A novel similarity index for better personalized recommendation

LING-JIAO CHEN[1,2], ZI-KE ZHANG[3], JIN-HU LIU[1,2], JIAN GAO[1,2] (a) and TAO ZHOU[1,2]

[1] *CompleX Lab, Web Sciences Center, University of Electronic Science and Technology of China - Chengdu 611731, PRC*
[2] *Big Data Research Center, University of Electronic Science and Technology of China - Chengdu 611731, PRC*
[3] *Institute of Information Economy, Hangzhou Normal University - Hangzhou 310036, PRC*

PACS 89.65.-s – Social and economic systems
PACS 89.20.Hh – World Wide Web, Internet
PACS 89.75.-k – Complex systems

Abstract –Recommender systems benefit us in tackling the problem of information overload by predicting our potential choices among diverse niche objects. So far, a variety of personalized recommendation algorithms have been proposed and most of them are based on similarities, such as collaborative filtering and mass diffusion. Here, we propose a novel similarity index named CosRA, which combines advantages of both the cosine index and the resource-allocation (RA) index. By applying the CosRA index to real recommender systems including MovieLens, Netflix and RYM, we show that the CosRA-based method has better performance in accuracy, diversity and novelty than the state-of-the-art methods. Moreover, the CosRA index is free of parameters, which is a significant advantage in real applications. Further experiments show that the introduction of two turnable parameters cannot remarkably improve the overall performance of CosRA.

Introduction. – The development of the Internet and e-commerce makes our lives more convenient as billions of products are available online [1]. Meanwhile, the problem of information overload plagues us everyday as it is much harder to dig out relevant objects than ever [2]. Thus far, personalized recommendation was thought to be the most promising way to efficiently solve the problem of information overload. Personalized recommendation benefits both buyers and sellers, and it is now playing an increasing role in our online social lives. Many online platforms (Amazon, eBay, AdaptiveInfo, Taobao, etc) have introduced personalized recommendation systems [3], which predict users' potential choices by analyzing historical behaviors of users, attributes of objects, and so on [4]. For example, Amazon.com recommends books by analysing users' purchase records [5], and AdaptiveInfo.com recommends news by using users' reading histories [6]. In recent years, personalized recommendation has found widely applications [7] in recommending movies [8], research articles [9], driving routes [10], locations [11,12] and so on.

So far, a variety of personalized recommendation algorithms have been proposed, among which user-based (UCF) and item-based collaborative filtering (ICF) are the most representative ones [13]. UCF and ICF are respectively based on the weighted combination of similar users' opinions and the similarity between items [14]. Recently, many diffusion-based algorithms are proposed by introducing some physical dynamics into the recommender systems, such as mass diffusion (MD) [15] and heat conduction (HC) [16]. The simplest version of MD can be considered as a two-step resource-allocation process in bipartite networks [17]. Later, Zhou *et al.* [18] and Jia *et al.* [19] proposed two algorithms by giving new strategies in the initial resource distribution, Zhou *et al.* [20] proposed a hybrid method that combines both MD and HC, Lü *et al.* [21] proposed a preferential diffusion method by considering node weights in redistributing resources, and Liu *et al.* [22] proposed a weighted heat conduction algorithm by considering edge weighting. Reviews of previous literatures can be found in Refs. [23] and [24].

Essentially, the aforementioned collaborative filtering and diffusion-based methods are similarity-based methods [25]. In collaborative filtering, the most commonly used index is cosine similarity [26]. However, it strongly tends to recommend popular objects, resulting in accurate yet less-diverse recommendations [27]. In diffusion-based methods, the diffusion is indeed a resource allocation process, and thus the node similarity is characterized

(a) E-mail: gaojian08@hotmail.com

by the resource allocation (RA) index [28]. The RA index gives high priority to assign resources to large-degree nodes, which also leads to high accuracy but low diversity of MD [29]. In fact, the cosine index and RA index are complementary to each other, and thus to combine the two can possibly improve the overall performance. How to design a suitable similarity index for personalized recommendation is still an open issue and such index can also be applied in characterizing many network structures and functions.

In this letter, we propose a novel similarity index for better personalized recommendation. The similarity index, named CosRA, combines advantages of both the cosine index and the RA index. Further, we propose a personalized recommendation algorithm based on the CosRA index. Extensive experiments on three real data sets suggest that the CosRA-based method performs better in accuracy, diversity and novelty than the state-of-the-art methods. To do a more systematic analysis, we extend the CosRA index to a general form by introducing two turnable parameters. Interestingly, results suggest that the original CosRA index is almost optimal, and its effectiveness cannot be remarkably improved by adjusting the parameters. Such feature is significant since a parameter-free index is more applicable than a parameter-dependent index. Our work sheds lights on the importance of a suitable similarity index in enhancing the overall performance of personalized recommendation.

Method. – A recommender system can be naturally described by a user-object bipartite network $G(U, O, E)$, where $U = \{u_1, u_2, \ldots, u_m\}$, $O = \{o_1, o_2, \ldots, o_n\}$ and $E = \{e_1, e_2, \ldots, e_z\}$ are sets of users, objects and links, respectively [30]. To distinguish object-related and user-related indices, we respectively use Greek and Latin letters for them. The bipartite network can be naturally represented by a adjacency matrix A, whose element $a_{i\alpha} = 1$ if there is a link connecting node U_i and node O_α, *i.e.*, user i has collected object α, otherwise $a_{i\alpha} = 0$. The main purpose of recommendation algorithms is to provide a target user with a ranking list of his uncollected objects. For user i, the recommendation list with length L is denoted as o_i^L, that is to say, o_i^L is a set of L objects with the highest recommendation scores for user i.

Firstly, two widely used similarity indices in recommendation algorithms are introduced, namely, the cosine index and the RA index. Taking objects α and β as an example, the cosine index between them is defined as

$$S_{\alpha\beta}^{Cos} = \frac{1}{\sqrt{k_{o_\alpha} k_{o_\beta}}} \sum_{i=1}^{m} a_{i\alpha} a_{i\beta}, \tag{1}$$

where k_{o_α} and k_{o_β} are respectively the degrees of objects α and β. For objects α and β, the RA index is defined as

$$S_{\alpha\beta}^{RA} = \sum_{i=1}^{m} \frac{a_{i\alpha} a_{i\beta}}{k_{u_i}}, \tag{2}$$

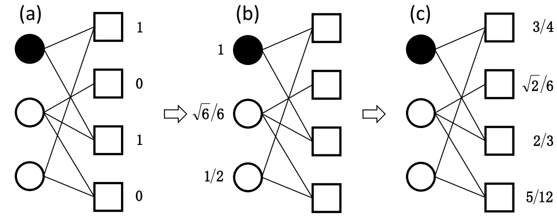

Fig. 1: Illustration of the CosRA-based method. (a) Initially, for a target user (colored black), the resources of objects are initialized by eq. (4). (b) Then, objects distribute the resources to the users who have collected them. (c) Finally, users redistribute the resources to the objects that they have collected. The processes in panels (b) and (c) are characterized by eq. (5).

where k_{u_i} is user i's degree. Indeed, the RA index is the element of the transformation matrix in the MD process.

The CosRA index can be seen as a combination of the cosine index and the RA index. For objects α and β, the CosRA index is defined as

$$S_{\alpha\beta}^{CosRA} = \frac{1}{\sqrt{k_{o_\alpha} k_{o_\beta}}} \sum_{i=1}^{m} \frac{a_{i\alpha} a_{i\beta}}{k_{u_i}}. \tag{3}$$

Based on the CosRA index, we further propose a personalized recommendation algorithm, which works as follows: Fistly, for user i, the resource of object α is initialized as

$$f_\alpha^{(i)} = a_{i\alpha}. \tag{4}$$

Secondly, the resources are redistributed via

$$f'^{(i)} = S^{CosRA} f^{(i)}, \tag{5}$$

where $f^{(i)}$ is an n-dimensional vector recording all objects' initial resources, given i the target user, and $f'^{(i)}$ is the final resource vector. At last, all objects are sorted by their resources $f'^{(i)}$ and the top-L uncollected objects are recommended to user i. An illustration of the CosRA-based method is shown in fig. 1.

Data. – Three commonly studied data sets, namely, MovieLens, Netflix and RYM, are used to test the performance of different methods. MovieLens data set is provided by the GroupLens project at University of Minnesota (www.grouplens.org). The data set uses a 5-point rating scale from 1 to 5 (*i.e.*, worst to best). When building the bipartite network, we only consider the links with ratings ≥ 3. After coarse graining, the data set contains 82520 links. Netflix is a huge data set that released by the DVD rental company Netflix for its Netflix Prize contest (www.netflixprize.com). The ratings are also given on a 5-point scale. Analogously, only links with ratings ≥ 3 are reserved. Then, we extract a small sampling data set by randomly choosing 10000 users and taking the associated 701946 links. RYM data set is publicly available on the music ratings website (rateyourmusic.com). The ratings are given on a 10-point scale from 1 to 10 (*i.e.*, worst to best). Here, only links with ratings ≥ 6 are considered,

Table 1: Basic statistics of the three real recommender systems.

Data	Users	Objects	Links	Sparsity
MovieLens	943	1574	82520	5.56×10^{-2}
Netflix	10000	5640	701946	1.24×10^{-2}
RYM	33197	5234	609792	3.51×10^{-3}

and thus the final data contains 609792 links. The basic statistics of the data sets are summarized in table 1.

Metrics. – In order to test the algorithmic performance, we randomly divide one data set into two parts: the training set and the testing set with ratios 0.9 and 0.1, respectively. For each user, we provide the recommendation list based on the training set, and test the performance on the testing set. To quantify the performance of recommendation, we here apply six widely used metrics, including three accuracy metrics (AUC, Precision and Recall), two diversity metrics (Inter-similarity and Intra-similarity), and one novelty metric (Popularity). In the following, we will briefly introduce these metrics.

Accuracy is one of the most important metric in evaluating the quality of recommendation algorithms. We fist introduce AUC (area under the ROC curve) [31]. Given the ranks of objects in the testing set, AUC value can be interpreted as the probability that a randomly chosen collected object is ranked higher than a randomly chosen un-collected object. To calculate AUC, at each time, a pair of collected and un-collected objects are selected to compare their resources. After N times independent comparisons, if there are N_1 times the collected object has more resources and N_2 times their resources are the same, the average value of AUC for all users is defined as [28]

$$AUC = \frac{1}{m}\sum_{i=1}^{m}\frac{(N_1 + 0.5N_2)}{N}. \tag{6}$$

Larger AUC value means higher algorithmic accuracy.

Then we introduce two L-dependent accuracy metrics, namely, Precision and Recall [32]. Precision is defined as the ratio of the number of recommended objects appeared in the testing set to the total number of recommended objects. Mathematically, for all user, the average value of Precision is defined as

$$P(L) = \frac{1}{m}\sum_{i=1}^{m}\frac{d_i(L)}{L}, \tag{7}$$

where $d_i(L)$ is the number of common objects in the testing set and the recommendation list with length L. Recall is defined as the ratio of the number of recommended objects appeared in user's recommendation list to the total number of objects in the test set. Mathematically, for all user, the average value of Recall is defined as

$$R(L) = \frac{1}{m}\sum_{i=1}^{m}\frac{d_i(L)}{D(i)}, \tag{8}$$

where $D(i)$ is the number of objects in the testing set. Larger Precision and Recall mean higher accuracy.

Diversity is an important metric in evaluating the variety of objects that recommended by personalized recommendation algorithms. One of the diversity metrics is Inter-similarity. To quantify the Inter-similarity, we use Hamming distance [18]. The average value of Hamming distance for all users is defined as

$$H(L) = \frac{1}{m(m-1)}\sum_{i=1}^{m}\sum_{j=1}^{m}(1 - \frac{C(i,j)}{L}), \tag{9}$$

where $C(i,j) = |o_i^L \cap o_j^L|$ is the number of common objects in user i's and j's recommendation lists. Larger value of Hamming distance corresponds to higher diversity.

Another diversity metric is Intra-similarity [33], which is defined as the similarity between objects appeared in target user's recommendation list. Mathematically, for all users, the average value of Intra-similarity is defined as

$$I(L) = \frac{1}{mL(L-1)}\sum_{i=1}^{m}\sum_{o_\alpha, o_\beta \in o_i^L, \alpha \neq \beta} S_{\alpha\beta}^{Cos}, \tag{10}$$

where $S_{\alpha\beta}^{Cos}$ is the cosine similarity between objects α and β in user i's recommendation list o_i^L with length L. Smaller value of Intra-similarity means higher diversity.

Novelty [23] is an important metric aiming to quantify the ability of an algorithm to generate novel (*i.e.*, unpopular) and unexpected results. Here, we use the average Popularity of the recommended objects to quantify the novelty, which is defined as

$$N(L) = \frac{1}{mL}\sum_{i=1}^{m}\sum_{o_\alpha \in o_i^L} k_{o_\alpha}, \tag{11}$$

where k_{o_α} is the degree of object α in user i's recommendation list o_i^L. Smaller value of Popularity indicates higher novelty and potentially better user experience.

Results. – We apply the CosRA-based method to the three real data sets. By comparison, some benchmark methods are also considered, including global ranking (GR), user-based collaborative filtering (UCF), item-based collaborative filtering (ICF), mass diffusion (MD) and heat conduction (HC). In GR, all objects are sorted in the descending order of their degrees and those with the largest degrees are recommended [17]. In UCF, the target user will be recommended the objects collected by the users sharing similar tastes [34]. Analogously, in ICF, the target user will be recommended objects similar to the ones he preferred in the past [33]. We adopt the cosine similarity to quantify the user and object similarity in UCF and ICF, respectively. MD and HC both can be considered as resource allocation processes on the user-object bipartite network [20, 35]. Nevertheless, they has several distinguishing characteristics. The total amount of resources is conserved in MD instead of in HC. The transformation

Table 2: Values of the six evaluation metrics after applying different recommendation algorithms on the three data sets. The length of recommendation list is set as $L = 50$. The results are averaged over 10 independent realizations. For each data set and each metric, the best result is emphasized by bold.

Mov.L.	AUC	P	R	H	I	N
GR	0.863	0.061	0.346	0.393	0.406	254
UCF	0.887	0.074	0.458	0.548	0.394	241
ICF	0.887	0.077	0.476	0.671	0.412	210
MD	0.897	0.079	0.509	0.616	0.354	229
HC	0.842	0.021	0.122	**0.855**	**0.054**	**22**
CosRA	**0.909**	**0.087**	**0.562**	0.722	0.334	203
Netflix	AUC	P	R	H	I	N
GR	0.932	0.044	0.369	0.355	0.374	2404
UCF	0.939	0.049	0.408	0.405	0.375	2373
ICF	0.936	0.053	0.425	0.557	0.374	2055
MD	0.947	0.050	0.423	0.425	0.368	2358
HC	0.887	0.001	0.019	**0.797**	**0.004**	**11**
CosRA	**0.950**	**0.053**	**0.447**	0.481	0.361	2287
RYM	AUC	P	R	H	I	N
GR	0.855	0.011	0.139	0.129	0.188	2161
UCF	0.919	0.015	0.417	0.759	0.168	1123
ICF	0.932	0.017	0.445	0.914	0.177	656
MD	0.941	0.017	0.471	0.789	0.155	1089
HC	0.932	0.013	0.358	**0.948**	**0.056**	**212**
CosRA	**0.952**	**0.019**	**0.482**	0.879	0.144	817

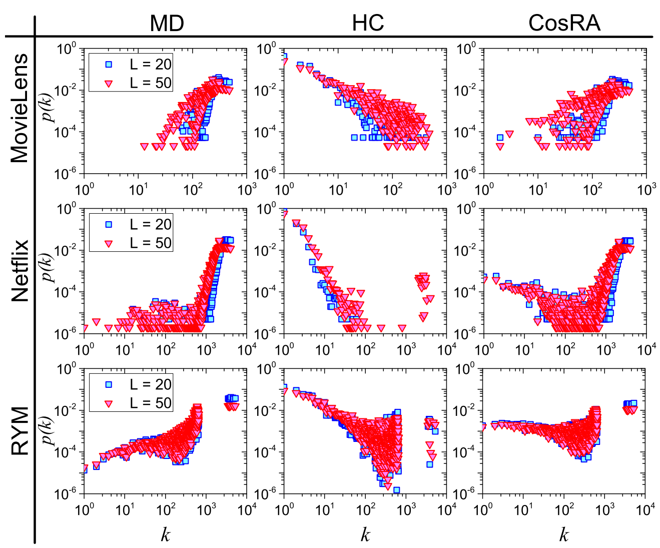

Fig. 2: Degree distribution of the recommended objects after applying MD, HC and CosRA-based methods on the three data sets. Results are shown for one realization in log-log plot. Blue squares and red triangles correspond to results under $L = 20$ and $L = 50$, respectively.

matrices in MD and HC are mutually transposed as the matrix is normalized by column in MD and by row in HC. More details in implementing the five benchmark methods can be found in the survey paper [23].

Results of the six evaluation metrics are shown in table 2. When focusing on the accuracy, CosRA-based method has remarkable advantage towards the other five methods on all data sets, as indicated by the higher values of AUC, Precision and Recall. The AUC values in CosRA-based method is 0.909, 0.950 and 0.952 for MovieLens, Netflix and RYM, respectively. Meanwhile, ICF and HC have poor performance, as indicated by the generally smaller values of accuracy metrics, especially for Recall. When focusing on diversity, on the one hand, the values of Inter-similarity (Hamming distance) in CosRA-based method are much larger than those in GR, UCF and MD and not far behind of those in ICF and HC. On the other hand, the values of Intra-similarity in CosRA-based method are smaller than those in GR, UCF and MD. These results suggest that CosRA-based method has advantage in diversity. When focusing on novelty, CosRA-based method remarkably outperforms GR, UCF and MD as indicated by the smaller Popularity values, although ICF and HC have the best performance. Based on these observations, it can be concluded that CosRA-based method has better accuracy, well diversity and novelty in personalized recommendation.

To better understand the mechanism of CosRA-based method, we show the degree distribution of the recommended objects for all users in fig. 2. To make a comparison, MD and HC are also studied. In MD, there is a high probability for large-degree objects being recommended (see the first column of fig. 2), whereas HC prefers to recommend small-degree objects (see the second column of fig. 2). The two strong trends of MD and HC both have disadvantages, resulting in poor diversity and novelty of MD and low accuracy of HC. Fortunately, CosRA-based method finds a balance among accuracy and diversity by recommending both large-degree and small-degree objects without any strong bias (see the last column of fig. 2).

For a more systematic analysis on the CosRA index, we extend it to a more general form by introducing two turnable parameters, η_1 and η_2. Mathematically, the generalized CosRA index is formulated as

$$S_{\alpha\beta}^{CosRA*} = \frac{1}{(k_{o_\alpha} k_{o_\beta})^{-\eta_2}} \sum_{l=1}^{m} \frac{a_{l\alpha} a_{l\beta}}{(k_{u_l})^{-2\eta_1}}. \qquad (12)$$

Notice that the original CosRA index is a special case when $\eta_1 = \eta_2 = -0.5$. By varying η_1 and η_2, we study how the similarity index affects the performance of recommendation. As shown in fig. 3, the generalized CosRA-based method achieves its best performance when both η_1 and η_2 are around -0.5. Specifically, when focusing on accuracy, the values of AUC, Precision and Recall reach their maximum when η_1 and η_2 are around -0.5, as marked by vertical and horizontal dash lines in the first three columns of fig. 3. The accuracy metrics perform best at almost the same parameters on all data sets, which is a strong evidence that the optimal parameters, $\eta_1 = -0.5$

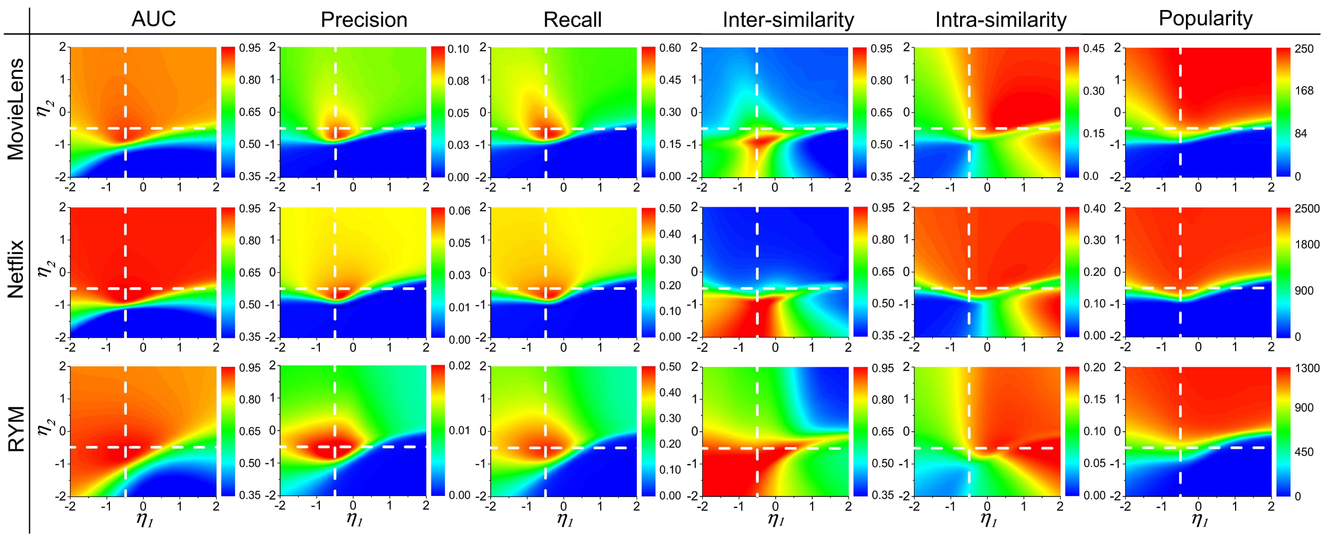

Fig. 3: (Color online) Performance of the generalized CosRA-based method after tested on the three data sets. The parameters η_1 and η_2 are varying from -2 to 2. Vertical and horizontal dash lines correspond to $\eta_1 = -0.5$ and $\eta_2 = -0.5$, respectively. The length of recommendation list is set as $L = 50$ and the results are not sensitive to the value of L. The results are averaged over 10 independent realizations.

and $\eta_2 = -0.5$, for the generalized CosRA index are universal.

When focusing on diversity, the generalized CosRA-based method has better performance when η_1 and η_2 are smaller than -0.5, as indicated by the larger values of Inter-similarity (Hamming distance) and the smaller values of Intra-similarity in the fourth and fifth columns of fig. 3, respectively. When η_1 and η_2 exceed -0.5, the diversity of the generalized CosRA-based method largely decreases. When focusing on novelty, the diagrams are almost divided into two parts by $\eta_2 \approx -0.5$ and the generalized CosRA-based method has remarkably lower Popularity (*i.e.*, higher novelty) when $\eta_2 < -0.5$ as shown in the last column of fig. 3. That's mainly because smaller η_2 benefits small-degree (*i.e.*, unpopular) objects in receiving resources. After a comprehensive consideration, it can be concluded that the original parameters, $\eta_1 = -0.5$ and $\eta_2 = -0.5$, are almost optimal and the effectiveness of the generalized CosRA index cannot be remarkably improved by adjusting the two parameters.

Conclusions and discussion. – In summary, we have proposed a novel similarity index for better personalized recommendation, which combines advantages of both the cosine index and the resource-allocation index. Based on the proposed index, we further propose a personalized recommendation algorithm. Extensive experiments on real data sets suggest that the proposed algorithm has better accuracy and well diversity and novelty compared with the state-of-the-art methods. To further understand how the similarity index works, we show the degree distribution of the recommended objects for all users. Results suggest that the proposed method does not have strong bias on objects' degrees compared with other benchmark methods. Indeed, the similarity index finds a balance among the three important evaluation metrics and improves the overall algorithmic performance. Further, we extend the similarity index to a more general form, however, results suggest that the original similarity index is almost optimal. That is to say, the similarity index is free of parameters, which is a significant advantage in real applications.

Our work highlights the importance of the similarity index in personalized recommendation and suggests that the adoption of suitable similarity index can enhance the algorithmic performance. By introducing the novel similarity index into the personalized recommendation, not only the accuracy is improved, but also the well diversity and novelty are achieved. Nevertheless, how to balance the accuracy, diversity and novelty in recommender systems is still an open issue. Our work just provides a promising way to deal well with the three metrics by proposing a novel parameter-free similarity index.

Further more, pairwise vertex similarity is a fundamental index for many network functions and dynamics [36]. That is to say, the proposed similarity index can find applications in solving many network-related problems, such as link predication [37, 38], community detection [39–41], spreading activation [42], network evolution [43], web searching [44], microarray data clustering [45], and gene ranking [46] . As future works, we could consider designing more suitable similarity indices for networks [36], studying their effects on the evolution of recommender systems [47], and introducing reputation systems into the personalized recommendation to improve its robustness in resisting spamming attacks [48–50].

* * *

The authors acknowledge Hai-Xing Dai for useful discussions. This work was partially supported by the National Natural Science Foundation of China under Grant Nos. 11222543 and 61433014. TZ acknowledges the Program for New Century Excellent Talents in University under Grant No. NCET-11-0070 and the Special Project of Sichuan Youth Science and Technology Innovation Research Team under Grant No. 2013TD0006.

REFERENCES

[1] Laudon K. C. and Traver C. G., *E-commerce* (Addison-Wesley, Boston) 2011.
[2] Zeng A., Vidmer A., Medo M. and Zhang Y.-C., *EPL*, **105** (2014) 58002.
[3] Rokach L., Shapira B. and Kantor P. B., *Recommender Systems Handbook* (Springer, New York) 2011.
[4] Gualdi S., Medo M. and Zhang Y.-C., *EPL*, **101** (2013) 20008.
[5] Linden G., Smith B. and York J., *IEEE Internet Comput.*, **7** (2003) 76.
[6] Billsus D., Brunk C. A., Evans C., Gladish B. and Pazzani M. J., *Commun. ACM*, **45** (2002) 34.
[7] Schafer J. B., Konstan J. A. and Riedl J., *Data Min. Knowl. Disc.*, **5** (2001) 115.
[8] Liu J.-H., Zhou T., Zhang Z.-K., Yang Z., Liu C. and Li W.-M., *PLoS ONE*, **9** (2014) e113457.
[9] Bogers T. and Van den Bosch A., in *Proceedings of the 2008 ACM Conference on Recommender Systems (RecSys)* (ACM Press) 2008 pp. 287-290.
[10] Ge Y., Xiong H., Tuzhilin A., Xiao K., Gruteser M. and Pazzani M., in *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (SIGKDD)* (ACM Press) 2010 pp. 899–908.
[11] Lian D., Ge Y., Zhang F., Yuan N. J., Xie X., Zhou T. and Rui Y., in *15th IEEE International Conference on Data Mining (ICDM)* (IEEE Press) 2015.
[12] Lian D., Xie X., Zhang F., Yuan N. J., Zhou T. and Rui Y., *IEEE Data Eng. Bull.*, **38** (2015) 35.
[13] Sarwar B., Karypis G., Konstan J. and Riedl J., in *Proceedings of the 10th International Conference on World Wide Web (WWW)* (ACM Press) 2001 pp. 285–295.
[14] Goldberg D., Nichols D., Oki B. M. and Terry D., *Commun. ACM*, **35** (1992) 61.
[15] Zhang Y.-C., Medo M., Ren J., Zhou T., Li T. and Yang F., *EPL*, **80** (2007) 68003.
[16] Zhang Y.-C., Blattner M. and Yu Y.-K., *Phys. Rev. Lett.*, **99** (2007) 154301.
[17] Zhou T., Ren J., Medo M. and Zhang Y.-C., *Phys. Rev. E*, **76** (2007) 046115.
[18] Zhou T., Jiang L.-L., Su R.-Q. and Zhang Y.-C., *EPL*, **81** (2008) 58004.
[19] Jia C.-X., Liu R.-R., Sun D. and Wang B.-H., *Physica A*, **387** (2008) 5887.
[20] Zhou T., Kuscsik Z., Liu J.-G., Medo M., Wakeling J. R. and Zhang Y.-C., *Proc. Natl. Acad. Sci. USA*, **107** (2010) 4511.
[21] Lü L. and Liu W., *Phys. Rev. E*, **83** (2011) 066119.
[22] Liu J.-G., Guo Q. and Zhang Y.-C., *Physica A*, **390** (2011) 2414.
[23] Lü L., Medo M., Yeung C. H., Zhang Y.-C., Zhang Z.-K. and Zhou T., *Phys. Rep.*, **519** (2012) 1.
[24] Bobadilla J., Ortega, F., Hernando A. and Gutiérrez A., *Knowl.-Based Syst.*, **46** (2013) 109.
[25] Deshpande M. and Karypis G., *ACM Trans. Inf. Syst.*, **22** (2004) 143–177.
[26] Zhao Z.-D. and Shang M.-S., in *Proceedings of the 3rd International Conference on Knowledge Discovery and Data Mining (WKDD)* (IEEE Press) 2010 pp. 478-481.
[27] Pan X., Deng G. and Liu J.-G., *Phys. Procedia*, **3** (2010) 1867.
[28] Zhou T., Lü L. and Zhang Y.-C., *Eur. Phys. J. B*, **71** (2009) 623.
[29] Liu J.-G., Zhou T. and Guo Q., *Phys. Rev. E*, **84** (2011) 037101.
[30] Shang M.-S., Lü L., Zhang Y.-C. and Zhou T., *EPL*, **90** (2010) 48006.
[31] Hanley J. A. and McNeil B. J., *Radiology*, **143** (1982) 29.
[32] Herlocker J. L., Konstan J. A., Terveen L. G. and Riedl J. T., *ACM Trans. Inf. Syst.*, **22** (2004) 5.
[33] Zhou T., Su R.-Q., Liu R.-R., Jiang L.-L., Wang B.-H. and Zhang Y.-C., *New J. Phys.*, **11** (2009) 123008.
[34] Liu R.-R., Jia C.-X., Zhou T., Sun D. and Wang B.-H., *Physica A*, **388** (2009) 462.
[35] Ou Q., Jin Y.-D., Zhou T., Wang B.-H. and Yin B.-Q., *Phys. Rev. E*, **75** (2007) 021102.
[36] Leicht E. A., Holme P. and Newman M. E. J., *Phys. Rev. E*, **73** (2006) 026120.
[37] Lü L. and Zhou T., *Physica A*, **6** (2011) 1150.
[38] Lü L., Pan L., Zhou T., Zhang Y. C. and Stanley H. E., *Proc. Natl. Acad. Sci. USA*, **112** (2015) 2325.
[39] Newman M. E. J., *Nat. Phys.*, **8** (2012) 25.
[40] Xiang B., Chen E.-H. and Zhou T., in *Complex Networks* (Springer Berlin Heidelberg) 2009 pp. 73-81.
[41] Pan Y., Li D.-H., Liu J.-G. and Liang J.-Z., *Physica A*, **389** (2010) 2849.
[42] Thiel K. and Berthold M. R., in *10th IEEE International Conference on Data Mining (ICDM)* (IEEE Press) 2015 pp. 1085-1090.
[43] Papadopoulos F., Kitsak M., Serrano M. Á., Boguñá M. and Krioukov D., *Nature*, **489** (2012) 537.
[44] Blondel V. D., Gajardo A., Heymans M., Senellart P. and Van Dooren P., *SIAM Rev.*, **46** (2004) 647.
[45] Sawa T. and Ohno-Machado L., *Comput. Biol. Med.*, **33** (2003) 1.
[46] Zhu C., Kushwaha A., Berman K. and Jegga A. G., *BMC Syst. Biol.*, **6** (2012) S8.
[47] Zhao D.-D., Zeng A., Shang M.-S. and Gao J., *Chin. Phys. Lett.*, **30** (2013) 8901.
[48] Zhou Y.-B., Lei T. and Zhou T., *EPL*, **94** (2011) 48002.
[49] Gao J., Dong Y.-W., Shang M.-S., Cai S.-M. and Zhou T., *EPL*, **110** (2015) 28003.
[50] Gao J. and Zhou T., arXiv:1509.00594.