Friend Recommendation Method of Weighted Networks Based on Value and Match

Simian Wu, Fuyi Wei¹, Zhirui Huang, Jiezi Wang, Juanli Zheng, Yitao Jiang and Longhui Lin
South China Agricultural University, College of Mathematics and Information, Guangzhou, China
¹Corresponding author, Email: weifuyi@scau.edu.cn

Abstract. To improve the accuracy of the recommendation in the knowledge social network, the paper proposes a friend recommendation algorithm based on user information to measure the value and the match between the users. On the one hand, it establishes the directed graph model of the value by ‘the number of answers’ and ‘the number of likes’, to calculate the users’ value scores. And it uses the material diffusion principle to predict the value scores which are zero in the model. On the other hand, it calculates the users’ similarity through ‘topic of concern’ and ‘industry’ simulates the similarity of user pairs by using probability distribution and establishes likelihood function. After using ‘EM’ algorithm to estimate the function parameters, it gets the users’ matched and mismatched probability to calculate their match scores and establish undirected graph model. Finally, it builds the comprehensive score formula between the value and the match. Combined with the directed and undirected graph model, the Top-N friend recommendation algorithm is recommended. Through the experiments, it shows that the method is better than the recommendation algorithm based on collaborative filtering in Precision, Recall and F1-measure on a large scale recommendation.

1. Introduction
With the development of science and technology, the online social network attracts lots of users. Not only do the users transfer the real relationships to the network, but also they establish singly online relationships unrelated to the offline friendships. On the social network, they build up the platform where they can communicate and share information with each other. As people's learning needs increase, some representative knowledge social platforms have appeared. For example, ‘Zhihu’ is the largest knowledge social network in the Chinese Internet.

In the many services offered by the social networks, there is an important personalization service called ‘recommend friends for users’. The service greatly extends the user's network of relationships and promotes the users’ communication in thought. What's more, it promotes the rapid development of social networks. In this paper, using the resources of ‘Zhihu’, we improve the traditional friend recommendation algorithms and get the friend recommendation method of weighted networks.

2. Related Work
Research on personalized recommendation technology and system, people has made rich theory and achievements, including recommendation algorithms based on user information and topology [1] [2].

Based on user information, the filter method includes user content and collaborative filtering [3]. The content-based filtering method is to match the same preferences or attributes among users from their information. Based on collaborative filtering, the method uses the users' rating of the project to find their neighbours whose preferences are like them. According to their neighbours’ other
preferences, it can predict the possible preferences of the target users. The level of the rating represents the users' preference for the project. The closer the user score is, the more similar the user's interest in the field is [4] [5]. By calculating the similarity between users based on the characteristics of interest, Xie Xing and other people proposed a generic recommendation framework for potential friends [6]. Li and other people incorporated the trust relationship in social networks into collaborative filtering and selected the users' neighbours by using trust of trust networks. The method improves the accuracy of the recommendation [7]. Tang Ying and other people proposed a clustering based on users and projects. Then they established the probabilistic model with the help of scores. The method calculates the similarity method based on scores between users [8]. Qin Jiwei put forward a collaborative recommendation algorithm, which is combined by scores and trust. Considering the similarity and trust, the method improves the recommendation accuracy and score coverage [9].

Based on user information, the recommendation method faces sparsity. Zhang Xuesheng proposed the collaborative filtering recommendation algorithm with the feature of controversy [10]. By using the user's rating of the project and the principle of material diffusion, Wang Qian calculated the similarity between users and solved the sparsity problem [11]. Xie proposed grey prediction model [12]. To infer more similarity of users, Hu proposed a hybrid personalized random walk algorithm [13].

Many friend recommendation methods are based on the trust and the similarity. And they usually face the sparsity. In this paper, the value is considered. Combined with the value and the match, we propose a new algorithm based on the user information. ‘Zhihu’ is a platform for learning rather than maintaining relationships, so the trust is unsuited to it. Only by taking full consideration of the influence about the value, the accuracy of the recommendation can be improved. For the problem of sparsity in the value, this paper uses the method of Wang Qian with the principle of material diffusion.

In terms of value, we use the ‘number of answers’ and ‘the number of likes’ to calculate users’ value scores and solves the sparsity combined with the principle of material diffusion. In terms of match, Ee-Peng Lim and other people proposed the method of collective network linkage across heterogeneous social platforms. The method models the similarity of user pairs to accommodate heterogeneous attributes, using different probability distributions of their similarity values [14]. In this paper, we calculate the similarity of the ‘topic of concern’ and ‘industry’ between users. Then we use the probability distribution to simulate the similarity of user pairs and measure the user’s match from matched probability and mismatched probability between users. Finally, combined with the value score and match score, a friend recommendation list is generated for the target user.

3. The Weight Directed Graph of User’s Value

‘Zhihu’ is an online community where people can ask and answer. The users share their knowledge, experiences and opinions. So the value of the users plays an important part in the recommendation. In this paper, the user’s value score is calculated by using the ‘the average number of like to answer’.

3.1. The Weighted Formula of User’s Value

We suppose that D is a directed graph, vertex set of users is \( U = \{ u_1, u_2, \ldots, u_y, \ldots, u_y, \ldots, u_z, \ldots, u_n \} \) and a set of directed arc is \( E_D = \{ (u_y, u_x) | u_x, u_y \in U \} \). Among them, \((u_y, u_x)\) is the directed arc, which is from the vertex \( u_y \) to the vertex \( u_x \). \( \omega(u_y, u_x) \) is the weight of the directed arc.

There are lots of topics in the ‘Zhihu’. Under the topics, users can ask and answer the questions. If the user thinks the answer is right, he can give it a like. For the reason, we define the indexes. \( Q_{u_y \rightarrow u_x} \) is an answer set of the user \( u_y \) under the topics which are focus by user \( u_x \). The number of the answers is \( |Q_{u_y \rightarrow u_x}| \). \( K_{u_y \rightarrow u_x} \) is a like set of the user \( u_y \) are under the topics which are focus by user \( u_x \). The number of the likes is \( |K_{u_y \rightarrow u_x}| \). So we define that ‘the average number of like to answer’ is \( \frac{|K_{u_y \rightarrow u_x}|}{|Q_{u_y \rightarrow u_x}|} \). If \( |Q_{u_y \rightarrow u_x}| = 0 \), we define \( \frac{|K_{u_y \rightarrow u_x}|}{|Q_{u_y \rightarrow u_x}|} = 0 \).

The weighted formula about value of \( u_y \) to \( u_x \):
Among them, the minimum of “the average number of like to answer” in U is \( \min_{u \in U} \), and the maximum is \( \max_{u \in U} \). To prevent the denominator from zero, we set \( \varepsilon = 0.001 \).

If user \( u_y \) doesn’t answer the questions under the topics which are focus by user \( u_x \), his answers doesn’t get likes or “the average number of like to answer” is the least among all users, the value score of \( u_y \) to \( u_x \) is zero. But in fact, \( u_y \) may be valuable to \( u_x \). In this paper, we can find \( u_k \). User \( u_y \) and user \( u_x \) are valuable to \( u_k \). Based on the principle of material diffusion, we can get the similarity between \( u_y \) and \( u_x \). Then we distribute the resources to \( u_y \), to predict the value of \( u_y \) to \( u_x \).

First, calculate the value quotas that \( u_x \) is willing to assign to \( u_y \). That is, the value similarity between \( u_y \) and \( u_x \):

\[
S(u_y, u_x) = \frac{1}{d^+(\omega_{u_x})} \sum_{k=1}^{n} \frac{\omega(u_y, u_k) \omega(u_x, u_k)}{d^+(\omega_{u_y})}
\]

Among them, \( d^+(\omega_{u_x}) = \sum_{k=1, k \neq x}^{n} \omega(u_x, u_k) \) is the sum of the value scores, which is \( u_x \) to others. \( d^-(\omega_{u_y}) = \sum_{l=1, l \neq y}^{n} \omega(u_l, u_k) \) is the sum of the value scores, which is others to \( u_k \). If \( d^+(\omega_{u_x}) \) or \( d^-(\omega_{u_y}) \) is zero, we define that the value similarity is zero.

If the value score is zero, we predict it:

\[
\overline{\omega}(u_y, u_x) = \frac{\sum_{z=1, z \neq x}^{n} S(u_y, u_z) \times \omega(u_x, u_z)}{\sum_{z=1, z \neq x}^{n} S(u_y, u_z)}
\]

In a word, if user \( u_y \) doesn’t answer the questions under the topics which are focus by user \( u_x \), his answers doesn’t get likes or “the average number of like to answer” is the least among all users, the paper use formula (3) to predict. That is, if the value scores of the formula (1) are zero, we predict the scores with formula (3).

The weighted formula about value of \( u_y \) to \( u_x \):

\[
\overline{\omega}(u_y, u_x) = \left\{ \begin{array}{ll}
\frac{K_{u_y-u_x} \cdot \min_{u \in U} K_{u_y-u_x} \cdot \min_{u \in U} K_{u_y-u_x} + \varepsilon}{\max_{u \in U, u_x \in U} Q_{u_y-u_x} - \min_{u \in U, u_x \in U} Q_{u_y-u_x} + \varepsilon}, & K_{u_y-u_x} \neq 0 \\
0, & K_{u_y-u_x} = 0
\end{array} \right.
\]

3.2. The Weight Directed Graph Model of the User’s Value

Because ‘the number of the answers’ and ‘the number of likes’ are directional, the directed graph model is established. In the model, the point represents the user and the directed arc \((u_y, u_x)\) represents the value of \( u_y \) to \( u_x \). The weight of the directed arc represents the user’s value score. \( \omega(u_y, u_x) \) or \( \overline{\omega}(u_y, u_x) \) represents the weight of the directed arc, which is form the point \( u_y \) to \( u_x \). Using the formula (1) or (3), we can get the value score of user \( u_y \) to target user \( u_x \).

4. The Weight Undirected Graph of the User’s Match

In this section, we use the ideas of Ee-Peng Lim and others to calculate the friend similarity in isomorphic social platforms. We establish the undirected graph model of friend similarity. The similarity function is used to measure the similarity between users in ‘topic of concern’ and ‘industry’. Then we use the probability distribution to simulate the similarity of user pairs.
4.1. Establish the Likelihood Function, Estimate the Parameters by Using the Em Algorithm

We suppose that $\alpha$ is ‘topic of concern’, $\beta$ is ‘industry’ and $r_j (j = 1, 2, \cdots, \frac{n^2-n}{2})$ is the use pairs between the users. $\alpha_\gamma$ is the set of topics by focus on user $u_\gamma$, and $\alpha_\delta$ is the set of topics by focus on user $u_\delta$. Using Jaccard similarity coefficient, we can calculate the similarity about ‘topic of concern’ and ‘industry’ between the users.

\[
S^\alpha_j = \frac{|\alpha_\gamma \cap \alpha_\delta|}{|\alpha_\gamma \cup \alpha_\delta|} \quad (4)
\]

\[
S^\beta_j = \frac{|\beta_\gamma \cap \beta_\delta|}{|\beta_\gamma \cup \beta_\delta|} \quad (5)
\]

Among this, $S_j^i (i = \alpha, \beta)$ is the similarity function of $r_j$ in $i$. $|\alpha_\gamma \cap \alpha_\delta|$ represents the number of topics, which are the intersection between the users’ ‘topic of concern’. And $|\alpha_\gamma \cup \alpha_\delta|$ represents the number of topics, which are the union between the users’ ‘topic of concern’. $|\beta_\gamma \cap \beta_\delta|$ is the number of the industries, which are the intersection between the users’ ‘industry’. And $|\beta_\gamma \cup \beta_\delta|$ is the number of the industries, which are the union. If the two users’ similarity reaches the threshold $q$, they are considered as match; otherwise, they are considered as mismatch. The set of matched pairs is called a matched set, denoted as $M$. A collection of mismatched pairs is called a mismatched set, denoted as $M^c$. For the each $S_j^i$, we model the similarity values of user pairs using two different probability distribution functions, one for matched user pairs and another for mismatched user pairs:

\[
P(S_j^i | r_j \in M, \eta) \sim f_{1,i}(S_j^i; \theta_{1,i})
\]

\[
P(S_j^i | r_j \in M^c, \eta) \sim f_{2,i}(S_j^i; \theta_{2,i})
\]

Among them, $\eta = \{p, \theta_{1,\alpha}, \theta_{1,\beta}, \theta_{2,\alpha}, \theta_{2,\beta}\}$. $f_{1,i}(S_j^i; \theta_{1,i})$ and $f_{2,i}(S_j^i; \theta_{2,i})$ are the probability density function of the exponential family distribution function in match and mismatch. $\theta_{1,i}$ and $\theta_{2,i}$ are unknown parameters of the function. Function to meet:

\[
f(x; \theta) = h(x) e^{\theta^T \tau(x) - z(\Theta)}
\]

The literature proposed the following likelihood function [14]:

\[
L(\theta) \propto \sum_{j=1}^{n} (l_j, 1 - l_j) \left[ \sum_{i=\alpha, \beta} \theta_{1,i}^T T_{1,i}(S_j^i), \sum_{i=\alpha, \beta} \theta_{2,i}^T T_{2,i}(S_j^i) \right]^T
\]

\[- \sum_{j=1}^{n} (l_j, 1 - l_j) \left[ \sum_{i=\alpha, \beta} z_{1,i}(\theta_{1,i}), \sum_{i=\alpha, \beta} z_{2,i}(\theta_{2,i}) \right] \right]^T + \sum_{j=1}^{n} (l_j, 1 - l_j) [\log p, \log (1 - p)]^T \quad (6)
\]

Among them, $l_j = \begin{cases} 1, & r_j \in M \\ 0, & r_j \in M^c \end{cases}$, $p = P(r_j \in M | \eta), \quad 1 - p = P(r_j \in M^c | \eta)$. Because of the implicit variable $l_j$, we use EM algorithm to estimate $\eta$. First, given $\eta_t$, we find the user’s matched probability $p_t$ as the expectation of $l_j$. Second, we maximize the log-likelihood function to estimate the parameter $\eta_{t+1}$, and get the matched probability of user $p_{t+1}$. The above two steps are repeated until the specified number of iterations is reached or converged. Finally, we can obtain $\hat{\eta}$ which is the parameter estimation of $\eta$, the user’s matched probability $\hat{p}$ and the mismatched probability $1 - \hat{p}$. 
4.2. The Weight Undirected Graph Model of the User’s Match

We calculate the matched probability \( P(\bar{r}_j \in M^\ell | \bar{s}_j, \hat{\eta}) \) and mismatched probability \( P(\bar{r}_j \in M^\ell | \bar{s}_j, \hat{\eta}) \).

Let \( s_j = \frac{1}{2} \left( \frac{P(\bar{r}_j \in M^\ell | \bar{s}_j, \hat{\eta})}{P(\bar{r}_j \in M^\ell | \bar{s}_j, \hat{\eta})} + \frac{P(\bar{r}_j \in M^\ell | \bar{s}_j, \hat{\eta})}{P(\bar{r}_j \in M^\ell | \bar{s}_j, \hat{\eta})} \right) \), we calculate the match score of the user pair \((u_x, u_y)\):

\[
\omega'((u_x, u_y)) = \frac{s_j - \min(s_j)}{\max(s_j) - \min(s_j) + \varepsilon}
\] (7)

Among them, \( j = 1, 2, \cdots, \frac{n^2-n}{2} \). For each user pair, we calculate their match score. To prevent the denominator from zero, we set \( \varepsilon = 0.001 \).

Since the match between the users have no direction, an undirected graph model is established. The points represent users. And the sides represent the match between users. \( \omega'((u_x, u_y)) \) represents a weight to measure a match score between the user \( u_x \) and \( u_y \).

5. Friend Recommendation of Comprehensive Weighted Model Based on User’s Value and Match

This paper proposes a recommendation algorithm, which is the weighted sum of the two dimensions. The match and the value are given different weights, which are determined by the training results in experiment. The user \( u_y \)’s comprehensive score for the target user \( u_x \) is as follows:

\[
\omega^*(u_x, u_y) = \begin{cases} 
\lambda \omega(u_y, u_x) + (1 - \lambda) \omega'(u_y, u_x), & \omega(u_y, u_x) \neq 0 \\
\lambda \omega(u_x, u_y) + (1 - \lambda) \omega'(u_x, u_y), & \omega(u_x, u_y) = 0 
\end{cases}
\] (8)

Among them, The greater \( \omega^*(u_x, u_y) \), the more recommended between the users.

The dataset is divided into a training set and a test set. According to the formula (8), we calculate the comprehensive score of other users to target user \( u_x \). And Top-N users are selected as potential friends to recommend the target user \( u_x \). This paper’s friend recommendation algorithm is as follows:

According to the formula (1), we can calculate the value scores of other users to the target user \( u_x \). If the scores are zero, turn to step (b). Otherwise, turn to step (c).

According to formula (3), we can recalculate the value scores. Then turn to step (c).

Using formulas (4) and (5), we can calculate the similarity between \( u_x \) and other users in the ‘topic of concern’ and ‘industry’.

Through training dataset, we can make an experiment to determine the threshold of similarity q, then use the threshold to determine the matched and mismatched sets.

According to the formula (6), EM algorithm is used to calculate the parameter \( \hat{\eta} \), which maximizes the likelihood function.

Using \( \hat{\eta} \), we can determine the matched probability and mismatched probability between other users and the target user \( u_x \). Then their match scores are calculated by using formula (7).

Through experiment, we can calculate the optimal weight \( \lambda \) in the formula (8).

According to the formula (8), we can calculate the comprehensive scores of the other users to \( u_x \). Then we sort the users by descending order and select N users who get the highest score to recommended to \( u_x \).

Among them, Step (d) (e) is done using training dataset and the remaining steps make use of the test dataset.

6. Experimental Results and Analysis

Through experiments, it is calculated that the optimal value of the similarity threshold q is 80, and the optimal value of \( \lambda \) is 0.4. Then we compare our algorithm (VM algorithm) with the friend recommendation algorithm based on collaborative filtering (CF algorithm) [15] by using three indicators.
6.1. Selection of Dataset
The data is from 500 users in ‘Zhihu’, including 1) users’ information, which are focus by users, including the topics and users. 2) Users’ other attributes, including the industries, the number of answers and the number of likes about their answers. The 3/5 data is training set, and the rest is a test set.

6.2. The Weight Directed Graph Model of the User’s Value
To evaluate the accuracy of recommendation, VM algorithm is compared with CF algorithm, using the Precision, Recall and the F1-measure. The three indicators of the evaluation are:

\[
\text{Precision} = \frac{|N_D \cap N_L|}{|N_L|}
\]

\[
\text{Recall} = \frac{|N_D \cap N_L|}{|N_D|}
\]

\[
F1 - \text{measure} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}
\]

\(N_D\) is the friend's set in the test set. \(N_D\) is the number of friends in the test set. \(N_L\) is the recommended user set. \(|N_L|\) is the number of recommended users. \(N_D \cap N_L\) represents the set that the recommended user has become the target user’s friends. \(|N_D \cap N_L|\) represents the number of users that the recommended user has become target user’s friends.

6.3. Data Processing
We use Jaccard similarity coefficient to calculate users’ similarity. The topic similarity is fitted into continuous data, and the EM algorithm is used to calculate the user's match score. The industry similarity is discrete data, so the industry match score is regarded as industry similarity.

Because of the limitations of the Jaccard similarity coefficient, it is possible that the similarity of users with more ‘topics of concern’ is higher than that of users with fewer ‘topics of concern’. But in fact, the user pairs, who’s the union of the ‘topics of concern’ is small but the intersection is big, have more significance to recommend.

To make better recommendations, the data is processed before the EM algorithm. If the number of topics which are the intersection between the users’ ‘topic of concern’ is less than the threshold, the original similarity \(S_0\) will be increased. Based on the study of data distribution, the paper selects the appropriate threshold and the extent of similarity improvement to achieve the best indicators. The threshold of the topic \(q\) is taken as 40, 60, 80, 100, and the similarity is increased by 0.1, 0.2, 0.3.

| range (value, similarity) | Precision | Recall | F1       |
|--------------------------|-----------|--------|----------|
| (40, \(S_0 + 0.1\))     | 0.021053  | 0.054099 | 0.029882 |
| (40, \(S_0 + 0.2\))     | 0.018713  | 0.044670 | 0.026377 |
| (40, \(S_0 + 0.3\))     | 0.021053  | 0.055723 | 0.030560 |
| (60, \(S_0 + 0.1\))     | 0.031579  | 0.070409 | 0.040138 |
| (60, \(S_0 + 0.2\))     | 0.016374  | 0.043969 | 0.023862 |
| (60, \(S_0 + 0.3\))     | 0.031579  | 0.048150 | 0.032087 |
| (80, \(S_0 + 0.1\))     | 0.042105  | 0.076257 | 0.036139 |
| (80, \(S_0 + 0.2\))     | 0.042105  | 0.060468 | 0.030058 |
| (80, \(S_0 + 0.3\))     | 0.063158  | 0.060840 | 0.033547 |
| (100, \(S_0 + 0.1\))    | 0.042105  | 0.072828 | 0.039296 |
| (100, \(S_0 + 0.2\))    | 0.042105  | 0.065197 | 0.035744 |
| (100, \(S_0 + 0.3\))    | 0.052632  | 0.065197 | 0.034049 |

From the experiment, we chose that \(q\) is 80 and the similarity is increased by 0.1. That is, the number of topics which are the intersection between the users’ ‘topic of concern’ is less than 80, the similarity increased by 0.1. Finally, we calculate the value score and match score between the users.
6.4. Experimental Results
The experiment of our algorithm contains an unknown parameter $\lambda$. At first, the paper finds out the value of $N$ when the experimental results are optimal in CF algorithm. Under the $N$, we seek the $\lambda$, which has an optimal result in VM algorithm. Finally, we select the value of the $\lambda$ to compare the three indicators of the two algorithms under different $N$.

In CF algorithm, we construct user’s similarity matrix and value scoring matrix. The similarity matrix is an $n \times n$ symmetric matrix, which takes the sum of similarities between ‘topic of concern’ and ‘industry’ as an element. The value scoring matrix is an $n \times n$ matrix, which takes the value score between the users as an element. By multiplying the two matrices, a recommendation matrix is obtained. Using the experimental data in CF algorithm, we can get the following recommended results:

| N   | Precision | Recall | F1-measure |
|-----|-----------|--------|------------|
| 1   | 0.042105  | 0.013158 | 0.020050   |
| 2   | 0.026316  | 0.018421 | 0.021672   |
| 3   | 0.021053  | 0.021053 | 0.021053   |
| 4   | 0.018421  | 0.023158 | 0.020522   |
| 5   | 0.023158  | 0.041871 | 0.029822   |
| 6   | 0.021053  | 0.043187 | 0.028307   |
| 7   | 0.022556  | 0.049678 | 0.031026   |
| 8   | 0.021053  | 0.050994 | 0.029802   |
| 9   | 0.019883  | 0.051871 | 0.028747   |
| 10  | 0.018947  | 0.053187 | 0.027941   |

It can be seen from Table 1 that when the $N$ is equal to 7, the indicators of evaluation is optimal. Considering the same conditions, this experiment finds the best value of $\lambda$ at $N = 7$, which optimizes the indicators.

![Figure 1. The three indicators in different $\lambda$.](image1)

![Figure 2. The Precision in different $N$.](image2)

From Figure 1, we can see that when the $N$ is equal to 7, we should make the $\lambda=4$ to maximize the three indexes. To compare the validity of the two algorithms, we calculate the Precision, Recall and F1-measure of the two algorithms under different $N$, according to the recommended results of CF algorithm and VM algorithm. The result is as follows:
From Figure 2, Figure 3 and Figure 4, we can find that VM algorithm and CF algorithm are similar in Precision, Recall and F1-Measure when the N is equal to 6 or less than 6. When N is greater than 7, VM algorithm is slightly better than CF algorithm. According to the above results, when N is small, there is no obvious difference between the two algorithms, and VM algorithm is slightly inferior to CF algorithm. However, with the increase of N, VM algorithm has higher Precision, Recall and F1-measure than CF algorithm, which shows that VM algorithm has more obvious advantages when it is used for accurate recommendation and large-scale recommendation (when N ≥ 7).

7. Summary and Prospect
‘Zhihu’ is a knowledge-based social platform, where people gather for learning and discussion. For the reason, the recommendation algorithm on the platform is different from other platforms, such as micro-blog, Renren and so on. People prefer to know some people who are similar to them and provide effective value for them. Therefore, a friend recommendation algorithm based on match and value is proposed in the paper. Experiments show that the proposed VM algorithm is more robust than CF recommendation algorithm from Precision, Recall and F1-measure. That is, VM algorithm is better than CF algorithm in precision recommendation and large-scale recommendation. However, our paper has some shortcomings. First, it is easy to recommend ‘big V’ (users who have many fans) to other users. Therefore, how to measure the weight of the value and match is a question that needs to be weighed in many experiments. And the data on the platform is relatively small and sparse, and it may bring some errors to the experiment. Second, we should add more attributes to more fully evaluate the user’s value. Third, we should use questionnaires to evaluate and improve the algorithm.

8. Acknowledgments
S M Wu, F Y Wei, Z R Huang, J Z Wang, J L Zheng, Y T Jiang, L H Lin thanks “National Undergraduate Scientific and Technological Innovation Project” and “Demonstration of joint training base about South China Agricultural University and Shenzhen VANDA Computer Software Co. Ltd” for their funding and support.

9. References
[1] Wu. H and Liu. D S Study on friends recommendation methods in social networks 2015 New Technology of Library and Information Service 35 59- 65
[2] Song B W Research on friend recommendation algorithm in online social network 2016 Journal of Taiyuan University of Technology 60 67
[3] Wang Y D and Huang T B. Weibo friend recommendation algorithm based on social networks and collaborative filtering 2016 Journal of Beijing Jiaotong University 42 70-75
[4] Zhu. M and Wei H Q Collaborative filtering algorithm based on item attribute preference 2017 Journal of Computer and Modernization 33 23 -26
[5] Li T L Research on friend recommendation technology in social networks 2014 Dalian Maritime University 58 54
[6] Xie X 2010 Potential friend recommendation in online social network Proceedings of the 3rd ACM International Conference on Syber, Physical and Social Computing (New York: ACM Press) 2010 pp 831-835
[7] Li Y M, Wu C T and Lai C Y A social recommender mechanism for ecommerce: combining similarity, trust, and relationship 2013 Decision Support Systems 55 740 -752
[8] Tang Y, Zhong N J and Fang J An improved friend recommendation algorithm combined with user rating information 2016 Computer Science 43 111-115
[9] Qin J W, Zheng Q H and Zheng D L Collaborative rating and trust recommendation algorithm 2013 Journal of Xi’an Jiaotong University 54 100-104
[10] Zhang X S, Chen C and Zhang Y F Collaborative filtering recommendation algorithm based on convergence dispute features 2012 Microcomputer Systems 33 694 - 699
[11] Wang Q and Duan S Y An improved recommendation algorithm based on two-graph network structure 2013 Application Research of Computers 30 771-774
[12] X F, C Z and Shang J Grey forecast model for accurate recommendation in presence of data sparsity and correlation 2014 Knowledge-Based Systems 69 179-190
[13] Hu Y, Peng Q M and Hu X H 2014 A time-aware and data sparsity tolerant approach for Web service recommendation Proceedings of the 2014 IEEE International Conference on Web Services (Washington, DC: IEEE Computer Society) pp 33-40
[14] Ling E P, Luo D W and Zhu F D 2015 CNL: Collective Network Linkage across Heterogeneous Social Platforms The IEEE International Conference on Data Mining pp 757-62
[15] Liu P , Ma B Q, Cao G Huang Y T and Qin M Optimization of friend system based on collaborative filtering recommendation for SNS website 2011 China Digital Television 8 65- 67