International Journal of Machine Learning and Cybernetics (2019) 10:1003–1024
https://doi.org/10.1007/s13042-017-0778-1

ORIGINAL ARTICLE

Friend recommendation in social networks based on multi-source information fusion

Shulin Cheng[1,2] · Bofeng Zhang[1] · Guobing Zou[1] · Mingqing Huang[1] · Zhu Zhang[1]

Received: 30 December 2016 / Published online: 23 February 2018

Abstract

Friend recommendation (FR) in social networks has been widely studied in recent years, which mainly focuses on social relationships and user interests. Friend of Friend method is one representative. However, the disadvantage is that most of existing solutions ignored other valuable information, such as user profile, location, influence and indirect trust. In fact, being friends among users is either determined by one or two dominant factors that originate from varying information sources, or the results of multiple main factors gaming. Motivated by the observations above, we propose a scalable FR framework in social networks, where multiple sources have been integrated based on improved D-S evidence theory. More specifically, we first analyzed 7 valuable information sources and categorized them into three classes, including Personal Features, Network Structure Features and Social Features. Furthermore, we also propose a fusion recommendation framework based on D-S evidence theory which embodies the minimal conflicts among evidences. In the proposed method, we first optimize the framework by importance degree and reliability of evidence based on original D-S evidence theory. Then, we designed a novel BPA evidence function by quantifying the evidence, where each evidence measures the relevance of forming friends among users. Finally, we describe the fusion FR algorithm plugged into our recommendation framework. The experiments on real-world dataset show that our proposed approach outperforms the other state-of-the-art algorithms on five evaluation metrics. The experimental results demonstrate the effectiveness of fusing multi-source information for FR in social networks.

Keywords Recommender system · Social networks · Friend recommendation · Information fusion · D-S evidence theory

1 Introduction

In the last few years, social networks [1] have achieved great attentions in multiple aspects and emerged many social media platforms, such as Twitter,[1] Facebook,[2] Google+,[3] LinkedIn,[4] Sina Weibo[5] and Tencent Weibo.[6] To a large extent, they have changed and influenced the way of communications among users. These achievements attracted many researchers to study the corresponding interesting social problems and technical issues [1], such as advertising marketing [2], system risk supervision [3], information diffusion [4], community application [5], item recommendation and friend recommendation [6]. People in social networks have obtained sufficient knowledge experiences and operating privileges like searching information, making friends, posting topics and sharing information. However, a large amount of information and new users appear in different social networks every day, so that people could get lost easily when acquiring interesting information and like-minded friends. Hence, recommendation technique [6, 7] plays an important role in addressing the issues in order to facilitate the user experiences in social networks.

✉ Bofeng Zhang
bfzhang@shu.edu.cn

✉ Guobing Zou
guobingzou@gmail.com

1 School of Computer Engineering and Science, Shanghai University, Shanghai 200444, China

2 School of Computer and Information, Anqing Normal University, Anqing 246133, China

[1] https://twitter.com/.

[2] http://www.facebook.com/.

[3] http://plus.google.com/.

[4] https://www.linkedin.com/.

[5] http://weibo.com/, a famous Chinese microblog web site which has far more than one hundred million users.

[6] http://t.qq.com/, another well-known Chinese microblog web site analogue to Sina Weibo.

Recommendation techniques have been successfully applied on the Internet like in Amazon [8] and E-learning [9], which utilize the correlative information and user behaviors or/and item content to produce recommendations in traditional recommender systems without using social information [7, 8, 10]. In social networks, recommendation techniques take full advantage of social information, such as friendships, user interactions, influence and trust relationship, so as to promote recommender system to improve its recommendation effectiveness [11–13]. Social recommendation consists of item recommendation (e.g., recommending news information, topics, pictures and videos) and friend recommendation finding similar users with high interest similarity or like-minded users.

In social networks, predicting the missing links or potential links could be valuable in the future, which corresponds to recommending friends to active users, could be an interesting research issue. FR is to mine potential link relations in the future or discover the missing link relations among users [14], which not only can increase the connections among users, but also improve the user's loyalty. Although it is similar to link prediction in the complex network, there has its own specific features [15] and it could be influenced by many kinds of information. In a social network, there are large number of nodes with many social features, as well as massive interactions and relations among users. More specifically, a social network could be viewed as a heterogeneous network no less than a traditional complex network, where followship/friendship and interactions (like review and retweet) could imply the underlying information of relevance. The role of each relationship varies along with user's preferences. Users in a social network sharing common or similar features and interests have high similarities or relevance in several aspects. Thus, they can be converted into linkage relevance from one user to another one, which leads to the friendship between the two users. As the saying goes, "birds of a feather flock together". As a result, for any user in social networks, he or she is more likely to be a friend of those who have larger relevance, "a feather", with him or her.

In FR, most of the current researches employed single or few information sources, like friendship [16], user profile [17] and trust [11, 13, 18, 19]. Nevertheless, in social networks, being friends could be affected by many factors, including user personal features, network structure, social relations and interactions among users. With the consideration of applying user personal features for FR, they are seldom used by itself, commonly combined with other information like friendships [17, 20, 21]. Regarding the network structure, the degree-based influence analysis among network nodes is more frequently estimated for FR. However, the existing methods like analogous Pagerank algorithms are relatively time consuming [22], and some influence computation did not embody true influences of the user. With regard to social relations and interactions, trust that extracted from them are widely leveraged to recommend friends [18, 23]. Notwithstanding, there are still challenging issues in the traditional methods, such as binary measurements with 1 trust and 0 distrust, few transmission layers, without considering interactions and so on. In addition, researchers also found that the binary measurement was not reasonable. To avoid these problems, they have built a nonbinary trust degree based on two kinds of ontologies and fuzzy linguistic modeling [24]. However, the nonbinary trust degree is still unsuitable for our work. The reason is that our trust measurement utilizes not only the friendship but also the interactions among users. Currently, some research work combined social relationships with other information including comment similarity, tag similarity and "like" similarity to model user relationship strength for FR in Instagram [25]. Unfortunately, they missed crucial information like location and their proposal, which did not embody mutual effects among different relations. More importantly, most of these FR approaches are lack of comprehensive strategy for fusing multiple key factors.

To solve these challenging research issues, depending on the multi-source information derived from the three aspects, in this paper we propose a scalable fusion framework for FR. First, we widely investigated and analyzed crucial information sources that may influence the selection of user's friends. Second, we propose a novel unified framework for FR with a fusion algorithm based on D-S theory, taking into account all the desired information sources with corresponding Basic Probability Assignment (BPA) functions. The resulting relevance among users after fusing these factors is viewed as the basis of recommending friends by *top-k* idea. Experimental results on the real-world dataset of Tencent Weibo demonstrate that our proposed recommendation framework can significantly improve the effectiveness compared to the traditional FR methods in terms of *Precision, Recall, NDCG, MRR* and *MAP* evaluation metrics. We validate from the experiments that information fusion of multiple sources is of vital importance and indispensable for FR in social networks. The contributions of this paper can be summarized as below.

- Proposing a scalable fusion FR framework based on D-S evidence theory which reflects the minimal conflicts among evidences. The recommendation framework can be scalable regardless of the number of evidences.
- To better improve the recommendation performance, we optimize the original D-S evidence theory by combining importance degree and reliability of evidence.
- Modifying the measurement methods by quantifying the relevance among users with several indices, including influence, direct trust and indirect trust.

- Conducting extensive experiments on real-world dataset of Tencent Weibo social network to demonstrate the effectiveness of fusing multi-source information for FR. Our method recommends both acquaintances and like-minded friends.

The remainder of the paper is organized as follows. We review the related work in Sect. 2 and formalize the FR problem in Sect. 3. The scalable unified framework and the algorithm for FR are presented by original and revised D-S evidence fusion theory based on multi-source information in Sect. 4. The experimental results are analyzed and illustrated in Sect. 5. Finally, we conclude the paper and discuss the future work in Sect. 6.

2 Related work

In recent years, social recommendations have attracted many researchers in social networks. As a significant part, FR mainly focuses on distinct information sources such as user interests and friendships, to find out some new or potential friends for a user. In this part of our paper, we will review the related work in three aspects in accordance with the needs of our work, i.e., user features, network structure and trust that reflects user's social friendships and behaviors.

2.1 Friend recommendation based on user features

User as the key object in social networks has many features, i.e., demographic features included in user profile. User's features are very significant, which could be leveraged to help recommend friends to a target user. User's features are generally combined with other information like user's interests and common neighbors. In the early days, Pazzani et al. [20] utilized user's demographic information to identify the types of users by the age, gender, education and so on. Said et al. [21] had made a comparison in recommendation results when using different demographic features including age, location and gender. These features are used to select high quality neighbors in collaborative filtering recommendation. The corresponding experimental results showed the positive impacts on recommendation.

Tang et al. [18] have addressed user interests, interactions and user features like a nickname, gender and location. The authors have computed the linear combination similarities between users and leveraged them to recommend friends for target users in micro-blog scenario. By utilizing both user feature similarity and interaction intensity, Agarwal et al. [17] have put forward a collaborative filtering framework for FR in social networks. The authors took gender, hometown, religion, educational status etc. into account and learned the weights of features by Genetic Algorithm. The experiment conducted on a synthetic dataset showed the effectiveness of results in the case of considering both user's features and interactions. Zhang et al. [26] introduced a FR system using the user's total features which is based on the Law of Total Probability, in which the probabilities of features are obtained by the statistical results according to the information of user's friends and friends' friends. Compared to other methods including Common-Neighbors, Adamic/Adar and Jaccard coefficient, the authors found that the method based on user's features performs is better on total, especially when the number of user's friends is less than 100.

In terms of the existing researches, though the users with similar features are not indispensable to be friends, the features of a user can be mostly positive role of recommending friends. Due to the relative difficulties to capture the full information of features, in this paper, we not only leverage the user's features, but also consider the reliability of feature information as a coordination factor when fusing it with other information sources.

2.2 Friend recommendation based on network structure

The FR problem can be viewed as link prediction in social networks in essence. Link prediction [27] attempts to estimate the likehood of existence of the link between two nodes by network structure, which have no link in current (or link maybe appear in the future). Link prediction has been widely addressed in complex network [27–30]. For many networks, such as biological networks like protein–protein interaction networks, food webs, science cooperation networks and social networks, detecting links is extremely significant since blindly checking each possible relation or interaction is not realistic. In pure complex network research, link prediction mainly makes full use of the information of network structure features such as degree and path. The mainstream algorithms are classified into three categories [27], namely similarity-based algorithms, maximum likelihood methods and relational modes which are based on probability. These algorithms provide some references for FR which can help users to find new friends and enhance their loyalties to the web sites related to social network. In FR system, the methods also called graph-based approaches by using network structure mainly comprise degree-based methods and path-based methods.

The relatively early and widely used degree-based method is FoF [31] which manifests two users share a lot of common friends who will be more likely to become friends in the future. The famous application of the FR system is on Facebook, which suggests a list of "people you may know" [26]. The Jaccard coefficient with Adamic and Adar (AA) methods [16] are two improved variants based on FoF, in which the number of common friends of users are utilized to estimate their similarities. Liben-Nowel et al. [32] are the first who discussed these related degree-based FR methods in a given social network. On the other hand, the information and algorithms related to degree are also combined with other information for recommending friends like user interests [18, 33] and content spreading [34].

The main aim of path-based method is to find the shortest path from a target user to a potential with the maximum probability for linkage, which is called analogous to Pagerank algorithm [35] which was proposed by Brin and Page. Random Walk algorithm is the most used with local or global structural features of social networks, which is based on Markov Chain Model and could calculate the steady-probability matrix of linkages [30, 36, 37]. Random walk algorithm manifests relatively promising performance [30, 37], especially local random walk method [30]. Generally, random walk algorithm is time consuming, and local ones like LRW and SRW [30] overcome it to some extent by sacrificing a very little accuracy sometimes. However there are two problems need to be verified in LRW and SRW, which are the number of walking steps and the times of superposing.

The information related to the network structure is greatly significant for recommending friends. But relying on this information alone is insufficient. In other words, this information needs to be combined with other information.

2.3 Friend recommendation based on trust in social network

In our daily life, people often resort to their friends for some suggestions of doing something like purchasing and travelling, which shows their trusts on their friends. So does in social networks. Currently, trust-based FR mainly leverages the information of friendships [19] and interactions [23, 38] which are significant social features. In most cases, trust information is utilized to alleviate the sparsity of user-item matrix and enhance neighborhood set in calculation [39, 40], while few researches directly utilized it to recommend friends. Due to no explicit trust information, the authors in [41] have calculated the trust value by rating difference between two users and described the trust propagation, which were combined with user feature information to generate FR for a target user in their work. Ma et al. [42] defined an un-weighted trust network according to the relationships between the users and the trust value of each direct friend of a user is equally assigned in terms of his/her neighbor number. The authors have combined the trust information, which was used to divide the communities, with the user's interested topics to recommend friends with high quality. Agarwal and Bharadwaj [43] argued trust was a subjective expectation a partner had about another's future behavior which is based on the history of their encounters, and reputation was a user's character or standing, which both affected the friend selection and should be considered together to recommend friends. The issues have been verified in their experimental results.

The trust information is useful to improve the performance of recommendation in terms of existing researches. However, most of these researches are considered the trust as binary value 1 for trust and 0 for distrust. While in reality it is a fuzzy value between 0 and 1, not absolute trust or distrust, just like the gradual trust proposed by Victor [44]. The existing methods are using trust information depend mainly on the friendships and their propagations. In this work, we have verified that if the trust information is enhanced by more other information like user's interests, interactions and influence, the better performance could be obtained.

2.4 Problem analysis and motivation based on related work

Recently, some researchers have achieved some notable successes in FR by distinct information sources. However, they have focused on either single information source or simple linear combination with few ones [18]. In fact, being friends is complicated and influenced by many factors such as profiles, interests, interactions, locations, social influence and network structure. For instance, if user B is with high similarity to user A based on their profiles, and they have different interests, then B will be not suitable to be recommended to A. As another example, in the same cases for two users B and C, if B is with more influence than C, then B is more likely to be recommended to A. Again, in the neighborhood-based model of FR with 2-hop scheme in social networks neglects the multi-level propagation of friendships. In this paper, we have carefully selected the multiple information sources which generally, play roles all together.

As we know, the formation of friendships are reflecting the relevance among users is either determined by one or two dominant factors represented varying information sources or the result of multiple main factors gaming. Similar to the intuitive rules that was proposed in [15], all the kinds of factors contribute to the total user relevance. For instance, two users sharing more features of profiles or interests have greater relevance. While, two users with more interactions are closer and relevant. Hence, the relevance of a target user on other users with no links between them is one of the most crucial metrics for predicting potential linkage of them. Which the more relevant is the more probable. In this paper, We have made great efforts to obtain total relevance on target user for recommending friends by fusing multiple information sources which is based on the scalable D-S evidence theory.

Table 1 The factors namely information sources influencing relevance between two user nodes

Factors/information sources	Category	Usage	Linked users	Potential Users
Profile	PF	Similarity	Direct	Direct
Location	PF	Similarity	Direct	Direct
Interest	PF	Similarity	Direct	Direct
User(node) position	NF	Relevance	Direct	Direct
Degree-based influence	NF	Attraction	Direct	Direct
Common friends	SF	Trust	Direct	Indirect
Interactions	SF	Trust	Direct	Indirect

PF, NF and SF denote personal features, network structure features and social features of nodes respectively

3 Problem formalization

According to the aforementioned analysis, social network is a heterogeneous network, in which there are several types of nodes and multiple relations among nodes. Given a social network, we denote it by $G = (V, E)$, V is the set of all kinds of nodes, E is the set of all edges generated by all relations. V comprises several types of nodes including user node and related feature nodes, namely $V = \{V_j^{(i)} | 1 \le i \le K_V, 1 \le j \le k_i\}$, and there are multiple relations in E which is represented by $E = \{E_j^{(i)} | 1 \le i \le K_E, 1 \le j \le k_i\}$, where i denotes type of nodes or edges, j denotes the ordinal of the nodes of $V^{(i)}$ or links/edges of $E^{(i)}$. G is a directed social graph in which some relations are directed like retweet, review and others are undirected relations. The undirected relations are converted into bidirectional relations composed of two directed edges, like similarities based on interests. Therefore in our work, the resulting total relevance of a user on another one is also directed, namely asymmetric. Assume that the first type of node is user node, the rest ones are item nodes used to extract user interest, feature nodes like profile nodes and location nodes. Here we view location information as one single class of nodes other than a feature of user profile, since the location is more important for FR compared to other features in user profile, especially in LBSN [45].

Hence, the relevance-based FR task can be expressed as: Given a user node v in $V^{(1)}$, find out a ranked list of some user nodes which could be connected to v and ranked with the relevance on v by descending order. Certainly, the existing linked nodes of v are removed in recommendation list.

The following Table 1 presents the factors, i.e., the information sources which affect the relevance between two users. They are categorized into three classes: personal features for calculating similarity, network structure features for computing structure relevance and social features for measuring trust, all of which form the total relevance together in FR.

Now given user nodes v, v' and v'', according to the following analysis, node v is more likely to be a friend with node v' rather than node v''.

1. *Profile* Node v' shares more same personal information and therefore has high similarity with node v than node v'', v is more likely to be friend of v' due to its greater relevance on v', which is generated by profile similarity. Profile features reflect the homophily [46] which represents the trend of one user links to another similar user.
2. *Location* According to [45], generally, users always give more trust to the people around them than those far away from them. For instance, it could be more likely for the persons in same campus or city to be friends. That is to say closer geographic location provides higher probability to be friends.
3. *Interest* Two users with more similar interests can be more likely to be friends, since they always have the shared topics that could be discussed. More relevance could be generated for them based on the similar interests.
4. *Position* Refer to the position of node in social network. The potential friend in social networks may be located close to each other. For example, node v' is 2-hop from node v, while node v'' is more than 4-hop away from node v, then v is more likely to be friend of v' than v''.
5. *Social influence* If user v' has a large number of fans and exceeds the ones of user v'', it can be more attractive than v'', and v is more likely to connect with v'.
6. *Common friendship* Embody the overlapping degree of friends between two users, the more friends they share, the more likely they could be friends. e.g., v and v' share more than 50 common friends, but v and v'' have only two common friends.
7. *Interactions* Like review, thumb up, retweet etc., also embody the trust between two users, the higher frequency and more quantity of interactions of them, then the greater trust between them.

The trust generated by common friendship and interactions can be transmitted in social network. Certainly, it decays a little in each time of transmitting. In this paper, we limit the times of transmitting less than 6 in terms of "Six Degrees of Separation" theory [47] which is a social theory and suggests that everyone is considered to be six or fewer steps away from others.

The above analysis is based on the single factor which denotes one kind of information source. But it maybe not embodies the reality in social. Take social influence for example, if v' has more social influence than v'', but shares less even no interests with v comparing to v'', then an interesting result is that v could be more likely to connect v'' other than v'. Therefore, it is very necessary to make a compound considering for all the factors, i.e., fusing them in terms of some rules. In this paper, we intend to fuse all the factors by the scalable D-S evidence theory based on the *Minimum Conflict Principle* [48, 49] to measure the total relevance between two users for FR.

For a target user in a social network, the user nodes are categorized into two classes, namely linked nodes and potential nodes that could be linked with target user node in the future. In Table 1, the social information sources of Common Friends and Interactions together generate the trust for linked users and the trust is transmitted along with some paths to other unlinked users (or potential users). For each linked node, we can directly measure its total relevance on the target node by these multiple information sources, while for each potential node, its total relevance on the target node can be directly measure by the former five information sources and indirectly by the last two information sources where the trust among users are generated and be transmitted. The more total relevance a potential node has, the more likely it could be linked to the target node in the future. That is to say the user with more relevance will be more possible to be the target user's friend.

4 FR framework based on multi-source information fusion

4.1 FR framework

In social network, especially in heterogeneous one, there exists abundant information like the factors in Table 1, which affects user's selection of friends. For instance showed in Fig. 1 in which other nodes and relations are omitted except user nodes and friendships, assume *Mary* is the common friend of *Alice, Bob* and *Carol*, but *Bob* and *Carol* are not friends of *Alice*. *Alice* and *Bob* are close in age, graduated from same university and have similar interests, then *Alice* is more likely to be friend with *Bob* than *Carol* who is older, from different university and has no or few common interests

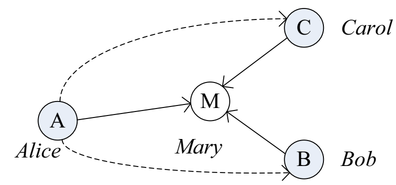


Fig. 1 A simplified social network sample with only user nodes and friendships. The dashed lines denote no linkages from A to B and C. It demonstrates *Bob* and *Carol* who will be more likely selected by *Alice* as her friend

with *Alice*. Therefore, it is indispensable to take multiple information sources together into account to measure the total relevance among users in FR. Also, analyzing the factors of Table 1, FR in social networks is influenced by all these information sources. Any individual factor among them is not enough to be leveraged to generate FR. Obviously, it is an uncertainty problem and needs to fuse all of the information sources to build the solution. D-S theory is a classic and famous inference framework about uncertainty information fusion problem. Hence, we propose a scalable FR framework based on the D-S evidence theory showed in Fig. 2 in this paper. When needed, it can be easily extended according to the number of information sources. D-S evidence theory proposed by Professor Dempster is a kind of extension of Bayesian method [50]. But its demands are less serious than Bayesian method. It depends on the calculation of belief function and combination rules based on the strength of conflicts between evidences. D-S theory reduces the uncertainty by enhancing the complements between evidences to build a new kind of information for solving the target problem, which cannot be archived by single information source. Therefore, we utilize D-S theory as our fusion proposal to solve the FR problem in social networks. To the best of our knowledge, we are the first to apply the D-S theory into the recommendation field.

As showed in Table 1, the information sources are classified into three categories which are personal features for computing user similarity, network structure features for measuring node relevance (or structure relevance) and social features for the measurement of trust that could be transmitted along with some paths. Here, we carefully select k ($k = 7$) information sources showed in Table 1 from the three categories as the evidences supporting the total relevance of target user to other potential user. Each type of evidence in D-S provides the related belief measurement of relevance between two user nodes. All the belief measurements contribute together to the total relevance by the combination in terms of *Minimum Conflict Principle* in D-S theory [48, 50]. Finally, the total relevance is leveraged to generate FR in accordance with *top-k* idea.

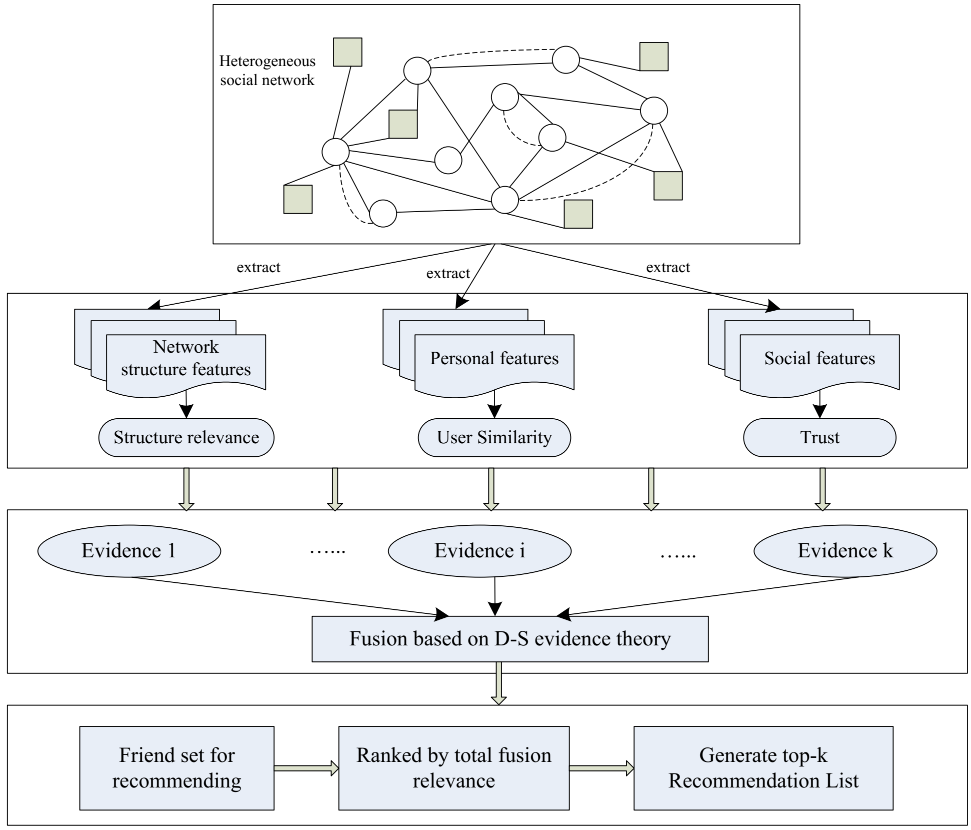


Fig. 2 The framework for FR by multi-sources based on D-S evidence theory. Different relations are denoted by varying lines like solid and dashed lines. Circle denotes user node and square represents abstract feature node which is not detailedly categorized in this figure

4.2 Multi-source information fusion based on D-S evidence theory

4.2.1 D-S theory in FR

D-S evidence theory [50], also called belief function theory, was proposed by Dempster and developed by his student Shafer [48], which is widely applied in uncertainty reasoning and information fusion [49, 51, 52]. D-S evidence theory is an axiom system like probabilistic system but with more weak conditions and more flexible in the case of distinguishing uncertainty and evidences collecting. The basic and initial D-S theory can be referred to the literatures of [48, 50]. In our work, we first use basic D-S theory to define our proposal and then discuss its improvements. Applying D-S theory for information fusion, commonly three steps are needed. First, determine the Frame of Discernment of the specific problem and evidence sources. Then design the Basic Probability Assignment (BPA) of each evidence. Lastly, combine all the BPAs by some rules.

Firstly, in our FR problem, given H_1 represents the hypotheses proposition that two users are relevant with some degree, and H_2 denotes the contrary proposition, namely irrelevance. H_1 and H_2 are mutually exclusive and exhaustive and form the Frame of Discernment [48, 49] $\Theta = \{H_1, H_2, \dots, H_N\}$ which is the universe of discourse with N (here $N=2$) finite elements, let $P(\Theta)$ be the power set of Θ containing all possible subsets of Θ, and $|P(\Theta)| = 2^N - 1$. All the propositions in frame of discernment needs related evidences to support them. Here, we analyze the social networks and carefully select seven evidences sources showed in Table 2. The details of each evidence source have been demonstrated in subsequent sections.

Secondly, under the frame of discernment Θ, define the function of m [49] as follows:

$$m : P(\Theta) \to [0, 1],$$

where $m(\emptyset) = 0$, $\sum_{A \subseteq \Theta} m(A) = 1$, m is utilized to express the belief and plausibility measures of propositions in Θ, the function is called *Basic Probability Assignment or mass function* [48, 50].

Where $\emptyset$ represents null set, and$A \subseteq \Theta$, the value $m(A)$ of $A$is assigned to A by the Basic Probability Assignment function. $m(A)$ manifests the credibility of supporting A. The detailed form of BPA depends on the specific problem and has no fixed formula, commonly determined by the fuzzy system theory. Two important concepts for understanding D-S theory are defined as follows:

Definition 1 (*Focal element* [48, 50]) For any $A \in P(\Theta)$, if $m(A) >\ 0$, A is the focal element of m.

Table 2 The corresponding BPA of each evidence source, $m_i(H_2) = 1 - m_i(H_1)$ due to the mutex feature in the Frame of Discernment Θ

Evidence sources	BPA of relevance	BPA of irrelevance
Profile	$m_1(H_1)$	$m_1(H_2)$
Location	$m_2(H_1)$	$m_2(H_2)$
Interests	$m_3(H_1)$	$m_3(H_2)$
Node position	$m_4(H_1)$	$m_4(H_2)$
Degree-based influence	$m_5(H_1)$	$m_5(H_2)$
Common friends and interactions	$m_6(H_1)$ for direct trust, $m_7(H_1)$ for indirect trust	$m_6(H_2)$ for direct distrust, $m_7(H_2)$ for indirect distrust

Definition 2 (*Belief function* [48, 50]) Given the frame of discernment Θ and corresponding BPA m, then belief function is denoted as:

$$Bel(A) = \sum_{B \subseteq A} m(B), \quad \forall A \subseteq \Theta,$$

where B is any subset of A, $Bel(A)$ represents the total belief degree of A on Θ. The belief function belongs to $Bel{:}2^A \rightarrow [0, 1]$. For instance, assume A contains two elements of X_1 and X_2, the belief function $Bel(A) = m(X_1) + m(X_2) + m(\{X_1, X_2\})$. In our problem, $Bel(A) = m(A)$ due to single element in A which is either relevance or irrelevance from Θ, and the combination of relevance and irrelevance is impossible.

Lastly, the fusion based on D-S theory is conducted by the following combination rules, also called orthogonality rules:

$$m(C) = m_i(A) \oplus m_j(B) = \begin{cases} 0 & A \cap B = \emptyset \\ \frac{\sum_{A \cap B = C, \forall A, B \subseteq \Theta} m_i(A) m_j(B)}{1 - K_{i,j}} & A \cap B \neq \emptyset, \end{cases} \tag{1}$$

where A and B denote focal elements respectively, $m_i(A)$ of evidence i on A and $m_j(B)$ of evidence j on B are combined to yield a new $m(C)$ which manifests the information fusion of two evidences i and j on the frame of discernmentΘ. $K_{i,j}$ denotes the conflict coefficient defined by the following formula and ranges in [0, 53]:

$$K_{i,j} = \sum_{X \cap Y = \emptyset, \forall X, Y \subseteq \Theta} m_i(X) m_j(Y). \tag{2}$$

$K_{i,j}$ is a normalization factor and measures the conflict degree between m_i and m_j. $K_{i,j} = 0$ and $K_{i,j}$=1 correspond to the no conflict and complete contradiction, respectively.

4.2.2 Improved D-S theory

Although basic D-S theory provides a good idea of information fusion, it has some drawbacks which occasionally results in illogical combination results such as weak evidence obtains strong supporting [54]. Therefore, in order to avoid the drawbacks, we propose the related improved method by considering the reliability and importance of evidence. The reliability refers to the integrity degree of evidence containing information, which depends on specific evidence and is demonstrated in following Sect. 4.6. The importance of evidence manifests the how much the evidence is important in all evidences, which is calculated by building a fuzzy similarity matrix M_E. Assume there are K_E evidence sources, namely K_E evidences, then a $K_E \times K_E$ fuzzy similarity matrix $(M)_{K_E \times K_E}$ is built as follows:

$$M = \begin{bmatrix} S(\mathrm{E}_1,\mathrm{E}_1) & S(\mathrm{E}_1,\mathrm{E}_2) & \cdots & S(\mathrm{E}_1,\mathrm{E}_{K_E}) \\ S(\mathrm{E}_2,\mathrm{E}_1) & S(\mathrm{E}_2,\mathrm{E}_2) & \cdots & S(\mathrm{E}_2,\mathrm{E}_{K_E}) \\ \vdots & \vdots & & \vdots \\ S(\mathrm{E}_{K_E},\mathrm{E}_1) & S(\mathrm{E}_{K_E},\mathrm{E}_2) & \cdots & S(\mathrm{E}_{K_E},\mathrm{E}_{K_E}) \end{bmatrix}_{K_E \times K_E}.$$

For clarity, M can be simplified as:

$$M = \begin{bmatrix} 1 & S_{1,2} & \cdots & S_{1,K_\mathrm{E}} \\ S_{2,1} & 1 & \cdots & S_{2,K_\mathrm{E}} \\ \vdots & \vdots & & \vdots \\ S_{K_\mathrm{E},1} & S_{K_\mathrm{E},2} & \cdots & 1 \end{bmatrix}, \tag{3}$$

where the diagonal elements of M equal to 1 and $S_{i,j} = S_{j,i}$. $S_{i,j}$ and $S_{j,i}$ denote the similarities between evidence i and j, which can be calculated by the evidence distance between evidences [54]. Evidence i is absolutely similar to itself, i.e., $S_{i,i} = 1$. Obviously, the elements of ith row totally represent the measurement of supporting ith evidence by other evidences. Therefore the supporting degree [54] of evidence i is defined as:

$$Sup(E_i) = \sum_{j=1, j \neq i}^{K} S_{i,j}(E_i, E_j) = \sum_{j=1, j \neq i}^{K} S_{i,j}. \tag{4}$$

The higher similarity one evidence is to others, the bigger supporting degree and higher credibility it has, vice versa.

Hence, the importance degree of evidence can be obtained by normalizing the supporting degree.

$$I(E_i) = \frac{Sup(E_i)}{\sum_{j=1}^{K} Sup(E_j)}. \tag{5}$$

Assume that v_i^{R} is reliability of evidence i, the modified BPA is denoted as:

$$m_i^{\mathrm{N}}(X_j) = \frac{1}{K_{\mathrm{c}}} v_i^{\mathrm{R}} * I(E_i) * m_i(X_j), \tag{6}$$

where $K_{\mathrm{c}} = \sum_{i=1}^{K} v_i^{\mathrm{R}} * I(E_i)$ is coordination factor, X_j is focal element. Therefore, the improved BPA is utilized to revise original D-S theory.

4.3 BPAs based on information source of user features

4.3.1 BPA of user profile

Generally, two users with similar profiles are relatively easy to form some relation. For instance, *Alice* and *Bob* are adjacent in age and study in same university. They could generate a certain relationship in social network since they can always discover some common things in some aspects like life, study and interest. They could be more likely to be friends. Certainly, it is not absolutely. But we emphasize the likelihood being friends. Therefore, profile similarity between users is very critical to measure the relevance of them. Many features could be applied to measure the similarity between users. Due to deriving the full features of users with much difficulty, Agarwal and Bharadwaj [17, 43] simulated the similarity calculation based on users' features in a synthetic dataset and obtain good effectiveness. But in social networks environment for preserving privacy, many features of users are missing. However, we still argue that the role of user features is significant in measuring the similarity between users. Therefore, we believe that more wisely chosen set of features of user profile would be an indicative of high measurement quality of relevance. Here, five prominent components of user profile that we carefully considered comprise age, gender, study organization (like university, institute or high school), speciality and career. Study organization and speciality reflect the education feature and career embodies employment feature, which are two mostly powerful features for forming links between users according to previous work [55]. We separately measure each component similarity between two users and then combine them since they are not real-valued variables. It is straight that the importance of each component varies. We suggest a weighted scheme that reflects the importance of them and the weights could be learned by using real-valued genetic algorithm (GA) proposed in [17] in an offline learning process. The total profile similarity between target user u and potential user v, $sim_{\mathrm{P}}(u, v)$, is calculated by combing all the similarities of 5 components as follows:

$$sim_{\mathrm{P}}(u, v) = \vec{S}_{\mathrm{P}} \vec{W}_{\mathrm{P}}^{T}, \tag{7}$$

where $\vec{\mathrm{W}}_{\mathrm{P}} = (\omega_1, \omega_2, \omega_3, \omega_4, \omega_5)$ denotes weight vector and $\sum \omega_i = 1$, and five similarities of the components in profile form the vector of $\vec{\mathrm{S}}_{\mathrm{P}} = (s(u_{A1}, v_{A1}), s(u_{A2}, v_{A2}), s(u_{A3}, v_{A3}), s(u_{A4}, v_{A4}), s(u_{A5}, v_{A5}))$. The similarity of ith component of profile between user u and user v, $s(u_{Ai}, v_{Ai})$, is determined as follows:

1. *Component A1: age* Generally, age is between 0 and 100. For a nature person registering in social networks, we assume he or she is greater than 10 years old. Otherwise, the information is unreliable and neglected. We can utilize the transformation related to integer age distance as the similarity.

$$s(u_{A1}, v_{A1}) = \frac{1}{1 + \log(1 + d(u_{A1}, v_{A1}))}, \tag{8}$$

where $d(u_{A1}, v_{A1})$ is the age distance equal to the absolute value of age difference. The purpose of using logarithm is to reduce the velocity of decrease of similarity.

2. *Component A2: gender* Gender is composed of male and female. It is not a good selection for similarity equal to 1 with same gender, otherwise equal to 0. We argue it is reasonable for user in social network to select friends based on the gender inclination. That is to say male users maybe not always select male friends; sometimes on the contrary they are inclined to select female friends. Hence, we design a gender vector, $\vec{G} = (\mathrm{g}_1, \mathrm{g}_2, \mathrm{g}_3)$, to measure the gender similarity, in which the components denote gender, male friend ratio and female friend ratio. The ratio of male or female friends of user u can be computed by using corresponding friend number to divide the total number of user $u's$ friends. Then we can take advantage of cosine similarity to calculate $s(u_{A2}, v_{A2})$ between user u and user v based on the gender vector $\vec{G}$.
3. *Component A3: study organization* In this paper it refers to the entity of user accepting the education, which is categorized into three classes (1) higher education entity like Shanghai University, (2) secondary education entity like No.1 high school of Shanghai, (3) and training institution. For two users of u and v, if the study organizations are same, $s(u_{A3}, v_{A3})$ equals to 1; if they are not same organizations but belong to same category of

education entity, $s(u_{A3}, v_{A3})$ equals to 0.5; otherwise they equals to 0.

4. *Component A4: speciality* Two users with same or similar speciality will have more topics to talk about. Here we compute the semantic distance of speciality directory tree (could be built from BaiduBaike[7] WordNet[8] or WikiPedia[9] determined by specific application scenario) to measure the speciality similarity. The smaller the distance, the greater the similarity. Clearly, the similarity between two specialties is related to their depths and the length of path in the tree. Therefore, we leverage the method [56] to calculate the speciality similarity, which is effective for measuring the similarity between two nodes in a vocabulary tree when $\alpha = 0.2$ and $\beta = 0.45$, as follows:

$$s(u_{A4}, v_{A4}) = \begin{cases} 1 & u_{A4} = v_{A4} \\ e^{-\alpha l} * \frac{e^{\beta h} - e^{-\beta h}}{e^{\beta h} + e^{-\beta h}} & u_{A4} \neq v_{A4}, \end{cases} \tag{9}$$

where l denotes the length of path from one node to another one, and h represents the depth of the first common ancestor node of two nodes in the speciality directory tree.

5. *Component A5: career* Similar to speciality, we build a career tree by BaiduBaike and adopt the same calculation method as speciality similarity.

So far, we deliberately have fined each component similarity in terms of corresponding components from profiles of two users. Then we can further use the real-valued GA [17] to learn the weights of the component similarities in profile similarity $sim_{\mathrm{P}}(u, v)$. The profile relevance between two users is proportional to their similarity which is normalized. Thus for simplicity, the BPA function of profile evidence in this paper is as follows:

$$m_1(H_1) = sim_{\mathrm{P}}(u, v). \tag{10}$$

4.3.2 BPA of location

The abundant information accumulated in social networks enables a variety of applications like recommendations of users and media. Location is one of the most significant components defining a user's context and affects the user's selection of friends [45]. And users who live close to each other are more likely to be friends [57], which is similar to the case in real life. Liben-Nowell et al. [58] found that more than 2/3 of friendships are generated by users' locations in an online social network. The proximity of location between users facilitate not only being friends online but also taking part in activities offline [59]. Scellato et al. [60] demonstrated the fact of about 40% connections in social networks is within 100 km by analyzing the data from Foursquare,[10] a location-based social network, in which many users shared the offline activities. Wu et al. [61] proposed an approach recommending friends with similar location preference, demonstrated the rationality and verified its effectiveness by conducting experiments in a real dataset called Gowalla, from which we also can understand the location information is significant for FR in social networks. Thus we also leverage the user location information in our work. Here, for simplicity we consider only the current place of residence or hometown which are viewed as the same meaning and represented by a unified phrase of *User Location or Location* in our work. As we all know, the hierarchical location information [45] spans multiple scales with varying granularities. For instance, a district belongs to a city; a city belongs to a province or state, and so on. Currently, using the hierarchical tree-based method to measure the similarities of locations is the mainstream method [45, 62–64], in which the similarity is computed by dynamic location sequences in location-based social networks. In our paper, user location information which does not focus on trace is relatively static and not frequently changed. So the similarity measurement based on tree nodes is suitable for location similarity computation. The hierarchical location tree is analogue to a simplified ontology tree. Hence, we leverage the method of Eq. (11) which is analogous and refers to Eq. (9) to measure the similarity between users based on the location information by the Chinese Location Library, a hierarchical location tree, built according to the data from Wiki.[11] Clearly, the location library can be extended to other one in terms of the corresponding country's geographic information. The user similarity based on location is as follows:

$$sim_{\mathrm{L}}(u, v) = \begin{cases} 1 & L_u = L_v \\ e^{-\alpha l} * \frac{e^{\beta h} - e^{-\beta h}}{e^{\beta h} + e^{-\beta h}} & L_u \neq L_v, \end{cases} \tag{11}$$

where L_u and L_v are user locations of u and v respectively, other parameters are same as the ones in Eq. (9).

Compared to a large volume of users in social networks, the number of a target user's friends is very small. Even in the same location for two users, it is still difficult to determine whether they are friends [45], that is to say only considering the location information source is insufficient, which needs to be combined with other information source. Thus, we argue information fusion is a wise selection for

7 http://baike.baidu.com/.

8 http://wordnet.princeton.edu/.

9 https://www.wikipedia.org/.

10 https://foursquare.com/.

11 https://en.wikipedia.org/wiki/China#Geography.

FR. The BPA function based on the location similarity is calculated as follows:

$$m_2(H_1) = sim_{\mathrm{L}}(u, v). \tag{12}$$

4.3.3 BPA of user interest

In social networks users can not only post or share some information but also interact mutually, which reflect users' interests to some extent. According to the experience of life and social selection theory, user is more easily to select like-minded persons with similar interests as his/her friends. Obviously, the greater interest similarity, the more relevance between two users. Thus interest similarity which utilizes the technique of text analytics [65] is undoubtedly significant for FR of a user in social networks. Here we focus on the social networks with regard to the information sharing like Twitter, Weibo (Chinese micro-blog) and constructing user interest ontology which is proposed as a popular definition of an explicit specification of a conceptual model by Gruber [66]. Many researchers have studied the construction of domain knowledge ontology by Open Project Directory or Wikipedia or Chinese BaiduBaike, which is also applied in this paper. The process of ontology construction [66, 67] is out of the scope of this paper and we take advantage of the method proposed by Zheng [68] to calculate the user interest degree of a subject. We can extract some user interested subjects from user's history information in social networks, which are mapped in domain ontology tree. User interest model could be represented by these subjects and corresponding interest degrees which comprise content interest degree and semantic subject coverage degree.

$$I_u(s) = \frac{2 \times Cid_u(s) \times Sid_u(s)}{Cid_s(s) + Sid_u(s)}, \tag{13}$$

where $Cid_u(s)$ and $Sid_u(s)$ denote content interest degree of user u on subject s and semantic coverage of s, respectively. For a specific subject, the more number of related information which user has posted or shared, and the greater semantic coverage degree of the subject, then the more the user is interested in this subject. Thus, the similarity of two users can be measured by the following Eq. (14):

$$sim_{\mathrm{I}}(u, v) = \frac{\sum_{s \in S_u \cap S_v} (I_u(s) - \bar{I}_u)(I_v(s) - \bar{I}_v)}{\sqrt{\sum_{s \in S_u \cap S_v} (I_u(s) - \bar{I}_u)^2} \sqrt{\sum_{s \in S_u \cap S_v} (I_v(s) - \bar{I}_v)^2}}, \tag{14}$$

where $S_u \cap S_v$ is the intersection set of subjects both u and v. $I_u(s)$ and $\bar{I}_u$ denote the personal interest degree on subject s and average interest degree of user u in user interest model respectively, so do $I_v(s)$ and $\bar{I}_v$ for user v. As we know, the greater similarity, the more likely two users are friends. The relevance between two users is also proportional to their interest similarity. Thus we can get the BPA function of interest similarity as follows:

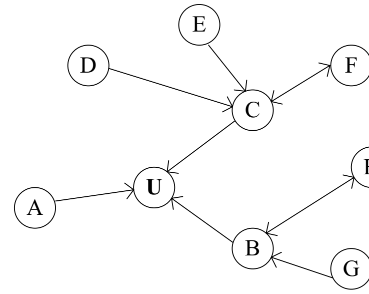


Fig. 3 A local sample of social graph of user u. It omits other relations except follow relation

$$m_3(H_1) = sim_{\mathrm{I}}(u, v). \tag{15}$$

4.4 BPAs based on information source of network structure

4.4.1 BPA of node position

The potential friends are likely to be located close to each other in social networks [15]. For example, for target user u, user u_1 is two steps away from u while user u_2 is five steps apart, then u_1 could be a friend of u with more possibility than u_2. For simplicity, we view the reciprocal of length of the shortest path from target user as the BPA of node position. So:

$$m_4(H_1) = L_{\mathrm{P}}(u, v)^{-1}. \tag{16}$$

4.4.2 BPA of user influence

In terms of the definition of Merriam-Webster dictionary, influence is "the power or capacity of causing an effect in indirect or intangible ways". In this paper, we argue that the influence of a user in social networks refers to the power or degree of resulting in some effect on another user like generating certain action. User influence is measured by several ways such as Pagerank value and the number of followers. According to the empirical analysis on Twitter dataset [69], indegree, retweet and mention are three key aspects affecting user influence. Indegree, on one hand, represents the role of a node according to the social network structure. On the other hand, it denotes the popularity of the corresponding user. But we should understand in social networks some users follow a user due to etiquette such as a polite feedback for the user's following, or just acquaintances in real life. So we intend to find out the true followers based on our opinion of user influence generated by their behaviors in social network. The behaviors we consider comprise retweet, review and mention. What's more, to our best knowledge, user influence is also enhanced by his/her followers. For instance showed in Fig. 3, assume user C

has great influence and follows user U, it is straightforward that U's influence is enhanced by user C, at least not bad. That is, the contribution called indirect influence of a user to its parent user is also important for user influence measurement. Kempe [22] put forward that the child nodes with the depth less than or equal to two are viewed as the most valuable ones for parent node's influence. Thus in this paper, we also refer to this idea in our work when calculating user influence. Considering all the above factors, we proposed a modified method to measure the user influence by Eq. (17):

$$Influence(u_i) = \log\left(\sum_{j=1}^{N_{u_i}}\left(f(u_i,u_j) + \zeta_j * \sum_{k=1}^{N_{u_j}} f(u_j,u_k)\right)\right), \tag{17}$$

where N_{u_i} represent the follower number of user u_i, $f(u_j,u_i)$ denotes an equivalent function of computing the effectiveness of user u_j as a true follower, which is calculated by Eq. (18) in terms of the quantified behaviors on its parent node's shared information.

$f(u_j,u_i)$ is equal to the ratio of the behavior number of user u_j on the shared information of user u_i over the total number of user u_i shared information.

$$f(u_i,u_j) = \frac{Behaviors(u_i,u_j)}{1 + SharedNum(u_i)}. \tag{18}$$

ζ_j is influence coefficient implying the weight or contribution degree of the child node u_j to the target node u_i, which is determined by Eq. (19). As for node u_j, the more influence of itself, the more influence contribution for its friend node u_i, and it would be assigned greater coefficient value.

$$\zeta_j = \frac{f(u_i,u_j)}{\sum_{k=1}^{N_{u_i}} f(u_k,u_j)}. \tag{19}$$

The user influence is composed of two parts, namely the degree of true direct followers and their contributions. Our measurement of influence embodies not only the role of user node's indegree but also the role of interactions between users. The purpose of using logarithm is to prevent the low computation precision in fusion due to the large difference between high influence and low influence.

For the target user, a potential with great influence is more attractive and more likely to be his/her friend. That is to say the selection probability is large. Thus the BPA of user influence is positive proportional to user influence and can be calculated as:

$$m_5(H_1) = K_{Inf} * Influence(u_i), \tag{20}$$

where K_{Inf} is a normalized coefficient which transforms the value of influence into the interval of 0 and 1.

4.5 BPAs based on information source of social features

In this section, we focus on analyzing and modeling the trust for FR in social networks. We first define the concept of trust based on the behaviors, which is computational and easy to be measured.

Definition 3 (*Trust*) According to the idea of Golbeck [19, 70] about trust,in social networks, trust of user A on user B is a certain dependence on B based on a belief or inherent personal opinion that A's future actions(like discussing some topics, favoring a certain activity or even spreading information) A will obtain good positive outcome due to being influenced by B.

The outcome maybe benefit from sharing similar features, interests, experience and so forth between A and B. For example, A trusts B means that A believes or favors B and adopts some actions on something. Moreover, the results of the actions did not generate negative effect on A, but if B posts a mendacious or negative information that makes A uncomfortable, A will decrease the trust on B, and even little by little, A distrusts B.

Similar to reality, in social networks, A trusts B and B trusts C, then A could trust C to some extent. That is to say trust can be transmitted. So trust can be categorized into two parts: one is direct trust and the other is indirect trust [19]. Moreover, A trust B did not mean B must trust A. That means trust is directional and asymmetric. Hence, trust has three properties of transitivity, asymmetry and personalization [19].

Definition 4 (*Direct Trust*) If user A follows user B, then trust measurement of A on B is called direct trust.

Definition 5 (*Indirect Trust*) If user A dose not follows user B, and A can reach to B in less than or equal to 6 steps, then trust measurement of A on B in terms of trust transitivity property is called indirect trust.

Definition 6 (*Trust Transitivity Hop*, $k_l - Hop$) Let $G_T = (U, E_T)$ be the directed network derived from social network G, which is formed by the trust relations among users, U be the set of user nodes, E_T be the set of directed edges of trust. If there are k_l vertexes on the path from user u_i to user u_j, we define it is $k_l - Hop$ of trust transitivity from u_i to u_j and $1 \le k_l \le 6$ which is determined by the "*Six Degrees of Separation*" theory [47].

k_l less than 6 is statistical and average value that one user reaches to any other user in social networks according to the "Six Degrees of Separation" theory. In real social network,

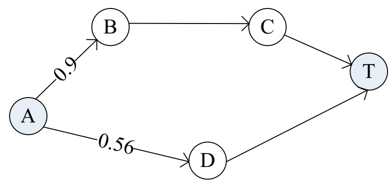


Fig. 4 An example of trust transitivity from node A to node T. There are two paths from A to T. The path of $A \rightarrow B \rightarrow C \rightarrow T$ is 3-hop and another path is 2-hop. The trust of A on T can be calculated with a certain transitive method according to trust transitivity in the two paths. The resulting trust of A on T is the greater value of them

there are also many users to which must be reached by more than six vertexes. We neglect these vertexes since trust transitivity through them decays much more and can be ignored for FR. A simple example of trust transitivity is showed in Fig. 4.

Also, we know that in social networks from one user to another user like in Fig. 4, maybe, there exists more than one path linking them. We should compute the final trust value of user A on user T by aggregating these trust value on different paths by certain mechanism [71]. Here, we propose a new concept "Best Trust Transmission Hop" defined as following Definition 7 to determine the final trust value.

Definition 7 (*Best Trust Transmission Hop*,$k_B - Hop$) Let $T_l = f(u_i, u_j, path_l)$ be the trust measure of user u_i on user u_j by $k_l - Hop$ trust transmission on path l, the best trust transmission $k_B - Hop$ satisfies that T_l is the maximum value [72].

4.5.1 Direct trust

In social networks, the user directional relations of following or being followed form the fans and idols, respectively, which reflect the trust among users. Idols are often called user's friends and trusted by fans. If a user follows another one, the previous work like [13, 16] argues that he or she totally trusts his or her idols. But different from them, we argue that this kind of trust is with some degree other than being unconditional. Again, sometimes one user follows another one who also follows him or her out of etiquette [73]. From the superficial form, they are friends each other, but in fact only user's active following generates the real friend linkage. To refine the trust measurement, we consider not only the friendships but also the interactions among them like retweet, review, which are consistent with and implied in Definitions 3 and 4. Hence, we measure the direct trust of user u on v by the following Eq. (21):

$$T^{\mathrm{D}}(u, v) = \min\left\{ \frac{1}{2} + \frac{N(u, v)}{1 + \sum_{v' \in FS(u)} N(u, v')}, 1 \right\}, \tag{21}$$

where $N(u, v)$ is the number of behaviors of user u on user v, $FS(u)$ is the friend set of user u, which expresses the feature of "Common Friends". Amongst all the friends of user u, the greater the value of $T^{\mathrm{D}}(u, v)$, the more direct trust. The direct trust $T^{\mathrm{D}}(u, v)$ is a decimal value in the interval of 0 and 1, while in the previous work it is equal to 1.

Empirically, user u linked to user v means that there is basic trust, so we set $\frac{1}{2}$ as basis in the measurement of direct trust. The rest part of Eq. (21) manifests the role of user's behaviors which could enhance the trust. And its maximum value is not greater than 1.

4.5.2 Indirect trust

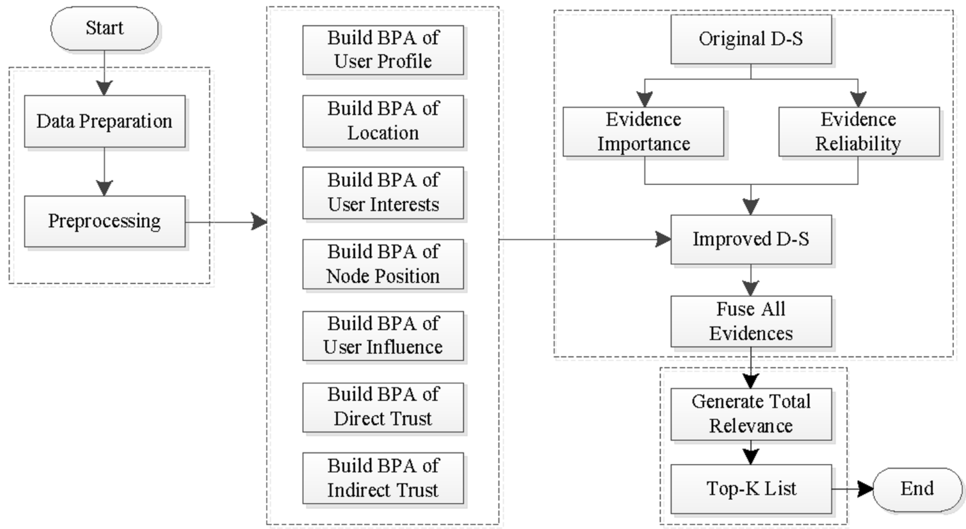
In reality, friends of friend are likely to be friends that proposed by Adamic [16], since direct trust on friends could be transmitted to the next ones. But it will decay in the process of transmitting along with the propagation path. That is to say trust does not transmit without any loss through distance. Due to this, Li [13] etc. proposed a concentric circle model to measure the indirect social trust for transmitting, which manifests good effectiveness. The number of the concentric circle's hierarchy from the center node to the outermost layer nodes is less than six. They viewed integer 1 as the original direct trust value, while our direct trust is a trust degree, a decimal value. Moreover, they aim at finding out the shortest path to the target user (node) in which the indirect trust value is achieved by the transmission value $\frac{1}{n+1}$, while we make efforts to obtain the Best Trust Transmission Hop, $k_B - Hop$, corresponding to the best path to target user on which the indirect trust of the target user is maximum [72] amongst all the paths from the target user to an end user. The length of paths is a critical factor in trust transmission since longer paths generally contain less liable information [74]. Hence, according to the "Six Degree Separation" theory [47], the number of trust transmission hop is less than or equal to 6. If any users are out of 6 hops, we set the trust of A on it is a fixed threshold or a value close to zero.

Assume it is reachable from user u to user v, but v is not u's friend, there exists $K^{\mathrm{P}}_{u,v}$ paths from u to v. In each path, u is root with level equal to 0, and all its direct friends' levels are equal to 1 and so on. Then the indirect trust of u on v can be calculated as:

$$T^{\mathrm{I}}(u, v) = \max_{j=1,\dots,K^P_{u,v}} \prod_{i=0}^{SN(j)-1} T^{\mathrm{D}}(v_i, v_{i+1}), \quad s.t. \ \forall v_i, v_{i+1} \in SN(j),$$
$$1 \le j \le K^P_{u,v}, \quad v_0 = u, v_{|SN(j)|} = v, \tag{22}$$

where j denotes the jth path from u to v, $SN(j)$ represents the set of nodes(users) on the jth path, v_i is ith node, $T^{\mathrm{D}}(v_i, v_{i+1})$ represents the direct trust of v_i on his or her friend v_{i+1}. The

Fig. 5 Fusion method flow of FR based on the improved D-S theory



resulting corresponding path is called the maximum indirect trust path, namely the Best Transmission Hop. And the maximum value is selected as the resulting indirect trust [72] of u on v in terms with the idea of Random Walk.

4.5.3 BPAs related to trust

In our work trust is categorized into two classes, namely direct trust and indirect trust corresponding to user trust on friend and on potential user along with a certain reachable path, respectively. Despite of which one, they manifest the trust degree or connection probability of a user on another one. They are positively proportional to user relevance. Therefore, we can deem the BPA functions of direct trust and indirect trust can be calculated as Eqs. (23) and (24), respectively.

$$m_6(H_1) = T^{\mathrm{D}}(u, v), \tag{23}$$

$$m_7(H_1) = T^{\mathrm{I}}(u, v). \tag{24}$$

4.6 Reliability of BPA function

The total relevance of one user on another one is generated by aforementioned BPA functions derived from corresponding information sources. BPA represents the linkage probability of one user connecting to another one. In an ideal case, if the information derived from its corresponding source is integrated and sufficient, its BPA is reliable. But in realistic cases, it is not always getting the ideal information we need. Therefore, we add a degree of reliability as a coefficient for tuning the value of BPA. The seven BPAs from three classes PF, NF and SF form a reliability vector, $V^{\mathrm{R}} = ({v^{\mathrm{R}}}_1, {v^{\mathrm{R}}}_2, {v^{\mathrm{R}}}_3, {v^{\mathrm{R}}}_4, {v^{\mathrm{R}}}_5, {v^{\mathrm{R}}}_6, {v^{\mathrm{R}}}_7)$, in which the default value of each component is equal to 1 except the components of user profile and location, whose reliabilities depend on the specific cases. In our work, the user profile is composed of five components. If lacking one component, the degree of its reliability will be decreased by 20%. As for location, if it is missing, its BPA value is assigned 0 otherwise 1.

4.7 Friend recommendation

In this section, we present the fusion method flow and algorithm of FR based on the improved D-S evidence theory, which are showed in Fig. 5 and Table 3, respectively.

The flow comprises four parts which are preparing data, building BPAs of evidences, fusing BPAs by improved D-S theory and generating recommendation list. The first part mainly includes data preparation and preprocessing, like collecting dataset and eliminating dirty data. The BPAs of all evidences are built and calculated in the second part. In the third part we focus on the improved D-S theory by the importance and reliability of evidence and fuse all the BPAs. Then the total relevance is obtained to generate the *top-k* recommendation list. BPAs are built by either similarity or relevance between evidences. We mainly proposed and improved some BPAs' calculation methods, which are user influence, direct trust and indirect trust. Some of BPAs refer to the current literatures including user interest similarity, location similarity. Others are experience formulas like age distance, gender similarity, and node position measurement. Due to the drawbacks of original D-S theory discussed in Sect. 4.2, we proposed the improved method by importance and reliability of evidence.

The FR algorithm is the core of our proposal, in which the general steps of the algorithm of are showed in Table 3. Firstly, the BPAs are calculated in line with the corresponding information sources. Secondly, the total fusion relevance of each potential user on target user is computed by the

Table 3 FR algorithm of fusing multi-source information in social networks

Algorithm: FR algorithm based on improved D-S evidence fusion theory
Input: u (target user), U_1 (potential user set), U_2 (friend set of user u), SR (social network), k (*top-k* value)
Output: L_u (the *top-k* FR list for the target user u)
Step 1: for each potential user $v \in U_1, v \notin U_2, v \neq u$ calculate BPA of each evidence source $m_i(\mathrm{H}_1)$ and $m_i(\mathrm{H}_2)$ in line with each information source user v and user u, where i is from 1 to 7.
Step 2: calculate the fuzzy similarity degree between evidences $S_{i,j}$ based on the evidence distance.
Step 3: build the evidence fuzzy similarity matrix $(M)_{7\times7}$ by $S_{i,j}$.
Step 4: calculate the supporting degree $Sup(E_i)$ by M and Eq. (4).
Step 5: calculate the importance degree $I(E_i)$ of each evidence by Eq. (5).
Step 6: determine the evidence reliability vector $V^{\mathrm{R}} = (v_1^{\mathrm{R}}, v_2^{\mathrm{R}}, v_3^{\mathrm{R}}, v_4^{\mathrm{R}}, v_5^{\mathrm{R}}, v_6^{\mathrm{R}}, v_7^{\mathrm{R}})$ based on the information integrity and sufficiency of each evidence source.
Step 7: calculate the total fusion relevance $R_{u,v}$ of user v on user u.
Step 7-1: revise each BPA of each evidence $m_i(\mathrm{H}_1)$ and $m_i(\mathrm{H}_2)$ by evidence importance $I(E_i)$ and reliability v_i^{R}.
Step 7-2: fuse all of the BPAs in terms of the combination rules by Eq. (1) to get the total relevance $R_{u,v}$ of user v on user u.
Step 8: Go to Step 1 until completing the fusion calculation of all the potential users in U_1.
Step 9: return the *top-k* recommendation list L_u based on the total fusion relevance of each potential user on target user u.

improved D-S evidence theory. Lastly, the FR list is obtained based on top-k principle.

5 Experimental analysis

5.1 Dataset

We choose the dataset derived from Tencent microblogging platform[12] which is an online Chinese social networking service analogous to Twitter and launched by Tencent Co. Ltd. Now time, the number of daily active users has exceeded 100 million. Furthermore, Tencent microblogging platform offers a free and open API for developers to gather the global user data. Given the significant advantages above, Tencent microblog is chosen as the experimental platform for the performance evaluation of FR.

To get the dataset and well conduct the experiments, we developed an experiment system, which is implemented by Java 1.8 and Eclipse Neon Release (4.6.0). All the data are stored in MySql database. The OS is Windows 7 (64 bits), RAM 4.00 GB, and Intel(R) Core (TM) i5-5200U CPU 2.20 GHz. We did not conduct the experiments in distributed environment, since we just validate our approach. But it can be paralleled if applied in real application.

[12] http://t.qq.com/.

In the process of collecting the dataset we did not select some users stochastically, since it cannot express accurately the social features of the users. Therefore, we begin from a specified user who is assigned as the initial node of the directed network, and all followers of this user are then added to the network. After that, all the followers of these new members are recruited in the same way, and we repeat this process until the number of nodes satisfies the quantity demand. We collected the information needed, including user features, shared information, social relationships, interaction records and spatio-temporal information. Besides, the individuals who have followers or friends up to more than 1000 are excluded, since they are referred to as stars or friend abusers. Finally, we got a small network with 12,761 users. We collected the two batches of data with the same 12,761 users which are acquired in June and August 2015 for prediction and verification, respectively. The statistics information of the dataset is summarized in Tables 4 and 5.

Table 4 Statistics of social relationships

Statistics	Batch 1		Batch 2	
	Fans per user	Idols per user	Fans per user	Idols per user
Max. num.	193	149	193	179
Min. num.	1	1	1	1
Avg. num.	7.911	6.582	8.056	6.757

Table 5 Statistics of dataset

Statistics	Batch 1	Batch 2
User num.	12,761	12,761
Social edge num.	83,809	86,018
Message num.	569,671	577,925
Average degree	6.57	13.492
Net diameter	17	20
Average path length	5.983	5.963

Table 6 Statistics of preliminary dataset

Item	Statistics
User num.	1615
Social edge num.	2601
Max. num. of fans	84
Min. num. of fans	1
Max. num. of idol	106
Min. num. of idol	1
Message num.	76,236
Average degree	3.221
Net diameter	12
Average path length	4.067

5.2 Evaluation metrics

We use two classic metrics in information retrieval field, namely average precision (AP), average recall (AR) and three popular metrics about the quality measurement of recommendation list, namely the average normalized discounted cumulative gain (A-NDCG), mean reciprocal rank (MRR) and mean average precision (MAP) to measure our approach's performance.

The metrics of precision and recall are often leveraged to evaluate the recommendation performance [75]. We recommend friends for n users in each *top-k* recommendation. The higher value of AP and AR means better performance.

$$\mathrm{AP} = \frac{1}{n}\sum_{i=1}^{n}\frac{N_1^i}{N_1^i + N_2^i}, \tag{25}$$

$$\mathrm{AR} = \frac{1}{n}\sum_{i=1}^{n}\frac{N_1^i}{N_1^i + N_3^i}. \tag{26}$$

The metrics of NDCG, MRR and MAP are leveraged to evaluate the quality of recommendation list [76, 77]. In NDCG the position of each recommendation item (here is friend) is discounted by logarithm function. The higher the NDCG, the better performance the recommendation list. MRR and MAP are also two precision metrics related to the items' positions in recommendation list [77]. We always hope the most related items are recommended to target user and the recommendation list has the highest relevance on target user in total with the most possibilities. In our *top-k* FR, the NDCG, MRR and MAP are defined as follows.

$$\mathrm{NDCG}@k = \frac{1}{Z}\sum_{i=1}^{k}\frac{2^{x_i}-1}{\log(i+1)}, \tag{27}$$

where $Z$is a normalized factor [78], x_i is the flag of item ranked at position i, which, in our binary FR, is equal to 1 when the item is accepted by target user and equal to 0 when the item is rejected by target user. In every experiment, we calculate the value of NDCG of each target user in each *top-k* recommendation. Finally, we compute the average value of all the target users is calculated as the final result of A-NDCG for each *top-k* recommendation.

$$\mathrm{MRR}@k = \frac{1}{k}\sum_{i=1}^{k}\frac{1}{i} * \mathrm{rel}(q_i), \tag{28}$$

$$\mathrm{MAP}@k = \frac{1}{k}\sum_{i=1}^{k}\frac{\mathrm{P}(q_i)}{i} * \mathrm{rel}(q_i), \tag{29}$$

where q_i denotes the ith item in recommendation list, $\mathrm{rel}(q_i)$ and $\mathrm{P}(q_i)$ represent binary relevance function and item ordinal function in the list composed with all the relevant items, respectively. If q_i is relevant to target item, $\mathrm{rel}(q_i)$ is equal to 1, otherwise 0.

5.3 Effectiveness of information source fusion

In the experiment of this section, we aim at verifying the effectiveness of our proposed pure D-S Fusion (PDSF) method fusing two information sources, namely friendships and user interest, by comparison with FoF method that also utilized friendships. Since it is little scale and preliminary experiment, we randomly generated a small preliminary dataset. Firstly, we randomly selected 200 users which were viewed as the fans or idol in social network and obtained the related data of social relationships from batch 1 dataset. Finally, we got 1615 user nodes and 2601 social edges. The statistics information of the preliminary dataset is summarized in Table 6.

The preliminary dataset is divided into training data and testing data with 80 and 20% ratio respectively. The experiments were carried out by tenfold cross validation. The results of performance testing are shown in Fig. 6.

We can observe that the PDSF method outperforms the traditional FoF on AP and AR on the whole. The trends of AP and AR of the two methods are similar respectively.

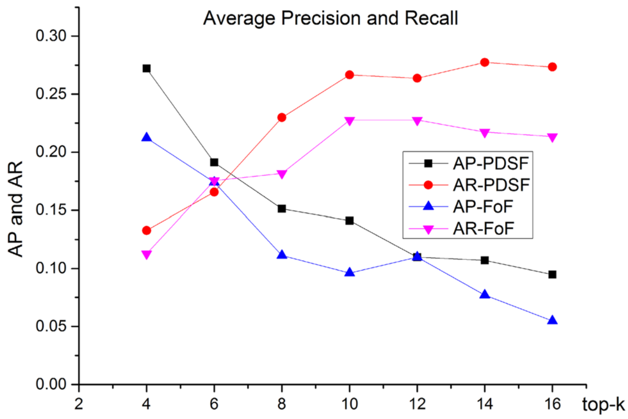


Fig. 6 The performance of AP and AR on preliminary dataset (PDSF and FoF represent Initial fusion method with D-S evidence theory and Friend of Friend method, respectively)

When the size of *top-k* is small, AP is relatively better than those with the bigger size of *top-k*. AP and AR of PDSF method is higher than those of FoF method regardless of the size of *top-k*. That means the PDSF method is more effective. Therefore, using the multiple information sources would indeed do better than using single information source in traditional methods. But AP and AR are both low about from 0.1 to 0.27, since there are a lot of users with 1 or 2 friends in the preliminary dataset.

5.4 Comparisons

In order to better evaluate the performance of our proposed fusion method, we here compare it with other methods including MPopular, FoF and LCIT. Amongst them, MPopular method is baseline recommendation approach which is built by our life experience based on user's popularity. FoF [31] is a mostly applied method and has relatively better effectiveness based on social graph. These two methods leveraged one or two information sources to generate FR. Our proposed fusion FR is, in some sense, a hybrid recommendation method. Therefore, we chose LCIT a hybrid FR approach as one of comparison method based on the idea extracted from Ref [31]. Also, in comparisons we applied our proposed methods into two forms, namely PDSF by initial D-S theory and IDSF by improved D-S theory. The details of each comparison approach are as following:

- *MPopular* This method sorts all the users in dataset based on its fans number viewed as the popularity. The popularity of users determines the order of the recommendation list. This method is viewed as the benchmark method, since most people often pay attention to celebrities or famous users and items in social networks.

Table 7 Statistics of comparison dataset

Item	Statistics
User num.	604
Social edge num.	9481
Max. num. of fans	105
Min. num. of fans	1
Max. num. of idol	140
Min. num. of idol	1
Message num.	56,924
Average degree	31.394
Net diameter	9
Average path length	3.495

- *FoF* It is a widely used FR method in many social networks such as Facebook web site, and has good performance. This method generates the recommendation list based on the fact that a friend of friend may be friend to target user. This algorithm focuses on two aspects. One is friend of friend and the other is the number of common friends.
- *LCIT* This method is a linear combinational recommendation with interest and social trust relationship weighted by 0.5 respectively. We argue that the two aspects in social networks are very reasonable and significant for recommending friends, since user is always inclined to make friends with similar interests and trust the friend's recommendation. Also, the methods leveraging content similarity are, sometimes, strong at discovering new friends with similar interests [31] in social networks. Hence, the combination of them is worthy for comparison.
- *PDSF, IDSF* They denote pure D-S fusion method and improved D-S fusion method, respectively. There may be some drawbacks in PDSF approach in line with our argument, which is discussed in Sect. 4.2. The approach of IDSF is improved by importance degree and reliability of evidence for overcoming the shortages of pure D-S evidence theory.

So far, we have proved the effectiveness of fusing multiple sources according to the results showed in Fig. 6, since the fusion method makes good use of the useful information derived from the corresponding source. But the AP and AR are relatively low in preliminary experiment. The main reasons we found are two aspects. One is that much of users with the friend number from 1 to 2 dominate the preliminary dataset. The other is that the user friend sets of many users change little. To make better comparisons and prove the advantages of our fusion approaches, we carefully screen out a comparison dataset, in which the friend number of most users from the Batch 1 data to Batch 2 data changes more or less. Then we could recommend the friends to target

Table 8 Performance of comparisons with different methods on comparison dataset

top-k	PDSF		IDSF		LCIT		FoF		MPopular	
	AP-PDSF	AR-PDSF	AP-IDSF	AR-IDSF	AP-LCIT	AR-LCIT	AP-FoF	AR-FoF	AP-MP	AR-MP
4	0.38542	0.20638	**0.42593**	0.21982	0.41256	**0.24762**	0.39128	0.22989	0.17269	0.03118
6	0.34852	0.20093	**0.35967**	**0.29455**	0.30182	0.24587	0.28470	0.22630	0.12327	0.03851
8	0.35863	0.23918	**0.36688**	**0.31601**	0.19882	0.25701	0.18182	0.23646	0.11554	0.04395
10	0.31885	0.25311	**0.33148**	**0.33940**	0.16741	0.26496	0.16465	0.24509	0.10609	0.05063
12	**0.27797**	0.31488	0.27226	**0.33326**	0.15288	0.27921	0.14825	0.25099	0.09662	0.06376
14	**0.25609**	0.31837	0.25557	**0.33559**	0.14814	0.26377	0.14351	0.24473	0.08168	0.06376
16	0.23526	0.32751	**0.24022**	**0.32944**	0.13898	0.27341	0.13171	0.25291	0.06670	0.06376

In each row data (*top-k* recommendations), bold values mean better results than other recommendation methods according to the corresponding metrics

Table 9 Comparisons of three methods on A-NDCG, MRR, and MAP with multiple sources

top-k	IDSF			PDSF			LCIT		
	A-NDCG	MRR	MAP	A-NDCG	MRR	MAP	A-NDCG	MRR	MAP
4	**0.67268**	**0.41071**	**0.64540**	0.61542	0.23668	0.28915	0.52364	0.13619	0.18866
6	**0.64685**	**0.29988**	**0.53458**	0.57697	0.19492	0.21348	0.49825	0.12571	0.16063
8	**0.56085**	**0.23958**	**0.47427**	0.46981	0.16359	0.17751	0.41139	0.07857	0.10476
10	**0.47374**	**0.19167**	**0.37942**	0.44835	0.12530	0.18656	0.43900	0.07682	0.12571
12	**0.39797**	**0.16683**	**0.35885**	0.38983	0.07708	0.15547	0.37044	0.07275	0.13968
14	0.40295	**0.14779**	**0.34112**	**0.40873**	0.09562	0.16386	0.31165	0.06236	0.11972
16	**0.40723**	**0.13323**	**0.32977**	0.40598	0.08454	0.16948	0.29887	0.05980	0.13095

In each row data (*top-k* recommendations), bold values mean better results than other recommendation methods according to the corresponding metrics

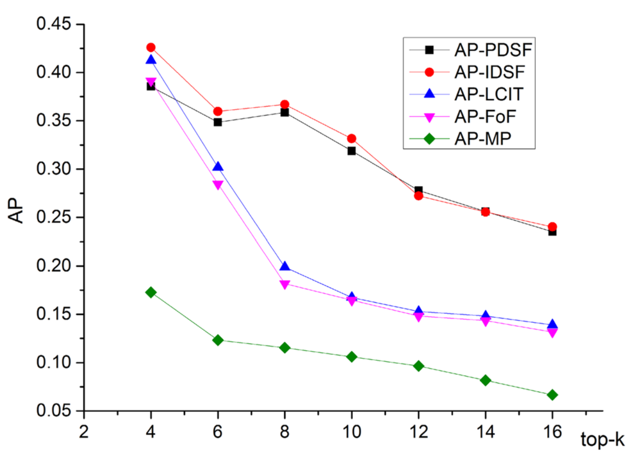


Fig. 7 Comparisons of 5 methods on AP

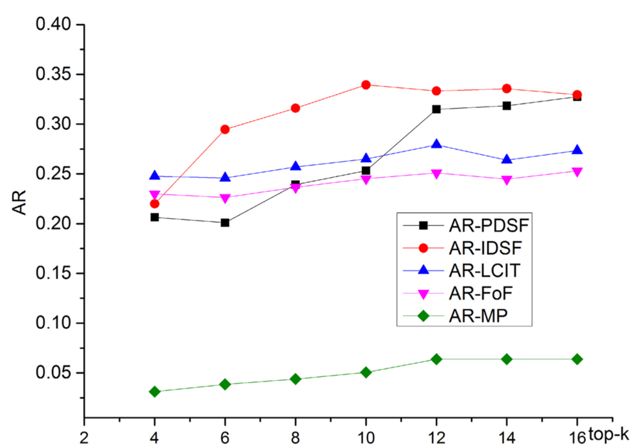


Fig. 8 Comparisons of 5 methods on AR

user and investigate the recommendation performance conveniently. The related information of comparison dataset is shown in Table 7 and the experimental results are shown in Tables 8 and 9 and Figs. 7, 8 and 9.

Table 8 and Figs. 7 and 8 show the evaluation results on AP, AR. We can observe that our fusion methods of PDSF and IDSF outperform the other methods in terms of precision and recall metrics. All the methods have the similar trends on the two metrics. AP decreases progressively and AR increases with *top-k* increasing. The greater the *top-k*, the lower AP is. When the *top-k* is greater and equal about to 10, AR is close to stabilization. The reason about this phenomenon can be explained by analyzing the computation formula $AR = \frac{1}{n}\sum_{i=1}^{n}\frac{N_1^i}{N_1^i+N_3^i}$. For a target user,

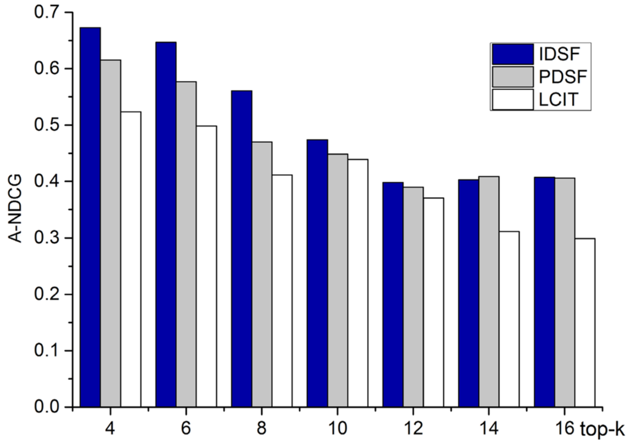


Fig. 9 A-NDCG histograms of comparison results of three methods with multiple sources

Table 10 The meanings of N_1^i, N_2^i and N_3^i

	Related	Unrelated
Recommended	N_1^i	N_2^i
Unrecommended	N_3^i	N_4^i

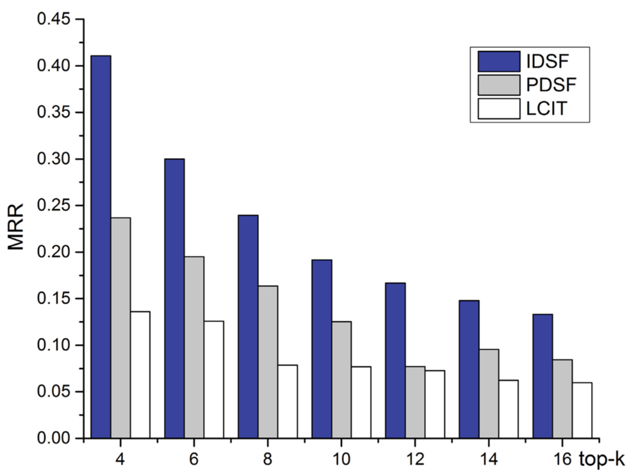


Fig. 10 MRR histograms of comparison results of three methods with multiple sources

the denominator is basically fixed, which denotes the total friend number of all "recommended and unrecommended" as showed in Table 10. So AR is determined by the numerator N_1^i. When top-k reaches around 10, the number of recommended friends is close to the best value that is almost unchanged. Therefore, AR is close to the stabilization. We argue that the top-k number that is close to stabilization depends on the specific dataset or social network. PDSF method fusing multiple sources performs better than

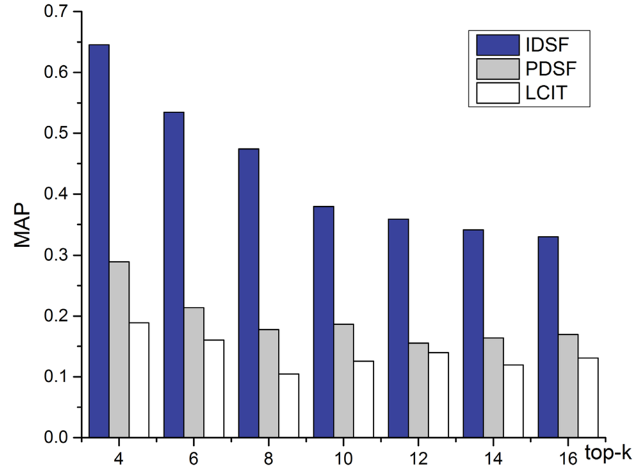


Fig. 11 MAP histograms of comparison results of three methods with multiple sources

LCIT, FoF and MPopular methods, but worse than IDSF method, since IDSF not only fuses multiple sources but considers the importance degree and reliability of evidence. FoF method utilized by Facebook site also has reasonable performance, although it is simple. And when it is combined with user interest to generate LCIT method, it is enhanced a little. This expresses that user interest is important in similarity measurement between friends, although some user's friends are acquaintance.

Table 9 and Figs. 9, 10 and 11 illustrate how the value of three methods all of which leverage multiple sources changes along with *top-k* value on the metrics of A-NDCG, MRR and MAP. The three metrics measure the quality of recommendation list and have the same trends. The higher they are, the better the recommendation list is. That means recommendation list with higher NDCG, MRR and MAP is more relevant to target user and user interested friends are as possible as at top positions. The difference between three methods on A-NDCG is little, while it is larger on MRR and MAP, since A-NDCG is discounted by logarithm function. Obviously, according to the three metrics, our fusion methods outperform linear combination method. IDSF method is the best, especially at the lower value of *top-k*.

So far, we have observed the comprehensive performance of IDSF method is the best on all five evaluation metrics, which could be benefit from the fusion of multiple sources and improvements of D-S evidence theory by applying the reliability and importance degree of evidence. Also, we can find the suitable *top-k* value of recommendation is about 8, at which AP and AR reach a relatively good level and A-NDCG, MRR and MAP work well. By our daily experience, this is consistent with the actual situation that user would like to select the interested users in FR list with the number about 8, since it does not make user uncomfortable and indeed helps user find out the interested users and its layout space in web pages is small.

5.5 Discussion

From the experimental results, we can observe that our proposed approach obtains expected effectiveness of friend recommendation in social networks. Although we cannot prove that it is the best solution among state-of-the art FR recommendation methods. Obviously, the effectiveness of our proposed FR framework benefits from the multi-source information fusion. During the process of friend recommendation, each information source provides its own useful information and positive influence.

We can understand users being friends in social networks are influenced by many factors like those information sources we have studied in this work. User profile expresses the demographic attributes. Users are more probably to be friends with similar profile users. In our real life, most of our friends are around us or near to us. Therefore, location-aware context information should be positive when finding out active user's potential friends [58]. We have proved the hypothesis in our proposed FR framework. In addition, we always predict possible friends with the same similar interests. User interests have high correlations when finding like-minded friends. That has been proved by many researchers as well as in our proposed FR framework.

FR is a hot research topic discussed in *social networks*. And a social network is like a net with social relations. The node/user position and social relations are two important factors for searching friends. Users close to each other in social networks are more probably to be friends, since the near node position is more reachable for target user. Furthermore, user influence also affects a target user to select friends. Users want to get useful and authoritative information from high influential friends. Finally, we cannot neglect the trust information among friends, which can be propagated from a user to another one. For example, friends of friends are friends with high possibility. All the information sources are theoretically combined by applying the fusion in D-S theory based on the *Minimum Conflict Principle* [48]. Consequently, the information sources in our paper are carefully selected. In this work, we carefully selected seven kinds of information sources that are categorized into three classes, and used them to improve the quality of recommendation results. In the scenario of other social networks, the number of selected related information sources depends on the specific demands. The number of information sources could be more or less, not be limited to predefined specific number. Our proposed fusion recommendation framework is still applicable. The main idea of our proposed method is how to find out the valuable information sources and how to select the corresponding indexes to quantify them. The experimental results demonstrate that the recommendation effectiveness is better than several mainstream friend recommendation methods by reasonably fusing multiple information sources. However, in this paper we did not conduct relevant experiments by removing some of them. It is still unknown that what information source would be more important for FR. That is what we will analyze and study in our future work.

Another two important aspects to be discussed in our proposed FR framework are algorithm efficiency measured by complexity and real application. Assume that n_1 denotes the number of users in a social network and n_2 represents the number of information sources. The complexity of our proposed approach consists of two parts that derive from D-S fusion algorithm and the BPA computation of information sources. The time complexity of D-S fusion algorithm is $O(n_2^2)$. The maximum complexity of seven information sources is $O(n_1^2)$. Totally, the overall complexity of our approach for friend recommendation is $K * O(n_2^2) * O(n_1^2)$, where K is a coefficient. Note that $O(n_2^2)$ is close to constant when $n_2 = 7$ is satisfied. Therefore, the complexity of our proposed approach is $O(n_1^2)$. In real world application scenarios, the key computations, namely all the BPAs, with the complexity of $O(n_1^2)$, can be done offline and implemented in parallel in distributed running environment. Thus, the real online calculation consumption is responsible for D-S fusion algorithm whose complexity is $O(n_2^2)$. Therefore, our proposed approach is valuable and feasible for friend recommendation in real applications in social networks.

6 Conclusion and future work

In this work, we focus on friend recommendation in social networks by fusing information derived from multi-sources. Our proposed fusion recommendation framework is based on the D-S evidence fusion theory. First, we investigated and analyzed the valuable information source and how they affect the friend selection of target user in social networks. After that, we proposed the fusion recommendation framework based on D-S theory with the BPA functions of all evidence. To further improve the recommendation performance, we observed the disadvantages of original D-S evidence theory and optimized it by incorporating importance degree and reliability of evidence in our proposed FR framework. Experimental results on real world dataset demonstrate that our proposed fusion recommendation method for friend recommendation in social networks can produce better results than several mainstream recommendation approaches in terms of five classic evaluation metrics. Although our method is proposed in Twitter-like social networks, it is still general and can be easily extended to other popular social networks like Facebook and LinkedIn.

In future work, we mainly focus on two aspects. On one hand, we will apply another fusion method and compare it to more combination recommendation methods like artificial neural network. On the other hand, we will estimate the importance of each information source to determine whether it should be fused in our recommendation framework and conduct experiments on more real world datasets.

Acknowledgements We would like to thank all of the anonymous reviewers for their insightful comments and useful suggestions that must lead to a much higher quality of our manuscript. This work was partially supported by the Outstanding Young Talents Program in colleges and universities of Anhui Province (Grant No. gxyqZD2018060).

References

1. Burke M, Marlow C, Lento T (2010) Social network activity and social well-being. In: Proceedings of the SIGCHI conference on human factors in computing systems. ACM, New York, pp 1909–1912
2. Trusov M, Bucklin RE, Pauwels K (2009) Effects of word-of-mouth versus traditional marketing: findings from an internet social networking site. J Mark 73(5):90–102
3. Mochón MC (2016) Social network analysis and big data tools applied to the systemic risk supervision. IJIMAI 3(6):34–37
4. Bakshy E, Rosenn I, Marlow C et al (2012) The role of social networks in information diffusion. In: Proceedings of the 21st international conference on World Wide Web. ACM, New York, pp 519–528
5. He C, Li H, Fei X et al (2016) A topic community-based method for friend recommendation in large-scale online social networks. Practice and Experience, Concurrency and Computation
6. Tang J, Hu X, Liu H (2013) Social recommendation: a review. Soc Netw Anal Min 3(4):1113–1133
7. Adomavicius G, Tuzhilin A (2005) Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. IEEE Trans Knowl Data Eng 17(6):734–749
8. Linden G, Smith B, York J (2003) Amazon.com recommendations: Item-to-item collaborative filtering. Internet Comput IEEE 7(1):76–80
9. Sanjuan-Martinez O, G-Bustelo BCP, Crespo RG et al (2009) Using recommendation system for e-learning environments at degree level. Int J Interact Multimed Artif Intell 1(2):67–70
10. Ricci F, Rokach L, Shapira B (2011) Introduction to recommender systems handbook. Recommender systems handbook. Springer, New York, 1–35
11. Liu NN, He L, Zhao M (2013) Social temporal collaborative ranking for context aware movie recommendation. ACM Trans Intell Syst Technol (TIST) 4(1):15
12. Yang X, Guo Y, Liu Y et al (2014) A survey of collaborative filtering based social recommender systems. Comput Commun 41:1–10
13. Li W, Ye Z, Xin M et al (2015) Social recommendation based on trust and influence in SNS environments. Multimed Tools Appl:1–18
14. Gong NZ, Talwalkar A, Mackey L et al (2014) Joint link prediction and attribute inference using a social-attribute network. ACM Trans Intell Syst Technol (TIST) 5(2):27
15. Yin Z, Gupta M, Weninger T et al (2010) A unified framework for link recommendation using random walks. Advances in Social Networks Analysis and Mining (ASONAM), 2010 International Conference on. IEEE, pp 152–159
16. Adamic LA, Adar E (2003) Friends and neighbors on the web. Soc Netw 25(3):211–230
17. Agarwal V, Bharadwaj KK (2013) A collaborative filtering framework for friends recommendation in social networks based on interaction intensity and adaptive user similarity. Soc Netw Anal Min 3(3):359–379
18. Tang F, Zhang B, Zheng J et al (2013) Friend Recommendation Based on the Similarity of Micro-blog User Model. Green Computing and Communications (GreenCom), 2013 IEEE and Internet of Things (iThings/CPSCom), IEEE International Conference on and IEEE Cyber, Physical and Social Computing. IEEE, pp 2200–2204
19. Golbeck J, Hendler J (2006) Inferring binary trust relationships in web-based social networks. ACM Trans Internet Technol 6(4):497–529
20. Pazzani MJ (1999) A framework for collaborative, content-based and demographic filtering. Artif Intell Rev 13(5–6):393–408
21. Said A, Plumbaum T, De Luca EW et al (2011) A comparison of how demographic data affects recommendation. User Modeling, Adaptation and Personalization (UMAP)
22. Kempe D, Kleinberg J, Tardos É (2003) Maximizing the spread of influence through a social network. In: Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, pp 137–146
23. Nepal S, Paris C, Pour PA et al (2013) A social trust based friend recommender for online communities "invited paper". In: Collaborative Computing: Networking, Applications and Worksharing (Collaboratecom), 2013 9th International Conference on. IEEE, pp 419–428
24. Martinez-Cruz C, Porcel C, Bernabé-Moreno J et al (2015) A model to represent users trust in recommender systems using ontologies and fuzzy linguistic modeling. Inf Sci 311:102–118
25. Guo D, Xu J, Zhang J et al (2017) User relationship strength modeling for friend recommendation on Instagram. Neurocomputing 239:9–18
26. Zhang Z, Liu Y, Ding W et al (2015) Proposing a new friend recommendation method, FRUTAI, to enhance social media providers' performance. Decis Support Syst 79:46–54
27. Lü L, Zhou T (2011) Link prediction in complex networks: A survey. Phys A Stat Mech Appl 390(6):1150–1170
28. Getoor L, Diehl CP (2005) Link mining: a survey. ACM SIGKDD Explor Newsl 7(2):3–12
29. Clauset A, Moore C, Newman MEJ (2008) Hierarchical structure and the prediction of missing links in networks. Nature 453(7191):98–101
30. Liu W, Lü L (2010) Link prediction based on local random walk. EPL 89(5):58007
31. Chen J, Geyer W, Dugan C et al (2009) Make new friends, but keep the old: recommending people on social networking sites. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, New York, pp 201–210
32. Liben-Nowell D, Kleinberg J (2007) The link-prediction problem for social networks. J Am Soc Inf Sci Technol 58(7):1019–1031
33. Yang SH, Long B, Smola A et al (2011) Like like alike: joint friendship and interest propagation in social networks. In: Proceedings of the 20th international conference on World Wide Web. ACM, New York, pp 537–546
34. Yu Z, Wang C, Bu J et al (2015) Friend recommendation with content spread enhancement in social networks. Inf Sci 309:102–118
35. Brin S, Page L (2012) Reprint of: The anatomy of a large-scale hypertextual web search engine. Comput Netw 56(18):3825–3833
36. Konstas I, Stathopoulos V, Jose JM (2009) On social networks and collaborative recommendation. In: Proceedings of the 32nd

international ACM SIGIR conference on Research and development in information retrieval. ACM, New York, pp 195–202
37. Ahmed NM, Chen L (2016) An efficient algorithm for link prediction in temporal uncertain social networks. Inf Sci 331:120–136
38. Cui L, Dong L, Fu X et al (2017) A video recommendation algorithm based on the combination of video content and social network. Concurr Comput Pract Exp 29(14)
39. Jamali M, Ester M (2010) A matrix factorization technique with trust propagation for recommendation in social networks. In: Proceedings of the fourth ACM conference on Recommender systems. ACM, pp 135–142
40. Ma H, King I, Lyu MR (2009) Learning to recommend with social trust ensemble. In: Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval. ACM, New York, pp 203–210
41. Agarwal V, Bharadwaj KK (2011) Trust-enhanced recommendation of friends in web based social networks using genetic algorithms to learn user preferences. In: Trends in Computer Science, Engineering and Information Technology. Springer, Berlin, Heidelberg, pp 476–485
42. Ma Y, Yu Z, Ding JA (2014) Method of user recommendation in social networks based on trust relationship and topic similarity. Social media processing. Springer, Berlin, pp 240–251
43. Agarwal V, Bharadwaj KK (2014) Friends recommendations in dynamic social networks. Encycl Soc Netw Anal Min:553–562
44. Victor P, Cornelis C, De Cock M et al (2011) Practical aggregation operators for gradual trust and distrust. Fuzzy Sets Syst 184(1):126–147
45. Bao J, Zheng Y, Wilkie D et al (2013) A survey on recommendations in location-based social networks. ACM Transaction on Intelligent Systems and Technology, New York
46. Wang Y, Wang X, Zuo W-L (2014) Trust prediction modeling based on social theories. J Softw (Chin) 25(12):2893–2904
47. Milgram S (1967) The small world problem. Psychol Today 2(1):60–67
48. Shafer G (1976) A mathematical theory of evidence. Princeton university press, Princeton
49. Telmoudi A, Chakhar S (2004) Data fusion application from evidential databases as a support for decision making. Inf Softw Technol 46(8):547–555
50. Dempster AP (1967) Upper and lower probabilities induced by a multivalued mapping. Ann Math Stat 38:325–339
51. Khaleghi B, Khamis A, Karray FO et al (2013) Multisensor data fusion: a review of the state-of-the-art. Inf Fusion 14(1):28–44
52. Wei D, Deng X, Zhang X et al (2013) Identifying influential nodes in weighted networks based on evidence theory. Phys A Stat Mech Appl 392(10):2564–2575
53. Ellison NB (2007) Social network sites: Definition, history, and scholarship. J Comput Mediat Commun 13(1):210–230
54. Yong D, WenKang S, ZhenFu Z et al (2004) Combining belief functions based on distance of evidence. Decis Support Syst 38(3):489–493
55. Gong NZ, Xu W, Huang L et al (2012) Evolution of social-attribute networks: measurements, modeling, and implications using google+. In: Proceedings of the 2012 Internet Measurement Conference. ACM, pp 131–144
56. Li Y, McLean D, Bandar Z et al (2006) Sentence similarity based on semantic nets and corpus statistics. Knowl Data Eng IEEE Trans 18(8):1138–1150
57. DeScioli P, Kurzban R, Koch EN et al (2011) Best friends alliances, friend ranking, and the myspace social network. Perspect Psychol Sci 6(1):6–8
58. Liben-Nowell D, Novak J, Kumar R et al (2005) Geographic routing in social networks. Proc Natl Acad Sci USA 102(33):11623–11628
59. Sui X, Chen Z, Ma J (2015) Location sensitive friend recommendation in social network. Web technologies and applications. Springer International Publishing, New York, pp 316–327
60. Scellato S, Noulas A, Lambiotte R et al (2011) Socio-spatial properties of online location-based social networks. ICWSM 11:329–336
61. Wu M, Wang Z, Sun H et al (2016) Friend Recommendation Algorithm for Online Social Networks Based on Location Preference. Information Science and Control Engineering (ICISCE), 3rd International Conference on. IEEE, 2016: 379–385
62. Li Q, Zheng Y, Xie X et al (2008) Mining user similarity based on location history. In: Proceedings of the 16th ACM SIGSPATIAL international conference on Advances in geographic information systems. ACM, New York, p 34
63. Zheng Y, Zhou X (2011) Computing with spatial trajectories. Springer Science & Business Media, New York
64. Xiao X, Zheng Y, Luo Q et al (2014) Inferring social ties between users with human location history. J Ambient Intell Humaniz Comput 5(1):3–19
65. Moreno A, Redondo T (2016) Text analytics: the convergence of big data and artificial intelligence. IJIMAI 3(6):57–64
66. Gruber TR (1993) A translation approach to portable ontology specifications. Knowl Acquis 5(2):199–220
67. Natalya FN, Deborah LM (2001) Ontology development 101: a guide to creating your first ontology. Stanford Knowledge Systems Laboratory Technical Report KSL-01-05 and Stanford Medical Informatics Technical Report SMI-2001-0880
68. Zheng J, Zhang B, Yue X et al (2015) Neighborhood-user profiling based on perception relationship in the micro-blog scenario. Web Semant Sci Serv Agents World Wide Web 34:13–26
69. Cha M, Haddadi H, Benevenuto F et al (2010) Measuring User Influence in Twitter: The Million Follower Fallacy. ICWSM 10(10–17):30
70. Golbeck J (2005) Computing and applying trust in web-based social networks. University of Maryland, Maryland
71. Victor P, Verbiest N, Cornelis C et al (2013) Enhancing the trust-based recommendation process with explicit distrust. ACM Trans Actions Web (TWEB) 7(2):6
72. Luo H, Niu C, Shen R et al (2008) A collaborative filtering framework based on both local user similarity and global user similarity. Mach Learn 72(3):231–245
73. Avnit A (2009) The million followers fallacy. Pravda Media Group, Tel Aviv
74. Verbiest N, Cornelis C, Victor P et al (2012) Trust and distrust aggregation enhanced with path length incorporation. Fuzzy Sets Syst 202:61–74
75. Shani G, Gunawardana A (2011) Evaluating recommendation systems. Recommender systems handbook. Springer, New York, pp 257–297
76. Weimer M, Karatzoglou A, Le QV et al (2008) Cofi rank-maximum margin matrix factorization for collaborative ranking. In: Advances in neural information processing systems, pp 1593–1600
77. Chakrabarti S, Khanna R, Sawant U et al (2008) Structured learning for non-smooth ranking losses. In: Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, pp 88–96
78. Järvelin K, Kekäläinen J (2000) IR evaluation methods for retrieving highly relevant documents. In: Proceedings of the 23rd annual international ACM SIGIR conference on Research and development in information retrieval. ACM, pp 41–48