Research Article

Multiangle Social Network Recommendation Algorithms and Similarity Network Evaluation

Jinyu Hu,[1] Zhiwei Gao,[2] and Weisen Pan[1]

[1] Department of Biostatistics and Computational Biology, University of Rochester, Rochester, NY 14642, USA
[2] Faculty of Engineering and Environment, Northumbria University, Newcastle upon Tyne, NE1 8ST, UK

Correspondence should be addressed to Zhiwei Gao; zhiwei.gao@northumbria.ac.uk

Multiangle social network recommendation algorithms (MSN) and a new assessment method, called similarity network evaluation (SNE), are both proposed. From the viewpoint of six dimensions, the MSN are classified into six algorithms, including user-based algorithm from resource point (UBR), user-based algorithm from tag point (UBT), resource-based algorithm from tag point (RBT), resource-based algorithm from user point (RBU), tag-based algorithm from resource point (TBR), and tag-based algorithm from user point (TBU). Compared with the traditional recall/precision (RP) method, the SNE is more simple, effective, and visualized. The simulation results show that TBR and UBR are the best algorithms, RBU and TBU are the worst ones, and UBT and RBT are in the medium levels.

1. Introduction

In recent years social tagging systems have become increasingly popular as a means to classify large sets of resources on the web. These systems allow users to add metadata in the form of keywords to share resources [1]. Nevertheless, the rapidly growing data in these systems present new technical challenges involved with recommended resources. Collaborative filtering [2] approach provides a solution to overcome this challenge, which makes recommendations solely base on the preferred database. The Top-N recommendation method [3] tries to recommend the top-n ranked resources that users may be interested in.

Computation of the similarity plays a key role in the collaborative filtering, and there are many different ways to compute the similarities such as the Pearson correlation [4], constrained Pearson correlation [5], cosine-based similarity [6], adjusted cosine similarity [7], and Spearman rank correlation [8]. Currently, the user-based collaborative filtering algorithm mainly considers user similarity from the resource perspective [9]. Some papers did not discuss the resource similarity from the user perspective [10]. Mostly, the tag-based recommendation only takes into account the tag similarity based on the resource tag [11]. Some papers also

comprehensively consider the user similarity and resource similarity but do not consider the tag similarity [12]. Based on the pursuit of a recommendation system, the assessment indicators are employed to evaluate an algorithm, for instance, recall/precision, mean absolute error (MAE), mean average precision (MAP), and area under curve (AUC) [13]. For recommended results, there are some auxiliary assessment indicators [14], such as coverage (COV), diversity (DIV), and average popularity (AP).

In this paper, the concepts of the 1-mode network and 2-mode network in social network [15] are addressed for collaborative filtering. For movies, this paper gives recommended explanations of the six algorithms as follows:

> UBR: "Users who watched your favorite movies also watch …"
>
> UBT: "Users who used your favorite tags also watch …"
>
> RBU: "Users who watch this movie also watch …"
>
> RBT: "Tags which annotate this movie also annotate …"
>
> TBR: "Other Tags of this movie also annotate …"
>
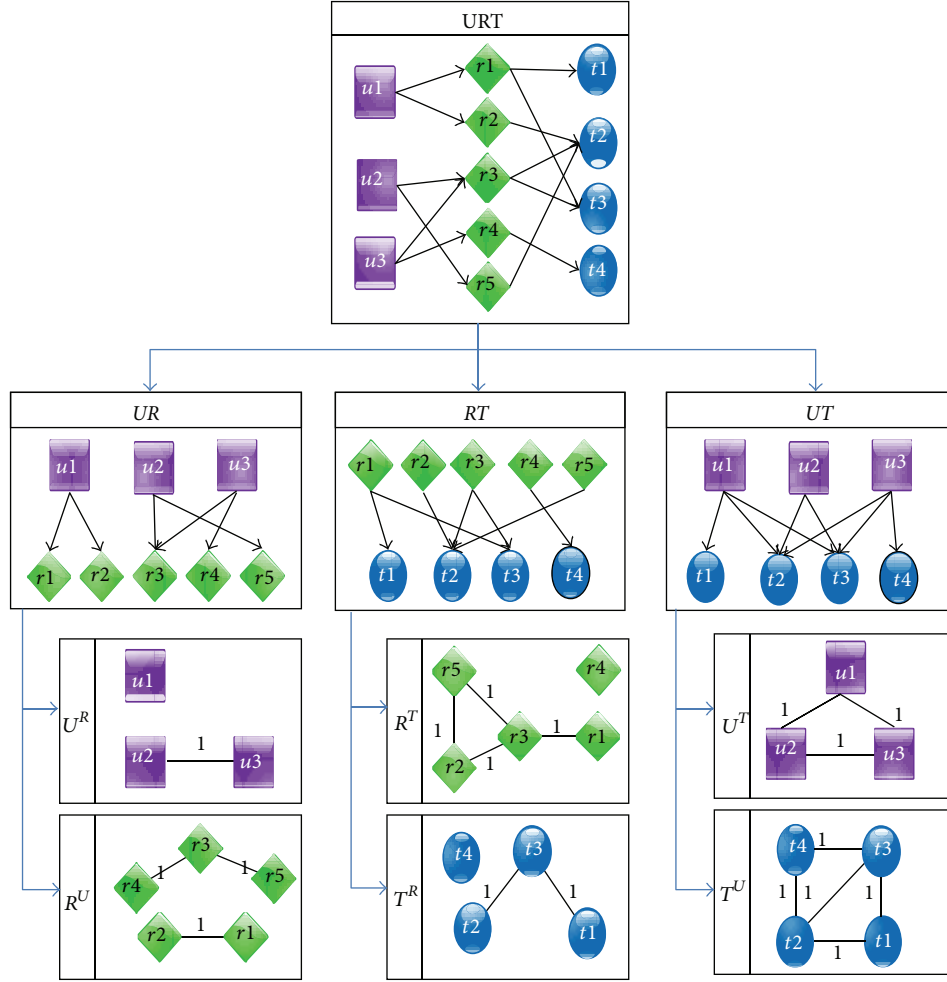> TBU: "Other Tags which you used also annotate …"

FIGURE 1: A process to obtain 2-mode networks and 1-mode networks.

On the basis of the similarity network evaluation (SNE), and comparison with the recall/precision, it is indicated that the results from the two assessment methods are consistent, but the SNE is more simple, visualized, and effective with less steps.

2. Multiangle Social Network Recommendation Algorithm (MSN)

2.1. Pretreatment of Social Network.
The triple (U, R, T) represents the raw data, which means the user U is given the tag T on the resource R. From this triple, one can deduce three 2-mode networks such as UR, TR, and UT. Each 2-mode network can deduce two 1-mode networks. For instance, UR can infer two 1-mode networks U^R and R^U; RT infers 1-mode networks R^T and T^R; and UT infers 1-mode networks U^T and T^U (see Figure 1). In the present paper, the symbol of A' denotes the transpose of A.

The matrix U^R shows the membership of each pair of users who like the common resources, described by $U^R = UR * UR'$. The matrix R^U represents the membership of each pair of resources which have the common users, described by $R^U = UR' * UR$. The matrix R^T is the membership of each pair of resources which have the same tags, described by $R^T = RT * RT'$. The matrix T^R indicates the membership of each pair of tags which have the common resources, with $T^R = RT' * RT$. The matrix U^T means the membership of each pair of users who annotate the common tags, with $U^T = UT * UT'$. The matrix T^U shows the membership of each pair of tags which have the common users, with $T^U = UT' * UT$.

2.2. User-Based Social Network Algorithm (USN).
There are two steps of the USN algorithm: firstly to find other users who are similar to the target user and then recommend the other users' favorite resources to the target user. The algorithm can be divided into the two algorithms UBR and UBT.

(1) *User-Based Algorithm from Resource Perspective (User-Based Resources, UBR).* The user similarity from resource (USR) perspective means that if two users like the same resources, they are similar. The element of the USR, denoted by usr_{ij} is defined as:

$$usr_{ij} = \frac{u_{ij}}{u_{ii} + u_{jj} - u_{ij}}, \quad u_{ij} \in U^R. \tag{1}$$

Based on the above similarity, the user-resource interest matrix UBR is described by:

$$UBR = UR' * USR. \tag{2}$$

(2) User-Based Algorithm from Tag Perspective (User-Based Tag, UBT). The user similarity from tag (UST) perspective indicates two users are similar if they prefer to use the same tags. The element of the UST, denoted by ust_{ij} is computed by

$$ust_{ij} = \frac{u_{ij}}{u_{ii} + u_{jj} - u_{ij}}, \quad u_{ij} \in U^T. \tag{3}$$

Based on the similarity, the interest matrix UBT is defined as:

$$UBT = UR' * UST. \tag{4}$$

2.3. Resource-Based Social Network Algorithm (RSN). The RSN algorithm also has two steps: firstly to find out the resources that the target user likes and then recommend other similar ones to the user. It is divided into two algorithms: RBU and RBT.

(1) Resource-Based Algorithm from User Perspective (Resources-Based User, RBU). The resource similarity from user (RSU) perspective means that if two resources are enjoyed by the same user, they are similar, where rsu_{ij} is defined by

$$rsu_{ij} = \frac{r_{ij}}{r_{ii} + r_{jj} - r_{ij}}, \quad r_{ij} \in R^U. \tag{5}$$

The interest matrix RBU is written as

$$RBU = RSU * UR'. \tag{6}$$

(2) Resource-Based Algorithm from Tag Perspective (Resources-Based Tag, RBT). The resource similarity from tag (RST) perspective implies that if two resources are enjoyed by the same user, they are similar. The element of the RST is defined as

$$rst_{ij} = \frac{r_{ij}}{r_{ii} + r_{jj} - r_{ij}}, \quad r_{ij} \in R^T. \tag{7}$$

The interest matrix RBT is described by

$$RBT = RST * UR'. \tag{8}$$

2.4. Tag-Based Social Network Algorithm (TSN). The TSN algorithm consists of three parts: the first step is to look for the frequently used tags of the target user, next to find other similar tags and merge both of them into a tag set, finally to recommend these tag-set corresponding resources to the target user.

(1) Tag-Based Algorithm from Resource Perspective (Tag-Based Resources, TBR). The tag similarity from resource (TSR)

perspective means that if two resources are enjoyed by the same user, they are similar. The element of the TSR is computed as

$$tsr_{ij} = \frac{t_{ij}}{t_{ii} + t_{jj} - t_{ij}}, \quad t_{ij} \in T^R. \tag{9}$$

The tag similarity matrix TSR, which is viewed from the resources, show that if two tags annotate the same resources, they are similar. One can have

$$TBR = RT * TSR * UT'. \tag{10}$$

(2) Tag-Based Algorithm from User Perspective (Tag-Based User, TBU). The tag similarity from user perspective (TSU) implies that if two resources are enjoyed by the same user, they are similar. The element of the TSU is defined as

$$tsu_{ij} = \frac{t_{ij}}{t_{ii} + t_{jj} - t_{ij}}, \quad t_{ij} \in T^U. \tag{11}$$

The interest matrix TBU can be written as:

$$TBU = RT * TSU * UT'. \tag{12}$$

3. Similarity Network Evaluation (SNE)

3.1. Basic Concepts

Definition 1 (similarity network (SN)). A connection matrix C is used to store similarity network (SN), whose element is c_{ij}, defined by

$$c_{ij} = \begin{cases} s_{ij}, & i > j, \ s_{ij} \geq T, \ 0 \leq T \leq 1, \\ 0, & i \leq j, \end{cases} \tag{13}$$

where s_{ij} is the similarity between the nodes i and j and $s_{ij} \in [0, 1]$. When $c_{ij} = 0$, there is no edge between nodes i and j; T is a threshold.

The definition of the similarity of the SN is actually borrowed from the definition of the gene community network (GCN) proposed by [16].

Definition 2 (types of the network). The similarity networks can be divided into six categories in terms of the intensity of similarities.

(1) Perfect correlation network (PCN): when $T = 1$, remove the edges whose weights are lower than 1, and the weights of all edges are 1.

(2) Very strong correlation network (VSN): when $T = 0.8$, delete the edges with the weights lower than 0.8.

(3) Strong correlation network (SCN): when $T = 0.6$, remove the edges with the weights lower than 0.6 to form a new network.

(4) Moderate correlation network (MCN): when $T = 0.4$, delete the edges whose weights are lower than 0.4 to form a similarity Network.
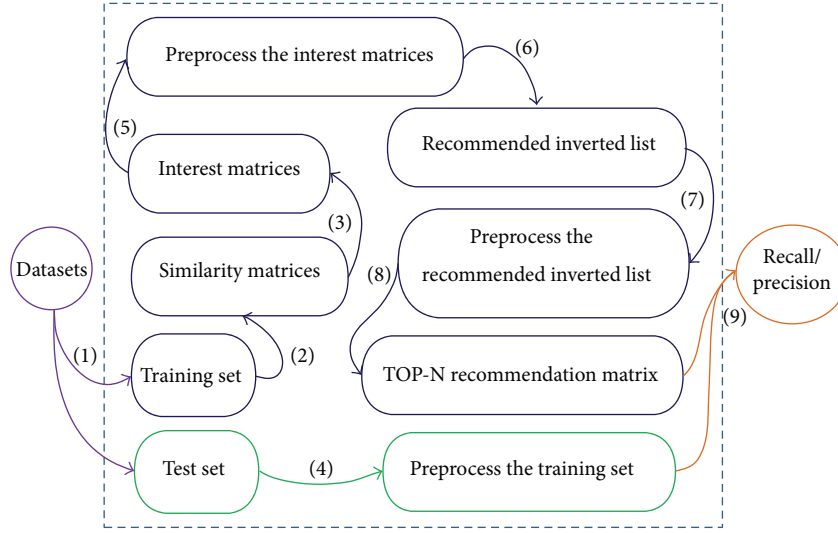
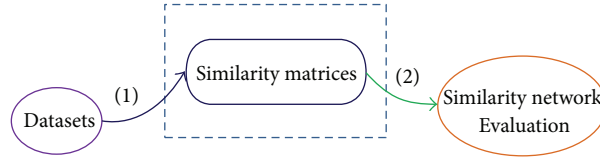FIGURE 2: The computing steps of Recall/Precision.

FIGURE 3: The computing steps of similarity networks evaluation.

(5) Weak correlation network (WCN): when $T = 0.2$, the values of all the edges are lower than 0.2 in the network.

(6) Uncorrelated network (UCN): when $T = 0$, all nodes are isolated nodes.

3.2. Algorithm Principle. The core of the collaborative filtering algorithm is to calculate the similarity. The SNE method can better evaluate the algorithm if under a certain threshold value there are fewer isolated nodes but more small communities in its similarity networks. It includes the following requirements.

(1) *A Certain Threshold Value.* Relative to the entire resource, these recommended resources are quite few. Therefore, when we evaluate an algorithm, the PCN network ($T = 1$) only needs to be considered. However, in order to have a comprehensive analysis, the VSN network ($T = 0.8$) and the SCN network ($T = 0.6$) should be considered together.

(2) *Fewer Isolated Nodes.* Recommendation algorithm is based on the similarity. The isolated node is not similar to any other nodes. In this case, the algorithm can not give a recommendation about these isolated nodes. Therefore, a good algorithm may produce a similarity network with fewer isolated nodes and more nonisolated ones.

(3) *More Communities.* Each community (nodes > 2) represents a different interest of users. The more the communities,

the more detailed features can they reflect, which makes the recommended results more in line with the user's taste.

(4) *Fewer Nodes in the Largest Community.* In the largest community, there are too many nodes to reflect the user's taste in detail. The nodes in the largest community should be as few as possible.

Taking into account the previous points, and also considering these different networks with different number of nodes, the score of the similarity network evaluation (SNE) is described by

$$ \text{SS}(T) = \frac{\text{NC} * \text{NI}}{\text{NA} * \text{NL}}, \quad \text{NA} \geq 1, \text{NL} \geq 1, \tag{14} $$

where SS is the score function of similarity network, T is the threshold, NA is the number of all nodes in the network, NI is the number of nonisolated nodes in the network, NC is the number of communities, and NL is the number of nodes in the largest community.

3.3. Steps of SNE. The recall/precision rate of an algorithm has nine calculation steps (see Figure 2).

The SNE method only needs to calculate similarity matrices; thus one can directly evaluate the algorithms after constructing similarity networks. It is a simple method, which only has one intermediate step (see Figure 3).
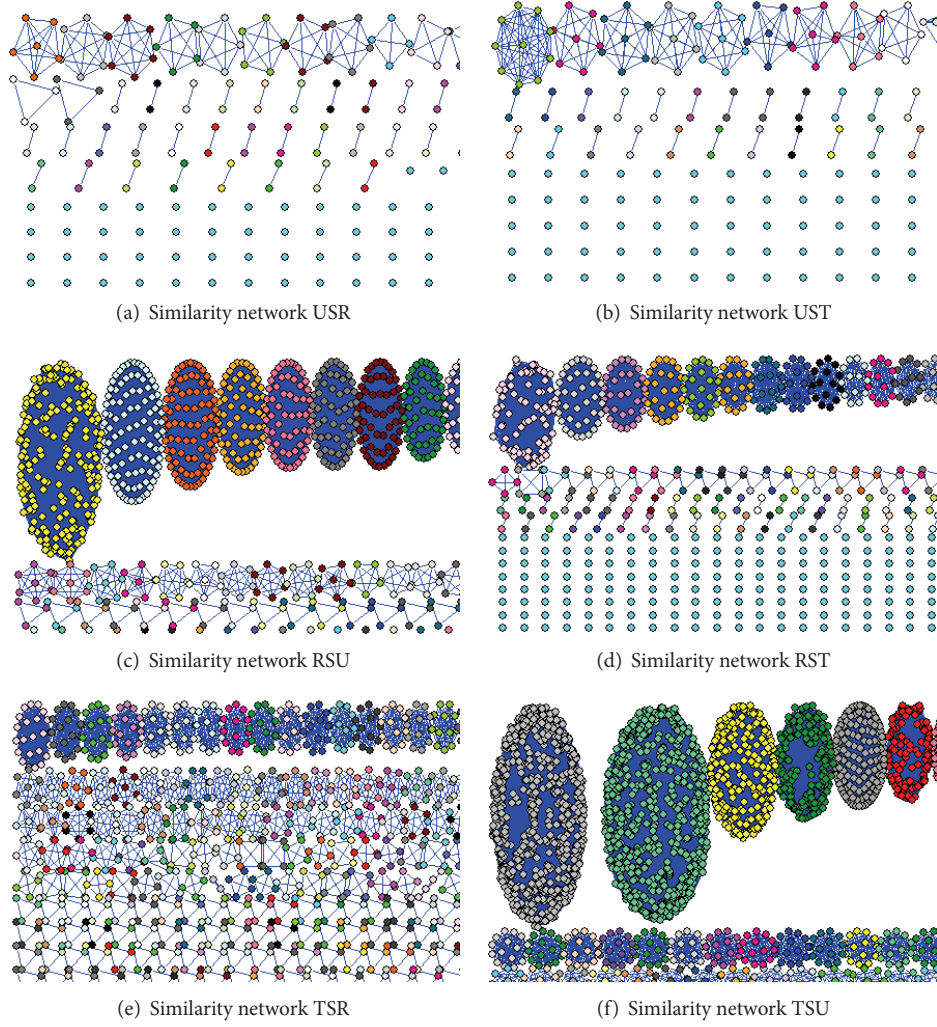
(a) Similarity network USR

(b) Similarity network UST

(c) Similarity network RSU

(d) Similarity network RST

(e) Similarity network TSR

(f) Similarity network TSU

FIGURE 4: The PCN networks of six algorithms.

TABLE 1: The score of the PCN.

SN	NA	NC	NL	NI	SS(T)
USR	2007	137	6	343	3.9
UST	2007	115	10	319	1.83
RSU	5640	315	201	2253	0.63
RST	5640	255	74	1081	0.66
TSR	8390	1146	23	4005	23.78
TSU	8390	544	611	5656	0.6

SS: the score function of similarity network; T: the threshold of similarity network; NA: the number of all nodes in the network; NI: the number of nonisolated nodes in the network; NC: the number of communities; NL: the number of nodes in the largest community.

TABLE 2: The score of the VSN.

SN	NA	NC	NL	NI	SS(T)
USR	2007	137	6	343	3.9
UST	2007	115	10	319	1.83
RSU	5640	321	201	2267	0.64
RST	5640	258	74	1089	0.67
TSR	8390	1150	23	4014	23.92
TSU	8390	546	611	5660	0.6

SS: the score function of similarity network; T: the threshold of similarity network; NA: the number of all nodes in the network; NI: the number of nonisolated nodes in the network; NC: the number of communities; NL: the number of nodes in the largest community.

4. Results and Discussions

The dataset is "hetrec2011-movielens-2k" from HetRec 2011 [17], which is an extension of MovieLens 10M dataset, with 2113 users, 10197 movies, 13222 tags, and 855598 ratings. In this paper, we focus on the users, resources, and tags and ignore the ratings.

4.1. Visual Analysis of the SNE. Visual similarity network analysis can be more intuitive to roughly evaluate various kinds of recommendation algorithms. We use the Pajek [18] to show the six PCN networks. To see more clearly, we just intercept 1/16 of the original screen in the upper-left corner (see Figure 4).
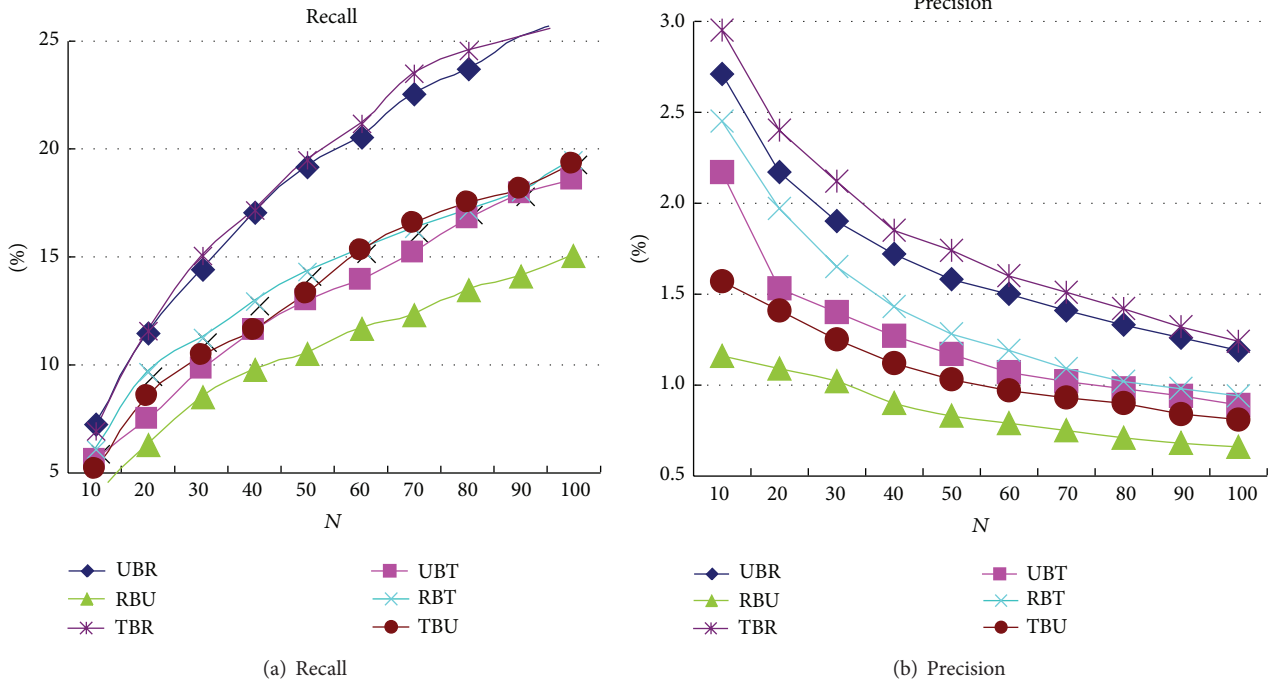
(a) Recall

(b) Precision

Figure 5: Recall/precision of the six algorithms.

Table 3: The score of the SCN.

SN	NA	NC	NL	NI	SS(T)
USR	2007	137	6	343	3.9
UST	2007	119	10	328	1.94
RSU	5640	288	799	2944	0.19
RST	5640	299	74	1263	0.9
TSR	8390	851	23	5518	24.33
TSU	8390	540	611	6041	0.64

SS: the score function of similarity network; T: the threshold of similarity network; NA: the number of all nodes in the network; NI: the number of nonisolated nodes in the network; NC: the number of communities; NL: the number of nodes in the largest community.

Table 4: Recall of the six algorithms.

Top-N	UBR	UBT	RBU	RBT	TBR	TBU
10	7.27%	5.65%	3.95%	6.18%	7.05%	5.18%
20	11.49%	7.55%	6.36%	9.71%	11.67%	8.57%
30	14.44%	9.80%	8.63%	11.34%	15.14%	10.44%
40	17.08%	11.61%	9.79%	12.95%	17.23%	11.65%
50	19.24%	12.96%	10.61%	14.39%	19.52%	13.29%
60	20.55%	13.95%	11.76%	15.43%	21.25%	15.35%
70	22.57%	15.20%	12.39%	16.39%	23.48%	16.56%
80	23.70%	16.77%	13.54%	17.21%	24.54%	17.51%
90	25.17%	17.94%	14.11%	18.07%	25.17%	18.21%
100	25.98%	18.54%	15.10%	19.51%	25.85%	19.35%

In the SNE evaluation, an algorithm with more small communities is better than the one with more large communities. In Figure 4, we can see the following:

Table 5: Precision of the six algorithms.

Top-N	UBR	UBT	RBU	RBT	TBR	TBU
10	2.71%	2.17%	1.16%	2.45%	2.95%	1.57%
20	2.17%	1.53%	1.09%	1.97%	2.40%	1.41%
30	1.90%	1.40%	1.02%	1.65%	2.12%	1.25%
40	1.72%	1.27%	0.90%	1.43%	1.85%	1.12%
50	1.58%	1.17%	0.83%	1.28%	1.74%	1.03%
60	1.50%	1.07%	0.79%	1.19%	1.60%	0.97%
70	1.41%	1.02%	0.75%	1.09%	1.51%	0.93%
80	1.33%	0.98%	0.71%	1.02%	1.42%	0.90%
90	1.26%	0.94%	0.68%	0.98%	1.32%	0.84%
100	1.19%	0.89%	0.66%	0.94%	1.24%	0.81%

(1) The UST's big communities are more obvious than the USR's; therefore, UBR is better than UBT, denoted by UBR > UBT.

(2) The RSU has more big communities than the RST; thus RBT > RBU, indicating RBT is better than RBU.

(3) There are very big communities in TSU. Conversely, TSR has fewer ones. Therefore, TBR > TBU, meaning TBR is better than TBU.

(4) Obviously, the TSU and RSU have the largest communities; thus the algorithms TBU and RBU are relatively poor.

(5) In the same way, the better algorithms are UBR and TBR, and the medium ones are UBT and RBT.

4.2. Quantitative Analysis of SNE. Based on score function of SNE, the scores of different algorithms under the threshold

TABLE 6: Comparison analysis of RP and SNE.

	Recall/precision (RP)	Similarity network evaluation (SNE)
Results of evaluation	Recall: UBR > TBR > RBT > UBT > TBU > RBU Precision: TBR > UBR > RBT > UBT > TBU > RBU Conclude from above: TBR > UBR > RBT > UBT > TBU > RBU	$T = 1$: TBR > UBR > UBT > RBT > RBU > TBU $T = 0.8$: TBR > UBR > UBT > RBT > RBU > TBU $T = 0.6$: TBR > UBR > UBT > RBT > TBU > RBU Conclude from above: TBR > UBR > UBT > RBT > TBU > RBU
Complexity	9 middle steps, high complexity	1 middle step, very simple
Visualization	No	Yes

values of 1, 0.8, and 0.6 are, respectively, displayed in Tables 1–3.

According to the similarity scores shown by Tables 1 and 2, one can give the order of the algorithms from best to worst as follows: TBR > UBR > UBT > RBT > RBU > TBU. And from Table 3, in order of the best algorithm, it goes as follows: TBR > UBR > UBT > RBT > TBU > RBU. Overall, RBU and TBU are the worst algorithms.

4.3. Comparison with Recall/Precision. The recall/precision rates can evaluate the accuracy of algorithms. Based on the training set, we recommend the top-N resources to users by using the six algorithms. Corresponding to the test set, one can obtain the Recall/Precision rates under different N values, respectively (see Tables 4 and 5).

The Recall and Precision rates of different algorithms and different N values are shown in Figure 5.

The comparative analysis of two evaluation indicators is given in Table 6.

From the comparison, one can see that the results of two evaluation algorithms are very agreeable, where the same results are UBR > UBT, RBT > RBU, and TBR > TBU. The similarity network visualization method can roughly evaluate algorithms. From the perspective of the complexity and visualization, the similarity network assessment algorithm is in an advantageous position.

5. Conclusion and Future Work

In the paper, not only the collaborative filtering algorithms are proposed based on social network, but also a new assessment method of similarity network evaluation is addressed. For these six algorithms, TBR and UBR are the best algorithms, RBU and TBU are the worst ones, and UBT and RBT are in the medium levels. From the recommended effects, we can conclude that UBR > UBT, RBT > RBU, and TBR > TBU. It is noted that, in the actual use of algorithms, the accuracy of the algorithm is not the unique factor to be considered; the complexity of the algorithm and the maintenance cost of the algorithm have to be taken into account as well.

Future work is encouraged for the three aspects: (1) recommended algorithm based on the current similarity calculation method is of interest; (2) hybrid algorithms based on two recommendation algorithms are of significance; (3)

the SNE assessment would be extended to evaluate other algorithms rather than the collaborative recommended algorithms.

References

[1] A. K. Milicevic, A. Nanopoulos, and M. Ivanovic, "Social tagging in recommender systems: a survey of the state-of-the-art and possible extensions," *Artificial Intelligence Review*, vol. 33, no. 3, pp. 187–209, 2010.

[2] A. K. Menon, K. Chitrapura, S. Garg, D. Agarwal, and N. Kota, "Response prediction using collaborative filtering with hierarchies and side-information," in *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '11)*, pp. 141–149, San Diego, Calif, USA, August 2011.

[3] X. Ning and G. Karypis, "SLIM: sparse linear methods for top-n recommender systems," in *Proceedings of the 11th IEEE International Conference on Data Mining (ICDM '11)*, pp. 497–506, Vancouver, Canada, December 2011.

[4] J. Bobadilla, F. Serradilla, and A. Hernando, "Collaborative filtering adapted to recommender systems of e-learning," *Knowledge-Based Systems*, vol. 22, no. 4, pp. 261–265, 2009.

[5] G. Chen, F. Wang, and C. Zhang, "Collaborative filtering using orthogonal nonnegative matrix tri-factorization," *Information Processing and Management*, vol. 45, no. 3, pp. 368–379, 2009.

[6] L. Candillier, F. Meyer, and M. Boullé, "Comparing state-of-the-art collaborative filtering systems," in *Proceedings of the International Conference on Machine Learning and Data Mining*, pp. 548–562, Leipzig, Germany, July 2007.

[7] K. Choi and Y. Suh, "A new similarity function for selecting neighbors for each target item in collaborative filtering," *Knowledge-Based Systems*, vol. 37, no. 1, pp. 146–153, 2013.

[8] D. Helic, "Managing collaborative learning processes in e-learning applications," in *Proceedings of the 29th IEEE International Conference on Information Technology Interfaces*, pp. 345–350, Dubrovnik, Croatia, June 2007.

[9] C. Cechinela, M. Siciliab, and S. Sánchez-Alonsob, "Evaluating collaborative filtering recommendations inside large learning object repositories," *Information Processing and Management*, vol. 49, no. 1, pp. 34–50, 2013.

[10] A. Rajput, M. Dubey, S. Thakur, and M. Gondgly, "Improved item based collaboration filtering using recommendation system," *Binary Journal of Data Mining and Networking*, vol. 1, no. 1, pp. 6–13, 2011.

[11] R. Jäschke, A. Hotho, F. Mitzlaff, and G. Stumme, "Challenges in tag recommendations for collaborative tagging systems: recommender systems for the social web," *Intelligent Systems Reference Library*, vol. 32, pp. 65–87, 2012.

[12] D. Zhang and C. Xu, "A collaborative filtering recommendation system by unifying user similarity and item similarity," in *Web-Age Information Management*, vol. 7142 of *Lecture Notes in Computer Science*, pp. 175–184, Springer, Berlin, Germany, 2012.

[13] Y. Shi, M. Larson, and A. Hanjalic, "Mining contextual movie similarity with matrix factorization for context-aware recommendation," *ACM Transactions on Intelligent Systems and Technology*, vol. 4, no. 1, article 16, 2013.

[14] L. Lü, M. Medob, C. Yeungb, Y. Zhang, Z. Zhang, and T. Zhou, "Recommender systems," *Physics Report*, vol. 519, no. 1, pp. 1–49, 2012.

[15] A. Mislove, B. Viswanath, K. P. Gummadi, and P. Druschel, "You are who you know: inferring user profiles in online social networks," in *Proceedings of the ACM 3rd International Conference on Web Search and Data Mining (WSDM '10)*, pp. 251–260, New York, NY, USA, February 2010.

[16] J. Hu and Z. Gao, "Modules identification in gene positive networks of hepatocellular carcinoma using Pearson agglomerative method and Pearson cohesion coupling modularity," *Journal of Applied Mathematics*, vol. 2012, Article ID 248658, 21 pages, 2012.

[17] I. Cantador, P. Brusilovsky, and T. Kuflik, "Second Workshop on Information Heterogeneity and Fusion in Recommender Systems," in *Proceedings of the 5th ACM conference on Recommender Systems (RecSys '11)*, pp. 387–388, Chicago, Ill, USA, October 2011.

[18] W. Nooy, A. Mrva, and V. Batagelj, *Exploratory Social Network Analysis with Pajek*, Cambridge University Press, Cambridge, UK, 2005.