# Impact of the Important Users on Social Recommendation System

Zehua Zhao[1,2], Min Gao[1,2(✉)], Junliang Yu[1,2], Yuqi Song[1,2],
Xinyi Wang[1,2], and Min Zhang[3]

[1] Key Laboratory of Dependable Service Computing in Cyber Physical Society, Chongqing University, Ministry of Education, Chongqing 400044, China
{zhzhao,gaomin,yu.jl,songyq,xywang}@cqu.edu.cn
[2] School of Software Engineering, Chongqing University, Chongqing 400044, China
[3] Changle Bowen School, Dezhou 262400, Shandong, China
zhangliuyiming@163.com

**Abstract.** Recommendation methods have attracted extensive attention recently because they intent to alleviate the information overload problem. Among them, the social recommendation methods have become one of the popular research fields because they are benefit to solve the cold start problem. In social recommendation systems, some users are of great significance, because they usually have decisive impacts on the recommendation results. However, it is still lack of research on how the important users make influence to recommendation methods. This paper presents three types of important users and utilizes three social frequently-used recommendation methods to analyze the influence from multiple perspectives. The experiments demonstrate that all the recommendation methods achieve better performance with important users, and the important neighbor users have the greatest impact on the recommendation methods.

**Keywords:** Recommendation systems · Social network
Important users

## 1 Introduction

Nowadays, people have entered an era of information overload [1,2]. To help users find the information they want, researchers have designed and developed the search engines and the recommender system. Unlike the search engine which has specific requirements for its input, the recommender system has become an inseparable part of people's daily life because of its automation, convenience and high efficiency.

The common recommender systems are divided into three categories: content based recommender systems, collaborative filtering [3] based recommender systems and hybrid recommender systems [4]. Existing collaborative filtering methods can be categorized into memory-based methods and model-based

© ICST Institute for Computer Sciences, Social Informatics and Telecommunications Engineering 2018
I. Romdhani et al. (Eds.): CollaborateCom 2017, LNICST 252, pp. 425–434, 2018.
https://doi.org/10.1007/978-3-030-00916-8_40

methods [5–7]. Model-based methods are very fast once the parameters of the model are learnt. The bottleneck for model-based methods is the training phase, while in memory-based methods there is no training, but the prediction (test) phase is slower [8]. So model based algorithm is more suitable for big data. In order to improve the efficiency of the experiment, the three recommendation methods used in this paper are the model-based method.

Although the recommendation methods have been greatly supplemented and improved, there are still a few serious problems. For example, when a user's rating vector is extremely sparse, it is difficult to recommend the product to the user accurately. This is "cold start problem". To alleviate this problem, researchers try to integrate social information into recommender algorithms that form social recommender algorithms such as SoRec [9] and TrustMF [3].

In social recommendation, an important step is to filter important users. The so-called "important users" means those users who play a crucial role in the recommender system and have a decisive impact on the sales of certain products. It can be seen that understanding the impact of important users on the social recommendation system will help us make better use of the important users and produce better recommendation results. But so far, there is little research in this area, so we choose comparative experiments to analyze the impact of important users on social recommender systems scientifically.

The contributions of this paper are as follows. (1) We propose three types of important users based on the reality of online social networks. (2) Compared the impact of different types of important users through different recommendation methods, and found that the important neighbor users have the greatest impact on the methods.

The rest of the paper is organized as follows: Some related work is discussed in Sect. 2. We introduce the methods of filtering important users in Sect. 3. Then, the experiments are reported in Sect. 4. Finally, we conclude the paper and present some directions for future work in Sect. 5.

## 2 Related Work

In this section, we will introduce the recommendation methods and the filtering methods for important users, especially three social recommendation methods which will be used in the experiments.

### 2.1 Recommendation Methods

Normally, the input of the social recommendation method includes a social network (as shown in Fig. 1) and a user-item rating matrix (as shown in Fig. 2). In the social network, it contains a set of nodes $u_i \in U$ which represent the users and a set of edges $s_{ik} \in (0, 1]$ which represent the trust weight from user $i$ to user $k$ (in this paper, we use the adjacency matrix of social trust graph to represents the social network, and called this adjacency matrix as user-user trust matrix). In the user-item rating matrix, it contains a set of users $u_i \in U$, a set of items

$v_j \in V$ and a set of rating records $r_{ij} \in R$ ($r_{ij}$ shows how much the user $i$ prefers the item $j$, in this paper, $r_{ij} \in [1, 6]$, where "1" represents the least preferred and "6" indicates the preference. If the user $i$ has no rated on the item $j$, then $r_{ij}$ is "?").

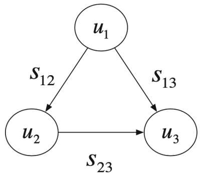



**Fig. 1.** Social trust graph

| | $v_1$ | $v_2$ | $v_3$ | $v_4$ |
|---|---|---|---|---|
| $u_1$ | $r_{11}$ | ? | $r_{13}$ | ? |
| $u_2$ | ? | ? | $r_{23}$ | ? |
| $u_3$ | ? | ? | ? | $r_{34}$ |

**Fig. 2.** User-item rating matrix

**RSTE: Recommend with Social Trust Ensemble** [10]**.** It is a probabilistic factor analysis framework, which naturally fuses the user's tastes and their trusted friends' favor together.

**SocialMF: A Matrix Factorization Based Model for Recommendation in Social Rating Networks** [8]**.** It is a model-based method which employing matrix factorization techniques. To improve previous work, they incorporate the mechanism of trust propagation into the model.

**SoReg: Recommender Systems with Social Regularization** [11]**.** SoReg coined the term Social Regularization to represent the social constraints on recommender systems. When recommending a product to the target user, it gives different weights to other users according to the similarity between the other users and the target users to distinguish the influence of different users on the target users.

### 2.2 Research Status of Important Users

As for the filtering methods of important users, new solutions have been put forward in recent years.

Shen [12] proposed a collaborative filtering method, incorporating both user-based method and item-based method. In their research, users' relationships are divided into two main categorizes: well-known people, institutions or other users of higher visibility, and their own friends and those who have low popularity.

Zhang [13] used the Bayesian network [14] model to compute user influence. Based on the location of nodes in the network and the number of edges connected by nodes, Wang [15] used the degree and betweenness to filter the important user. And Fang [16] used the degree centrality and the betweenness centrality to filter the center node and tested its effect on the recommendation system, which shows the center node plays a particularly important role in the recommendation methods.

On the basis of the above work, considering the role of users in the social network and the relationship between them, we proposed three new filtering methods of important user.

## 3 Important User Filtering

In social network, A user's following list usually contains three types of users: The first is users with high visibility in the entire social network, such as Sina V users in micro-blog[1]; the second is users with high reputation in the Interest tribe, such as music blogger; the last one is users who are less well-known but have a offline relationship with target user, such as target user's classmates and colleagues. Based on this phenomenon, we proposed three types of important users to analysis which kind of important users have the greatest impact on recommendation results. Some users may be considered as more than one types of important users when filtering important users.

The important users of these three types of users are listed as the important overall users, the important community users and the important neighbor users.

### 3.1 The Important Overall Users

One of the obvious features of the important overall users is that a lot of users follow them while they only follow few users. In social network graph, the ratio of the in degree to the out degree of these users will be high.

Based on this intuition, we proposed Algorithm 1 to filter the important overall users.

**Algorithm 1.** Filter the important overall users

**Input:** user-user trust matrix.
**Output:** the top-k important overall users.
1: **for** $u_i \in U$ **do**
2: Count the in degree $in_i$ of user $u_i$.
3: Count the out degree $out_i$ of user $u_i$.
4: Calculate the ratio $r_i$ of $in_i$ to $out_i$ of user $u_i$.
5: **end for**
6: Sort users in descending order according to $r$
7: Select the top-k users as the important overall users

In lines 1–5, the algorithm calculates the ratio of the in degree to the out degree for each user; then we sort users in descending order according to their ratio and select the top-k as the important overall users.

[1] http://www.weibo.com.

### 3.2 The Important Community Users

The important community users usually play an important role in an interest tribe, which means their preferences have a higher similarity with others.

Based on this intuition, we proposed Algorithm 2 to filter the important community users.

In lines 2–5, we calculate two users' preference similarity $sim_{i,k}$ by their user-item rating vector. The similarity is calculated by cosine similarity (Eq. (1)). After that, we set a threshold $\theta_2$ according to the specific situation of the dataset. If $sim_{i,k} > \theta_2$, we will add both of the users into the nearest neighbor [17] list $l$. Then we count the number of occurrences of each user in the list, sort users in descending order, and select the top-k users as the important community users [18].

$$sim_{i,k} = \frac{\sum_{j=1}^{m}(R_{u_i j} \times R_{u_k j})}{\sqrt{\sum_{j=1}^{m} R_{u_i j}^2} \times \sqrt{\sum_{j=1}^{m} R_{u_k j}^2} \tag{1}$$

**Algorithm 2.** Filter the important community users

**Input:** user-item rating matrix.
**Output:** the top-k important community users.

   Initialize the similarity list $l$
2: **for** $< u_i, u_k > (i \neq k, u_i \in U, u_k \in U)$ **do**
      Calculate the preference similarity $sim_{i,k}$ between $u_i$ and $u_k$ according to their rating vectors
4:    if $sim_{i,k} > \theta_2$ then add $< u_i, u_k >$ to list $l$
   **end for**
6: Count the number of occurences $o$ of each user in the list $l$
   Sort users in descending order according to $o$
8: Select the top-k users as the important community users

### 3.3 The Important Neighbor Users

The important neighbor users usually have offline relationships with the target user and have more common friends with the target user. In the social trust graph, their friends lists have a higher similarity with others.

Based on this intuition, we proposed Algorithm 3 to filter the important neighbor users for the target user.

Algorithm 3 is similar with Algorithm 2. The difference is the input. In Algorithm 2, the input is the user-item rating matrix while in Algorithm 3 it is the user-user trust matrix.

**Algorithm 3.** Filter the important neighbor users

**Input:** user-user trust matrix.
**Output:** the top-k important neighbor users.
    Initialize the similarity list $l$
    **for** $< u_i, u_k > (i \neq k, u_i \in U, u_k \in U)$ **do**
        Calculate the friends lists similarity $sim_{i,k}$ between $u_i$ and $u_k$ according to their trust vectors
        if $sim_{i,k} > \theta_3$ then add $< u_i, u_k >$ to list $l$
    **end for**
    Count the number of occurences $o$ of each user in the list $l$
    Sort users in descending order according to $o$
    Select the top-k users as the important community users

## 4 Experiments

### 4.1 Dataset

The data used in our experiments is Ciao [19]. Ciao is a product review website, with ratings ranging from 1 to 6. This dataset is extremely sparse and imbalanced containing 3951 users and 60552 items. The total number of trust relations between users is 40133, the number of ratings is 327120, and the average rating is 4.777. The densities of relations and ratings are 0.257% and 0.137%, respectively.

### 4.2 Experimental Setup

The recommendation methods used in this paper are RSTE, SocialMF, and SoReg. We used a 10-fold cross-validation for learning and testing. In each time we randomly selected 90% of data as training set and the rest 10% as test set.

We set $\alpha = 0.1$ in RSTE method, the parameter $\alpha$ controls how much do users trust themselves or their trusted friends; In Algorithm 2, $\theta_2 = 0.5$, it is a threshold, if the preference similarity $sim_{i,k}$ is greater than $\theta_2$, it means user $u_i$ and user $u_k$ have higher preference similarity and they are important community user for each other; Similar to $\theta_2$, $\theta_3$ is the threshold in Algorithm 3, and $\theta_3 = 0.5$, if the friends lists similarity $sim_{i,k}$ is greater than $\theta_3$, it means user $u_i$ and user $u_k$ are important neighbor user for each other.

First we filter the top-0, top-200, top-400, top-600, top-800 and top-1000 important users of the overall users, the community users and the neighbor users, respectively. Thus we get 18 different sets of important users. After that, one important users set is deleted from the original dataset each time, so we obtain 18 new datasets. Finally, we input these new datasets into the three algorithms and reevaluate the recommendation results.

### 4.3 Metrics

The evaluation metrics used in our experiments are mean absolute error (MAE) and root mean square error (RMSE) [20], which are defined as Eq. (2) and Eq. (3).

$$MAE = \frac{\sum_{t=1}^{N} |(predicted_t - observed_t)|}{N} \tag{2}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{t=1}^{N}(predicted_t - observed_t)^2} \tag{3}$$

In addition, we define a metric the rate of change ($\Delta$) to evaluate the impact of important users on the MAE and RMSE of different methods. As shown in Eq. (4).

$$\Delta = \frac{delete_{1000} - delete_0}{delete_0} \tag{4}$$

$delete_{1000}$ is the MAE/RMSE value after deleting the top-1000 important users, and $delete_0$ is the MAE/RMSE value without deleting important users.

### 4.4 Experimental Results

The experimental results are shown in Fig. 3 and Table 1. Figure 3 shows the MAE and RMSE of three kinds of recommendation methods without important overall users, important community users and important neighbor users, respectively. Table 1 shows the rate of change of MAE and RMSE.

**Table 1.** The rate of change of MAE and RMSE of different recommendation methods for different kinds of important user

| | | #overall user | #community user | #neighbor user |
|---|---|---|---|---|
| RSTE | MAE | 0.1295196 | 0.1258491 | 0.4498069 |
| | RMSE | 0.1238030 | 0.1146042 | 0.3450586 |
| SocialMF | MAE | 0.1444981 | 0.1057310 | 0.3822633 |
| | RMSE | 0.1000046 | 0.0834284 | 0.3178401 |
| SoReg | MAE | 0.1290263 | 0.1116101 | 0.2788628 |
| | RMSE | 0.1107464 | 0.0866980 | 0.2262674 |

We will analyze the results from two perspectives.

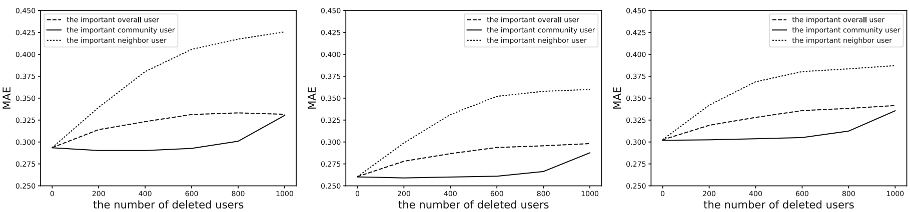
**From Metrics Perspective**

- MAE
  As can be seen from Eq. (2), MAE is sensitive to the accumulation of small errors. In Fig. 3, with the reduction of important users, MAE shows an upward trend, indicating that the reduction of important users increases the number of minor errors.

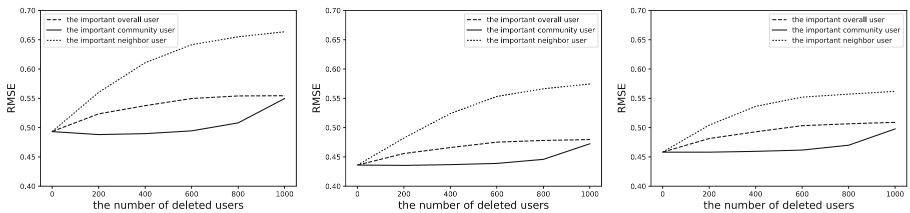
- RMSE
  It can be seen from Eq. (3) that RMSE has higher sensitivity to large errors. In Fig. 3, with the reduction of important users, RMSE is also showing an upward trend, indicating that the reduction of important users will increase the error of hot items.
- The rate of change
  All those recommendation methods will be affected by the deletion of the important users, among which RSTE usually has the largest rate of change as can be seen in Table 1.
  Among the three kinds of important users, the important neighbor users have the greatest impact on the rate of change, and the largest rate of change of MAE and RMSE are 0.4498 and 0.3451, respectively.

**From Important User Perspective.** Table 1 shows that the important neighbor users have the greatest impact on the methods compared to the important overall users and the important community users, which can be seen visually from Fig. 3.

In addition, in Fig. 3, as more and more important overall users and important neighbor users are deleted, RMSE and MAE are increasing gradually: the top600 users have obvious influence on the recommendation result, and then the influence of the important users is waning. This is in contrast to the important community users: the top600 users almost have no effect on the recommendation result or even to make a small decline in MAE and RMSE, and after that the effect of important users increased gradually. That may due to the noise in the data: there are some noise in the top600 important community users, causing




(a) MAE of RSTE (b) MAE of SocialMF (c) MAE of SoReg




(d) RMSE of RSTE (e) RMSE of SocialMF (f) RMSE of SoReg

**Fig. 3.** Analysis results for rating data in Ciao dataset

the MAE and RMSE to stay almost unchanged before top 600 important users are deleted.

As can be seen in Fig. 3, the effect of each type of important users on different recommendation methods is almost the same, which demonstrates that the important user filtering methods have good universality in different recommendation methods.

## 5 Conclusions and Future Work

This paper focuses on the impact of important users on recommendation results in social recommender systems. First, we classify users who are in the target user's following lists, and propose three types of users that are usually followed by others: the overall users, the community users and the neighbor users. Then, based on the characteristics of these three types of users, we design the filtering methods of important users. Finally, we analyze the impact of different types of important users on three social recommendation algorithms. The experimental results show that both the RMSE and MAE of the social recommendation methods show a downward trend after adding important users, which indicates that the recommended results are more accurate with the addition of important users. In addition, the important neighbor users have the greatest impact on recommender systems among the three important users.

In the future work, we will focus on the impact of the important users on the long tail items, to help businesses promote items better and users to find the more attractive items in the long tail to improve the diversity of recommendation results.

**Acknowledgements.** The work is supported by the Basic and Advanced Research Projects in Chongqing under Grant No. cstc2015jcyjA40049, and the Guangxi Science and Technology Major Project under Grant No. GKAA17129002.

## References

1. Levitin, D.J.: The organized mind: thinking straight in the age of information overload. Dutton (2014)
2. Lee, A.R., Son, S.M., Kim, K.K.: Information and communication technology overload and social networking service fatigue: a stress perspective. Comput. Hum. Behav. **55**, 51–61 (2016)
3. Yang, B., Lei, Y., Liu, J., Li, W.: Social collaborative filtering by trust. IEEE Trans. Pattern Anal. Mach. Intell. **39**(8), 1633–1647 (2017)
4. Tang, J., Xia, H., Liu, H.: Social recommendation: a review. Soc. Netw. Anal. Min. **3**(4), 1113–1133 (2013)
5. Goldberg, D., et al.: Using collaborative filtering to weave an information tapestry. Comm. ACM **35**(12), 61–70 (1992)
6. Breese, J.S., Heckerman, D., Kadie C.: Empirical analysis of predictive algorithms for collaborative filtering. In: Fourteenth Conference on Uncertainty in Artificial Intelligence, pp. 43–52 (1998)

7. Su, X., Khoshgoftaar, T.M.: A survey of collaborative filtering techniques. Adv. Artif. Intell. **2009**(12), 4 (2009)
8. Jamali, M., Ester, M.: A matrix factorization technique with trust propagation for recommendation in social networks. In: ACM Conference on Recommender Systems, pp. 135–142 (2010)
9. Ma, H., Yang, H., Lyu, M.R., King, I.: SoRec: social recommendation using probabilistic matrix factorization. In: ACM Conference on Information and Knowledge Management, pp. 931–940 (2008)
10. Ma, H., King, I., Lyu, M.R.: Learning to recommend with social trust ensemble. In: International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 203–210 (2009)
11. Ma, H., Zhou, D., Liu, C., Lyu, M.R., King, I.: Recommender systems with social regularization. In: Forth International Conference on Web Search & Web Data Mining, pp. 287–296 (2011)
12. Shen, C.W.: Research and analysis on the influence of micro blogging users based on micro blogging data. Ph.D. thesis, Beijing University of Posts and Telecommunications (2013)
13. Song, C., Zhang, X.K., Jia, J., et al.: Ranking importance of user node in microblog social network. Comput. Eng. Des. **37**(8), 2050–2056 (2016)
14. Chickering, D.M.: Learning equivalence classes of Bayesian network structures. In: Twelfth International Conference on Uncertainty in Artificial Intelligence, pp. 150–157 (1996)
15. Wang, H.X., Shan, Z., Liu, H.Y.: An importance analytical approach for online social network. J. Shanghai Jiaotong Univ. **47**(7), 1055–1059 (2013)
16. Wen, J.H., Fang, Q.Q., Liu, L., et al.: Impact of social relationship on model-based social recommender system. J. Front. Comput. Sci. Technol. (2017)
17. Weinberger, K.Q., Saul, L.K.: Distance metric learning for large margin nearest neighbor classification. J. Mach. Learn. Res. **10**(1), 207–244 (2009)
18. Cao, G., Kuang, L.: Identifying core users based on trust relationships and interest similarity in recommender system. In: IEEE International Conference on Web Services, pp. 284–291 (2016)
19. Lo, Y.Y., Liao, W., Chang, C.S.: Temporal matrix factorization for tracking concept drift in individual user preferences. Comput. Sci. (2015)
20. Chai, T., Draxler, R.R.: Root mean square error (RMSE) or mean absolute error (MAE)? Geosci. Model. Dev. **7**(1), 1525–1534 (2014)