

Individual Friends Recommendation Based on Random Walk with Restart in Social Networks

Jibing Gong^{1,2,3}, Xiaoxia Gao^{1,2(✉)}, Yanqing Song^{1,2},
Hong Cheng⁴, and Jingjing Xu^{1,2}

¹ School of Information Science and Engineering, Yanshan University,
Qinhuangdao City, China
gongjibing@gmail.com, gaouxysu@gmail.com, songyqysu@gmail.com,
xujjysu@gmail.com

² The Key Laboratory for Computer Virtual Technology and System Integration
of Hebei Province, Yanshan University, Qinhuangdao City, China

³ State Key Lab of Mathematical Engineering and Advanced Computing,
Wuxi, China

⁴ Department of Systems Engineering and Engineering Management,
The Chinese University of Hong Kong, Hong Kong, China
hcheng@se.cuhk.edu.hk

Abstract. In social networks, current friend/user recommendation methods are mainly based on similarity measurements among users or the structure of social networks. In this paper, we design a novel friend recommendation method according to a new individual feature *intimacy degree*. Intimacy degree reflects the degree of interaction between two users and further indicates how close two users pay attention to each other. Specifically, we first formally define this problem and perform a theoretical investigation of the problem based on random walk with restart model. And then we design an individual friend recommendation algorithm based on the social structures and behaviors of users. At last, we conduct experiments to verify the method on a real social data set. Experimental results show that the performance of friend recommendation outperforms the existing methods, and the proposed algorithm is effective and efficient in terms of PV Value, UV Value and Conversion Rate.

Keywords: Friend recommendation · Intimacy degree · Random walk model · Social network analysis

1 Introduction

Recommendation is an effective way to reduce the cost for finding information and also a powerful way to attract customers. The flourish of the dynamic social networks provides a new environment for validating the recommendation methods, at the same time brings new challenges, e.g., how to recommend friends according to interaction information?

This paper systematically investigates the friend recommendation problem, and proposes a novel method according to a new individual feature intimacy degree. *Intimacy degree* is defined as the total number of reviewing, forwarding, making comments, clicking a like, replying and making @relationship which two friends give each other in social networks. As we know, interaction relation is a kind of equivalent relation and thus interaction degree can be regarded as one new feature for friend recommendation among users in social networks [1]. According to this above idea (as illustrated in Fig. 1), we propose a novel method based on the social structures and behaviors of users for friend recommendation. Specifically, we first extract friend relationships data and their interacting activities data from social networks. And then we compute the intimacy degree between target users and candidate friends using random walk algorithm (RW) based on a bipartite graph. Finally, we build an individual friends recommendation model in order to provide intelligent recommending service in social networks.

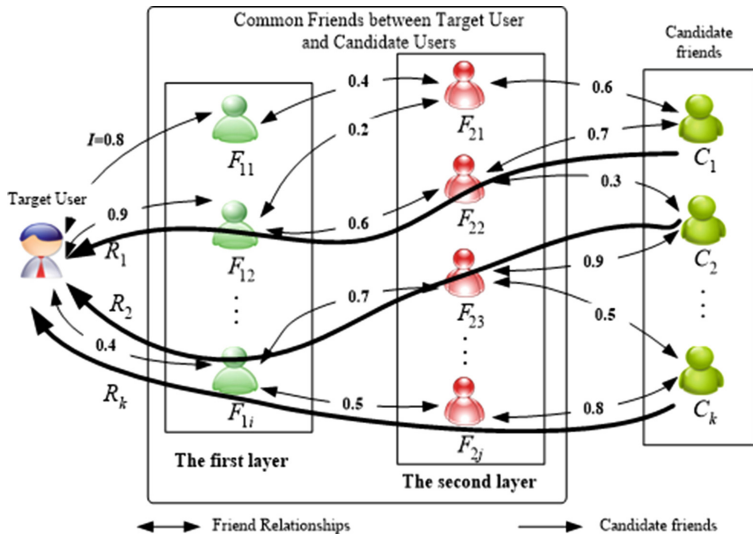


Fig. 1. The architecture of our method

In Fig. 1, I means the set of intimacy degree values, F_{1_i} denotes the i th user in the first layer of friends of Target User, C_k means the k th candidate user being recommended to Target User, F_{2_j} states the j th user in the second layer of friends of Target User, and R_k is the Ranking value of the k th candidate friend.

Individual friend recommendation is an important task in various social-network-based applications such as searching product information and rating, recommending advertisement and services, as well as public sentiment surveillance. But, some challenges still exist:

- Due to the network heterogeneity, it is difficult to find an appropriate method to model social networks in order to make friend recommendation.
- Due to the dynamics of users behaviors and the importance of users interaction, it is challenging to design an effective model to utilize the two properties for friend recommendation.
- It is non-trivial to find a dynamic united model considering the above factors to make individual and effective friend recommendation. In addition, it is very hard to obtain the large volume of real data to validate the proposed recommending methods.

The main contributions of this paper are as follows:

- Aiming at the shortcomings of the existing friend recommendation model of social networks, this paper presents an individual recommendation model based on the social structures and behaviors of users.
- This paper designs a random walk recommendation algorithm, gives algorithm scheme and conducts the implementation of the algorithm.
- We conduct experiments to verify the method on a real social data set. Experimental results show that the performance of friend recommendation outperforms the existing methods, and the proposed algorithm is effective in terms of PV Value, UV Value and Conversion Rate. All data and codes are publicly available.

The rest of our paper is organized as follows. Section 2 describes related work. Section 3 gives the proposed individual friend recommendation algorithm based on random walk (RW). Section 4 describes experimental details and validations of our results, and Sect. 5 offers concluding remarks.

2 Related Work

Most recommendation systems are based on Bipartite Graph, and model connections between two user sets. A bipartite graph (or bigraph) is a graph whose vertices can be divided into two disjoint sets U and V (that is, U and V are each independent sets) such that every edge connects a vertex in U to one in V . Generally, there are edges between two parts and there is not edge among the nodes in the same part. The recommendation system based on bipartite graph hides the characteristics of information between users, and makes the relationship between users abstracted into a mapping recommendation relationship between M users and N users. If user i chooses or has browsed user j , then an edge will exist in the bipartite graph, denoted as $e_{ij} = 1$, otherwise $e_{ij} = 0$ and in this way it will constitute an $M + N$ bipartite graph. The target of such algorithm is to establish a connection between i and the user which has not been selected, recommend Top N users to user i according to the users preferences, use w_{ij} representing user i 's preferences to user j [2].

Content-based recommendation technology is to analyze the target user information that is selected or browsed by the user [3]. The recommendation process

of collaborative filtering system is divided into two steps: the first step is using the existing users history information to calculate the similarity between users [4, 5], sorting the neighbor users with similarity degree, and selecting Top N users with a high degree of similarity [6, 7]. The second step is computing the score of Top N neighbor users to target user, the user with the higher score will be recommended to target user [8]. Recommendation system based on the structure has also been applied in heterogeneous social networks [9, 10]. It can be extended to the individual setting by combining the user behavior logs and user preference information. Additional information such as tags can eventually be integrated into the scope list to improve the accuracy of individual recommendation [11].

3 Individual Friend Recommendation Algorithm Based on RW

Compared to existing friend/user recommendation methods, our individual friend recommendation method is based on the social structures and behaviors of users. Thus, it no longer depends on the users profile information, and instead uses both historical friend and interaction relationships of users. This strategy can help the method effectively avoid the recommendation problems caused by missing users information. Theoretically, our solution should have better recommendation results than existing friend recommendation methods, and better help social network users to find more potential friends.

3.1 Problem Definition

A simple network relationship can be commonly defined as a relational graph, which consists of a set of users (called points), as well as correlation which is called edges between users (such as friend relationship, fans relationship and so on). It can be expressed as $G = \langle V, E \rangle$ based on graph theory, where V represents the set of all points, E represents the set of all edges on the relational graph. The problem discussed by this paper is: given a point a in V , how to find the most relevant point of a in V . In order to solve this problem, our solution is that we compute and obtain the set of neighbor points which have the largest similarities with a . According to the above way, we calculate all points in set V .

A recommendation relationship can be defined as a bipartite graph relationship formed by users. The bipartite graph network is commonly defined as $G = \langle U \cup I, E \rangle$, It is shown in Fig. 2.

In Fig. 2, U denotes a set of users, I means another set of users, E states a relationship set between U and I . In our proposed method, U is the set of target users that will receive candidate friends list with ranking values, I is the set of candidate friends that will be recommended to one target user, E means relationships among users, such as friend relationships, fans relationships, comment relationships, @ relationships, and so on.

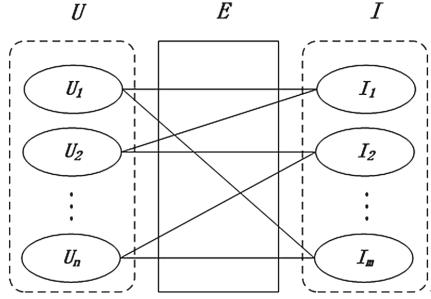


Fig. 2. Bipartite graph of User to User (U2U) degree

3.2 The Method

Random Walk Algorithm.

Aiming at the characters of social networks, this paper is based on the Random Walk algorithm, and introduces an individual friend recommendation method using interaction information to rate the friends of target user. Random Walk (RW) model uses first order Markov Model and thus the status of users probability at the time of $t + 1$ only can be influenced by the time of t on the graph structure, but not by other time. It is expressed by formula (1).

$$\Pr(X_{u,t+1} = i | X_{u,t}) \tag{1}$$

where i represents the set of candidate friends in the system, and m represents the size of the set of candidate friends. So a directed graph $G = \langle V, E \rangle$ which has m points can be established. Weight $P_{i,j}$ represents the probability moving from point i to point j on the graph. According to the description of formula (1), we get formula (2).

$$P_{i,j} = \Pr(X_{u,t+1} = j | X_{u,t} = i) \tag{2}$$

We can get a $m \times m$ state transition probability matrix P (also, **intimacy transition probability matrix**) according to formula (2). In formula (2), u means the set of all users in the recommending system, n states the size of the user set, $\bar{P}_{i,j}^k$ stands for the column vector of matrix P , $\bar{R}_{u,-}^*$ denotes the row vector of matrix P , as well as represents the rating value from user u to user m . According to the active information of users on the network, we can establish a $n \times m$ dimensional information matrix between target users and candidate friends. And then, we can obtain the probability transition matrix from target users to candidate friend j at the time of t .

$$\Pr(X_{u,t} = j) = \alpha \sum_{i=1}^m \Pr(X_{u,t-1} = i) P_{i,j} = \alpha^k \sum_{i=1}^m R_{u,i}^* P_{i,j}^k = \alpha^k \bar{R}_{u,-}^* \bar{P}_{-,j}^k \tag{3}$$

According to formula (3), we can deduce the total probability formula from target friends to user j , as shown in formula (4).

$$\Pr(X_u = j) = \frac{\sum_{k=1}^x \Pr(X_{u,k} = j)}{\sum_{k=1}^x \sum_{i=1}^m \Pr(X_{u,k} = i)} = c \sum_{k=1}^x \alpha^k \bar{R}_{u,-}^* \bar{P}_{-,j}^k \quad (4)$$

According to formula (4), the overall sequencing matrix of each user ratings for all users is given in formula (5).

$$\tilde{R} = \sum_{k=1}^x \alpha^k R P^k = R \alpha P (1 - \alpha P)^{-1} \quad (5)$$

where matrix \tilde{R} reflects the rating that target user to user, we can sequence the ratings that target user to all users according to this matrix, and we can receive the users Top N recommendation by using the result of sequencing.

Random Walk with Restart Algorithm Based on the Bipartite Graph.

Random Walk with Restart(RWR) algorithm calculates similarity between one point and point j according to formula (6).

$$R_t = (1 - c) \tilde{W} R_{t-1} + c e_j \quad (6)$$

where $c \in (0, 1)$ is restarting probability, \tilde{W} is adjacency matrix, e_j is an initial vector, and R_t is probability distribution vector after t times iteration.

The next key step is how to abstract the friend relation, fans relation, comment relation, retransmission relation and @ relation among users to a target user-user Bipartite Graph, establish grade mechanism among users, then introduce random walk with restart quick algorithm which is based on Bipartite Graph, and individually recommend friends to a target user.

As mentioned above, we can give the process of the **Random Walk with Restart Algorithm** based on the Bipartite Graph, as shown in Table 1.

4 Experimental Results

4.1 Experimental Setup

In this paper, we use Microblog social data from Sina Mobile Client Platform (shortly SMCP) as experimental datasets, and design three solutions to validate our methods on three groups of datasets. Every dataset includes 100,000 users. The detailed information of experimental settings is displayed in Table 2.

New Recommendation strategy: our presented recommendation strategy, New Algorithm: our proposed algorithm, Old Recommendation strategy: the existed recommendation strategy from SMCP, and Old Algorithm: the existed algorithm from SMCP.

In Table 2, the first solution consists of New Recommendation strategy and New Algorithm, the second solution includes New Recommendation strategy and Old Algorithm, and the third solution contains Old Recommendation strategy and Old algorithm.

Table 1. The description of the algorithm based on random walk with restart.

Input:	The normalized matrix \tilde{W} and the start vector r_0
Output:	Stationary distribution probability r
First:	Through \tilde{W} to construct the state transition matrix M
Second:	Segmentation matrix M is $M = M_1 + M_2$, among them M_1 contains all the internal relations, M_2 contains all the cross relations
Third:	Initializer $r_{(k)}^{(0)} = (r_1^0, r_2^0, \dots, r_n^0)$, $m = 1$
Fourth:	For $i = 1, 2, \dots, N$, calculating $\phi_i^{(m-1)} = \frac{r_i^{(m-1)}}{\ r_i^{(m-1)}\ _1}$
Fifth:	Construct the aggregation matrix $A^{(m-1)}$, among them $(A^{(m-1)})_{ij} = \phi_i^{(m-1)} M_{ij} e_j$
Sixth:	To solve the characteristic vector $\xi^{(m-1)} = \xi^{(m-1)} A^{(m-1)} + c_{1 \times N}$
Seventh:	For $i = 1, 2, \dots, N$, get W_i and ξ_i^m

Table 2. The detailed information of the experimental setting

Dataset no	Dataset size	Recommendation strategy	Algorithm
1	100000	New	New
2	100000	New	Old
3	100000	Old	Old

4.2 Evaluation Metrics

- **PV Value (Page View):** It is the number of page views or page exposure. The PV value of one webpage within a day is the total number of visiting this page this day.
- **UV Value (Unique Visitor):** It refers to page independent visiting. A page within a day of the UV value that is the day how many users access the page, a user logs in the page several times a day is only remembered once, that is, UV values of 1.
- **CR (Conversion Rate):** It is defined as the UV value divided by the PV value of the day. The conversion rate is an important indicator of the validation of a recommendation.

4.3 Recommendation Performance Analysis

At the end of the first round launching our experimental scheme, we obtained experimental data from SMCP servers and listed the PV and UV values of these three group experiments in Tables 3, 4, and 5.

In one experimental period, Conversion Rates of the three groups can be obtained and listed in Table 6. According to the result data, we can give the trend graph of Conversion Rate, as shown in Fig. 3.

Table 3. The PV and UV values of the first group experiment.

Time	Login PV	Login UV	Follow PV	Follow UV	Conversion Rate
2012/4/4	423183	31860	992	742	2.33 %
2012/4/5	385415	31050	1131	910	2.93 %
2012/4/6	394291	30708	893	643	2.09 %
2012/4/7	371477	29861	743	411	1.38 %
2012/4/8	373583	30106	528	434	1.44 %
2012/4/9	393173	32448	926	545	1.68 %
2012/4/10	396847	31651	765	608	1.92 %

Table 4. The PV and UV values of the second group experiment.

Time	Login PV	Login UV	Follow PV	Follow UV	Conversion Rate
2012/4/4	403972	30959	902	673	2.17 %
2012/4/5	360750	29859	976	749	2.51 %
2012/4/6	373079	29570	827	588	1.99 %
2012/4/7	342670	28688	568	364	1.27 %
2012/4/8	339181	29095	564	390	1.34 %
2012/4/9	354664	31212	791	477	1.53 %
2012/4/10	355508	30626	798	543	1.77 %

Table 5. The PV and UV values of the third group experiment

Time	Login PV	Login UV	Follow PV	Follow UV	Conversion Rate
2012/4/4	434946	31867	937	646	2.03 %
2012/4/5	389522	30801	932	746	2.42 %
2012/4/6	399585	30564	868	668	2.19 %
2012/4/7	373897	29440	598	408	1.39 %
2012/4/8	368926	29792	507	400	1.34 %
2012/4/9	392042	32107	769	464	1.45 %
2012/4/10	393412	31342	713	511	1.63 %

Table 6. Conversion Rates in one experimental period

Group No	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7
Group I	2.33	2.63	2.09	1.68	1.74 %	1.98	2.22
Group II	2.17	2.51	1.99	1.57	1.64 %	1.83	2.07
Group III	2.13	2.32	1.98	1.45	1.65 %	1.75	1.93

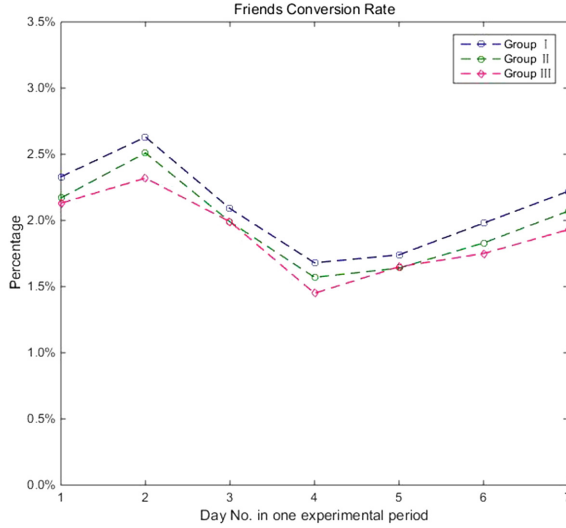


Fig. 3. The trend graph of Conversion Rates

From Fig. 3, the recommended conversion rate of the first group is greater than the second group, and the second group is greater than the third group. The results show that the recommendation effect based on the individual recommendation algorithm is better than the existing recommendation algorithm. At the same time, the experimental results also show that the user acceptance of a recommendation algorithm not only depends on the results of the algorithm itself, but also depends on a good recommendation strategy.

We utilize Sina Microblog Mobile Client Service Platform (SMCP) as the experimental environment. But, due to the limitations of the use of habits and mobile terminal platform, the proportion of users to add friends through the mobile terminal platform is relatively small.

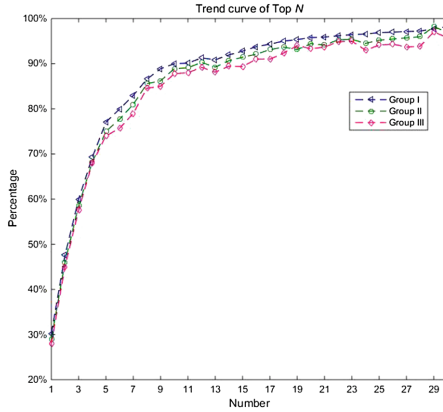
We randomly select three groups of users from the whole micro blog platform data. Although our proposed algorithm is a little better than the existing recommendation one, the advantage is not great. The reason is that the advantages of our algorithm had been scattered to the data of the entire platform when SMCP keeps on running.

4.3.1 Determining N Value of Top N in Recommendation Results

In this paper, our proposed individual recommendation algorithm adopted recommendation mode of Top N . So, how to determine the value of N became a key problem. In our experiments, we determine the range of N values by analyzing the situation of adding new friends of users from datasets during the experimental period.

Table 7. The number statistics of users having added new friends

Group No	First	Second	Third
Number	78419	77329	78159

**Fig. 4.** The trend curve of Top N

We selected users who had added new friends in the experiment period from these three groups of experimental data, and then computed the distribution of each group of the number of the users following behaviors. The number statistics of users that had added new friends or conducted following behaviors are given in Table 7.

From Table 7, the number of adding new friends from three groups in the experimental period had little difference. The reason is that the advantage of our individual recommendation algorithm is weakened by the massive data from entire SMCP. In addition, the distribution trend graph of the number of following new friends in very group of users is given in Fig. 4.

From Fig. 4, the attention/following number of three experimental groups is within 10 in terms of 86 % of users. During the experimental period, it is smaller than 30 in terms of 97 % of users. From the experimental results, it is acceptable that the Top N values were recommended to 0~30. If recommending results were updated more quickly, the value of N can be relatively smaller. Our experimental period is set as 7 days, so it is more reasonable that we took the value N as 30.

5 Conclusions

In this paper, we study a novel problem of individual friend recommendation in social networks, with the objective of finding an effective and dynamic model to utilize the social structures and behaviors of users for friend recommendation. We formally define this problem and perform a theoretical investigation

of the problem based on random walk with restart model. We design an individual friend recommendation algorithm and conduct experiments to verify the method on a real social data set. Experimental results on a real social dataset demonstrate the effectiveness and efficiency of the proposed method.

Acknowledgements. The research is supported by the National High Technology Research and Development 863 Program of China under Grant No. 2015AA124102, the Hebei Natural Science Foundation of China under Grant No. F2015203280, the Open Project Program of the State Key Laboratory of Mathematical Engineering and Advanced Computing, and the National Natural Science Foundation of China under Grant No. 61303130. © Springer-Verlag Berlin Heidelberg 2011.

References

1. Zhang, J., Tang, J., Liang, B.Y., Yang, Z., Wang, S.J., Zuo, J.J., Li, J.Z.: Recommendation over a heterogeneous social network. In: Proceedings of the 9th International Conference on Web-Age Information Management (WAIM 2008), pp. 309–316 (2008)
2. Deshpande, M., Karypis, G.: Item-based top-N recommendation algorithms. *ACM Trans. Inf. Syst.* **22**, 143–177 (2004)
3. Cai, Y., Leung, H., Li, Q., Tang, J., Li, J.Z.: Recommendation based on object typicality. In: Proceedings of the Nineteenth Conference on Information and Knowledge Management (CIKM 2010), pp. 1529–1532 (2010)
4. Xu, B., Bu, J., Chen, C., Cai, D.: An exploration of improving collaborative recommender systems via user-item subgroups. In: Proceedings of the 21st International Conference on World Wide Web, Lyon, France, pp. 21–30 (2012)
5. Cai, Y., Leung, H., Li, Q., Han, H., Tang, J., Li, J.Z.: Typicality-based collaborative filtering recommendation. *IEEE Trans. Knowl. Data Eng. (TKDE)* **26**(3), 766–779 (2014)
6. Tang, J., Wu, S., Sun, J., Su, H.: Cross-domain collaboration recommendation. In: Proceedings of the Eighteenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2012), pp. 1285–1293 (2012)
7. Hilmi, Y., Mukkai, S.K.: A random walk method for alleviating the sparsity problem in collaborative filtering. In: Proceedings of the 4th ACM Conference, pp. 131–138 (2008)
8. Zhang, M., Tang, J., Zhang, X.C., Xue, X.Y.: Addressing cold start in recommender systems: a semi-supervised co-training algorithm. In: Proceedings of the 37th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2014), pp. 73–82 (2014)
9. Chiang, M.F., Liou, J.J., Wang, J.L., Peng, W.C., Shan, M.K.: Exploring heterogeneous information networks and random walk with restart for academic search. *Knowl. Inf. Syst.* **36**, 1–24 (2013)
10. Dong, Y.X., Tang, J., Wu, S., Tian, J.L., Nitesh Chawla, Rao, J.H., Cao, H.H.: Link prediction and recommendation across heterogeneous social networks. In: Proceedings of 2012 IEEE International Conference on Data Mining (ICDM 2012), pp. 181–190 (2012)
11. Liang, H., Xu, Y., Li, Y., Nayak, R., Tao, X.: Connecting users and items with weighted tags for personalized item recommendations. *HT 2010*, pp. 51–60. *ACM* (2010)