Discovering Unknown But Interesting Items on Personal Social Network

Juang-Lin Duan, Shashi Prasad, and Jen-Wei Huang

Yuan Ze University,
135 Yuan-Tung Road, Chung-Li, Taiwan 32003, R.O.C.
{s986039,s989112}@mail.yzu.edu.tw,
jwhuang@saturn.yzu.edu.tw

Abstract. Social networking service has become very popular recently. Many recommendation systems have been proposed to integrate with social networking websites. Traditional recommendation systems focus on providing popular items or items posted by close friends. This strategy causes some problems. Popular items always occupy the recommendation list and they are usually already known by the user. In addition, items recommended by familiar users, who frequently communicate with the target user, may not be interesting. Moreover, interesting items from similar users with lower popularity are ignored. In this paper, we propose an algorithm, UBI, to discover unknown but interesting items. We propose three scores, i.e., Quartile-aided Popularity Score, Social Behavior Score, and User Similarity Score, to model the popularity of items, the familiarity of friends, and the similarity of users respectively in the target user's personal social network. Combining these three scores, the recommendation list containing unknown but interesting items can be generated. Experimental results show that UBI outperforms traditional methods in terms of the percentages of unknown and interesting items in the recommendation list.

Keywords: recommendation, social network, unknown but interesting.

1 Introduction

With the tremendous success of social networking websites nowadays, diverse social network services have been vigorous and much popular. Although many services related to social network websites exist, it is important to set up a standard which helps users to make a decision if it is worthy of them. In recent years, many recommendation systems have been proposed to integrate with social networking websites and been used in many different business applications such as movies, music, books, news, etc. Some popular e-commerce websites such as Amazon, eBay, and Netflix analyze the shopping behavior of users to build the recommendation list of products to their customers. The online shopping system recommends each customer the products bought by others who have similar shopping behavior in the past. In addition, some well-known social networking websites, Facebook, Myspace and Twitter provide users to establish

P.-N. Tan et al. (Eds.): PAKDD 2012, Part II, LNAI 7302, pp. 145–156, 2012.

their own personal communities or social networks based on friends. There are many services for personalized recommendations in social networking websites. For example, InSuggest[1] provides personalized recommendations of bookmarks originating from the social bookmarking site Delicious[2], and Outbrain[3] provides personalized blog recommendations from blogging services. The purpose of these recommendations are to adapt the contents of the websites to the specific needs of the individual user by presenting the most attractive and relevant items to users.

Traditional recommendation systems usually generate recommendation lists based on popularity of items and/or analyze the behavior of the target user and then make further recommendations. However, these systems focus on providing popular items or items posted by close friends. This leads to problems listed as follows:

1. Popular items always occupy the recommendation list and they are usually already known by the user.
2. Items recommended by familiar users, who frequently communicate with the target user, may not be interesting.
3. Interesting items from similar users with lower popularity are ignored.

Fig. 1 shows the personal social network of the target user U_1. Circular nodes represent users and the link between two users indicates the friend relationship. The number on the link denotes the number of direct communication between these two users and MF represents the number of mutual friends between U_1 and the other user. In addition, square nodes represent items posted by the connected user, where the number indicates the number of comments left by all users and the number of likes, which means other users are interested in the item. Traditional recommendation systems usually recommend items $I_4 : 150$ and $I_1 : 100$, since I_4 is the most popular item and I_1 is copied by many users. However, these items are easily found by the target user and should not occupy the recommendation list. On the other hand, $I_9 : 55$ from similar user U_3 with lower popularity is easily ignored. New paragraph to remedy these defects, we propose to discover unknown but interesting items, and design an algorithm to generate recommendation list on personal social networks. The personal social network contains the target user and his/her direct friends. We also include items posted by these users.

Our proposed algorithm not only considers the popularity of items and the similarity with friends, but also discovers unknown but interesting items through the target user's social behavior. We propose three scores to calculate Unknown But Interesting Score of each item in the target user's personal social network. The first score is Quartile-aided Popularity Score, which is based on the popularity adjusted by quartiles of items, to find out items with lower popularity. The

[1] http://insuggest.wordpress.com/
[2] http://www.delicious.com/
[3] http://www.outbrain.com/

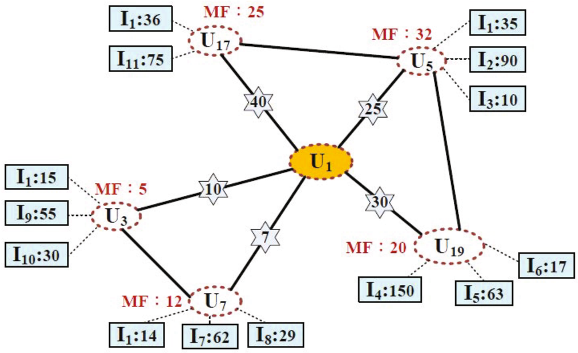

Fig. 1. The personal social network of the target user U_1

second score is Social Behavior Score, which depends on social interactions on social network websites and direct communication between users. The third score is User Similarity Score, which is based on interests between users, to model the similarity of the target user and his/her friends. Combining these three scores, we can generate the recommendation list with unknown but interesting items to the target user.

Finally, to evaluate our approach, we implement our system on a social networking website, and collect users' feedback to compare differences between traditional approaches and our proposed algorithm. The experimental results show that our proposed algorithm can successfully find unknown but interesting items and the satisfaction percentage of our system is higher than compared methods.

The remainder of this paper is organized as follows. In Section 2, a brief survey of related works is presented. The proposed algorithm is introduced in Section 3. Section 4 presents the experimental results. Finally, we conclude this work in Section 5.

2 Related Works

2.1 Social Networking

Social networking is the development of social collaborative technologies, and connected by one or more specific types of relationship, such as friendship, similar interest. In recent year, online social networking has been around in the world, therefore, many online social networking websites are being generated which allow users to establish their social network by adding other users to their friend lists. For example, many users of popular social networking sites such as Facebook and Twitter. Many research issues of social networking focused on development of information techniques and data processing [7], and then extend to social networking analysis [10], which is a set of methods to discover relations between nodes in a social network.

A number of measures in social networking analysis including network size, degree centrality, betweenness, density etc. are considered. Bird *et al.* [5] proposed a method to extract social networks from e-mail communication. Agrawal *et al.* [3] using web mining techniques to understand the behavior of users in news group, the proposed the behavior is meaning a newsgroup posting consists of one or more quoted lines from another posting followed by the opinion of the author. Adamic *et al.* [2] developed a method to discover the relationship of friends and neighbors in the web. Many social networking analysis approaches have propose similar ideas to find neighborhoods and paths with the social network [8], [9]. In our work, we extend the concept of social networking to discover the unknown but interesting items from social network site.

2.2 Recommendation Systems

Recommendation systems are widely used for personalized information filtering technology, always used to recommend items that are of interest to users based on customer demographics, features of items, or user preferences. Therefore, users should provide their interest profiles to recommendation systems in order to get recommendations. Then recommendation systems can utilize these interest profiles to estimate the ratings of the unrated items for users or predict that items to be liked by users. In general, recommendation systems are usually classified into the following three methods: content-based recommendation, collaborative filtering and hybrid approaches.

The first method is based on contents [12], which analyzes the contents of information products and user information to produce the recommended method. This method is mainly dependent on the data description of goods and users of consumer behavior in the past, for the two meta-analysis to calculate the characteristics of different commodities of the scores for the summary, identify the items for the user with a higher satisfaction scores in order to establish recommended.

The second method is based on collaborative filtering [14], which utilizes similarities of user's preferences to recommend items. Collaborative filtering is a set of similarity measure methods, as follows: Jaccard's coefficient of similarity, Cosine similarity [15], Pearson correlation-based similarities [13]. Many approaches employ the technique of collaborative filtering, for instance, Bell *et al.* [4] proposed novel algorithms for predicting user ratings of items by integrating complementary models that focus on patterns at different scales. Facebook has a feature, called "People You May Know", which recommends user to connect with based on a "friend of a friend" approach [1].

Finally, many recommendation systems use hybrid approaches by combining content-based methods and collaborative filtering [6], which helps to avoid certain limitations of content-based and collaborative systems. For example, TANGENT [11] focused on the "surprise me" query, in which a user may be bored with usual genre of items, and may recommend new genre of items. This research

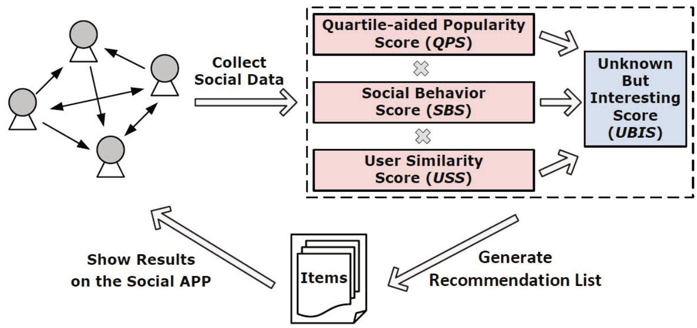

Fig. 2. The architecture of the system

closes to our belief, however, traditional recommendation systems always focus on high frequency of item with similarity of user. That gives us an inspiration that we can make use of the impact in our work.

3 Unknown But Interesting Recommendation System

In this section, we illustrate the system architecture and the details of our proposed algorithm, Unknown But Interesting algorithm.

3.1 System Architecture

We build a recommendation system on a social networking platform. The architecture of the system is shown in Fig. 2. First, when the target user logs in to our system, the target user's social data is collected and is used to generate the personal social network. The social data includes the target user's profile, which contains his/her posts, interests, friends list, and social interactions. From the friends list, the open information of the target user's friends is also collected, which includes the number of mutual friends and direct communication. We only consider 1-level friends, who have the direct connections to the target user. The reason is that users are usually not interested in the social behavior of friends of 1-level friends and friends of friends have little influence power to the target user. After the collection of the social data, we analyze the personal social network to calculate three different scores of each item and obtain unknown but interesting score by the proposed algorithm. Finally, we can generate the recommendation list of items show the results on the social networking website.

3.2 Unknown But Interesting Algorithm

Unknown But Interesting (UBI) algorithm focuses on the popularity of items, the similarity between users, and the social behavior of the target user. The proposed algorithm calculate three scores as follows:

iv	I_3U_5	I_1U_7	I_1U_3	I_6U_{19}	I_8U_7	$I_{10}U_3$	I_1U_5	I_1U_{17}	I_9U_3	I_7U_7	I_5U_{19}	$I_{11}U_{17}$	I_2U_5	I_4U_{19}
PS_{iv}	10	14	15	17	29	30	35	36	55	62	63	75	90	150
QS_{iv}	53	49	48	46	34	33	28	27	8	1	0	12	27	87
QPS_{iv}	0.40	0.44	0.45	0.48	0.61	0.63	0.68	0.69	0.91	0.99	1	0.86	0.69	0.01

Fig. 3. Example of PS_{iv}, QS_{iv}, and QPS_{iv}

1. Quartile-aided Popularity Score (QPS) of each item.
2. Social Behavior Score (SBS) by considering social interactions of users.
3. User Similarity Score (USS) of each friend of the target user.

Finally, we combine these three scores to obtain the Unknown But Interesting Score ($UBIS$), and provide the recommendation list to the target user. We explain the formulas and significance of each score as follows.

Quartile-Aided Popularity Score. In the social networking website, users can post messages, photos, videos, and links on their own pages. Other users are able to leave comments on the posted items or simply click "like" button to show their interests in the items. Therefore, in our system, the popularity score, PS_{iv}, is defined as the number of comments and likes of a certain item i posted by user v.

As we explained earlier, traditional recommendation systems always recommend popular items to users, but these items are well-known and easily noticed by users themselves. In order to determine a certain degree of popular items, we use the concept of quartile, $Q_r = |r\,(n+1)\,/4|$, where r determines which quartile, and n is total number of items. The items are sorted by their popularity score, PS_{iv}, ascendingly. The upper quartile, $Q_3(PS_{iv})$, represents the popularity of the item with the $\lfloor 3\,(n+1)/\,4 \rfloor$th rank in the list. In this way, we define quartile score, QS_{iv}, to be the popularity score minus the upper quartile.

$$QS_{iv} = PS_{iv} - Q_3\,(PS_{iv}) \tag{1}$$

Fig. 3 shows the PS_{iv} and QS_{iv} of each item posted by each user in Fig. 1. In this example, n is 14. Therefore Q_3 is the 11th lowest value of PS_{iv}. The respective QS_{iv} is shown in middle row. Furthermore, in order to find items which are not very popular but still have enough attention, we propose Quartile-aided Popularity Score, QPS_{iv}.

$$QPS_{iv} = 1 - \frac{QS_{iv}}{Max\,(QS_{iv}) + 1} \tag{2}$$

QPS_{iv} normalizes QS_{iv} by the maximum value and gives the upper quartile the highest credit. In this way, we can capture the popularity of items adjusted by the upper quartile.

Social Behavior Score. In addition to adjusting the popularity of items, UBI considers the social behavior of the target user to further discover unknown

uv	U_1U_3	U_1U_7	U_1U_{19}	U_1U_5	U_1U_{17}
$MF_{(uv)}$	5	12	20	32	25
$DC_{(uv)}$	10	7	30	25	40
$SBS_{(uv)}$	0.64	0.53	0.1	0.01	0.007

Fig. 4. Example of MF_{uv}, DC_{uv} , and SBS_{uv}

items. In the social networking website, users usually allowed to meet friends and make connections to one another. Users can easily get information from the friends they are familiar with. Therefore, UBI includes two factors from the social behavior, i.e., mutual friends and direct communication. We define mutual friend, MF_{uv}, be the number of mutual friends between user u and user v, and direct communication of users, DC_{uv}, be the number of direct communication between user u and user v. For the target user u, the more mutual friends u and v have the more likely it is that items are spread between those friends. In addition, the more direct communication there is between u and v, the more likely it is that items posted by user v are already known by the target user u. Therefore, in order to find unknown items, we define Social Behavior Score as follows.

$$SBS_{uv} = \left(1 - \frac{MF_{uv}}{\underset{v \in friends\ of\ u}{Max}(MF_{uv}) + 1}\right) \times \left(1 - \frac{DC_{uv}}{\underset{v \in friends\ of\ u}{Max}(DC_{uv}) + 1}\right), \tag{3}$$

where $Max(MF_{uv})$ represents the maximum value of MF_{uv} among all friends v of the target user u, and $Max(DC_{uv})$ is the maximum value of DC_{uv} among all v. SBS_{uv} represents the inverse probability that the items posted by user v are already known by the target user u. Fig. 4 shows some SBS_{uv} of users in Fig. 1.

User Similarity Score. UBI not only takes popularity of items and familiarity of users into consideration, but also includes the similarity of users to obtain interesting items. If the item is recommended by a similar user of the target user, it is more likely that the target user is interested in the item. At first, users can do different actions in the social networking website to show their interest in some items, e.g., posting a link, commenting on a photo, or liking a video. We give each action a worth value, WV, indicating how much a user is interested to some item by performing this action.

$$WV_j = \frac{\sum_{j \in all\ actions} times\ of\ action\ j}{times\ of\ action\ j} \tag{4}$$

For example, the number of articles is 100, the number of comments is 500, and the number of likes is 1000. We can get the sum of all action as 1600, and we can

	I_1	I_2	I_3	I_4	I_5
U_1	post+like	like+comment	like	comment	n/a
U_3	post	post+comment	comment	post+like +comment	post
U_{19}	n/a	like+comment	n/a	like	like

	I_1	I_2	I_3	I_4	I_5
U_1	17.6	4.8	1.6	3.2	0
U_3	16	19.2	3.2	20.8	16
U_{19}	0	4.8	0	1.6	1.6

Fig. 5. The score of user's behavior

calculate the worth value of posting an article to be 16, leaving a comment to be 3.2, and liking an item to be 1.6. Then, users may have a variety of behavior on the same item, as shown in Fig. 5. Therefore, we sum up the worth value of all actions performed on the same item i to get the interesting score, IS_i.

$$IS_i = \sum_{j \in all\ actions} WV_j \tag{5}$$

Finally, we can define the user behavior as

$$UB_u = \{IS_{I_1}, ..., IS_{I_n}\}, \tag{6}$$

where IS_i is the total interesting score of the user v to the item i. Accordingly, the User Similarity Score, USS_{uv}, between user u and user v is computed by the following equation.

$$USS_{uv} = \frac{UB_u \cdot UB_v}{\|UB_u\| \, \|UB_v\|} \tag{7}$$

From the user behavior listed in Fig. 5, user similarity score between U_1 and U_3 is 0.5, and USS between U_1 and U_{19} is 0.21.

Unknown But Interesting Score. Finally, we combine QPS, SBS, and USS to calculate the unknown but interesting score for each item on the personal social network of the target user u. Thus, we define Unknown But Interesting Score as follows.

$$UBIS_i = \sum_v (QPS_{iv} \times SBS_{uv} \times USS_{uv}) \tag{8}$$

where $\sum$ is the sum of same item i among all user v. Consequently, we can generate the recommended list based on $UBIS$. As shown in Fig. 6, we sort $UBIS$ and recommend the Top-k items to the target user.

i	I_9	I_{10}	I_7	I_8	I_5	I_1	I_6	I_2	I_3	I_{11}	I_4
$UBIS_i$	0.29	0.20	0.15	0.09	0.07	0.04	0.03	0.006	0.004	0.0006	0

Fig. 6. Example of $UBIS_i$

4 Experiments

In this section, the methodology and the performance evaluation are discussed. The experiment is conducted to measure the percentages of unknown but interesting items in the recommendation list. The methodology is discussed in Section 4.1. The performance evaluation is presented and discussed in Section 4.2.

4.1 Methodology

We implement recommendation system on a popular social networking website, Facebook, in order to compare our algorithm to traditional recommendation systems. We can obtain user's social information easily to discover unknown items. We generate three recommendation lists each on Facebook, traditional method, and our algorithm. First, Facebook recommendation list is based on latest updates from user's posting. Second, traditional method is based on popular items with user's preferences on Facebook. In other words, users usually focus on popularity of items with similarity among users. At last, our algorithm is based on $UBIS$ which recommends unknown but interesting items. We generate recommendation list which presents Top-20 items to the target user, as shown in the Fig.7, which is the interface of UBI recommendation system on Facebook, and we show one of the lists randomly. Furthermore, we show posted user's name, content of the item, and each list has two questions for each message, and questionnaires, which are as follows: unknown or known, and interesting. The question about unknown or known represents whether the message is unread or read by the target user respectively. The question about interesting denotes whether the target user is interested in the message. We can compare UBI algorithm with the other two methods based on our questionnaires.

4.2 Performance Evaluation

We randomly invited 355 users to participate in our experiment. Our experiments were conducted in the months starting from July through September of 2011, and 185 active users participated per month. At first, we compared three recommendation lists, and focused on unknown and interesting questions checked by users. In other words, these two indicators are used to determine the target user's satisfaction. Fig. 8 presents percentages of items in the recommendation lists. Facebook (FB) recommends unknown items better than others, because FB usually recommends latest items, but users are usually not interested in them. The percentage of unknown items of traditional method with popularity with

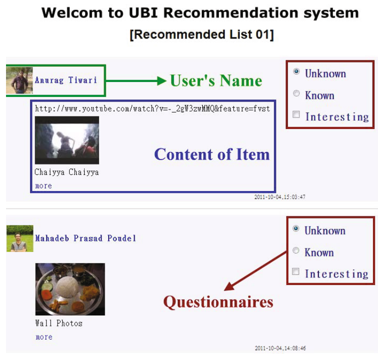

Fig. 7. System user interface for the recommended list

similarity (PS) is worst than that of other methods whereas the percentage of interesting items of traditional method with PS is better than that of FB. Because traditional method recommends items, which are usually already known to users, based on PS among users. Our UBI algorithm can discover unknown items almost same as FB does and interesting items is better than FB and PS. In terms of overall satisfaction with unknown and interesting questionnaires, our algorithm can recommend unknown but interesting items exactly.

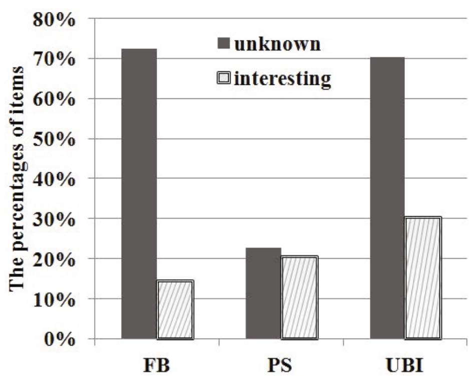

Fig. 8. The percentages of items in the recommendation lists

In addition, we recommend Top-20 items for each recommendation system, then we compared percentages of Top-5 to Top-20. We want to ensure good performance of satisfaction for each stage. Fig. 9 represents the percentage of options checked by users for Top-5 to Top-20. In Fig. 9(a), the UBI algorithm discovers unknown items almost the same as FB, and the percentage of interesting is higher than FB and PS. Besides, Fig. 9(b), (c), and (d) also show this trend.

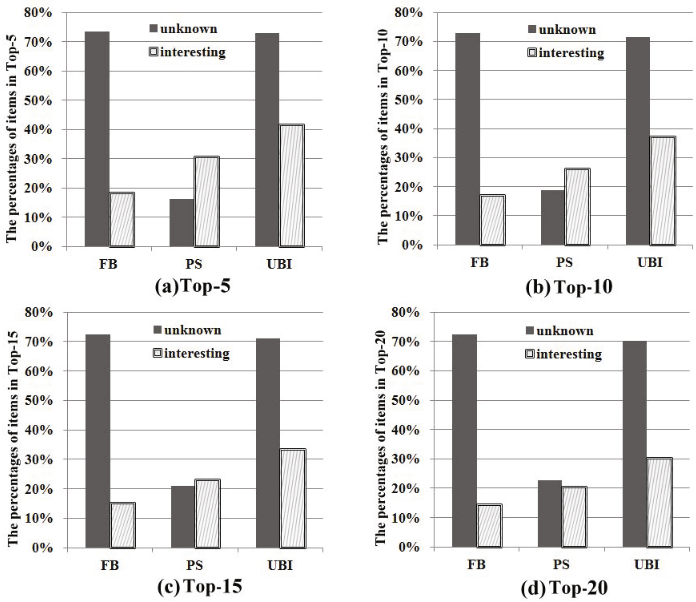

Fig. 9. The percentages of items in the Top-k list

This phenomenon represents not only Top-20 better UBI recommendations, but Top-5 to Top-15 also better than FB and PS. Therefore, we can obtain user's satisfaction from user's feedback is better than that of FB and PS. Finally, We found that our proposed algorithm to discover unknown but interesting items is better than Facebook and the traditional methods.

5 Conclusions

Traditional systems are based on similarity and popularity. This strategy leads to some problems. We proposed an algorithm which recommends unknown but interesting item by utilizing three scores: Quartile-aided Popularity Score, Social Behavior Score, and User Similarity Score. We focus on the communication among users and mutual friends, and discover the unknown but interesting item for user. In other words, we not only consider the similarity but also care about the user's social interaction. Experimental results show that the performance of UBI significantly outperforms that of traditional methods in terms of the percentages of unknown and interesting items in the recommendation list. Our future work could focus on information propagation in social networks, and friend of friend structure and utilize cloud computing techniques to improve the system performance.

References

1. Official facebook blog, `http://blog.facebook.com/blog.php?post=15610312130`
2. Adamic, L., Adar, E.: Friends and neighbors on the web. Social Networks 25(3), 211–230 (2003)
3. Agrawal, R., Rajagopalan, S., Srikant, R., Xu, Y.: Mining newsgroups using networks arising from social behavior. In: Proceedings of the 12th International Conference on World Wide Web, WWW 2003, pp. 529–535. ACM (2003)
4. Bell, R., Koren, Y., Volinsky, C.: Modeling relationships at multiple scales to improve accuracy of large recommender systems. In: Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD 2007, pp. 95–104. ACM (2007)
5. Bird, C., Gourley, A., Devanbu, P., Gertz, M., Swaminathan, A.: Mining email social networks. In: Proceedings of the 2006 International Workshop on Mining Software Repositories, MSR 2006, pp. 137–143. ACM (2006)
6. Claypool, M., Gokhale, A., Miranda, T., Murnikov, P., Netes, D., Sartin, M.: Combining content-based and collaborative filters in an online newspaper. In: Proceedings of ACM SIGIR Workshop on Recommender Systems (1999)
7. Fu, F., Liu, L., Wang, L.: Empirical analysis of online social networks in the age of web 2.0. Physica A: Statistical Mechanics and its Applications 387(2-3), 675–684 (2008)
8. Geyer, W., Dugan, C., Millen, D.R., Muller, M., Freyne, J.: Recommending topics for self-descriptions in online user profiles. In: Proceedings of the 2008 ACM Conference on Recommender Systems, RecSys 2008, pp. 59–66. ACM (2008)
9. Groh, G., Ehmig, C.: Recommendations in taste related domains: collaborative filtering vs. social filtering. In: Proceedings of the 2007 International ACM Conference on Supporting Group Work, GROUP 2007, pp. 127–136. ACM (2007)
10. Mislove, A., Marcon, M., Gummadi, K.P., Druschel, P., Bhattacharjee, B.: Measurement and analysis of online social networks. In: Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement, IMC 2007, pp. 29–42. ACM (2007)
11. Onuma, K., Tong, H., Faloutsos, C.: Tangent: a novel, 'surprise me', recommendation algorithm. In: Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD 2009, pp. 657–666. ACM (2009)
12. Pazzani, M., Billsus, D.: Content-Based Recommendation Systems. In: Brusilovsky, P., Kobsa, A., Nejdl, W. (eds.) Adaptive Web 2007. LNCS, vol. 4321, pp. 325–341. Springer, Heidelberg (2007)
13. Sarwar, B., Karypis, G., Konstan, J., Reidl, J.: Item-based collaborative filtering recommendation algorithms. In: Proceedings of the 10th International Conference on World Wide Web, WWW 2001, pp. 285–295. ACM (2001)
14. Schafer, J.B., Frankowski, D., Herlocker, J., Sen, S.: Collaborative Filtering Recommender Systems. In: Brusilovsky, P., Kobsa, A., Nejdl, W. (eds.) Adaptive Web 2007. LNCS, vol. 4321, pp. 291–324. Springer, Heidelberg (2007)
15. Steinbach, M., Karypis, G., Kumar, V.: A comparison of document clustering techniques. In: KDD-2000 Workshop on Text Mining, pp. 109–111 (2000)