on random walk [4] [7], but Sangkeun Lee et al. realized that it was difficult to make a breakthrough in the results if only the structural information of heterogeneous information networks was used to rank the nodes. So they tried to construct a heterogeneous information network model that contained both structural information and attribute information and proposed the PathRank algorithm [6] [8]. The algorithm is a flexible recommendation system that simulates recommendation based on collaborative filtering, content-based recommendation and context-based recommendation.

Based on the PathRank algorithm, this paper proposes the venture capital recommendation al-

gorithm of heterogeneous information networks. This algorithm aims at ranking the startups and recommending startups to investment companies.

### 3 The model of venture capital recommendation algorithm

#### 3.1 Venture capital network

**Definition 1** (Heterogeneous venture capital network): Venture capital network is a typical heterogeneous information network  $G = (V, E)$ , where  $V$  is a finite set of venture capital nodes,  $E \subseteq V \times V$  is a finite multi-set of venture capital edges. Venture capital nodes are VC firms, startups and the attributes of startups in heterogeneous venture capital network. Every venture capital edge links two venture capital nodes that belong to different venture capital node types.

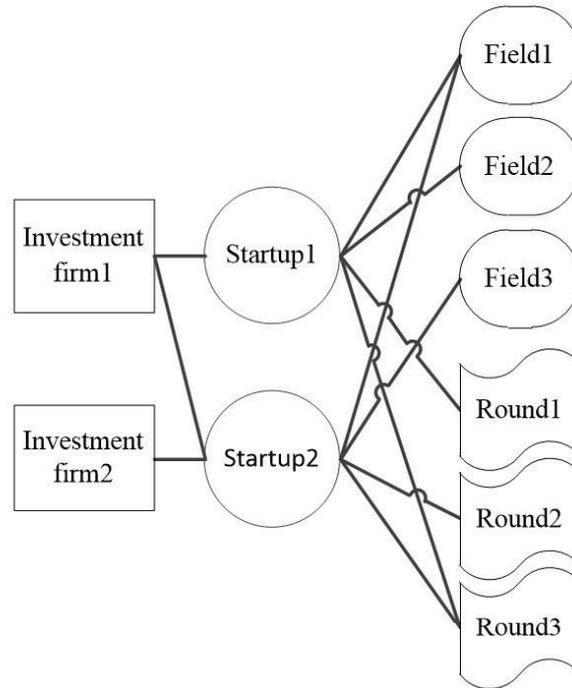


Figure 1: An example of heterogeneous venture capital network

Fig.1 is an example of heterogeneous venture capital network. Investment firm1, investment firm2, startup1, startup2, field1, field2, field3, round1, round2 and round3 are all venture capital nodes.  $Investment Firm1 \xrightarrow{investing} startup1$ ,  $startup1 \xrightarrow{invested\ by} Investment Firm1$ ,  $startup1 \rightarrow field1$ ,  $startup1 \rightarrow round1$ ,  $field1 \rightarrow startup1$ ,  $round1 \rightarrow startup1$  and so on are venture capital edges.

**Definition 2** (Venture capital network schema): Venture capital network schema, denoted as  $TG = (A, R)$ , is a meta template for heterogeneous venture capital network  $G = (V, E)$  with the object type mapping  $\Phi_V = V \rightarrow A$  and the link type mapping  $\Phi_E = E \rightarrow R$ , which is a directed graph defined over object types  $A$ , with edge types as relations from  $R$ . where  $A$  is a finite set of venture capital node types, and  $R$  is a finite set of venture capital edge types. Venture capital node type is abstracted from venture capital nodes and venture capital edge type is abstracted from venture capital edges.

Fig.2 is a venture capital network schema. It is abstracted from Fig.1. For example, investment firm1 and investment firm2 are abstracted as the venture capital node type of investment firms, startup1 and startup2 are abstracted as the venture capital node type of startups, field1, field2 and field3 are abstracted as the venture capital node type of fields, round1, round2 and round3 are abstracted as the venture capital node type of rounds.  $Investment Firm1 \xrightarrow{investing} startup1$  is abstracted as the venture capital edge type of  $Investment Firm \xrightarrow{investing} startup$ .

Different attributes of startup projects contribute to different venture capital network schemas. For example, if the attribute node type of a startup project's network model contains the entrepreneurial

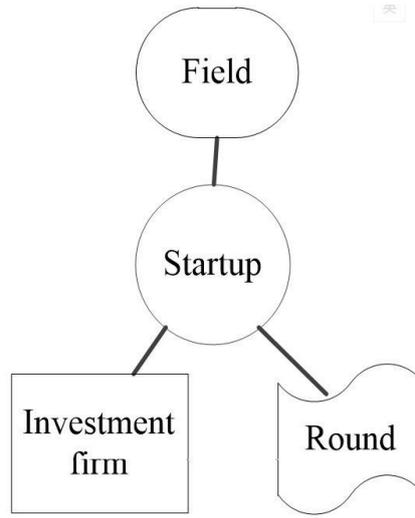


Figure 2: An example of venture capital network schema

location, another network model does not include the entrepreneurial location, the two network models are different network models.

**Definition 3** (Venture capital meta path): A venture capital meta path on the venture capital network schema  $TG = (A, R)$ , is defined as  $A_1 \xrightarrow{R_1} A_2 \xrightarrow{R_2} \dots A_n \xrightarrow{R_n} \dots \xrightarrow{R_l} A_{l+1}$ , representing the flexible relationships from  $A_1$  to  $A_{l+1}$ , where  $A = \{A_1, A_2, \dots, A_n, \dots, A_l, A_{l+1}\}$ ,  $R = \{R_1, R_2, \dots, R_n, \dots, R_l\}$ ,  $A_1$  is a VC firm,  $A_2$  and  $A_{l+1}$  are startups,  $A_3, \dots, A_l$  are not limited to startups or VC firms.

Because our work is to recommend startups to VC firms, the first node of the venture capital meta path must be a VC firm, the second and the last one must be startups.

**Definition 4** (Adjacency matrix): In heterogeneous venture capital network  $G = (V, E)$ ,  $R$  is the set of all venture capital edge types, adjacency matrix  $P_A$  is a  $|V| \times |V|$  matrix, where

$$P_A(i, j) = \begin{cases} 1 & \text{if } node_i \rightarrow node_j \text{ can be abstracted as an edge type belonging to } R \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

**Definition 5** (Transition matrix): Transition matrix  $P_R$  is a  $|V| \times |V|$  matrix, where

$$P_R(i, j) = \begin{cases} \frac{1}{\sum_j P_A(i, j)} & \text{if } node_i \rightarrow node_j \text{ can be abstracted as an edge type belonging to } R \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

**Definition 6** (Edge adjacency matrix): In heterogeneous venture capital network  $G = (V, E)$ , edge adjacency matrix  $E_{R_n}$  is a  $|V| \times |V|$  matrix under the venture capital edge type  $R_n$ , where

$$E_{R_n}(i, j) = \begin{cases} 1 & \text{if } node_i \rightarrow node_j \text{ can be abstracted as } R_n \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

**Definition 7** (Edge transition matrix): Edge transition matrix  $P_{R_n}$  is a  $|V| \times |V|$  matrix under the venture capital edge type  $R_n$ , where

$$P_{R_n}(i, j) = \begin{cases} \frac{1}{\sum_j E_{R_n}(i, j)} & \text{if } node_i \rightarrow node_j \text{ can be abstracted as } R_n \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

**Definition 8** (Path transition matrix): In heterogeneous venture capital network  $G = (V, E)$ , path transition matrix  $P_{mp}$  under meta path  $A_1 \xrightarrow{R_1} A_2 \xrightarrow{R_2} \dots A_n \xrightarrow{R_n} \dots \xrightarrow{R_l} A_{l+1}$  is a  $|V| \times |V|$  matrix, where  $P_{mp} = P_{R_1} \times P_{R_2} \times \dots \times P_{R_n} \times \dots \times P_{R_l}$ .

### 3.2 The specific introduction of VC-Recom

In this section, we explain how to compute the scores of startups for every VC firm under meta-paths.

Like PathRank, VC-Recom is the stationary probability of random walker's being at each node after enough number of random walk iterations. But the PathRank ranks all the nodes of all node types, VC-Recom recommends the startups to VC firms. So we need to rank all the nodes of startups.

In each iteration of VC-Recom, the random walker has three options:

Transition: With probability  $w_{trans}$ , move to one of the adjacent nodes.

Restart: With probability  $w_{re}$ , restart the random walk from the VC firms that we are going to recommend to.

Path Following: With probability  $w_{path}$ , follow one of venture capital meta paths.

Transition and Restart are inherited from Personalized Page-Rank, and they fulfill the roles to reflect global authorities of the invested startups and relevance to other startups. Path Following is to reflect the semantics of the given venture capital meta paths in the ranking score. Formally, VC-Recom vector is defined as follows under venture capital meta paths of  $mp_1, \dots, mp_c$ .

$$\vec{r} = w_{trans} \times P_R \times \vec{r} + w_{path} \times (w_1 P_{mp_1}^T + \dots + w_c P_{mp_c}^T) \times \vec{r} + w_{re} \times \vec{t}. \quad (5)$$

Let  $V$  be the set of all the venture capital nodes in heterogeneous venture capital network,  $\vec{r}$  is a  $|V| \times 1$  vector,  $P_R$  is a  $|V| \times |V|$  transition matrix describing the relationship among all the venture capital nodes and  $P_{mp_1}, \dots, P_{mp_c}$  are  $|V| \times |V|$  path transition matrices describing the relationship between nodes under venture capital meta path  $mp_1, \dots, mp_c$ .

$\vec{t}$  is a  $|V| \times 1$  restart vector, where

$$t_i = \begin{cases} 1 & \text{if the node}_i \text{ is a VC firm that we will recommend to} \\ 0 & \text{otherwise} \end{cases}. \quad (6)$$

Besides,  $w_1 + w_2 + \dots + w_c = 1$ .

For example, if we want to recommend startups to investment firm2 in Fig.1 under the venture capital meta path 'investment firm  $\xrightarrow{\text{investing}}$  startup project  $\xrightarrow{\text{has the attribute}}$  round  $\xrightarrow{\text{the attribute of startup project}}$ ', denoted as  $mp_1$ , we need to calculate transition matrix  $P_R$ , path transition matrix  $P_{mp_1}$  and  $\vec{t}$ .

Let  $R_1, R_2, R_3$  represent 'investment firm  $\xrightarrow{\text{investing}}$  startup project', 'startup project  $\xrightarrow{\text{has the attribute}}$  round', 'round  $\xrightarrow{\text{the attribute of startup project}}$ '.  $P_R$  is described in Tab.1,  $P_{mp_1}$  is equal to  $P_{R_1} \times P_{R_2} \times P_{R_3}$ , where  $P_{R_1}$  is the edge transition matrix of venture capital edge 'investment firm  $\xrightarrow{\text{investing}}$  startup project',  $P_{R_2}$  is the edge transition matrix of venture capital edge 'startup project  $\xrightarrow{\text{has the attribute}}$  round',  $P_{R_3}$  is the edge transition matrix of the venture capital edge 'round  $\xrightarrow{\text{the attribute of startup project}}$ ',  $P_{R_1}, P_{R_2}, P_{R_3}$  are described in Tab.2, Tab.3, Tab.4.  $\vec{t} = (0, 1, 0, 0, 0, 0, 0, 0, 0, 0)$ .  $I_1, I_2, S_1, S_2, R_1, R_2, R_3, F_1, F_2$  and  $F_3$  represent investment firm1, investment firm2, startup1, Startup2, round1, round2, round3, field1, field2 and field3.

The specific steps of VC-Recom:

Input:  $w_{trans}, w_{re}, w_{path}, \vec{t}, w_1, \dots, w_c$  transition matrix  $P_{mp_1}, \dots, P_{mp_c}$  and  $P_R$ , the number of recommended startups  $k$

Initialize  $\vec{r} = \text{ones}(|V|, 1) / |V|$ ,  $\vec{r}_1 = \vec{0}$ ;

While  $|\vec{r} - \vec{r}_1| > 0.0001$  do

$$\vec{r}_1 = \vec{r}, \vec{r} = w_{trans} \times P_R \times \vec{r} + w_{path} \times (w_1 P_{mp_1}^T + \dots + w_c P_{mp_c}^T) \times \vec{r} + w_{re} \times \vec{t}$$

End while

Return *topk* startup projects;

Output: *topk* startup projects.

Table 1: The description of  $P_R$

|       | $I_1$ | $I_2$ | $S_1$ | $S_2$ | $R_1$ | $R_2$ | $R_3$ | $F_1$ | $F_2$ | $F_3$ |
|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| $I_1$ | 0     | 0     | 1/2   | 1/2   | 0     | 0     | 0     | 0     | 0     | 0     |
| $I_2$ | 0     | 0     | 0     | 1     | 0     | 0     | 0     | 0     | 0     | 0     |
| $S_1$ | 1/5   | 0     | 0     | 0     | 1/5   | 0     | 1/5   | 1/5   | 1/5   | 0     |
| $S_2$ | 1/6   | 1/6   | 0     | 0     | 0     | 1/6   | 1/6   | 1/6   | 0     | 1/6   |
| $R_1$ | 0     | 0     | 1     | 0     | 0     | 0     | 0     | 0     | 0     | 0     |
| $R_2$ | 0     | 0     | 0     | 1     | 0     | 0     | 0     | 0     | 0     | 0     |
| $R_3$ | 0     | 0     | 1/2   | 1/2   | 0     | 0     | 0     | 0     | 0     | 0     |
| $F_1$ | 0     | 0     | 1/2   | 1/2   | 0     | 0     | 0     | 0     | 0     | 0     |
| $F_2$ | 0     | 0     | 1     | 0     | 0     | 0     | 0     | 0     | 0     | 0     |
| $F_3$ | 0     | 0     | 0     | 1     | 0     | 0     | 0     | 0     | 0     | 0     |

Table 2: The description of  $P_{R_1}$

|       | $I_1$ | $I_2$ | $S_1$ | $S_2$ | $R_1$ | $R_2$ | $R_3$ | $F_1$ | $F_2$ | $F_3$ |
|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| $I_1$ | 0     | 0     | 1/2   | 1/2   | 0     | 0     | 0     | 0     | 0     | 0     |
| $I_2$ | 0     | 0     | 0     | 1     | 0     | 0     | 0     | 0     | 0     | 0     |
| $S_1$ | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     |
| $S_2$ | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     |
| $R_1$ | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     |
| $R_2$ | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     |
| $R_3$ | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     |
| $F_1$ | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     |
| $F_2$ | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     |
| $F_3$ | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     |

Table 3: The description of  $P_{R_2}$

|       | $I_1$ | $I_2$ | $S_1$ | $S_2$ | $R_1$ | $R_2$ | $R_3$ | $F_1$ | $F_2$ | $F_3$ |
|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| $I_1$ | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     |
| $I_2$ | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     |
| $S_1$ | 0     | 0     | 0     | 0     | 1/2   | 0     | 1/2   | 0     | 0     | 0     |
| $S_2$ | 0     | 0     | 0     | 0     | 0     | 1/2   | 1/2   | 0     | 0     | 0     |
| $R_1$ | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     |
| $R_2$ | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     |
| $R_3$ | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     |
| $F_1$ | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     |
| $F_2$ | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     |
| $F_3$ | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     |

## 4 Experiment

This section describes the experimental results on the evaluation of VC-Recom. With the development of recommendation systems, how to evaluate the advantages of a recommendation system has gradually become an important issue, which has become an independent research subject. At present, the commonly used evaluation indicators are: accuracy and recall. This paper mainly uses both accuracy and recall to evaluate the VC-Recom.

### 4.1 Experimental setup

We conducted several experiments to validate the proposed VC-Recom method using the dataset in Tab.5. According to the dataset, we built the heterogeneous venture capital network schema (Fig.2).

Table 4: The description of  $P_{R_3}$ 

|       | $I_1$ | $I_2$ | $S_1$ | $S_2$ | $R_1$ | $R_2$ | $R_3$ | $F_1$ | $F_2$ | $F_3$ |
|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| $I_1$ | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     |
| $I_2$ | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     |
| $S_1$ | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     |
| $S_2$ | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     |
| $R_1$ | 0     | 0     | 1     | 0     | 0     | 0     | 0     | 0     | 0     | 0     |
| $R_2$ | 0     | 0     | 0     | 1     | 0     | 0     | 0     | 0     | 0     | 0     |
| $R_3$ | 0     | 0     | 1/2   | 1/2   | 0     | 0     | 0     | 0     | 0     | 0     |
| $F_1$ | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     |
| $F_2$ | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     |
| $F_3$ | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     |

Table 5: Data description of venture capital

| Data Field       | Data Description                |
|------------------|---------------------------------|
| Investment firm  | Total 1119 investment companies |
| Startup project  | Total 8739 startup projects     |
| Investment event | Total 17985 investment events   |
| Field            | Total 31 kinds of fields        |
| Round            | Total 18 financing rounds       |

In the experiments, we wanted to recommend the startups to the VC firms, so we exploited several different venture capital meta paths in Tab.6. We use the capital of the first letter of each node type name as its abbreviation. Like PathRank, not only can VC-Recom exploit the semantics behind the different types of nodes and edges in a heterogeneous graph, but also it can emulate various recommendation semantics such as collaborative filtering, content based filtering, and their combinations. For example, ISXS is a mixed meta path of ISRS and ISFS. ISXS in this paper is equal to  $0.5ISFS+0.5ISRS$ .

Table 6: Meta paths and semantic analysis

| Meta Path | Semantic Analysis  | Same method                        |
|-----------|--|------------------------------------|
| ISIS      | Recommends startup projects with the same investment firms as the projects that the investment firm has invested in                  | User-based collaborative filtering |
| ISFS      | Recommends startup projects that are similar to the field of the projects that the investment firm has invested in                   | Content(field)-based filtering     |
| ISRS      | Recommends startup projects with the same investment rounds as the project that the investment firm has invested in                  | Content(round)-based filtering     |
| ISXS      | Recommends startup projects that are similar to startup projects that the investment firm has invested in based on fields and rounds | Mixed recommendation               |

## 4.2 Experimental results

For studying the recommendation performance of the VC-Recom algorithm, we need to choose some venture capital meta paths first, and then study the influence of the number of recommended startups  $k$  to accuracy and recall of the recommendation under the guide of venture capital meta paths. Furthermore, we compare the result of VC-Recom to the user-based collaborative filtering algorithm and HeteRecom.

Each experimental result is the average of 5 runs with the dataset divided into two parts, training set (80% of the dataset) and test set (20% of the dataset). And we set  $w_{trans} = 0.35$ ,  $w_{re} = 0.05$ ,  $w_{path} = 0.6$  to conduct experiments to get the best result after optimizing the parameters  $w_{trans}$ ,  $w_{re}$  and  $w_{path}$ .

#### 4.2.1 The effect of recommendation number $k$

Fig.3 and Fig.4 study the effect of recommendation number  $k$  on the accuracy and the recall rate of each path.

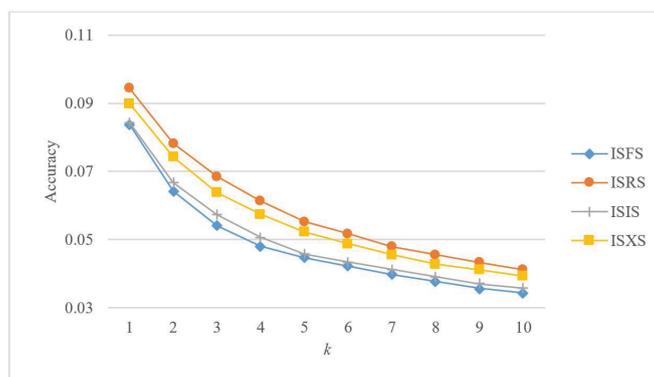


Figure 3: The influence on the accuracy of each path when  $k$  changes

From Fig.3, we can see that as the value of  $k$  increases continuously in the venture capital algorithm based on heterogeneous information networks, the recommendation accuracy under each path gradually decreases. Among all the meta-paths ISFS, ISRS and ISIS, the accuracy result of ISRS is best, showing that the venture capital node type round reflects the investment companies' preference mostly. ISFS has the lowest recommendation accuracy rate, indicating the venture capital node type field reflects the investment companies' preference less than other node types. The ISXS integrates the information of ISRS and ISFS, and has a high recommendation accuracy rate.

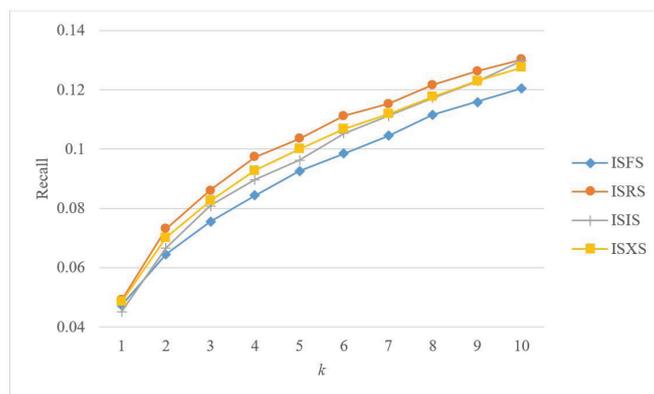


Figure 4: The influence on the recall of each path when  $k$  changes

As we can see from Fig.4, with the increase of  $k$ , the recall rate under each path gradually increases. The recommendation recall rate of meta path ISRS and ISXS is generally higher than that of other paths because of the larger association between the node type round and the investment companies in meta path ISRS. However, ISIS and ISFS have lower recall rates, but ISIS has a higher recall rate than ISFS because of the venture capital node type field having less association with investment companies.

#### 4.2.2 Comparison of VC-Recommend other algorithms

This part mainly compares VC-Recom with user-based collaborative filtering algorithm and Het-eRecom algorithm under the meta path. Because the venture capital meta path ISIS has the same

semantic to user-based collaborative filtering algorithm (UserCF), we choose the venture capital meta path ISIS to make comparisons.

(1) Comparison of recommended results

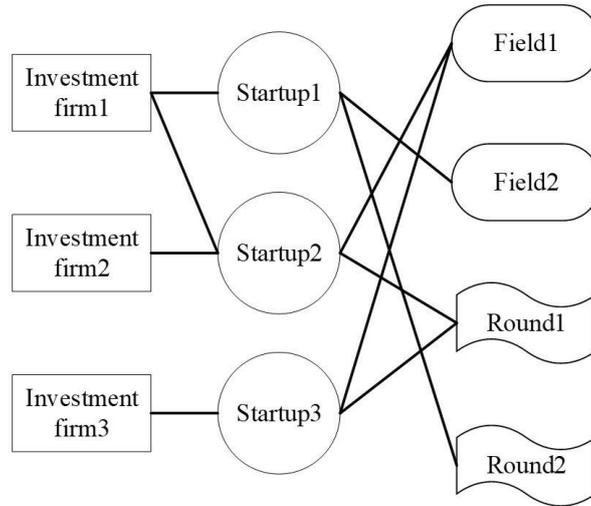


Figure 5: A heterogeneous venture capital network

Fig.5 is a heterogeneous venture capital network. Now we want to recommend a startup to investment firm2 by using VC-Recom, HeteRecom and UserCF according to Fig.5. And the result is in Tab.7.

Table 7: Recommendation result analysis

| VC-Recom | HeteRecom | UserCF   |
|----------|-----------|----------|
| Startup3 | Startup1  | Startup1 |

From Tab.7, we can see that if we use HeteRecom and UserCF algorithms, the recommended startup is startup1, but VC-Recom is startup3. What makes the difference is that HeteRecom and UserCF algorithms only consider the relationship between startups and investment firms, while the VC-Recom algorithm can use the information of any nodes in the heterogeneous information network by random walking. Startup3 has the same attributes with startup2, which are much closer to investment firm2's preferences. So, recommending startup3 to investment firm2 is more reasonable than startup1. That is to say, VC-Recom can recommend startups that are much more accordant with investment firms' preferences.

(2) Comparison in accuracy and recall

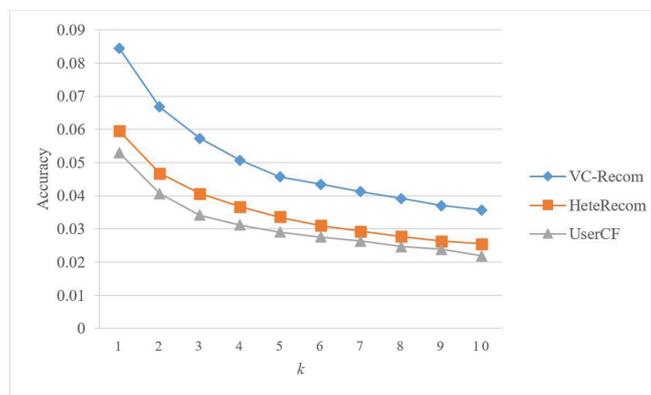


Figure 6: Comparison of the accuracy of the three algorithms

Fig.6 studies the difference of the accuracy among VC-Recom, HeteRecom and user-based col-

laborative filtering algorithm. It can be seen that the recommendation accuracy of VC-Recom and HeteRecom are decreasing as the recommendation number  $k$  increases. And their accuracy rates are higher than UserCF. The accuracy rate of the VC-Recom algorithm is higher than that of HeteRecom. That is to say, VC-Recom is more effective to recommend startups preferred by investment companies.

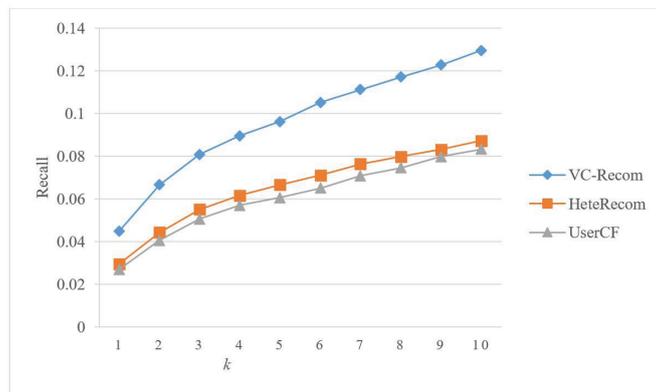


Figure 7: Comparison of the recall of the three algorithms

Fig.7 studies the difference of the recall rate among VC-Recom and the other two algorithms. It can be seen that the recall rate is increasing with the increase of the recommendation number  $k$ . VC-Recom's recall rate is higher than those of HeteRecom and UserCF. UserCF has the lowest recall rate.

## 5 Conclusions

This paper establishes a heterogeneous venture capital network model, applying the recommendation algorithm to the field of venture capital analysis, in which the venture capital entities and their attributes are transferred into network nodes and incorporated into the heterogeneous information network model. Then it proposes the venture capital recommendation algorithm VC-Recom based on the established heterogeneous information network model, and compares the proposed algorithm with user-based collaborative filtering algorithm and HeteRecom algorithm. The experiment results show that VC-Recom is more effective to recommend startups preferred by investment companies.

## Funding

This work is supported by National Natural Science Foundation of China (NSFC) under Grant No. 71971025, 71271027.

## Author contributions

The authors contributed equally to this work.

## Conflict of interest

The authors declare no conflict of interest.

## References

- [1] Chircop, J.; Johan, S.; Tarsalewska, M. (2017). Does religiosity influence venture capital investment decisions?, *Social Science Electronic Publishing*, 2017.
- [2] Dicks, D. L.; Fulghieri, P. (2016). Innovation waves, investor sentiment, and mergers, *Social Science Electronic Publishing*, 2016.

- [3] Feng, X. D.; Wu, S.; Tang, Z. W.; Li, Z. C. (2018). Sparse latent model with dual graph regularization for collaborative filtering, *Neurocomputing*, 284, 128–137, 2018.
- [4] Jeh, G.; Widom, J. (2003). Scaling personalized web search, In *International Conference on World Wide Web*, 47, 271–279, 2003.
- [5] Jiang, L. (2016). Application Analysis on Internal Rate of Return Rule for Investment Decision, In *International Conference on Humanity, Education and Social Science*, 2016.
- [6] Kahng; Minsuk; Lee; Sangkeun; Lee; Sanggoo. (2011). Ranking objects by following paths in entity-relationship graphs, In *ACM International Conference on Information and Knowledge Management*, 11–18, 2016.
- [7] Kleinberg, J. M. (1998). Authoritative sources in a hyperlinked environment, *Acm-Siam Symposium on Discrete Algorithms*, 46, 668–677, 1998.
- [8] Lee, S.; Park, S.; Kahng, M.; Lee, S. G. (2012). PathRank: a novel node ranking measure on a heterogeneous graph for recommender systems, In *ACM International Conference on Information and Knowledge Management*, 1637–1641, 2012.
- [9] Li, M.; Wu, C. (2016). A distance model of intuitionistic fuzzy cross entropy to solve preference problem on alternatives, *Mathematical Problems in Engineering*, 8, 1–9, 2016.
- [10] Page, L. (1998). The PageRank citation ranking : Bringing order to the web, *Stanford Digital Libraries Working Paper*, 9(1), 1–14, 1998.
- [11] Shi, C.; Zhou, C.; Kong, X.; Yu, P. S.; Liu, G.; Wang, B. (2012). HeteRecom: a semantic-based recommendation system in heterogeneous networks, In *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1552–1555, 2012.
- [12] Snellman, Carita. Lilian. (2015). University in Knowledge Society: Role and Challenges, *Journal of System and Management Sciences*, 5(4), 84–113, 2015.
- [13] Stone, T.; Zhang, W.; Zhao, X. (2013). An empirical study of top-n recommendation for venture finance, *ACM Press*, 1865–1868, 2013.
- [14] Sun, S. L.; Chen, V. Z.; Sunny, S. A.; Chen, J. (2018). Venture capital as an ecosystem engineer for regional innovation in an emerging market, *International Business Review*, 13–18, 2018.
- [15] Sun, Y.; Han, J.; Yan, X.; Yu, P. S.; Wu, T. (2011). PathSim: Meta Path-Based Top-K Similarity Search in Heterogeneous Information Networks, *Proceedings of the Vldb Endowment*, 4(11), 992–1003, 2011.
- [16] Xu, R.; Chen, H.; Zhao, L. J. (2017). Predicting Corporate Venture Capital Investment, *Association for Information Systems*, 8, 5841–5849, 2017.
- [17] Xu, S. Y.; Raahemi, B. (2016). A semantic-based service discovery framework for collaborative environments, *International Journal of Simulation Modelling*, 15(1), 83–96, 2016.
- [18] Zuo, Z.; Shi, F.; Tan, G.; Accounting, S. O. (2017). Does State-owned Venture Capital Play a "Crowding-in" Role? An Empirical Research Based on Certification Mechanism, *Journal of Finance & Economics*, 43(12), 17–29, 2017.



Copyright ©2020 by the authors. Licensee Agora University, Oradea, Romania.

This is an open access article distributed under the terms and conditions of the Creative Commons Attribution-NonCommercial 4.0 International License.

Journal's webpage: <http://univagora.ro/jour/index.php/ijccc/>



This journal is a member of, and subscribes to the principles of,  
the Committee on Publication Ethics (COPE).

<https://publicationethics.org/members/international-journal-computers-communications-and-control>

*Cite this paper as:*

Wu, S.; Li, H.; Liu, L. (2020). A Venture Capital Recommendation Algorithm based on Heterogeneous Information Network, *International Journal of Computers Communications & Control*, 15(1), 1009, 2020. <https://doi.org/10.15837/ijccc.2020.1.3779>