Social Filtering Using Social Relationship for Movie Recommendation

Inay Ha[1], Kyeong-Jin Oh[1], Myung-Duk Hong[1], and Geun-Sik Jo[2]

[1]Intelligent E-Commerce Lab. Department of Computer & Information Engineering, INHA University, 100 Inha-ro, Nam-gu, Incheon, Korea
{inay,okjkillo,hmdgo}@eslab.inha.ac.kr
[2] School of Computer & Information Engineering, INHA University, 100 Inha-ro, Nam-gu, Incheon, Korea
gsjo@inha.ac.kr

Abstract. Traditional recommendation systems provide appropriate information to a target user after analyzing user preferences based on user profiles and rating histories. However, most of people also consider the friend's opinions when they purchase some products or watch the movies. As social network services have been recently popularized, many users obtain and exchange their opinions on social networks. This information is reliable because they have close relationships and trust each other. Most of the users are satisfied with the information. In this paper, we propose a recommendation system based on advanced user modeling using social relationship of users. For the user modeling, both direct and indirect relations are considered and the relation weight between users is calculated by using six degrees of Kevin Bacon. From the experimental results, our proposed social filtering method can achieve better performance than a traditional user-based collaborative filtering method.

Keywords: Social Relation, Collaborative Filtering, User Modeling, Social Filtering, Recommendation.

1 Introduction

The emergence of personalized recommendation system allows the users to easily select products, movies and other information. The recommendation system provides appropriate information according to user preferences after analyzing user profiles and behaviors. Many web services employ the recommendation system to provide useful information to the users. MovieLens[1] recommends movies to users using rating histories. Some of the web sites like Musicovery[2] recommend new music to users with a visual approach. Amazon, which is an online retailer, recommends books based on past purchase behaviors. However, there is a limitation of the traditional recommendation systems. The systems only handle user profiles and history. Users usually

[1] http://www.movielens.org/
[2] http://musicovery.com/

N.-T. Nguyen et al. (Eds.): ICCCI 2012, Part I, LNAI 7653, pp. 395–404, 2012.

consider friend's opinion to purchase some products or choose movies. The contents or information recommendation from the friends are reliable because the recommendation is accepted based on trust in the friends. According to [1], [2], [5], [6], [13], item recommendation based on social network can improve the recommendation accuracy with trust. Social network, which is constructed based on trust element, presents relations between users using graph model [15].

In this paper, we propose a novel recommendation approach that applies advanced user modeling using relationships between social network users. In the user modeling, we analyze both direct and indirect social relationships and deal with the rating histories which already considered in traditional recommendation systems. The user relations are represented by six degrees of Kevin Bacon [14], [17]. Weight 'one' is given to those who have connections directly, and weight 'zero' is given to those who are over six steps away and have no relation to each other. According to a research of Kevin Bacon, everybody can know certain people through six depths of people. Otherwise, we can regard certain people as unrelated people. The remainder of the paper is organized as follows. The next section will describe~~s~~ research background of collaborative filtering and social network. Section 3 illustrates a proposed advanced user modeling. Section 4 will present the evaluation results of proposed approach. Finally, we conclude this paper and discuss future works in the last section.

2 Background Knowledge

Collaborative filtering is one of the widely used methods for personalized recommendation. The main idea is to predict the interest level of a target user on unrated items based on the available rating information from the nearest neighbors [7], [8].

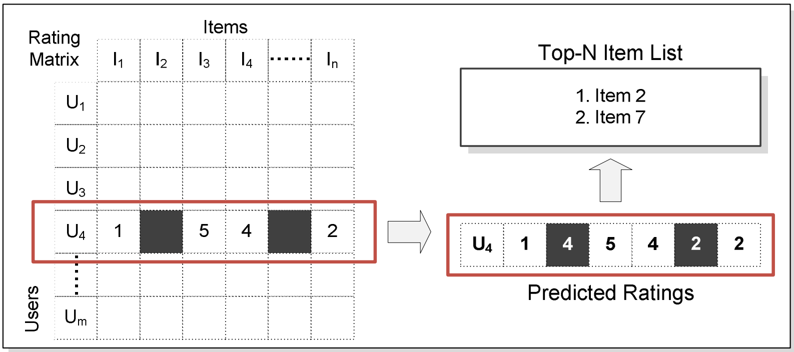


Fig. 1. Process of traditional collaborative filtering

Fig. 1 illustrates the traditional collaborative filtering process. Firstly, a system identifies the nearest neighbors of a target user. In other words, the system will identifies people who is closest to user's tendencies, and then recommends some items based on the analysis. In order to analyze user tendencies, the system collects user behaviors including rating histories. The system will predict preferences on items are not evaluated by a target user. Finally, the system chooses Top-N items that have high prediction values and recommends them to the target user.

The collaborative filtering methods can be classified into user-based collaborative filtering and item-based collaborative filtering. User-based collaborative filtering is a method to construct nearest neighbor considering users that prefer similar items and recommend items based on the neighbor. In item-based collaborative filtering, similarities between each item are calculated and appropriate items are recommended to users. In user-based collaborative filtering, predictions for unrated items are conducted based on rating information of nearest neighbors and items that have a high prediction score are recommended. In order to build the nearest neighbors that have similar tendencies, similarity between a target user and each user is calculated using Equation (1).

$$w_p(a,i) = \frac{\sum_{j \in \text{Commonly Rated Items}} (r_{aj} - \overline{r_a})(r_{ij} - \overline{r_i})}{\sqrt{\sum_{j \in \text{Commonly Rated Items}} (r_{aj} - \overline{r_a})^2 \sum_{j \in \text{Commonly Rated Items}} (r_{ij} - \overline{r_i})^2}} \tag{1}$$

In Equation (1), a and i denote a target user and one of the nearest neighbor, respectively. r_a means a score of an item that the target user rated and average of items that target user rated is represented using $\overline{r_a}$. A prediction for a certain item is performed using Equation (2) based on similarity values between target user and each user using Equation (1).

$$r_{aj} = \overline{r_a} + \frac{\sum_i w(a,i)(r_{ij} - \overline{r_i})}{\sum_i |w(a,i)|} \tag{2}$$

To predict an unrated item, rating of other user that evaluates the same item and the similarity value between a target user and the other user is used in Equation (2). After predicting preferences of all unrated items, Top-N items are recommended to the target user. Item-based collaborative filtering attempts to recommend items that have high prediction score using similarity between items that target user rated. Item-based collaborative filtering also considers co-rated items. Same equations are used like user-based collaborative filtering, but a and i denote a target item and another item.

In this paper, we apply user-based collaborative filtering to recommend items to users and consider social influence. Generally, we usually consult with acquaintances or close friends to purchase a product or choose a restaurant even a movie to watch. The opinions are valuable because of their experiences with the product and the recommendation is performed based on tendencies of the questioner.

3 Social Filtering Using Social Relationship

3.1 System Architecture

In this section, we describe a proposed recommendation system that recommends movies based on rating information for movies that users already watched and neighbors that have social relationship with the user. Fig. 2 shows the overall process of the proposed system.

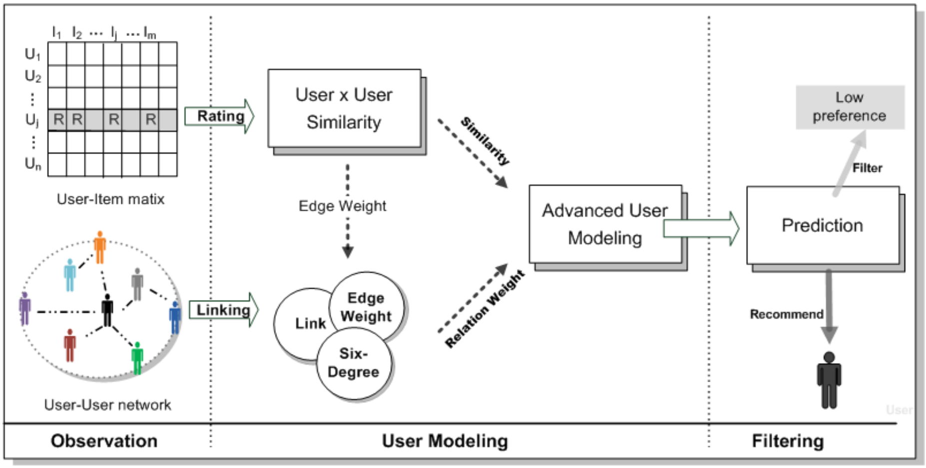


Fig. 2. Overall process of social filtering using social relationship

All movies are scaled from one to five to represent how the users liked the movies. The proposed system constructs User-Item matrix using rating information of each user for movies seen by the user. And User-User network is built based on link data that express explicit relations with friends. In order to identify users who are similar to a target user, the similarities are calculated using Equation (1), and then User-User similarity matrix is built. The values of all elements of Equation (1) are obtained from User-Item matrix.

3.2 Advanced User Modeling

This research is an expansion of traditional collaborative filtering system and considers social relationship between users under the premise that if users have a closer relationship with other users, the preferred items are similar. Users are normally influenced by opinions of reliable friends rather than reviews of corresponding web sites. Considering this point of view, we calculate relation weights between the users using six degrees of Kevin Bacon in Fig. 3. Six degrees of separation is a theory that everyone is on average six steps away from any other person.

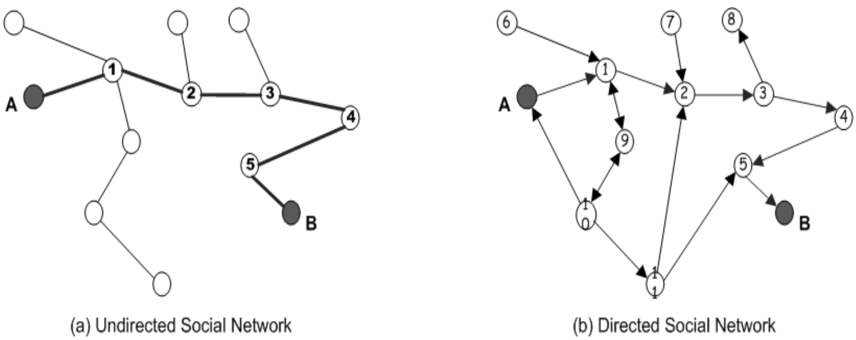


Fig. 3. User-User relationship in the social network

We defined a relation weight with the six degrees. If arbitrary two nodes are directly connected, we define the weight as 1.0 point and whenever another node is inserted into a path, the weight value is decreased by 0.2 point. If more than five nodes exist, the weight value is 0.0. The weight set of social relationship is comprised of six elements like $w_r = \{1.0, 0.8, 0.6, 0.4, 0.2, 0.0\}$.

Relationship between the users can be represented like the (a) and (b) of Fig. 3, which are undirected and directed graphs, respectively. Let's see the relation weight values of node A and B from each graph. In Fig. 3(a), node A and B are connected after passing the node 1, 2, 3, 4, and 5. In this case, there exist five nodes between node A and B. It means two people have no connection with each other and relation weight is 0.0. However, for node A and 1, they have a close relation and relation weight is 1.0 because they are connected directly. In Fig. 3(b), there are three paths connecting node A and B such as <A, 1, 2, 3, 4, 5, B>, <A, 1, 9, 10, 11, 2, 3, 4, 5, B> and <A, 1, 9, 10, 11, 5, B> with the directivity for the paths. The paths pass through 5 or 8 people. As a result, node A and B have relation weight 0.0 due to the number of nodes exist on the paths. According to user's scope of activities, several paths between the user and arbitrary user can exist. Each path can have a weight value differently, depending on which path is selected among the paths. In our research, we find all paths between two users and compare each value of the paths. If some paths have same weight values, then we calculate standard deviation about all users exist on each path. We use one path has the smallest standard deviation to build advanced user modeling. All standard deviations on the paths are calculated by using Equation (3).

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2} \tag{3}$$

In Equation (3), n denotes the number of all users on a path and x_i is an edge weight of neighboring two users. The edge weight means a user-user similarity calculated by using Equation (1). $\overline{x}$ represents an average value of all edge weights on a certain path. We choose a path that has the smallest standard deviation, and assign the value as a relation weight value to the path. If there are some paths that have same standard deviation, then we calculate evaluation function $f'(u_i, u_j)$ for each path and choose a path that has the smallest value from the function.

$$f'\left(u_i, u_j\right) = S\left(u_i, u_j\right) - \overline{x} \tag{4}$$

For advanced user modeling, a path has the smallest standard deviation is selected and advanced user-user similarity is calculated using Equation (5).

$$ASim\left(u_i, u_j\right) = Sim\left(u_i, u_j\right) + \left(1 - Sim\left(u_i, u_j\right)\right)\left(w_r\right) \tag{5}$$

Equation (5) is used to calculate advanced user similarity of user i and j. $Sim(u_i, u_j)$ is a similarity value based on the evaluation record of two users through User-User similarity. w_r is a social relation weight value of the two users. Our proposed method improves the user modeling using social relationships between users and we can yield more accurate recommendation using the advanced user modeling. After completing the advanced user modeling, our system performs a ~~predictions~~ process for unrated items of a target user using Equation (2). Finally, Top-N items are recommended to the user among the predicted items.

4 Experimental and Result

4.1 Experimental Evaluation Method

We perform a performance comparison between social network-based collaborative filtering (SCF) and traditional collaborative filtering (TCF) to verify our proposed method. In our experiments, we use movie evaluation data and link data between users from the epinion.com[3]. Each movie is rated using a discrete scale from one to five. Collected dataset consists of 981 users and 105,754 rated items. Link data contains relation information of 21,476 users and each link has directivity. The number 1 means that a user trusts another user. For example, 'A B 1' means that user A trusts user B. In order to evaluate accuracy of preference prediction, we use 'All But One' protocol on items that user already rated. In 'All But One' protocol, a system chooses one item among all items that a user rated, and predicts preference on the item using remaining dataset except the item [3].

We use data of Table. 1 to evaluate the correctness of decision support.

Table 1. Evaluate data of decision support accuracy

	Recommended	Not recommended	Range of preference
Relevant	N_{rs}	N_{rn}	3, 4, 5
Irrelevant	N_{is}	N_{in}	1, 2

Recall and precision are used to measure the prediction accuracy of the recommendation system. Equation (6) presents the precision and recall. Precision means how many items have 3, 4 or 5 preference points among the recommended items and recall indicates the number of recommended items among all items have high preference values. In Equation (6), N_{rs} is the number of items has the preference of 3, 4 or 5 points which is a higher preference among recommended items. N_{is} indicates the number of items has the preference of 1 or 2 points that mean lower preference. In other words, it means how many predictions are incorrect. N_{rn} represent the number of items do not recommend but real rating is 3, 4 or 5.

$$\Pr ecision = \frac{N_{rs}}{N_{rs} + N_{is}}, \quad \operatorname{Re} call = \frac{N_{rs}}{N_{rs} + N_{rn}} \tag{6}$$

Both precision and recall is appropriate to evaluate whether the recommendation system provides good recommendations to users. The bigger values of both are, the

[3] epinion.com: `http://www.epinion.com`

better the performance is. However, the precision value is normally in inverse proportion to the recall value. For example, if the N is increased, the value of recall is increased but the value of precision is decreased. F1-measure is the harmonic mean of precision and recall. If both precision and recall have high values, F1-measure also increased in value. Equation (7) denotes F1-measure.

$$F1-measure = \frac{2 \times \mathrm{Re}call \times \mathrm{Pr}ecision}{\mathrm{Re}call + \mathrm{Pr}ecision} \tag{7}$$

We also use the Hit Ratio (HR) evaluate recommendation quality using Equation (8). Through the HR measure, we evaluate how many items are appropriately recommended to users compared with ratio of the number of hits.

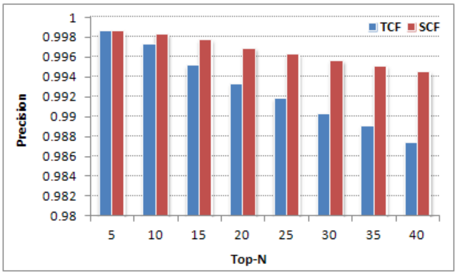
$$HR = \frac{|Test_u \cap TopN_u|}{|Test_u|} \tag{8}$$

In Equation (8), $Test_u$ is the movie list of user u in the test data and $TopN_u$ is a Top-N recommended movie list for user u. Although the recommendation system shows high accuracy for existing items, the system sometimes did not well perform with the new items. If the system recommends items that the users positively rated in practice, we conclude the recommendation is useful for the users. In other words, the higher the HR value is, the better the recommendation system is for the users. The HR thus is suitable for our experiment dataset to verify recommendation quality.

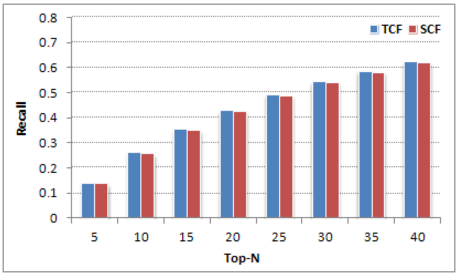
4.2 Experimental Result and Discussion

We performed comparative analysis to evaluate recommendation performance of proposed social filtering based on recall, precision and F1-measure. Our system predicts items that users have not yet rated, and then recommends Top-N items, which have higher prediction value than others, to each user.

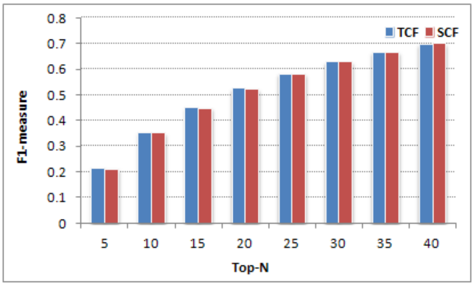
Fig. 4(a) presents the precision of the item recommendation with increasing Top-N. Both TCF and SCF show better recommendation precision when the number of recommended items is small. With the increasing number of N, SCF draws the better performance than TCF. From this, we verify that the more the number of items increased, recommendation using SCF is more correct than TCF. The comparison of recall of recommendation is shown in Fig. 4(b). As the number of N increases, recall values of TCF and SCF are improved. In contrast with the precision, TCF shows higher recall value as 0.01 on average. F1-measure graph shows similar pattern to recall case. With increasing the number of N, values of F1-measure is increased. Performance of TCF is better than SCF as 0.05 on average. Through performance experiments, we can exactly recommend items to user using social relationships of users when rated items are rare. We also expect that we can improve the recommendation performance if social relation information of users is enough.



(a) Precision



(b) Recall



(c) F1-measure

Fig. 4. Performance evaluate

To evaluate the quality of top-N recommendation, we measure the hit rate (HR) on TCF and SCF. Each HR is measured with increased number of recommended items. Overall, the recommendation quality of SCF is quite outstanding than TCF. In Fig. 5, HR of TCF sinks sharply with the number of N. For recommendation over twenty items, TCF could not recommend items that is appropriate to user preferences due to low HR value as 0.005. In contrast, HR of SCF is gradually going down. Until N of SCF is 25, SCF shows the better HR than TCF. HR of SCF for 40 items has a more high value than HR of TCF for 10 items.

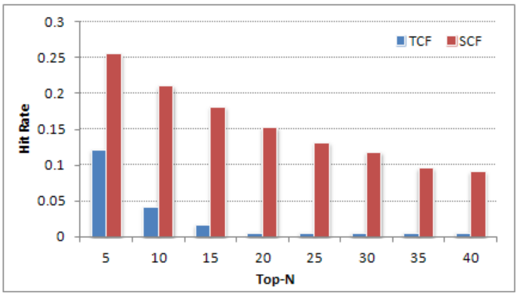


Fig. 5. Hit rate (HR) with increasing N (number of recommended items)

From the results of HR experiments, we verify that recommendation system using user modeling based on relation weight of users improves the recommendation quality. In other words, we can say that preferences of users who trust each other are similar. From all experimental results and analysis, overall recommendation performance of SCF is similar to TCF. However, SCF has improved 0.075 and 1.030 in accuracy and quality perspective, compared with TCF. As a result, we can conclude that our proposed method, which applied social filtering using social relationships of users, recommends more appropriate items to target users.

5 Conclusion and Future Work

We proposed a SCF method which employs an advanced user modeling using social relation weight for a movie recommendation system. In this method, we assign relation weight to social relationship of users using six degrees of Kevin Bacon by considering both direct and indirect relations between users. Those who have direct relation have weight value as one, and weight 0 is given to those who are over six steps away and have no relation to each other. In order to analyze the SCF performance, we compare the performance results between in SCF and TCF. F1-measure of SCF and TCF is almost comparable. From the recommendation quality perspective, the HR of SCF is better than TCF. SCF results quite improve the recommendation quality since HR value is gradually going down although the number of N is increased. From the experimental results, we verify that SCF is more effective than TCF and improve the recommendation quality. While proposed SCF improves recommendation quality, the recommendation performance is similar to TCF. In order to improve performance of SCF, we need to analyze item attributes such as movie genres, actor, screening year, and running time. We expect that if we use user behaviors such as comment, visit, trust information, etc, from social network services like [4], [9], [10], [11], [12], [16], in addition to social relationship, then the system can recommend items to user more accurately.

Acknowledgement. This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MEST) (No. 2012-0005500).

References

1. Bellogin, A., Cantador, I., Diez, F., Castells, P., Chavarriaga, E.: An Empirical Comparison of Social, Collaborative Filtering, and Hybrid Recommenders. ACM Transactions on Intelligent Systems and Technology, Special Issue on Context-Aware Movie Recommendation (in press, 2012)
2. Bian, L., Holtzman, H.: Online Friend Recommendation through Personality Matching and Collaborative Filtering. In: The 15th International Conference on Mobile Ubiquitous Computing, System, Services and Technologies, pp. 230–235. IARIA (2011)

3. Breese, J.S., Heckerman, D., Kadie, C.: Empirical Analysis of Predictive Algorithms for Collaborative Filtering. In: Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence, pp. 43–52. Morgan Kaufmann (1998)
4. Debnath, S., Ganguly, N., Mitra, P.: Feature Weighting in Content Based Recommendation System Using Social Network Analysis. In: Proceedings of the 17th International Conference on World Wide Web, pp. 1041–1042. ACM (2008)
5. DuBois, T., Golbeck, J., Kleint, J., Srinivasan, A.: Improving Recommendation Accuracy by Clustering Social Networks with Trust. In: Proceedings of the 3rd ACM Conference on Recommender Systems Workshop: Recommender Systems and the Social Web, New York (2009)
6. He, C., Tang, Y., Chen, G., Fu, C., Wu, L.: Collaboative Recommendation Model Based on Social Network and Its Application. Journal of Convergence Information Technology (JCIT) 7(2), 253–261 (2012)
7. Herlocker, J.L., Konstan, J.A., Borchers, A., Riedl, J.: An Algorithmic Framework for Performing Collaborative Filtering. In: Proceeding of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 230–237. ACM (1999)
8. Herlocker, J.L., Konstan, J.A., Terveen, L.G., Riedl, J.T.: Evaluating Collaborative Filtering Recommender Systems. ACM Transactions on Information Systems 22(1), 5–53 (2004)
9. Kuter, U., Golbeck, J.: SUNNY: A New Algorithm for Trust Inference in Social Networks Using Probabilistic Confidence Models. In: Proceedings of the 22nd National Conference on Artificial Intelligence, Vancouver, vol. 2, pp. 1377–1383 (2007)
10. Machanavajjhala, A., Korolova, A., Sarma, A.D.: Personalized Social Recommendations – Accurate or Private? In: Proceeding of the 37th International Conference on Very Large Data Bases, vol. 4(7), pp. 440–450. VLDB Endowment (2011)
11. Palau, J., Montaner, M., López, B., de la Rosa, J.L.: Collaboration Analysis in Recommender Systems Using Social Networks. In: Klusch, M., Ossowski, S., Kashyap, V., Unland, R. (eds.) CIA 2004. LNCS (LNAI), vol. 3191, pp. 137–151. Springer, Heidelberg (2004)
12. Procter, R., McKinlay, A.: Social Affordances and Implicit Ratings for Social Filtering on the Web. In: Proceedings of the 5th DELOS Workshop on Filtering and Collaborative Filtering, pp. 89–96. ERCIM Press (1997)
13. Said, A., De Luca, E.W., Albayrak, S.: How social relationships affect user similarities. In: Proceedings of the International Conference on Intelligent User Interfaces Workshop on Social Recommender Systems, Hong Kong (2010)
14. Shu, W., Chuang, Y.: The perceived benefits of six-degree-separation social networks. Internet Research 21(1), 26–45 (2011)
15. Song, J., Liu, W., Chen, S.: Relation Grid: A Social Relationship Network Model. In: Proceedings of the 1st International Conference on Semantics, Knowledge and Grid, pp. 248–255. IEEE (2005)
16. Tiroshi, A., Kuflik, T., Kay, J., Kummerfeld, B.: Recommender Systems and the Social Web. In: Ardissono, L., Kuflik, T. (eds.) UMAP Workshops 2011. LNCS, vol. 7138, pp. 60–70. Springer, Heidelberg (2012)
17. Zhang, L., Tu, W.: Six Degrees of Separation in Online Society. Proceedings of the Web Science 3(12), 1–5 (2009)