



Information Filtering on Coupled Social Networks

Da-Cheng Nie¹, Zi-Ke Zhang^{2,3,4*}, Jun-Lin Zhou¹, Yan Fu¹, Kui Zhang²

1 Web Sciences Center, School of Computer Science & Engineering, University of Electronic Science and Technology of China, Chengdu, People's Republic of China, **2** College of Communication Engineering, Chongqing University, Chongqing, People's Republic of China, **3** Alibaba Research Center for Complexity Sciences, Hangzhou Normal University, Hangzhou, People's Republic of China, **4** Alibaba Research Institute, Hangzhou, People's Republic of China

Abstract

In this paper, based on the coupled social networks (CSN), we propose a hybrid algorithm to nonlinearly integrate both social and behavior information of online users. Filtering algorithm, based on the coupled social networks, considers the effects of both social similarity and personalized preference. Experimental results based on two real datasets, *Epinions* and *Friendfeed*, show that the hybrid pattern can not only provide more accurate recommendations, but also enlarge the recommendation coverage while adopting global metric. Further empirical analyses demonstrate that the mutual reinforcement and rich-club phenomenon can also be found in coupled social networks where the identical individuals occupy the core position of the online system. This work may shed some light on the in-depth understanding of the structure and function of coupled social networks.

Citation: Nie D-C, Zhang Z-K, Zhou J-L, Fu Y, Zhang K (2014) Information Filtering on Coupled Social Networks. PLoS ONE 9(7): e101675. doi:10.1371/journal.pone.0101675

Editor: Peter Csermely, Semmelweis University, Hungary

Received: March 29, 2014; **Accepted:** June 10, 2014; **Published:** July 8, 2014

Copyright: © 2014 Nie et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability: The authors confirm that all data underlying the findings are fully available without restriction. All relevant data are within the paper and its Supporting Information files.

Funding: NSFC:11105024, 61103109, 1147015, 11301490 and 11305043; Zhejiang Talents Project: QJC1302001, Zhejiang Provincial Natural Science Foundation of China (No. LY14A050001), EU FP7 Project: 611272 (GROWTHCOM). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* Email: zhangzike@gmail.com

Introduction

In the past two decades, the rapid development of Internet has offered unlimited sources for us to search and find out what we need [1]. For instance, we now can enjoy plenty of TV channels as well as countless programs, while only few choices are available twenty years ago. Moreover, the *Internet* not only offers various games, but also becomes a versatile tool to change the lifestyle that we have kept constantly over centuries. For example, online shopping has become more and more popular due to the exponential growth of e-commerce services (e.g. *Amazon.com*, *Ebay.com*, *Taobao.com*, etc), which allow us to choose, compare and purchase goods with single clicks. In addition, there is a vast class of novel job opportunities arising with the emergence of web related applications, such as *SOHO* workers (working at home but communicating via Internet). However, everything has two sides. Although Internet has changed the world a lot and greatly improved our daily life through effectively and efficiently contacting with others, it also brings many side effects and some of which are becoming critically important and even disruptive to our day-to-day routines. One of the most significant dilemmas is the well-known problem of *Information Overload*. Let's take the aforementioned TV programs as an example. In despite of the fact that we indeed enjoy more choices than ever before, it is simultaneously surprising to see that it is even more difficult to find a proper program that satisfies us. That is to say, we are facing too many choices to be able to compare them and make the appropriate decisions.

Recently, researchers from various disciplines, including computer science, social science, physics, etc., have devoted much

effort to helping users avoid being drowned into the *Information Ocean* [2]. Among numerous applications, the most successful one is the *Search Engine* (SE) [3], whose emergence can be regarded as a milestone. It can help users locate targets by filtering irrelevant objects with designed keywords, hence has soon been widely applied on the Internet. Despite its great success in information filtering, the SE technology also has some apparent drawbacks which interferes its further application in modern human society. On one hand, SE does not consider the personalization of each user, and return exactly the same results for every query with same keywords, regardless of whatever they have searched before [4]. On the other hand, we need to know priori profiles of targets which, however, normally are not very clear for us when the searching is being performed. In addition, sometimes, it is difficult for users to explicitly describe and express their potential intentions in simple words or sentences. So it further increases the difficulty in predicting their underlying preferences. Moreover, SE can only when users proactive submit their queries [5], thus, it lacks the power of actively providing results based on users' searching histories and personalized preferences.

As a consequence, *Recommender Systems* (RS), focusing on mining users' potential options, is considered as a promising candidate to address the excessive sources problem in the information era [6,7,8,9,10]. RS has achieved a great success in the past few years because it can significantly help users find relevant and interesting items. A recommender system is able to automatically provide personalized recommendations based on the historical records of users' activities. These activities are usually represented by the connections in a user-object bipartite graph [11,12]. The majority of relevant works in this area can be generally classified into six

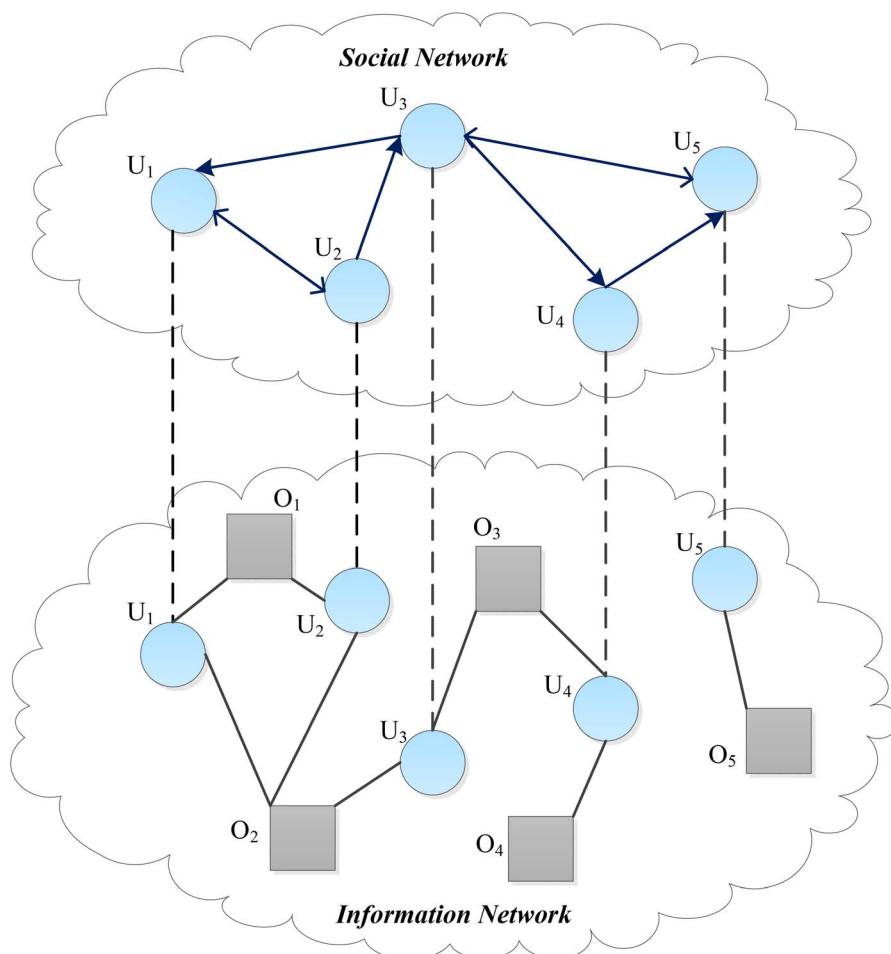


Figure 1. Illustration of a coupled social network with five users and five items, where circles denote users and squares represent objects. (upper layer) social network consists of five users; (lower layer) the information network consists of five objects and five users, while user nodes are the same in the social network.
doi:10.1371/journal.pone.0101675.g001

representative fields: i) Collaborative Filtering (CF) [13,14]; ii) Content Based Algorithms (CB) [15,16]; iii) Probability Based Models [17,18]; iv) Dimension Reduced Approaches [19]; v) Network Based Inference (NB); [12,20]; vi) Hybrid Algorithms [21,22]. CF tends to recommend to users with objects that people with similar tastes and preferences favored in the past. There are two categories respectively considering user-based [23] and object-based [14,24] factors, which should be alternatively applied in different online systems according to their own properties. For instance, *Amazon.com* is a well-known book service provider in

which the number of books is more stable than the rapid growth of readers, and thus object-based algorithms could achieve more reliable recommendation results [24]. Comparatively, *Delicious.us*(<http://www.delicious.com/>) is a typical user-driven social bookmarking platform [25], hence user-based algorithm is more suitable and effective [26]. Content based methods mainly use text mining techniques to automatically extract out meaningful content and then provide recommendations. Both probability and dimension reduced approaches require much more computational time to obtain the latent variables or vectors [27]. By contrast,

Table 1. Basic properties of the two datasets. $|U|$, $|I|$, N_R and N_S respectively represent the number of users, items, ratings and social activities. $S_r = \frac{R}{|U| \times |I|}$ and $S_p = \frac{S}{|U| \times (|U| - 1)}$ denotes the data sparsity of information and social networks respectively.

Data sets	$ U $	$ I $	N_R	N_S	S_r	S_p
<i>Epinions</i>	4,066	7,649	154,122	217,071	5.0×10^{-3}	1.3×10^{-2}
<i>FriendFeed</i>	4,188	5,700	96,942	386,804	4.1×10^{-3}	2.2×10^{-2}

doi:10.1371/journal.pone.0101675.t001

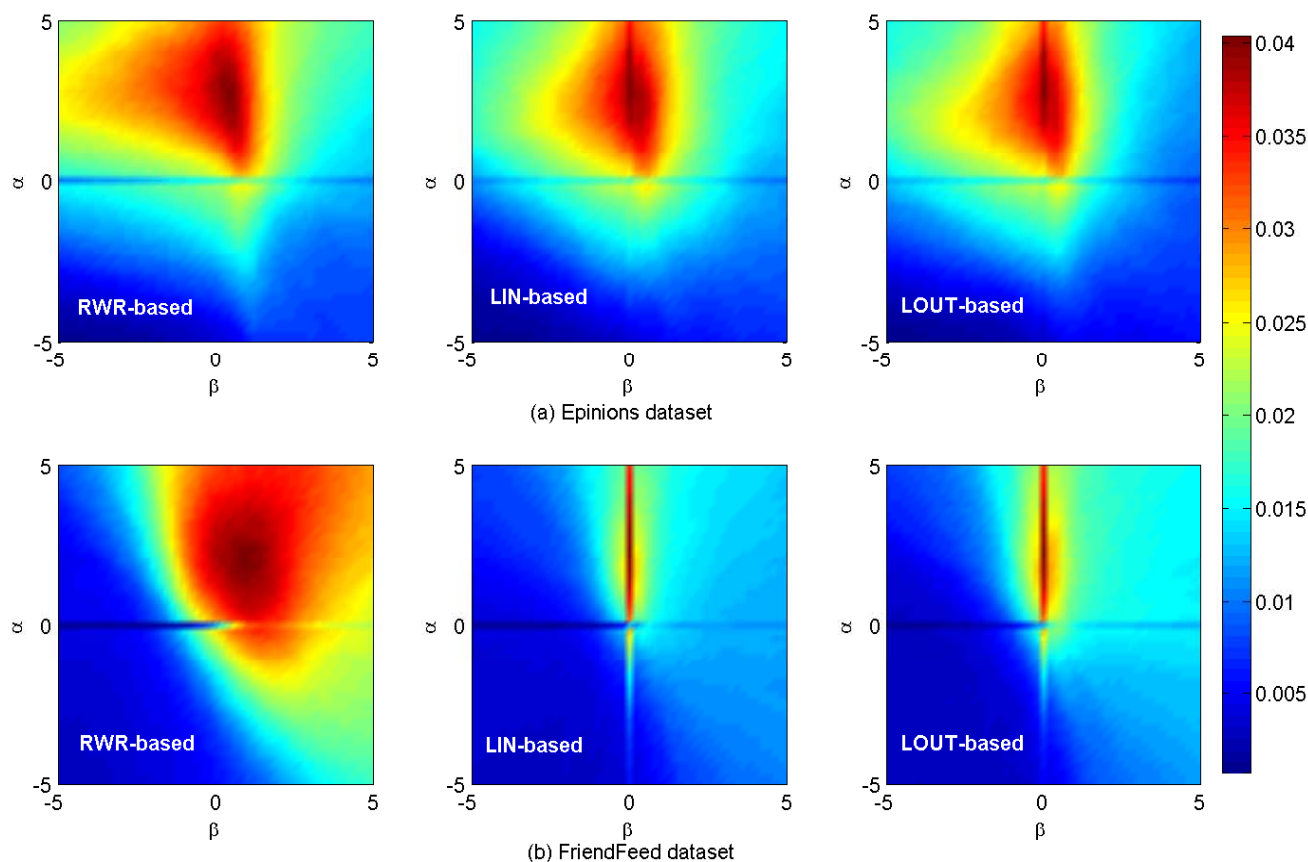


Figure 2. Precision results on Epinions and FriendFeed data sets. The length of recommendation list L is set as 10.
doi:10.1371/journal.pone.0101675.g002

network based models, making use of physical dynamics (e.g. random walk [28,29,30], heat conduction [20,31,32]), try to apply node diffusion process [33] to measure the likelihood of given pair of users and objects to be connected. Such methods would be adjusted to consider the effects of those small-degree (saying *cold*) objects [34,35] and are especially efficient for recommendation on sparse data sets [36]. Hybrid algorithms do not intend to design new methods but to introduce one or more tunable parameters to integrate different models [37,22].

Recently, *Social Networks* (SN) [38] have become a powerful tool to characterize various online social services emerging with various Web 2.0 applications [39] in evolutionary games [40,41], community detection [42] and medical science [43], etc. A great many websites have attracted millions of users active online daily. For example, *Twitter* has more than 1.7×10^8 users all over the world. *Facebook* has reported to have more than 900 million users registered within two years. *Sina Weibo*, the largest microblogging service provider in China, has been involved by almost 10% of the national population. Therefore, SN provides rich and meaningful social relations to weigh social similarities among users. Therefore, it is expected to be a very useful ingredient to generate more accurate, instructive and explainable recommendation results [44].

Coupled networks (CN), also known as interdependent networks [45], contain a joint two-layer network, such as electricity and Internet networks [46], airport and railway networks [47]. There is a kind of coupled nodes, such as cities in the two aforementioned networks, which play the roles of interconnection and maintenance between these two-layer networks [45,48]. Consequently,

those nodes are critically important for the robustness of whole networks [49]. Coupled social networks (CSN), similar with the interdependent networks, also contain such coupling nodes (saying users), which both make friends in the layer of social networks and collect favorites in the layer of information networks. Therefore, those users are especially vital to maintain the structure, connectivity and robustness of social and information networks. Fig. 1 shows an illustration of a simple CSN with five users and five objects. It can be seen that the value of similarity between user U_4 and user U_5 is zero since they do not collect the same object in the information network. So in the traditional complex network theory [50], the relationship between U_4 and U_5 might be considered as irrelevant. However, in fact U_4 and U_5 are friends and may have frequent contacts in the social network and they might have many common interests, such as making acquaintance with congenial friends and performing other mutual social activities. Therefore, a comprehensive consideration for the similarity for those two nodes should help improve the consequent recommendation performance. Based on users' distance from a fixed propagation horizon, Massa and Avesani [51] proposed a social propagation method which increased the recommendation coverage while preserving the quality of closeness. Some prior studies also brought social trust and distrust relations to the research of recommender systems [52,53]. For instance, Knapskog [54], the propagation approach was used to combine pairs of trust and distrust. Bhuiyan [55], the author discussed the definition of trust, and their results demonstrated the positive relationship between trust and interest similarity in online social networks. Crandall [56] proposed a

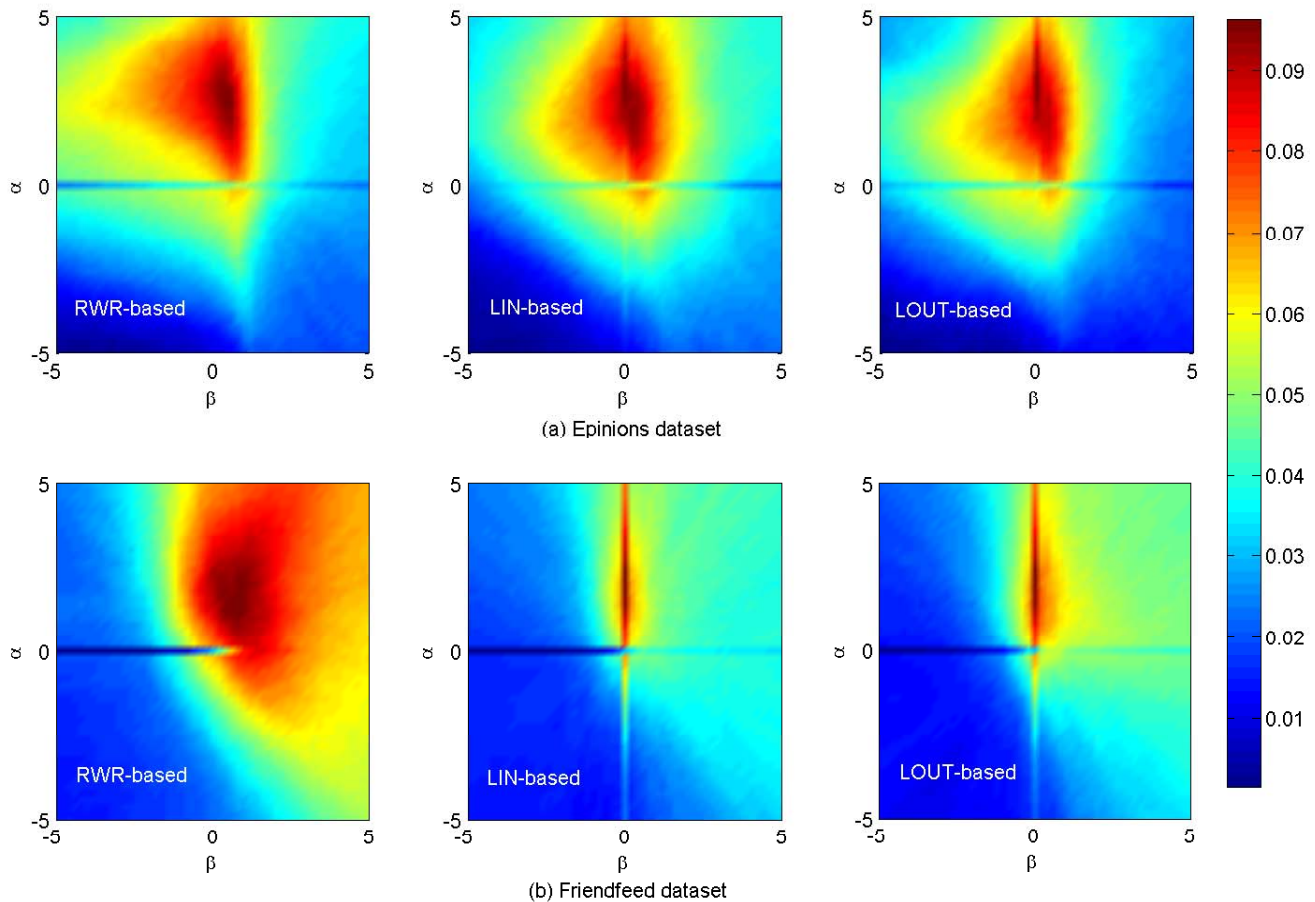


Figure 3. Recall results on *Epinions* and *FriendFeed* data sets. The length of recommendation list L is set as 10.
doi:10.1371/journal.pone.0101675.g003

feedback effect between similarity and social influence in online communities. Based collaborative filtering, Esslimani *et al.* [57] proposed a new information network and exploited navigational patterns and transitive links to model users, analyzed behavior similarities, and eventually explored missing links. As we can see, many relationships can constitute a social network such as trust, friendship, community, organizational structure, etc. And some relations are directed, like trust and follower-followee, while others are undirected such as friendship. By utilizing those social relations, we can obtain the strength of social relationship between users, and we can use this weighted social relationship to generate more accurate, explainable and acceptable recommendations through user behavioral information or profiles are unavailable.

The authors [58,59,60] have already demonstrated that recommendation performance can be improved by taking into consideration the effect of users' social network. However, how much the effect of social network will take when the social similarity and preference collaboratively work together on recommendation is still unclear. Massa *et al.* [58], the authors claimed that their purpose is to evaluate the possible contributions of trust-awareness to recommender systems and not to propose a combination technique that would require a dedicated evaluation. Walter *et al.* [59], the authors presented a model of a trust-based recommendation system on a social network. In their model, agents use their social networks to obtain information and their trust relationships to filter those useless information. However,

how to combine the social similarity and preference is still unknown. Zeng *et al.* [60], the authors designed a social diffusion recommendation algorithm that improves the performance of recommendations. Moreover, they proposed a linear combination of their method and the hybrid method [22]. In this paper, we quantitatively investigate the relationship between social similarity and personal preference for each pair of users through empirical analysis and use a nonlinear method to adjust the effects of them. Therefore, we proposed an algorithm based on CSN by considering the similarities both from social and information networks, and provided recommendations in the classical CF framework. Numerical experiments on two benchmark data sets, *Epinions* and *FriendFeed*, demonstrate that our method can offer more accurate recommendations than previous methods. In addition, extensive analyses show that the RWR-based social similarity can not only enhance the connections between small-degree and large-degree user pairs, but also reveal the large-distance user pairs which cannot be revealed by other direct metrics. As a consequence, a wider range of similar users, which cannot be discovered solely from information network, can be made use of to generate more reliable and more precise recommendations.

Methods

In this section, we start by introducing the approaches to respectively evaluating the social similarity and personalized preference between two users. Then, we integrate them to

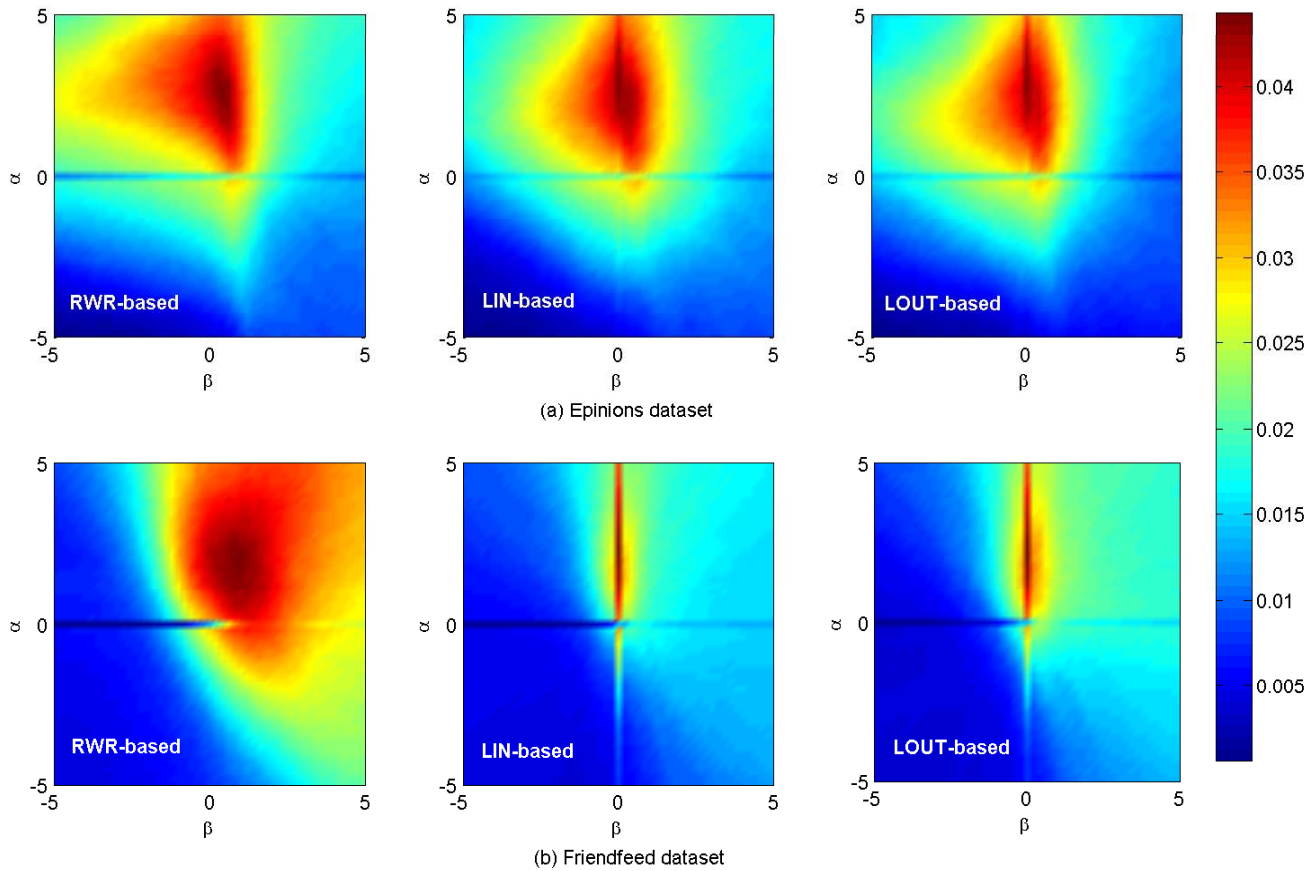


Figure 4. F-measure results on Epinions and FriendFeed data sets. The length of recommendation list L is set as 10. doi:10.1371/journal.pone.0101675.g004

measure the final similarity of each pair of users, and apply them in recommender systems. Generally, a recommender system consists of two sets, respectively of users $U = \{U_1, U_2, \dots, U_n\}$, and items $I = \{I_1, I_2, \dots, I_m\}$. Denote $R_{m \times n}$ as the adjacent matrix of the user-item bipartite network, of which each element $R_{ij} = 1$ if user U_i has collected item I_j , and $R_{ij} = 0$ otherwise. Analogously, $T_{m \times m}$ is an asymmetric matrix, denoting the directed social network, where $T_{ij} = 1$ if the user U_i has linked to user U_j , and $T_{ij} = 0$ otherwise.

1.1 Social Similarity

Firstly, we use the Random Walk with Restart (RWR) [61,62,63] method to evaluate the social similarity of directed networks. Consider a random walker starting at node i . At each step, it can move to i 's nearest neighbors via directed links with probability $c \in [0,1]$ or returns to node i with probability $1 - c$. And the final probability of each node at the stationary state will be considered as their respective peer-to-peer influence with node i . Denote A as the transition matrix of the directed network, where $A_{ij} = 1/k_i$ (k_i is the out-degree of node i if node i and j are linked). So, the final probability of i 's influence on others can be defined in a vector manner, s_i^{RWR} , as

$$\overrightarrow{s_i^{RWR}} = (1 - c)(1 - cA)^{-1} \overrightarrow{e_i}, \tag{1}$$

where $\overrightarrow{e_i}$ is a unit vector with dimension $m \times 1$, and m is the

number of users. Besides the RWR metrics, we also employ two typical local methods: *LIN* and *LOUT* to evaluate the social similarity, and use the adjusted *Jaccard* method, namely *Tanimoto* coefficient [64,65], to compute the social similarity between two users. They are defined as:

LIN:

$$s_{ij}^{LIN} = \frac{\sum_{k=1}^m T_{ki} T_{kj}}{\sqrt{\sum_{k=1}^m T_{ki}^2} + \sqrt{\sum_{k=1}^m T_{kj}^2} - \sum_{k=1}^m T_{ki} T_{kj}}, \tag{2}$$

LOUT:

$$s_{ij}^{LOUT} = \frac{\sum_{k=1}^m T_{ik} T_{jk}}{\sqrt{\sum_{k=1}^m T_{ik}^2} + \sqrt{\sum_{k=1}^m T_{jk}^2} - \sum_{k=1}^m T_{ik} T_{jk}}, \tag{3}$$

Then these metrics (Eq. (1)–Eq. (3)) will be used to quantify how much one user influences others. It can be seen that both s_{ij}^{LIN} and s_{ij}^{LOUT} only consider the local information. That is to say, only the common linked nodes of users i and j are taken into account.

Comparatively, $\overrightarrow{s_i^{RWR}}$, from the perspective of dynamic influence flow, considers both the local and global structure of directed networks. Therefore, it is expected to be a promising index to

Table 2. Performance of the recommendation algorithms on four metrics: precision (P), recall (R), f-measure (F), and AUC in *Epinions* and *FriendFeed* data sets, respectively. L is the length of recommendation list.

Method	Metrics	Epinions		FriendFeed	
		$L=10$	$L=20$	$L=10$	$L=50$
RWR	P	0.0526 (2.8, 0.4)	0.0402 (2.6, 0.4)	0.0425 (2, 0.8)	0.0231 (1.6, 1)
	R	0.0717 (2.8, 0.4)	0.1076 (2.2, 0.4)	0.1006 (1.4, 0.8)	0.2550 (1.6, 1)
	F	0.0512 (2.8, 0.4)	0.0503 (2.6, 0.4)	0.0469 (2, 0.8)	0.0370 (1.6, 1)
LIN	AUC	0.7755 (2.4, 0.2)		0.9053 (0, 2.2)	
	P	0.0503 (3, 0)	0.0393 (3.2, 0)	0.0403 (2.4, 0)	0.0221 (2, 0)
	R	0.0683 (3, 0)	0.1043 (2.6, 0)	0.0963 (2.2, 0)	0.2399 (2, 0)
LOUT	F	0.0489 (3, 0)	0.0487 (3.2, 0)	0.0443 (2.4, 0)	0.0352 (2, 0)
	AUC	0.7729 (2.2, 0)		0.8204 (2.4, 0)	
	P	0.0503 (3, 0)	0.0393 (3.2, 0)	0.0403 (2.4, 0)	0.0221 (2, 0)
	R	0.0683 (3, 0)	0.1043 (2.6, 0)	0.0963 (2.2, 0)	0.2399 (2, 0)
	F	0.0489 (3, 0)	0.0487 (3.2, 0)	0.0443 (2.4, 0)	0.0352 (2, 0)
	AUC	0.7729 (2.2, 0)		0.8208 (1.4, 0)	

doi:10.1371/journal.pone.0101675.t002

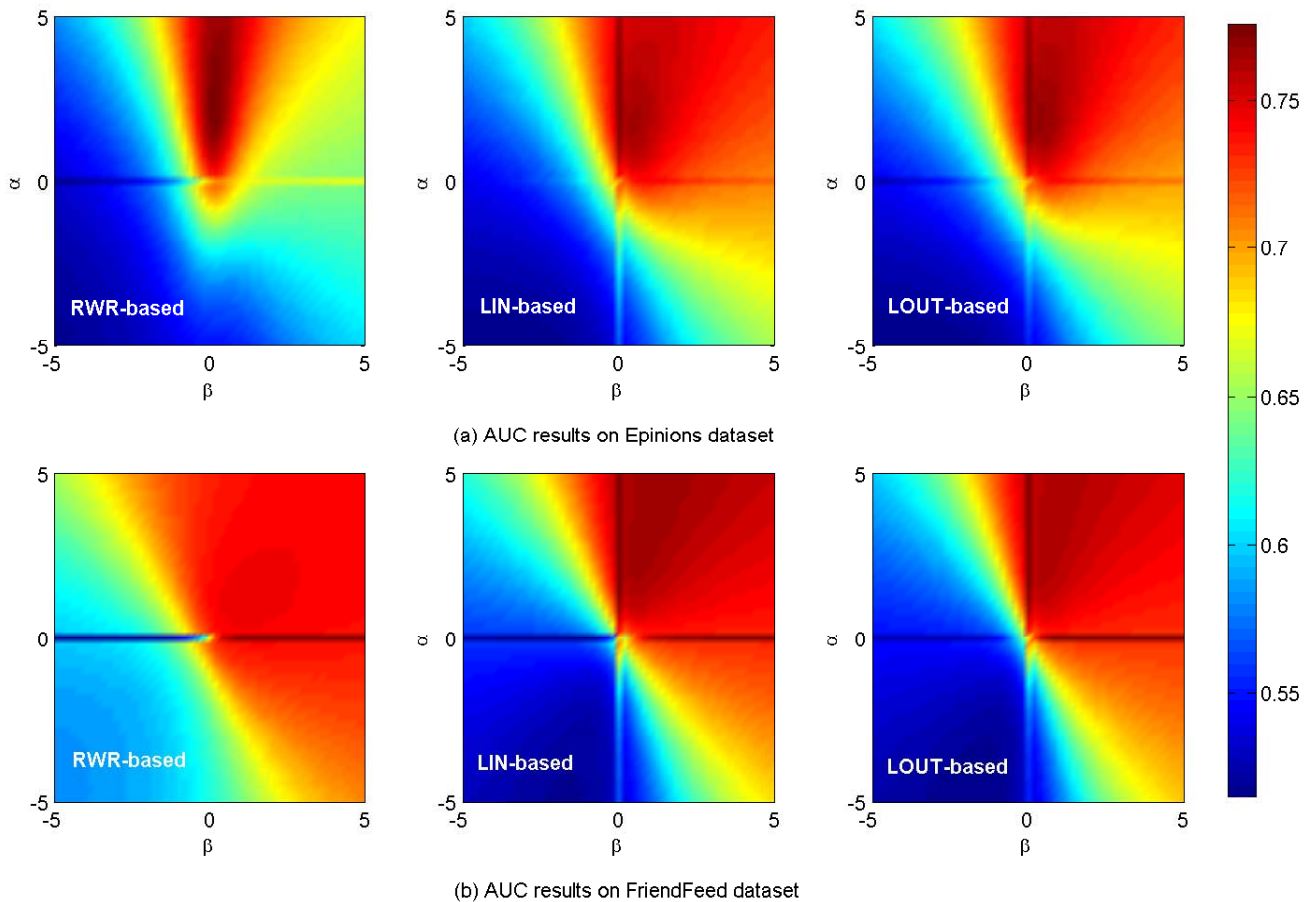


Figure 5. AUC results on Epinions and FriendFeed data sets.
doi:10.1371/journal.pone.0101675.g005

characterize the social similarity, hence it may provide better a recommendation performance. In addition, when use the Eq. (2)–Eq. (3), we remove the negative value and then normalize the social similarity.

1.2 Personalized Preference

There are many methods to compute the common preference between users or items in recommender systems, in which the cosine metric [66] is one of the most frequently used one [67,68]. It reads as follows:

$$p_{ij} = \frac{\sum_{k=1}^n R_{ik} R_{jk}}{\sqrt{\sum_{k=1}^n R_{ik}^2} \sqrt{\sum_{k=1}^n R_{jk}^2}}, \quad (4)$$

where p_{ij} is the examined common preference between nodes i and j .

1.3 Hybrid Algorithm

To fully make use of the effects made both by influence and preference of users, we adopt a nonlinear hybrid method to integrate them. The final similarity between users i and j , S_{ij} , is denoted as

$$S_{ij} = p_{ij}^\alpha * s_{ij}^\beta. \quad (5)$$

Data & Metrics

2.1 Data set

In this paper, we use two data sets (datasets are free to download as **Data S1**), *Epinions.com* [69] and *Friendfeed.com* [70], to evaluate the effect of the algorithm. *Epinions* not only allows users to rate items but also permits them to make social connections with others. *Friendfeed* is a microblogging service provider founded in 2007 and acquired by *Facebook* in 2009. To alleviate the sparse problem [71], we purify the two data sets by making sure that each user has at least twenty six out and in-links (2 for *Friendfeed*) in the social network, and that each user at least collects 7 items (8 items for the *Friendfeed* data set) that each item is collected at least 7 times (8 times for *Friendfeed*). Finally, we obtained a purified data set with 4,066 users, 7,649 items, 217,071 social links and 154,122 bipartite links for *Epinions*, and with 4,188 users, 5,700 items, 386,804 social links and 96,942 bipartite links for *Friendfeed*. Table 1 shows the basic statistics for two representative data sets).

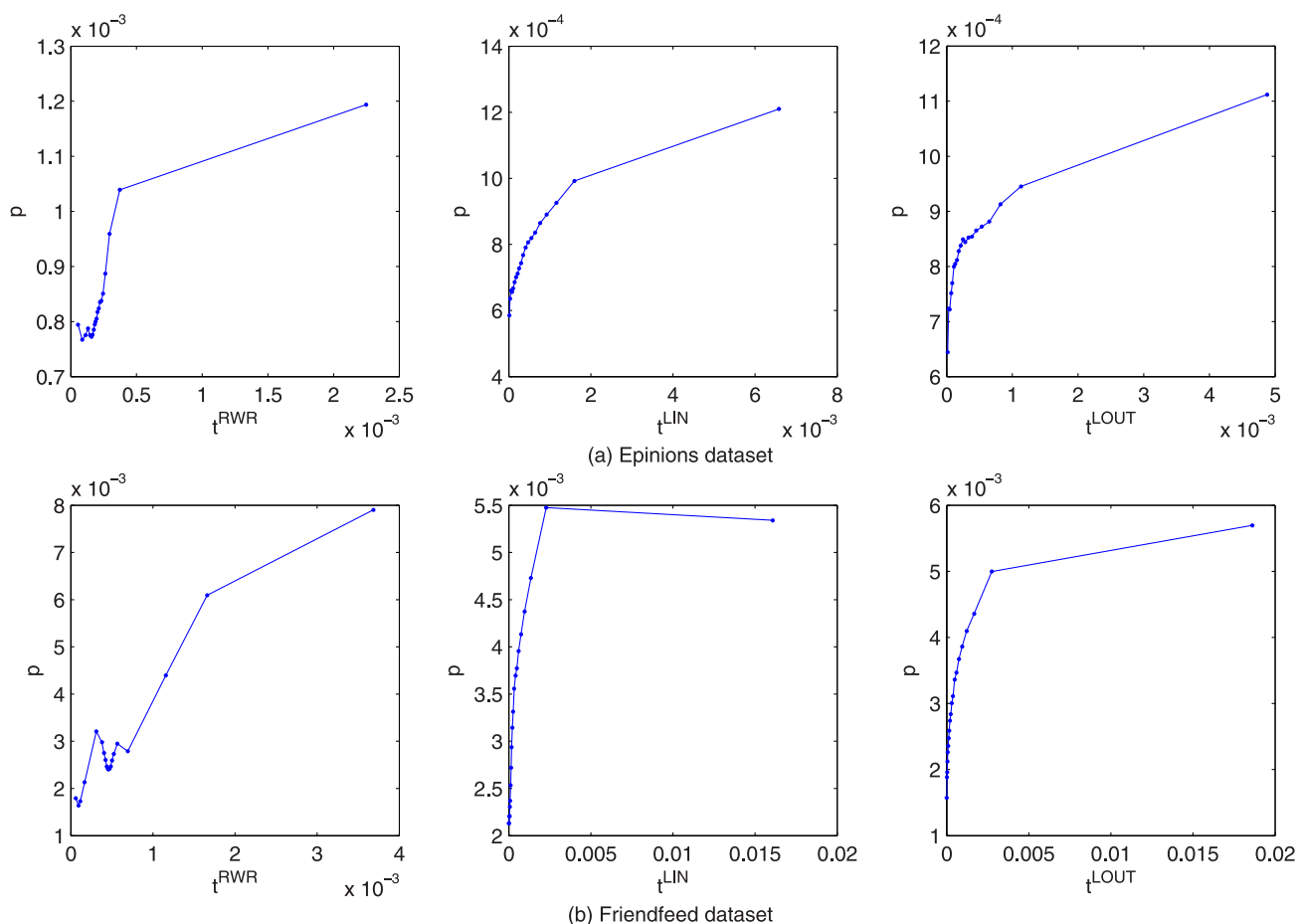


Figure 6. HD results on Epinions and FriendFeed data sets. The length of recommendation list L is set as 10. doi:10.1371/journal.pone.0101675.g006

2.2 Metrics

Every data set is randomly divided into two parts: the training set which is consisted of 80% of the entries and the testing set consisted of the remaining 20%. For a general recommendation process, the training set is treated as known information to run algorithms and generate corresponding recommendations, yet the information in testing set is unavailable while making recommendations. In addition, we use five metrics to do evaluation in order to fully explore the methods' performance, and we consequently employ five different metrics that characterize recommendation performance:

1. *Precision* [8].— *Precision* represents the probability to what extent a selected item is relevant in a given recommendation list, defined as:

$$P_i = \frac{N_{rs}^i}{L}, \tag{6}$$

where L represents the length of recommendation list, and N_{rs}^i is the number of truly recovered items for user i . We can obtain the precision of the whole recommender system by averaging over all individuals' precisions,

$$P = \frac{1}{m} \sum_{i=1}^m P_i, \tag{7}$$

where m represents the number of users. Obviously, a higher precision means that the algorithm is more accurate.

2. *Recall* [8].— *Recall* represents the probability that a relevant item will be picked from testing set, defined as:

$$R_i = \frac{N_{rs}^i}{N_p^i}, \tag{8}$$

where N_p^i is the number of items collected by user i in the testing set, and N_{rs}^i is the number of recovered items of user i . We then obtain the overall recall of the whole recommender system by averaging over all individuals,

$$R = \frac{1}{m} \sum_{i=1}^m R_i. \tag{9}$$

A higher recall means that the algorithm is more accurate.

3. *F-measure* [8] — The *F-measure* metric is a widely used metric for alleviating the sensitivity of sole usage of precision or recall, defined as,

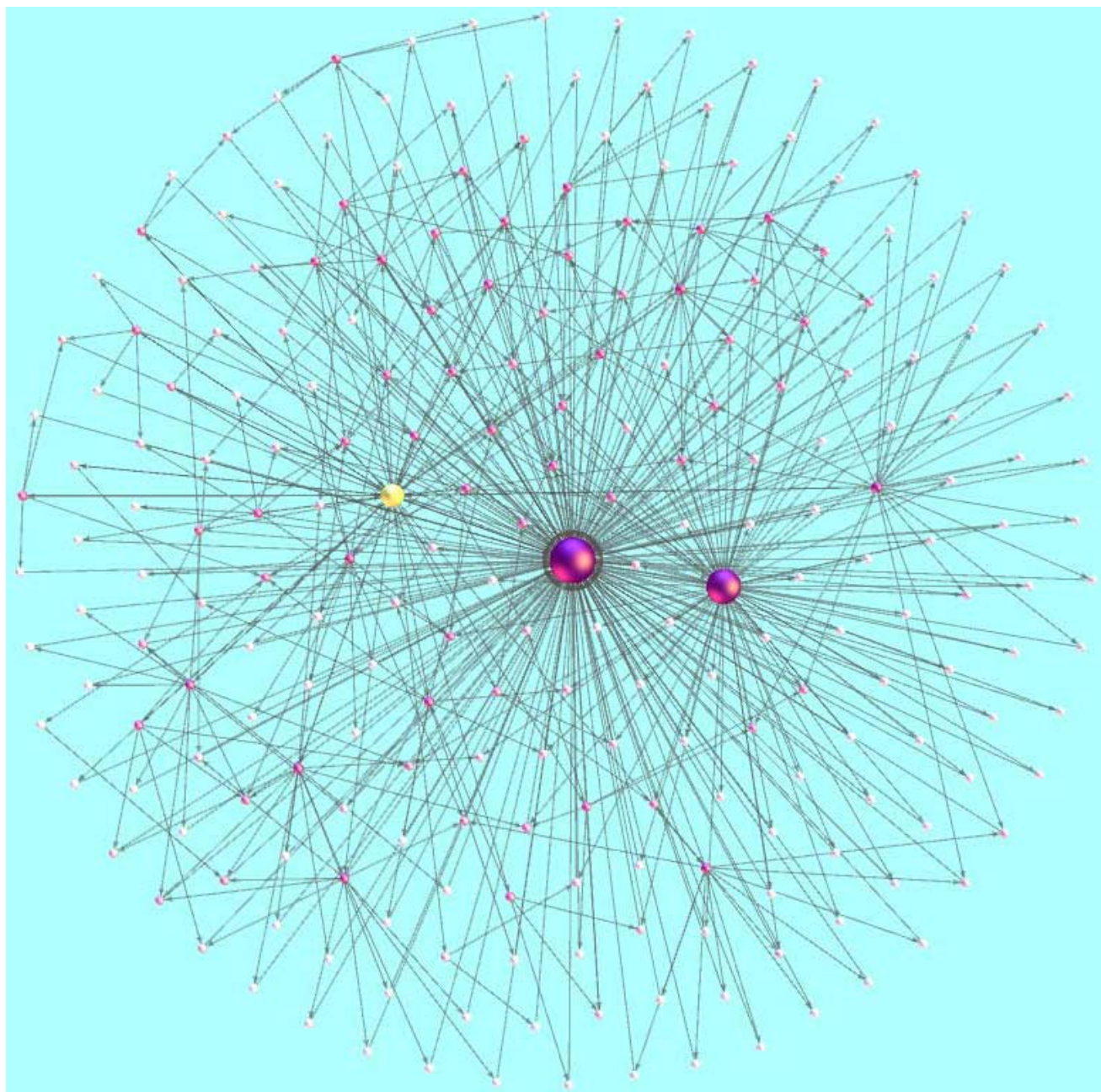


Figure 7. AUC results for HHP, BHC and PD methods on *Epinions* and *FriendFeed* data sets.
doi:10.1371/journal.pone.0101675.g007

Table 3. Performance of the MD with RWR-based methods obtained under the three-fold data division on *Epinions* data set. The recommendation list is set as 10.

Methods	precision	recall	f-measure	HD	AUC
MD	0.0275	0.0708	0.0344	0.5999	0.7757
RWR-based	0.0277	0.0723	0.0344	0.6545	0.7975

doi:10.1371/journal.pone.0101675.t003

Table 4. Performance of the MD with RWR-based methods obtained under the three-fold data division on *Friendfeed* data set. The recommendation list is set as 10.

Methods	precision	recall	f-measure	HD	AUC
MD	0.0254	0.0908	0.0331	0.9258	0.7902
RWR-based	0.0301	0.112	0.0397	0.9397	0.8437

doi:10.1371/journal.pone.0101675.t004

$$F_i = \frac{2P_iR_i}{P_i + R_i} \tag{10}$$

$$F = \frac{1}{m} \sum_{i=1}^m F_i \tag{11}$$

Anomalously, we can obtain the *F-measure* of the whole system by averaging over all individuals,

4. *AUC*— *AUC* (Area Under ROC Curve) is different from the above three metrics, for *AUC* evaluates the likelihood of all items instead of the *TOP L* recommendation, where ROC stands for the receiver operating characteristic [72,8]. It can be approached with a sampling method

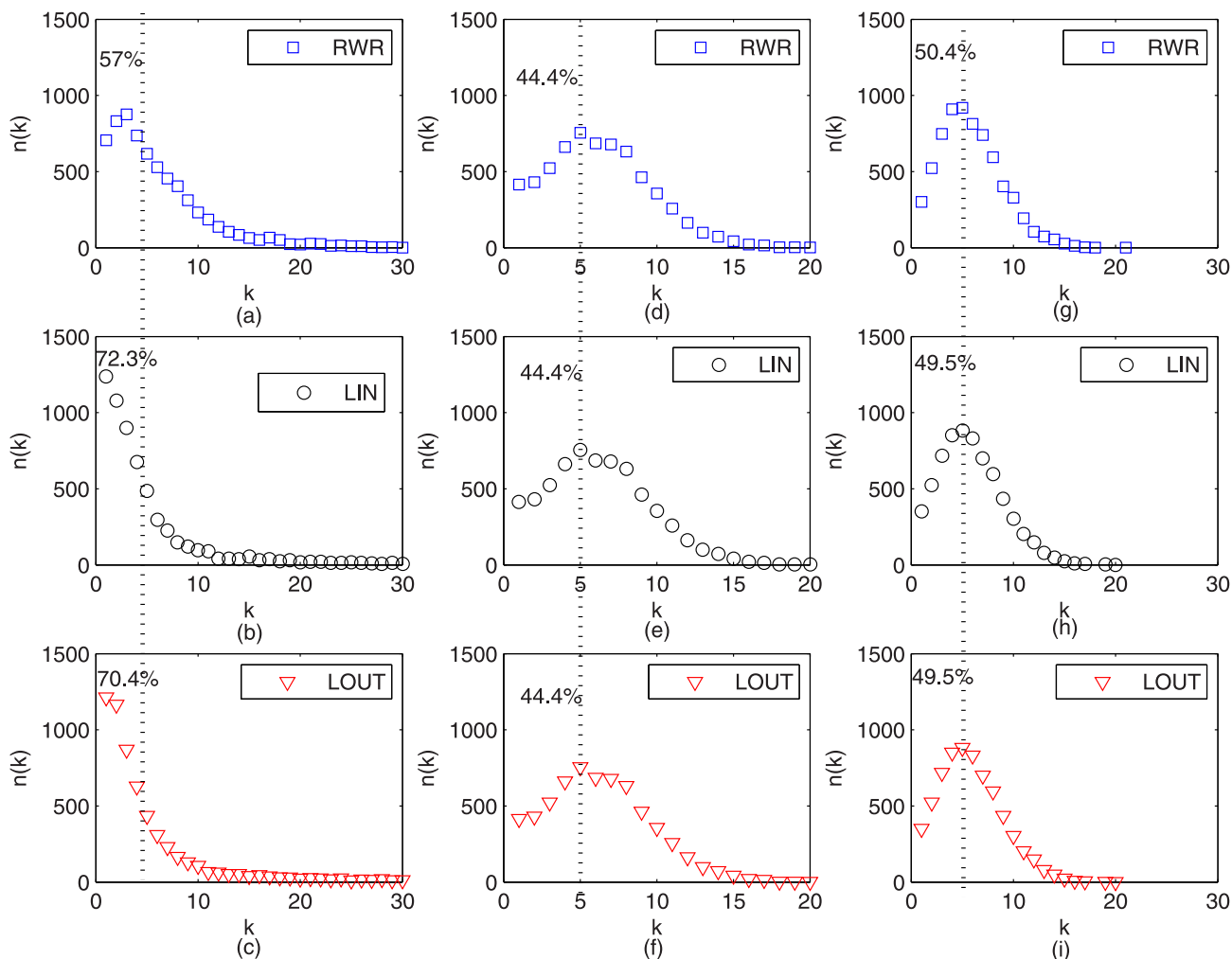


Figure 8. Mean personal preference versus social similarity for *Epinions* and *Friendfeed*, respectively. From left to right, the metrics are respectively RWR-, LIN-, LOUT-based social similarity. The personal preference is averaged according to each social similarity value.
doi:10.1371/journal.pone.0101675.g008

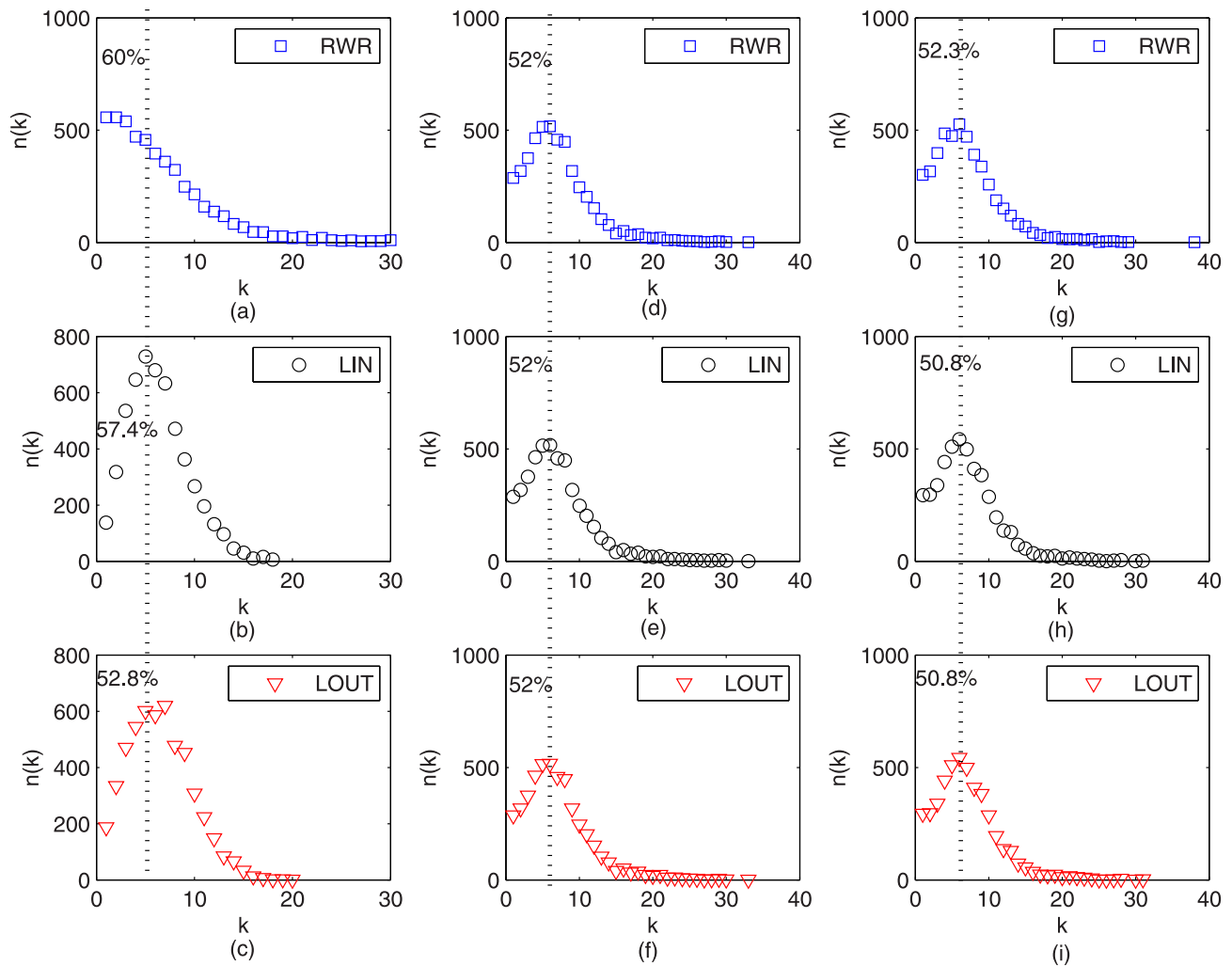


Figure 9. Illustration of a typical example of an ego network for a node with the largest social similarity value (the biggest size).
doi:10.1371/journal.pone.0101675.g009

$$AUC = \frac{n' + 0.5n''}{n}, \quad (12)$$

where n is the number of independent sampling, and n' is the number of the predicted score of target item which is higher than that of the randomly selected item, and n'' is the times of the target which is the same with random items'. If all the scores are generated from an independent and identical distribution, the then AUC should be 0.5. Therefore, how much the value of the AUC exceeds 0.5 indicates how much the algorithm performs better than a random prediction.

5. *Diversity (HD)*.— HD [22] considers the unique and different user's recommendation list. Given two users i and j , the difference between their recommendations lists can be measured by the Hamming distance.

$$HD_{ij}(L) = 1 - \frac{Q_{ij}(L)}{L}, \quad (13)$$

where $Q_{ij}(L)$ is the number of recommended items in the top- L places of both lists. Averaging over all pairs of users' $HD_{ij}(L)$, we

can obtain the diversity of the observed algorithm. Clearly, higher result (HD) means higher personalization of users' recommendation lists.

Results & Analysis

3.1 Experimental Results

Fig. 2–Fig. 4 show the algorithm results on *Epinions* and *Friendfeed* data sets. It can be seen that, for a given length of recommendation list L , the precision, recall, F-measure and AUC achieve the optimal accuracy for the same parameters for both the LIN-based and LOUT-based method (see also Table 2), which indicates that the local information of both in-flow and out-flow has the similar impact on information filtering. Comparatively, for a moderately small length of recommendation list $L = 10$, the precision, recall and F-measure values of RWR-based method reach their maximum value 0.0526, 0.0717 and 0.0512 for $(\alpha, \beta) = (2.8, 0.4)$, respectively. Moreover, the corresponding results are 0.0503, 0.0683 and 0.0489 for $(\alpha, \beta) = (3, 0)$ on *Epinions* data set whether LIN-based or LOUT-based. For *Friendfeed*, those metrics under RWR-based method have reached 0.0425, 0.1006 and 0.0469 for parameter set $(\alpha, \beta) = (2, 0.8)$, $(1.4, 0.8)$ and $(2, 0.8)$,

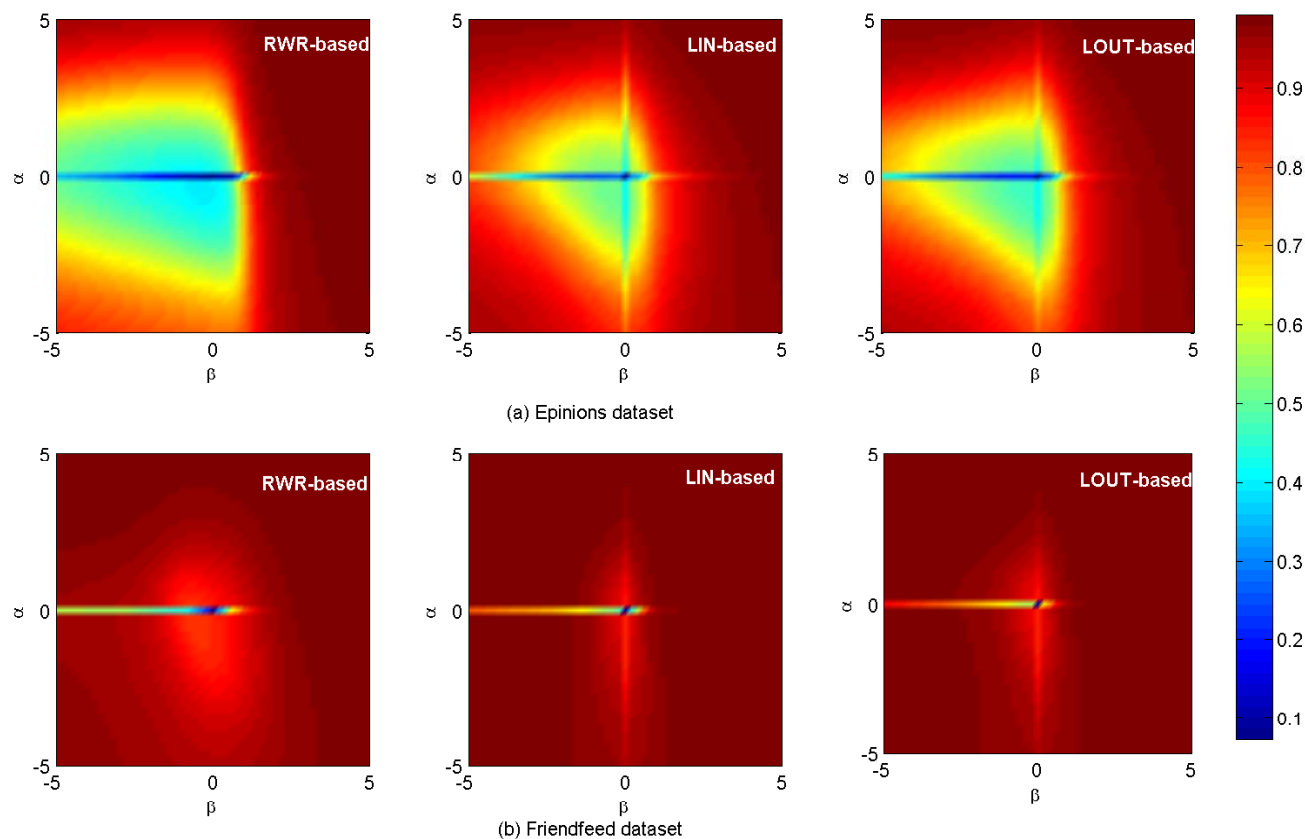


Figure 10. Number of recommended items versus degree on *Epinions* for $L = 10$. From left to right, the parameters (α, β) of Eq. (5) are set as (1,0), (0,1), and (α^*, β^*) given in Table 2, respectively. The dash line indicates the degree of 5, and the corresponding number shows the percentage of all the recommendation items.

doi:10.1371/journal.pone.0101675.g010

respectively. For LIN-based or LOU-based methods, when $(\alpha, \beta) = (2.4, 0)$, such metrics obtain their maximum value 0.0403, 0.0963 and 0.0443. Similar results can also be found for $L = 20$ and $L = 50$ (see Table 2).

Fig. 5 shows the *AUC* results. In Fig. 5(a), the maximum *AUC* values are respectively 0.7755, 0.7729 and 0.7729 for $(\alpha, \beta) = (2.4, 0.2)$, $(\alpha, \beta) = (2.2, 0)$ and $(\alpha, \beta) = (2.2, 0)$ on *Epinions* data set. In Fig. 5(b), the corresponding maximum values are respectively 0.9053, 0.8204 and 0.8208 for $(\alpha, \beta) = (0, 2.2)$, $(\alpha, \beta) = (2.4, 0)$ and $(\alpha, \beta) = (1.4, 0)$ on *Friendfeed*, respectively. A brief summary is given in Table 2. Fig. 6 shows the *HD* results on *Epinions* and *Friendfeed* data sets, respectively, and the length of the recommendation list is 10. For all the diversity, their maximum diversity lies in the same position $(\alpha, \beta) = (5, 5)$. In Fig. 6 (a), the maximum *HD* values are respectively 0.9864, 0.9817 and 0.9815 for RWR-based, LIN-based and LOU-based in *Epinions* data set. In Fig. 6 (b), the maximum *HD* with RWR-based, LIN-based and LOU-based, is 0.9928, 0.9923 and 0.9918 for *Friendfeed* data set, respectively. However, we can find that the diversity in the best *AUC* value's position is higher than that of only using the personal preference. For example, when the recommendation list $L = 10$ on *Epinions* data set, the *HD* values are 0.6944, 0.5297 and 0.4923 in the best *AUC* value's position, only using the personal preference and using the social similarity, respectively.

It is noticed that, for all aforementioned results two crossing lines can be obviously found for LIN- and LOU-based methods at $\alpha = 0$ or $\beta = 0$, while only a horizontal line is observed for RWR-based method at $\alpha = 0$. As we known, the cosine, LIN and LOU

are methods for computing similarity simply based on local information, while RWR-based method considers not only the local information, but also takes into account the global social structure. In addition, the behavior network and social network are sparse. Therefore, the personal preference matrix and the social similarity computed by LIN and LOU might be sparse but the matrix by RWR is full, i.e., there are many zero elements in those matrices that are computed by the cosine, namely LIN, LOU and RWR. When $\alpha = 0$, only the social similarity works. Since the personal preference is small, the final similarity will be much sparser. When $\beta = 0$, only personal preference works, and the final similarity matrix will be much sparser when using LIN and LOU methods, i.e., the LIN and LOU methods will filter the recommendation but the RWR method will supplement it. Thus, that is why it has horizontal lines in the figures and only LIN and LOU methods have vertical line. As shown in Table 1, the information network is much sparser than that of the corresponding social network, hence more items are possible to be discovered via social connections. In addition, the size of hot areas (corresponding to high performance) of RWR-based method is much larger than that of the other two methods, as it considers not only the nearest neighbors, but also integrates the effect of remote nodes which are not directly connected. Comparatively, the local based (LIN- and LOU-based) methods can only take into account the commonly direct neighbors, neglecting the global role of each individual. Furthermore, the hybrid case will achieve the best performance for both the observed data sets with optimal

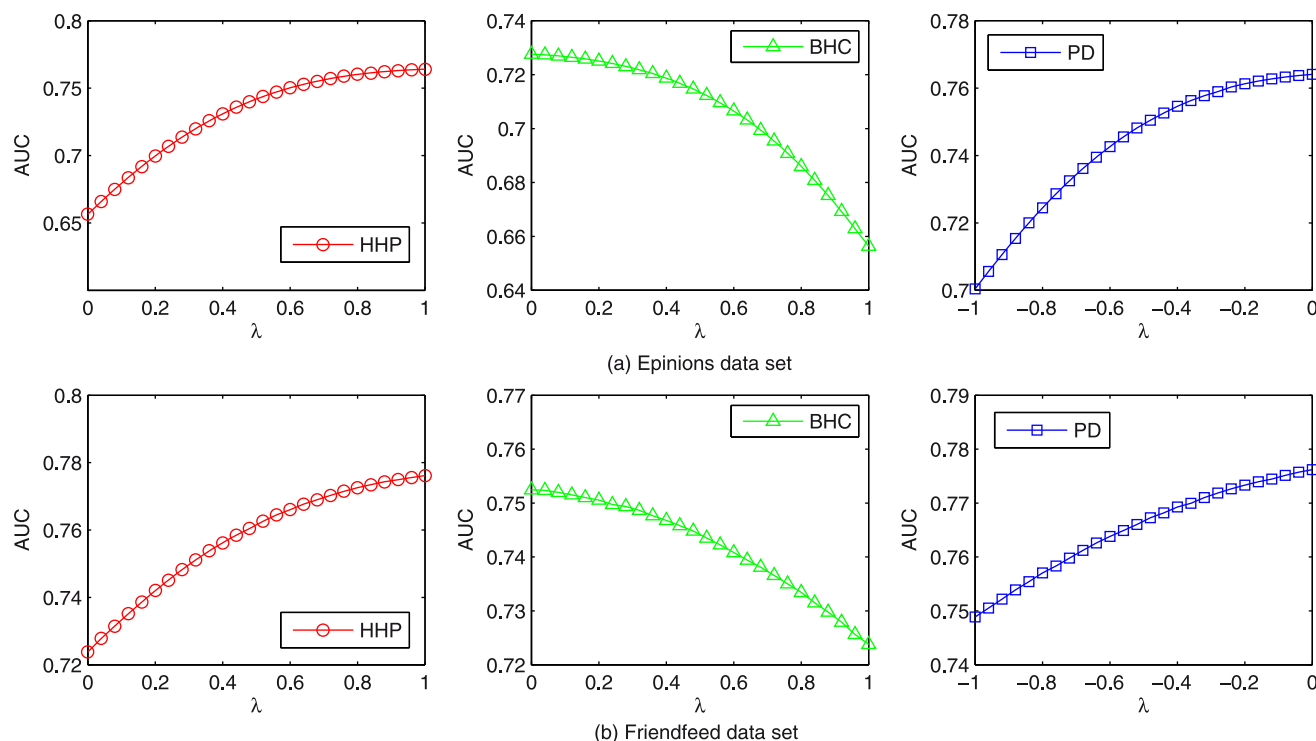


Figure 11. Number of recommended items versus degree on *FriendFeed* for $L=10$. From left to right, the parameters (α, β) of Eq. (5) are set as $(1,0)$, $(0,1)$, and (α^*, β^*) given in Table 2, respectively. The dash line indicates the degree of 5, and the corresponding number shows the percentage of all the recommendation items.

doi:10.1371/journal.pone.0101675.g011

parameters $\alpha^* > \beta^*$, which also proves that social reinforcement is more significant than individual behaviors in information filtering.

Fig. 7 shows that the *AUC* result with one baseline method [22] (HHP for short) and its two variants, [31] (BHC for short) and [30] (PD for short) on *Epinions* and *Friendfeed* data sets, respectively. It can be seen that the *AUC* value of HHP method changes monotonously with λ [17], i.e., the HHP method degenerates to pure Mass Diffuse (MD for short) method when $\lambda=1$. We find that the *AUC* of both HHP and PD methods increase with λ , while that of BHC decreases with λ (When $\lambda=1$, HHP degenerates to the pure MD method, and BHC degenerates to the pure Heat Conduction (HC) method. When $\lambda=0$, PD degenerates to pure MD method). Generally, the MD method has higher accuracy but lower diversity, while the HC method has higher diversity but lower accuracy. For a better recommendation algorithm, it should ensure higher accuracy principally, thus users might continue to use the system and enlarge their vision by its diverse functions. Therefore, we additionally compare our method with MD. In order to avoid the over-fitting problem [73], we use the three-fold data division [74] to validate our method (see Table 3 and Table 4), where we use 80% of the data as training set, and obtain the optimal parameter value with 10% of the data. We then use the remaining 10% to validate it. It can be seen that the proposed method outperforms the MD algorithm on all the five different metrics.

3.2 Empirical Analysis

To better understand how the different layers of coupled networks interact with each other, in this section, we empirically investigate the relationship between social similarity and personal preference from micro and macro perspectives. Fig. 8 described

that the relationship between social similarity and personal preference for each pair of users. The result shows that, generally, social similarity are positively correlated [55] with personal preference at both local and global measures, indicating that the mutual reinforcement principle [66] also applies to online social activities.

In Fig. 9, we also find that a typical example of an ego network [75] for a node with the largest social similarity value (with the biggest size). It can be seen that it connects to a node of relatively large social similarity yet small similarity (yellow one), suggesting the rich-club phenomenon [76] of social interests activities. That is to say, users with high social impact tend to interact with users of high social similarity, even if they lack common activities. Furthermore, we also find that the degree distribution of successfully recommended items in Fig. 10 and Fig. 11 for *Epinions* and *Friendfeed*, respectively. In Fig. 10(a–c) and Fig. 11(a–c), the parameters of Eq. 5 are set as $\alpha=0$ and $\beta=1$, of which only the social similarity takes effect in the recommendation process. It shows that the local measures (LIN and LOU) are more likely to find small-degree items (the degree is smaller than 5) than the RWR metric (around 57%). Similarly, for another extreme case of Eq. 5, (α, β) is set as $(1,0)$, implying that only the personal preference will work for information filtering, hence all the results are identical in Fig. 10(d–f) and Fig. 11(d–f), respectively. In addition, the number of recommended small-degree items is fewer than that of social based method. Comparatively, in Fig. 10(g–i) and Fig. 11(g–i), the parameter (α, β) is set as the optimal case given in Table 2. Since both the social similarity and personal preference are integrated, the hybrid algorithm not only can find those *cold* items [34,26] (where the social similarity primarily works), but also can push some popular

items (which is largely because of the personal preference). Therefore, it finally can achieve a better performance for information filtering. In addition, the novelty [10] of recommender systems refers to how different the recommended objects are from what the users have already seen before. The simplest way to quantify the ability of an algorithm to generate novel and unexpected results is to measure the average popularity of the recommended objects. The lower the average objects's degree in the recommendation list, the better the novelty of the system. From Fig. 10 and Fig. 11, we can see that the number of recommended small-degree items is larger than that of only using personal preference and fewer than that of the social based method, i.e., our method has higher novelty than that of only using personal preference.

Conclusions & Discussion

In this paper, we have proposed a hybrid information filtering algorithm based on the coupled social networks, which considers the effects of both social similarity and personalized preference. We apply three metrics, *LN*, *LOUT* and *RWR*, to evaluate the asymmetrically social similarity, and use the cosine similarity to measure the symmetrically personalized preference. In addition, we integrate them with two tunable parameters in order to obtain better recommendation results. Experimental results show that the

hybrid pattern can not only provide more accurate recommendations, but also enlarge the recommendation coverage while adopting global metric (*RWR*). Further empirical analyses demonstrate that the mutual reinforcement can also be extended to coupled networks where the same individuals occupy the core position of the entire online society. However, this article only provides a simple start for making use of both behavior and social information, while a couple of issues remain open for future study. Especially, the underlying mechanism driving the interaction of social and information networks is of particular importance to deeply understand how coupled social networks work, as well as its potential applications.

Supporting Information

Data S1 Datasets.

(ZIP)

Author Contributions

Conceived and designed the experiments: DCN YF ZKZ. Performed the experiments: DCN JLZ. Analyzed the data: DCN JLZ KZ ZKZ. Contributed reagents/materials/analysis tools: DCN KZK. Contributed to the writing of the manuscript: DCN JLZ ZKZ YF KZ.

References

- Faloutsos M, Faloutsos P, Faloutsos C (1999) On power-law relationships of the internet topology. *Comput Commun Rev* 29: 251–262.
- Froomkin AM (1995) Flood control on the information ocean: Living with anonymity, digital cash, and distributed databases. *Journal of Law and Commerce* 15: 395.
- Brin S, Page L (1998) The anatomy of a large-scale hypertextual web search engine. *Computer networks and ISDN systems* 30: 107–117.
- Morita M, Shinoda Y (1994) Information filtering based on user behavior analysis and best match text retrieval. In: *Proc. 17th Ann. Intl. ACM SIGIR Conf. Research Develop. Infor. Retr.* Springer-Verlag New York, Inc., pp. 272–281.
- Lee D (2008) To find or to be found, that is the question in mobile information retrieval. In: *Proce. SIGIR 2008 Workshop on Mobile Infor. Retr.* pp. 7–10.
- Resnick P, Varian H (1997) Recommender systems. *Commun ACM* 40: 56–58.
- Burke R (2002) Hybrid recommender systems: Survey and experiments. *User modeling and user-adapted interaction* 12: 331–370.
- Herlocker JL, Konstan JA, Terveen LG, Riedl JT (2004) Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems* 22: 5–53.
- Adomavicius G, Tuzhilin A (2005) Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering* 17: 734–749.
- Lü L, Medo M, Yeung CH, Zhang YC, Zhang ZK, et al. (2012) Recommender systems. *PhysRep* 519: 1–49.
- Huang Z, Chen H, Zeng D (2004) Applying associative retrieval techniques to alleviate the sparsity problem in collaborative filtering. *ACM Trans Info Syst* 22: 116–142.
- Zhou T, Ren J, Medo M, Zhang YC (2007) Bipartite network projection and personal recommendation. *Physical Review E* 76: 046115.
- Breese J, Heckerman D, Kadie C (1998) Empirical analysis of predictive algorithms for collaborative filtering. In: *Proc. 4th Conf. Uncertainty Artif. Intel.* Morgan Kaufmann Publishers Inc., pp. 43–52.
- Sarwar B, Karypis G, Konstan J, Riedl J (2001) Item-based collaborative filtering recommendation algorithms. In: *Proc. 10th Intl. Conf. WWW.* ACM, pp. 285–295.
- Pazzani M, Billsus D (2007) Content-based recommendation systems. *The adaptive web*: 325–341.
- Liu JH, Zhang ZK, Yang C, Chen L, Liu C, et al. (2014) Gravity effects on information filtering and network evolving. *PLoS ONE* 9: e91070.
- Kumar R, Raghavan P, Rajagopalan S, Tomkins A (1998) Recommendation systems: A probabilistic analysis. In: *Foundations of Computer Science, 1998. Proceedings. 39th Annual Symposium on.* IEEE, pp. 664–673.
- Krestel R, Fankhauser P, Nejdl W (2009) Latent dirichlet allocation for tag recommendation. In: *Proceedings of the third ACM conference on Recommender systems.* ACM, pp. 61–68.
- Sarwar B, Karypis G, Konstan J, Riedl J (2000) Application of dimensionality reduction in recommender system—a case study. Technical report, DTIC Document.
- Zhang YC, Blattner M, Yu YK (2007) Heat conduction process on community networks as a recommendation model. *Physical Review Letters* 99: 154301–154304.
- Burke R (2002) Hybrid recommender systems: Survey and experiments. *User modeling and user-adapted interaction* 12: 331–370.
- Zhou T, Kuscsik Z, Liu JG, Medo M, Wakeling JR, et al. (2010) Solving the apparent diversity-accuracy dilemma of recommender systems. *Proc Natl Acad Sci USA* 107: 18803–18808.
- Resnick P, Iacovou N, Suchak M, Bergstrom P, Riedl J (1994) Grouplens: an open architecture for collaborative filtering of netnews. In: *Proc. 1994 ACM Conf. Comput. Supported Cooperative work.* ACM, pp. 175–186.
- Linden G, Smith B, York J (2003) Amazon.com recommendations: Item-to-item collaborative filtering. *IEEE Internet Comput* 7: 76–80.
- Zhang ZK, Liu C (2010) A hypergraph model of social tagging networks. *J Stat Mech* 2010: P10005.
- Zhang ZK, Liu C, Zhang YC, Zhou T (2010) Solving the cold-start problem in recommender systems with social tags. *EPL* 92: 28002–28007.
- Zhang ZK, Zhou T, Zhang YC (2011) Tag-aware recommender systems: A state-of-the-art survey. *Journal of Computer Science and Technology* 26: 767–777.
- Wang F, Landau D (2001) Efficient, multiple-range random walk algorithm to calculate the density of states. *Phys Rev Lett* 86: 2050–2053.
- Leicht E, Holme P, Newman M (2006) Vertex similarity in networks. *Phys Rev E* 73: 026120.
- Lü L, Liu W (2011) Information filtering via preferential diffusion. *Physical Review E* 83: 066119.
- Liu JG, Zhou T, Guo Q (2011) Information filtering via biased heat conduction. *Physical Review E* 84: 037101.
- Liu JG, Shi K, Guo Q (2012) Solving the accuracy-diversity dilemma via directed random walks. *Physical Review E* 85: 016118.
- Sun D, Zhou T, Liu JG, Liu RR, Jia CX, et al. (2009) Information filtering based on transferring similarity. *Phys Rev E* 80: 17101.
- Qiu T, Chen G, Zhang Z, Zhou T (2011) An item-oriented recommendation algorithm on cold-start problem. *EPL* 95: 58003.
- Chen G, Qiu T, Zhang Z (2012) Promotional effect on cold start problem and diversity in a data characteristic based recommendation method. *arXiv:12052822*.
- Zhou T, Su RQ, Liu RR, Jiang LL, Wang BH, et al. (2009) Accurate and diverse recommendations via eliminating redundant correlations. *New Journal of Physics* 11: 123008.
- Zhang ZK, Zhou T, Zhang YC (2010) Personalized recommendation via integrated diffusion on user-item-tag tripartite graphs. *Physica A* 389: 179–186.
- Freeman L (1979) Centrality in social networks conceptual clarification. *Social networks* 1: 215–239.
- Fu F, Liu L, Wang L (2008) Empirical analysis of online social networks in the age of web 2.0. *Physica A* 387: 675–684.
- Nowak MA (2006) Five rules for the evolution of cooperation. *Science* 314: 1560–1563.

41. Szabó G, Fáth G (2007) Evolutionary games on graphs. *Phy Rep* 446: 97–216.
42. Fortunato S (2010) Community detection in graphs. *Phys Rep* 486: 75–174.
43. Kamel Boulos M, Wheeler S (2007) The emerging web 2.0 social software: an enabling suite of sociable technologies in health and health care education1. *Health Information & Libraries Journal* 24: 2–23.
44. Kautz H, Selman B, Shah M (1997) Referral web: combining social networks and collaborative filtering. *Commun ACM* 40: 63–65.
45. Radicchi F, Arenas A (2013) Abrupt transition in the structural formation of interconnected networks. *Nature Physics* 9: 717–720.
46. Buldyrev S, Parshani R, Paul G, Stanley H, Havlin S (2010) Catastrophic cascade of failures in interdependent networks. *Nature* 464: 1025–1028.
47. Givoni M, Banister D (2006) Airline and railway integration. *Transport Policy* 13: 386–397.
48. De Domenico M, Solé-Ribalta A, Cozzo E, Kivela M, Moreno Y, et al. (2013) Mathematical formulation of multilayer networks. *Phys Rev X* 3: 041022.
49. Gao J, Buldyrev S, Stanley H, Havlin S (2011) Networks formed from interdependent networks. *Nature Phys* 8: 40–48.
50. Lü L, Jin CH, Zhou T (2009) Similarity index based on local paths for link prediction of complex networks. *Phys Rev E* 80: 46122.
51. Massa P, Avesani P (2007) Trust-aware recommender systems. In: *Proceedings of the 2007 ACM conference on Recommender systems*. ACM, pp. 17–24.
52. Guha R, Kumar R, Raghavan P, Tomkins A (2004) Propagation of trust and distrust. In: *Proceedings of the 13th international conference on World Wide Web*. ACM, pp. 403–412.
53. Abdul-Rahman A, Hailes S (2000) Supporting trust in virtual communities. In: *System Sciences, 2000. Proceedings of the 33rd Annual Hawaii International Conference on*. IEEE, pp. 9–pp.
54. Knapskog S (1998) A metric for trusted systems. In: *Proceedings of the 21st National Security Conference*. Citeseer, pp. 16–29.
55. Bhuiyan T (2010) A survey on the relationship between trust and interest similarity in online social networks. *Journal of Emerging Technologies in Web Intelligence* 2: 291–299.
56. Crandall D, Cosley D, Huttenlocher D, Kleinberg J, Suri S (2008) Feedback effects between similarity and social influence in online communities. In: *Proceeding of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, pp. 160–168.
57. Esslimani I, Brun A, Boyer A (2009) From social networks to behavioral networks in recommender systems. In: *Intl. Conf. Adv. Social Netw. Anal. Mining (ASONAM'09)*. IEEE, pp. 143–148.
58. Massa P, Avesani P (2004) Trust-aware collaborative filtering for recommender systems. In: *On the Move to Meaningful Internet Systems 2004: CoopIS, DOA, and ODBASE*, Springer. pp. 492–508.
59. Walter FE, Battiston S, Schweitzer F (2008) A model of a trust-based recommendation system on a social network. *Autonomous Agents and Multi-Agent Systems* 16: 57–74.
60. Zeng W, Zeng A, Shang MS, Zhang YC (2013) Membership in social networks and the application in information filtering. *The European Physical Journal B* 86: 1–7.
61. Tong H, Faloutsos C, Pan J (2006) Fast random walk with restart and its applications. In: *Proceedings of the Sixth International Conference on Data Mining, 2006. ICDM '06*. Ieee, pp. 613–622.
62. Shang MS, Lü L, Zeng W, Zhang YC, Zhou T (2009) Relevance is more significant than correlation: Information filtering on sparse data. *EPL* 88: 68008.
63. Yu AW, Mamoulis N, Su H (2014) Reverse top-k search using random walk with restart. *Proceedings of the VLDB Endowment* 7.
64. Anderberg MR (1973) *Cluster analysis for applications*. Academic Press.
65. Mild A, Reutterer T (2002) An improved collaborative filtering approach for predicting crosscategory purchases based on binary market basket data. *Journal of Retailing and Consumer Services* 10: 123–133.
66. Kleinberg JM (1999) Authoritative sources in a hyperlinked environment. *J ACM* 46: 604–632.
67. Ziegler CN (2013) *On recommender systems*. In: *SocialWeb Artifacts for Boosting Recommenders*, Springer. pp. 11–20.
68. Liu C (2014) Discriminant analysis and similarity measure. *Pattern Recognition* 47: 359–367.
69. Massa P, Avesani P (2006) Trust-aware bootstrapping of recommender systems. In: *ECAI 2006 Workshop on Recommender Systems*, Riva del Garda, Italy. Citeseer, pp. 29–33.
70. Celli F, Di Lascio F, Magnani M, Pacelli B, Rossi L (2010) Social network data and practices: The case of friendfeed. *Advances in Social computing*: 346–353.
71. Zeng W, Zeng A, Shang MS, Zhang YC (2013) Information filtering in sparse online systems: recommendation via semi-local diffusion. *PLoS one* 8: e79354.
72. Hanley JA, McNeil BJ (1982) The meaning and use of the area under a receiver operating characteristic (roc) curve. *McNeil, Radiology* 143: 29–36.
73. Abu-Mostafa YS, Magdon-Ismail M, Lin HT (2012) *Learning from data*. AMLBook.
74. Zeng A, Vidmer A, Medo M, Zhang YC (2014) Information filtering by similarity-preferential diffusion processes. *EPL* 105: 58002.
75. Mednick SC, Christakis NA, Fowler JH (2010) The spread of sleep loss influences drug use in adolescent social networks. *PLoS ONE* 5: e9775.
76. Zhou S, Mondragón RJ (2004) The rich-club phenomenon in the internet topology. *Commun Lett* 8: 180–182.