frontiers | Frontiers in Physics

TYPE Original Research
PUBLISHED 17 October 2022
DOI 10.3389/fphy.2022.1018781

OPEN ACCESS

EDITED BY
Jianguo Liu,
Shanghai University of Finance and Economics, China

REVIEWED BY
Xiaolong Ren,
University of Electronic Science and Technology of China, China
Zhuoming Ren,
Hangzhou Normal University, China

*CORRESPONDENCE
Yixiu Kong,
yixiu.kong@bupt.edu.cn
Yizhong Hu,
kuai@bupt.edu.cn

SPECIALTY SECTION
This article was submitted to Social Physics,
a section of the journal
Frontiers in Physics

RECEIVED 13 August 2022
ACCEPTED 20 September 2022
PUBLISHED 17 October 2022

CITATION
Kong Y, Hu Y, Zhang X and Wang C (2022), Structural centrality of networks can improve the diffusion-based recommendation algorithm.
Front. Phys. 10:1018781.
doi: 10.3389/fphy.2022.1018781

Structural centrality of networks can improve the diffusion-based recommendation algorithm

Yixiu Kong[1]*, Yizhong Hu[1]*, Xinyu Zhang[2] and Cheng Wang[1]

[1]School of Science, Beijing University of Posts and Telecommunications, Beijing, China, [2]Tsinghua Education Foundation, Tsinghua University, Beijing, China

The recommendation system has become an indispensable information technology in the real world. The recommendation system based on the diffusion model has been widely used because of its simplicity, scalability, interpretability, and many other advantages. However, the traditional diffusion-based recommendation model only uses the nearest neighbor information, which limits its efficiency and performance. Therefore, in this article, we introduce the centralities of complex networks into the diffusion-based recommendation system and test its performance. The results show that the overall performance of heat conduction algorithm can be improved by 184%–280%, using the centrality of complex networks, reaching almost the same accuracy level as the mass diffusion algorithm. Therefore, the recommendation system combining the high-order network structure information is a potentially promising research direction in the future.

KEYWORDS

recommendation system, centrality, complex network, diffusion model, collaborative filtering

1 Introduction

With the development of information technology, people are overwhelmed by an increasing amount of information. Although the development of technological innovations has made our lives easier, meanwhile, overloaded information consumes our time and efforts when we are searching online. Therefore, in response to this demand, searching systems and recommendation systems are evolving accordingly, and they both are the technologies that have been developed to deal with information overload. The searching systems solve the problem of directed search, while the recommendation systems can predict the possible preferences and interests of users based on the previous data. Recommendation systems have been developed for decades, and each part has been gradually improved and developed toward more multi-level and applicable models.

Collaborative filtering [1–3] is one of the most widely used, least computationally complex, and most effective information-filtering algorithms. The CF algorithm provides personalized recommendations for each user based on the user's past purchase history database and product search records. Breese et al. [2] classified CF into two broad categories, namely, memory-based and model-based approaches. Memory-based methods predict information and recommend products based on a measure of similarity between the user

and the product [3–5]. Model-based algorithms use a collection of user and object information to generate information-filtering models [10, 11] through clustering [6], Bayesian approach [7], matrix factorization [8, 9], and other machine learning methods.

Different from computer science, the application of physics in the field of interdisciplinary science has also obtained some successful complex network theories, and various classical physical processes have provided some new insights and solutions for the active field of information filtering in recent years [1, 12, 13]. For example, the diffusion process like the heat conduction process on a dichotomous complex network [14], the principles of dynamic resource allocation in dichotomous complex networks [15], opinion diffusion [16], and gravity [17] have been applied in information filtering.

Overall, CF and other diffusion-based recommendation algorithms have been successfully applied to many well-known online e-commerce platforms. Meanwhile, in recent years, a lot of research studies, such as the heat conduction, mass diffusion, or hybrid method [20, 21], biased heat conduction [22, 23], multi-channel diffusion [24], preferential diffusion [25, 26] based on the CF direct random walk method [27], hypergraph models with social labels [28, 29], and multilinear interactive matrix factorization [30],are devoted to studying the two variations of the algorithm. These algorithms will further improve the efficiency of information-filtering systems. In addition, multiple explorations [31, 32] on information filtering considering external constraints have also been made.

Meanwhile, we should note that the compressed network structural information, including the information core and information backbone, provides some enlightenment for the in-depth understanding of information filtering [18, 19]. Some network structural centralities, such as PageRank and eigenvector centrality, can indicate the structural properties of networks in one-dimensional metrics, while traditional network structure statistics, such as degree, can only contain first-order structure information (nearest neighbor information). We can conclude that it is highly probable to obtain richer structural information about the network by replacing the influence of first-level nearest neighbors with one-dimensional complex network structural centralities, such as coreness and PageRank. This structural information considers the long-range correlations in the network structure, which is expected to help improve the accuracy of recommendation systems and improve the application prospect in real scenarios. In the following part of this article, we will show that the adoption of network structural information can greatly improve the performance of some recommendation algorithms.

2 Materials and methods

2.1 Dataset description

We obtained seven commonly used datasets in the research of recommendation systems. The datasets are listed in Table. 1.

2.2 Evaluation metrics

In the dataset, we know some of the items that users collect, and we need to recommend other items for users. Therefore, to compare the recommendation performance of recommendation algorithms [33], we divide the dataset into two categories: training set and test set. The training set is used for each user-recommended item, and then the test set is used to evaluate recommendation algorithm. When calculating the evaluation metrics, we will regard the selection of the user from the user's recommendation lists as a positive example, and otherwise, a negative example. For each user, in accordance with the actual results and predicted results of each item, 1 is expressed as a positive case, and 0 is expressed as a negative case. We can define the following three indicators: precision, recall, and F1-score.

Precision and recall are commonly used evaluation indices. Precision is defined as the proportion of data in the dataset that the label predicts correctly, and recall is the proportion of data that we predict to be correct in a certain class. They are calculated as

$$Recall = \frac{TP}{TP + FP},$$

But for each user, what we obtain is a list of users, namely, the TOP-K recommendations. So, we assume that the number of items recommended to the user is K, the precision is the number of recommended items in the test set, and the recall rate is the correct number of recommended items in the test set. For a target user, the precision and recall are defined as

$$P_i = \frac{d_i(K)}{K},\ R_i = \frac{d_i(K)}{D_i},$$

where $d_i(K)$ is the degree of the test set in the user i's TOP-K recommended list and $D_i$is the length of the test set for the ith user. After calculating the precision and recall of the TOP-K recommendation list of all users, we can obtain the precision and recall of the algorithm by calculating the average value of all users.

When recommending TOP-K items for users, what need to be considered are the precision and recall rate of the reality (whether the user likes it or not) and the predicted label (recommendation list) of each item. Therefore, F1-score, also called balanced F-score, is an index commonly used to express the precision rate of binary classification problems in statistical data analysis. F1-score can be regarded as the harmonic average result of two indices, and the value range is [0, 1]. F1-score is defined as

$$f_1 - \text{score} = \frac{2 \cdot precision \cdot recall}{precision + recall}.$$

2.3 The general Markovian form of diffusion-based recommendations

In this section, we briefly introduce the three algorithms adopted in this article. When they are applied to calculate the

TABLE 1 Introduction of datasets.

Dataset	Number of users	Number of items	Number of records	Sparsity
Movielens_latest small	610	9,784	100,836	1.70%
Movielens_100K	943	1,682	100,000	6.30%
Movielens_90K	2,113	10,109	855,598	4.00%
Movielens_1M	6,040	3,883	1,000,209	4.26%
Movielens_10M	69,878	10,681	10,000,054	1.33%
Last.fm	1,892	12,523	186,479	0.79%
Netflixdataset	5,967	16,977	1,261,097	1.24%

recommendation list of users on the dataset, it is necessary to calculate the probability transition matrix on each training set, but the time complexity is higher if the calculation is carried out directly through the formula. Therefore, through observation, it is found that the calculation form of the probability transition matrix is to multiply the corresponding elements after row normalization and column normalization of some two columns in the adjacent matrix. Therefore, the calculation formula can be simplified into a matrix form [34, 35] to reduce and improve the calculation efficiency of the algorithm. For example, the calculation formula of the probability transition matrix of the heat conduction algorithm is as follows:

$$W_{\alpha\beta}^{H} = \frac{1}{k_\alpha}\sum_{i=1}^{u}\frac{a_{i\alpha}\cdot a_{i\beta}}{k_i} = \sum_{i=1}^{u}\frac{a_{i\alpha}\cdot a_{i\beta}}{k_\alpha\cdot k_i}.$$

Here, $a_{i\alpha}$ and $a_{i\beta}$ are the elements of the ith row in two columns of the matrix a and β, respectively. The meaning of the aforementioned expression is to first normalize the elements in some two columns of the bipartite graph adjacency matrix, then matrix multiplication is performed, and the vector elements obtained are added after the multiplication step. Therefore, the formula of the probability transition matrix can be simplified to the following matrix form:

$$W_{\alpha\beta}^{H} = D_o^{-1}\cdot A^T\cdot D_u^{-1}\cdot A.$$

where

$$D_u^{-1} = \begin{pmatrix} \frac{1}{d_1} & 0 & \cdots & 0 \\ 0 & \frac{1}{d_2} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \frac{1}{d_u} \end{pmatrix}, D_o^{-1} = \begin{pmatrix} \frac{1}{d_1} & 0 & \cdots & 0 \\ 0 & \frac{1}{d_2} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \frac{1}{d_o} \end{pmatrix}.$$

Here, D_u^{-1} and D_o^{-1} represent a diagonal matrix formed by calculating the reciprocal of the node degree values of the two categories of users and items, respectively. If the node degree is 0, the derivative value is not necessarily to be 0. d_i is the degree of the ith user or item node. Similarly, the formulas of the other two algorithms are analyzed and can be simplified into the following matrix expressions:

$$W_{\alpha\beta}^{M} = \mathrm{A}^T\cdot D_u^{-1}\cdot A\cdot D_o^{-1},$$
$$W_{\alpha\beta}^{H+M} = D_o^{-1}\lambda\cdot A^T\cdot D_u^{-1}\cdot A\cdot D_o^{-1}1-\lambda.$$

If a network graph corresponding to a dataset is defined as $G = (V, E)$ and the adjacency matrix of the corresponding bipartite graph is defined as $A_{u\times o}$, then we can calculate the probability transition matrix of the users and regard the user's favorite vector in the adjacency matrix as the initial heat or resource amount for the item, so that we can calculate the initial heat or resource amount for the item according to the following expression:

$$result = A\cdot W^T,$$

where result is a matrix of size $u \times o$ and the ith row vector of the matrix represents a score list of all items for the user i. It was assumed that the set of all items in the dataset is V_{Item}, the set of items collected by the user is V_{collect}, and the uncollected set is

$$V_{notCollect} = V_{Item} - V_{collect}.$$

According to the final rating result, some items that are not collected by the user yet are sorted according to the score. Since the TOP-K recommendation with the highest score is selected in the dataset, it is the recommendation list of the user i that is recorded as a matrix of result_K. The ith row of the matrix is the K items recommended for the user i. According to the recommendation list and the test set, the evaluation metrics for the user i can be calculated.

2.4 A brief introduction to the centralities of networks

2.4.1 Closeness centrality

The closeness [36] indicates the degree of difficulty in arriving at other nodes from a certain node, and the larger the value is, the farther the distance from other nodes is; a

lower value indicates a closer distance to other nodes. It is calculated using the following equation:

$$Closeness(V) = \frac{1}{\sum_{q-1}^{n-1} d(p,q)},$$

where p represents the node that is to be calculated, q represents other nodes in the network, and $d(p,q)$ is the shortest path length from node p to node q. The idea of this closeness is that the closer the nodes are to the network center, the more quickly they can reach other nodes. Therefore, the importance of each node obtained through closeness not only is influenced by the number of adjacent nodes but also reflects the minimum average shortest path to other nodes by utilizing the characteristics of the whole network.

2.4.2 Eigenvector centrality

Eigenvector centrality [37] is a metric to measure the impact of nodes on the network. This metric is used when nodes with the same number of links are present. A high score for eigenvector centrality means that the node is connected to a few nodes that have high scores themselves. For a given network with an adjacent matrix A, if two nodes i and j are not directly connected, then $A_{i,j} = 0$, and otherwise $A_{i,j} = 1$. The eigenvector of A must satisfy the following expression:

$$A\boldsymbol{x} = \lambda \boldsymbol{x},$$

and eigenvector centrality is given by the eigenvector corresponding to the largest eigenvalue of A.

2.4.3 Katz centrality

Like eigenvector centrality, Katz centrality [38] also measures the importance of nodes. The difference is that it considers the nodes that have an in-degree 0 by adding a decay coefficient α and a bias term β. Also, with the help of the adjacency matrix A, for a node i, Katz centrality x_i is calculated as follows:

$$x_i = \alpha \sum_{j \subseteq G} a_{i,j} x_j + \beta.$$

In practice, the attenuation coefficient $\alpha < 1/\lambda$ is usually chosen to ensure that the matrix is invertible and that centrality can be obtained.

2.4.4 PageRank

PageRank [39] algorithm was originally a calculation method for calculating weights to solve the ranking between web pages, which was developed based on eigenvector centrality. Although the method is proposed to solve the problem of a directed graph, this method can be used in any graph, and now it is often applied to the analysis of the importance of various networks.

For a directed graph $G = (V, E)$, we can define a Markovian process that has the probability transition matrix $M = (m_{i,j})$ in the graph, and the normalized initial centrality of all nodes is R_0. Thus, we can obtain the value after one-step transition as MR_0 and then proceed in turn until t-step probability transition:

$$R_0, MR_0, M^2R_0, \ldots, M^tR_0, \ldots$$

If this series converges, the final vector R represents the stationary distribution of the Markov chain and satisfies MR =R. Therefore, the value of the vector R is the PageRank value of the nodes in the network. In practice, a random distribution term E is added to ensure that the node with zero in-degree can also receive incoming links, and the weight of this random term is usually set to 0.15.

Finally, we can obtain the following expression:

$$\text{PageRank} = 0.85 \times \text{R} + 0.15 \times \frac{\text{E}}{n},$$

where n is the total number of nodes and E is a matrix whose elements are all equal to 1.

3 Results

The degree is the most commonly used metric in the study of a network structure model, which is defined as the number of neighboring nodes of a given node, thus reflecting the importance of a node in the network. In the diffusion-based recommendation algorithm, the process of heat conduction or mass diffusion of each node is completed governed by the degree of each node. But this idea is, in fact, too simple to be applied in the real world. In a social network, if a member B knows only one influential member A, although the node has a only degree that equals to 1, B's influence will increase due to the higher importance of node A. From this point of view, there is a room for improvement in the use of the network structure as an indicator for user recommendation.

Therefore, the method proposed in this article is to improve the recommendation algorithm based on the traditional diffusion process by replacing the degree with the network structural centralities like closeness, eigenvector centrality, PageRank, and Katz centrality. These four types of network structural centralities are used in this article.

3.1 The selection of test sets

We first randomly divide the dataset into 90% of the training set and 10% of the test set according to the conventional practice and use the algorithm to recommend items for users. However, since the size of the dataset of each user is not average, in real life, the length of the recommendation list required by each user is not the same. Therefore, to compare the difference of the recommendation performance among different algorithms, we

TABLE 2 Recommendation evaluations on seven datasets based on random selection of test sets.

Dataset	Algorithm	Precision (%)	Recall (%)	F1-score
ml_latest small	**Heats**	0.92	7.12	1.63%
	Probs	7.57	33.23	1.23%
	Hybrid	8.03	35.73	1.31%
ml_90K	**Heats**	0.12	0.57	0.20%
	Probs	15.84	23.16	18.8%
	Hybrid	16.16	23.93	19.3%
ml_100K	**Heats**	11.37	15.53	13.1%
	Probs	26.76	42.03	32.7%
	Hybrid	28.84	45.70	35.4%
ml_1M	**Heats**	5.10	18.88	8.04%
	Probs	7.94	28.12	12.4%
	Hybrid	8.99	33.27	14.14
Netflix	**Heats**	0.03	0.16	0.04%
	Probs	8.06	24.70	12.1%
	Hybrid	8.71	27.82	13.3%
ml_10M	**Heats**	5.58	29.84	9.40%
	Probs	7.51	36.68	12.5%
	Hybrid	7.94	39.40	13.2%
Last.fm	**Heats**	0.08	0.63	0.14%
	Probs	0.75	7.10	0.14%
	Hybrid	0.72	5.86	0.13%

only consider the TOP-K recommendation list of each user, if the length of the recommendation list is $K = 50$. The hybrid algorithm needs to set a weighting coefficient to determine the weight of the two algorithms, and the weighting coefficient is set to $\lambda = 0.5$. We obtain the performance of the data recommendation system under random segmentation, as shown in Table 1.

However, in the application of the recommendation algorithm, the random selection of the test set is inconsistent with the application of the recommendation system because it neglects the temporal information and can violate causality. Therefore, in this article, the test set can be selected in a temporal way. As all datasets are sorted in accordance with the timestamp in this article, we select the time of the latest 10% of the data as the test set of the data and the rest of the time before as the training set.

3.2 Results of datasets with temporal test sets

On the datasets, after selecting the test set with a temporal sequence, the results are as follows:

For the test set selected in a temporal sequence, several indicators of the performance are a bit lower than those of the results of random selection (shown in Table 2 and Table 3). This is mainly because of the following reason: if the dataset is small and the timestamps of the users are not likely to be evenly distributed, then it is likely that the user's activity will not be selected into the test set. This will reduce the number of samples in the training set; these users will be less connected to the whole system. This will also affect the algorithm for these users of the prediction results, and the calculation of the precision of the algorithm only considers this subset of the user's TOP-K recommended list. So, the accuracy of the four indicators will decline. However, in this way, we obtain the training set and the test set which are closer to the reality, and the result of the algorithm is more meaningful and applicable than that of random selection of test sets.

3.3 Results incorporating centralities

In the aforementioned Materials and methods section, it was shown that for the original algorithm, calculation is directly carried out through a probability transition matrix calculation

TABLE 3 Recommendation evaluations on seven datasets with temporal selection of test sets.

Dataset	Algorithm	Precision (%)	Recall (%)	F1-score (%)
ml_latest small	**Heats**	0.70	0.14	0.23
	Mass	4.78	3.61	4.12
	Hybrid	5.22	3.95	4.48
ml_90K	**Heats**	0.78	0.48	0.59
	Mass	6.31	6.53	6.42
	Hybrid	6.41	6.69	6.59
ml_100K	**Heats**	4.66	5.56	5.07
	Mass	8.81	14.80	11.0
	Hybrid	8.99	14.65	11.1
ml_1M	**Heats**	5.22	5.16	5.19
	Mass	18.95	15.72	17.2
	Hybrid	19.23	16.52	17.8
Netflix	**Heats**	0.15	0.22	0.18
	Mass	4.68	7.56	5.78
	Hybrid	4.55	8.15	5.84
ml_10M	**Heats**	2.40	2.02	2.20
	Mass	4.17	3.29	3.68
	Hybrid	4.13	3.28	3.66
Last.fm	**Heats)**	0.20	0.61	0.30
	Mass	1.74	5.10	2.60
	Hybrid	1.43	4.25	2.14

TABLE 4 Recommendation results adopting centralities.

	Precision (%)	Recall (%)	F1-score (%)
Degree (heats)	0.70	0.14	**0.23**
Degree (mass)	4.78	3.61	4.12
Degree (hybrid)	5.22	3.95	4.50
Closeness (heats)	4.51	3.32	**3.83**
Closeness (mass)	4.62	3.38	3.90
Closeness (hybrid)	4.54	3.31	3.82
Eigenvector (heats)	0.54	0.08	**1.32**
Eigenvector (mass)	4.95	3.88	4.35
Eigenvector (hybrid)	5.08	4.14	4.56
PageRank (heats)	4.24	3.28	**3.70**
PageRank (mass)	4.68	3.65	4.10
PageRank (hybrid)	4.70	3.70	4.14
Katz (heats)	4.35	3.27	**3.74**
Katz (mass)	4.38	3.29	3.76
Katz (hybrid)	4.41	3.30	3.78

formula, so that the time complexity is high; however, the formula can be converted into a matrix form through simplification, namely, the weight of heat or resource allocation can be changed, so that the performance of the algorithm can be improved. We calculate the five network structural metrics, and the results of the movielens_latest small dataset are as follows:

From the aforementioned Table 4, the following conclusions can be obtained between different network structural centralities and different algorithms:

1) The results of various network structural centralities basically meet the common knowledge that the precision of the hybrid algorithm is greater than that of the mass diffusion algorithm and is greater than that of the heat conduction algorithm. However, for the results of PageRank or eigenvector centrality, it can be seen that the results of mass diffusion algorithm are better than those of hybrid algorithm, mainly due to the influence of the fixed weighting coefficient of the hybrid algorithm.
2) More interestingly, by changing the network structural centrality, the precision of the algorithm is improved, especially the precision of the heat conduction algorithm is greatly improved by 5.73–16.7 times, with the F1-score measure, and almost reaches the level of mass diffusion. The results of HC are marked in bold font.

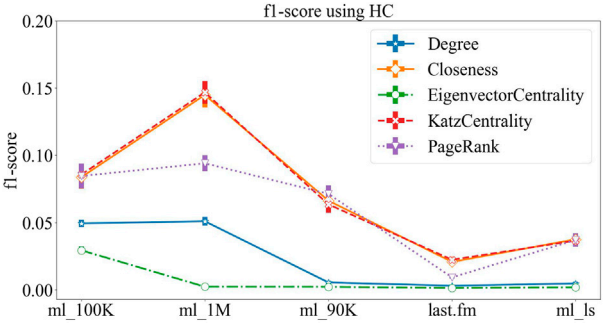

FIGURE 1
Results of different datasets using heat conduction.

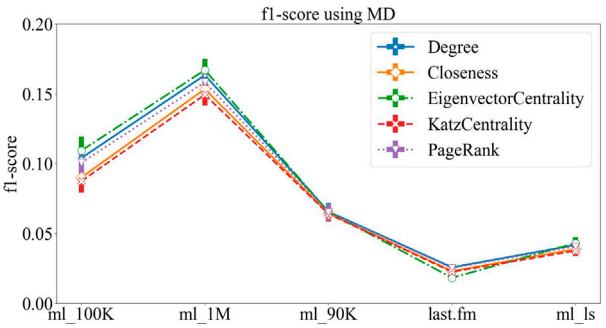

FIGURE 2
Recommendation results of five datasets using mass diffusion.

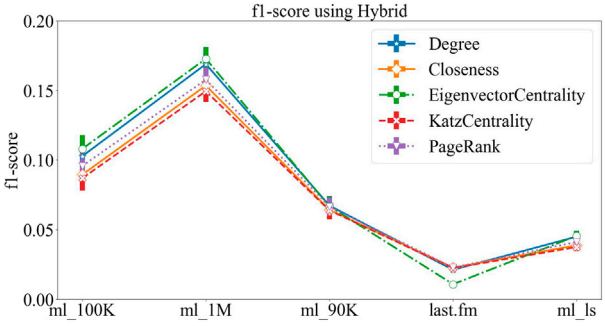

FIGURE 3
Recommendation results of five datasets using hybrid algorithm.

For further analysis, we choose the more comprehensive index F1-score to compare the differences of recommendation results obtained by different centralities on five datasets. As in the aforementioned dataset, the F1-score results obtained by the heat conduction algorithm using five network structural indicators are shown in Figure 1.

From the aforementioned figure, we can see that the method of introducing the network structural metrics improves the performance when using the HC algorithm. Among these metrics, Katz centrality and closeness centrality are the most effective, but the eigenvector centrality greatly reduces the precision of the original method. Moreover, we can see that the F1-score of the algorithm is improved by about 280% by using Katz centrality and closeness centrality on the movielens_1M dataset, and the F1-score of the algorithm is improved by 184% by using PageRank.

The results of MD and hybrid algorithms are shown in Figure 2 and Figure 3, respectively. The figures show that the results of the two algorithms are relatively similar, and the recommendation results when different structural centralities are applied are basically similar; also, there are a few small differences in different datasets. But one can see that for each dataset, when using the three metrics, the results of degree, eigenvector centrality, and PageRank are better, while the results of the other two metrics are slightly worse, but the differences between each metric are generally within 5%. On some datasets, such as movielens_100K and movielens_1M, we can see that the recommendation results obtained by using eigenvector centrality are marginally better than those obtained by using the degree.

4 Discussion

In this study, we mainly focus on the traditional collaborative filtering (CF)-based recommendation systems to improve the information-filtering technique by utilizing the network structural centrality. This article focuses on the recommendation systems that are based on the diffusion processes: heat conduction algorithm and mass diffusion algorithm, which are based on physics theories. However, these two algorithms have their own focus in terms of accuracy and diversity. Combining the two algorithms by a weighting coefficient λ can lead to a hybrid algorithm that can obtain better performance in both the metrics [20].

The recommendation results of the aforementioned three algorithms that adopt degree information as the input of the recommendation system are obtained on the movielens and other datasets commonly used. Through the analysis, we find that the degree contains network structural information of nearest neighbors, which only reflects the number of neighbors of each node. Therefore, this article proposes an improved method that contains the information of higher-dimensional network structural information to improve the recommendation algorithm, by using metrics such as closeness centrality, eigenvector centrality, Katz centrality, and PageRank These metrics not only take the number of neighboring nodes into account but also contain the

importance of neighboring nodes as well as their structural information.

The method proposed in this article shows that applying different network structural centralities improves the recommendation performance. The results of the HC algorithm obtained with all the tested network structural centralities show an improvement in performance in the range of 184%–280% (the only exception is that using eigenvector centrality causes a decrease in accuracy). For the MD algorithm, the differences in the results obtained after applying different metrics are small. Among them, the optimal results were obtained for three metrics: degree, eigenvector centrality, and PageRank. Meanwhile, the hybrid algorithm has overall better prediction results, but the results of some metrics are likely to be influenced by the weighting factor λ, and even better results could be obtained after adjusting the parameters.

Overall, in this article, we show that the centrality of networks contains higher-order structural information of the network topology than the traditional adoption of degree. Surprisingly, the recommendation algorithms incorporating such centralities, especially the heat conduction algorithm will have a significantly improved performance, almost comparable to that of the mass diffusion algorithm.

Data availability statement

Publicly available datasets were analyzed in this study. These data can be found at Stanford Large Network Dataset Collection: https://snap.stanford.edu/data/.

Author contributions

YK designed the research; YH and XZ collected the data; Cheng Wang, YK, and YH analyzed the data; and XZ performed the visualization. All authors wrote the manuscript.

Funding

This work was supported by the Research Funds for the Central Universities from Beijing University of Posts and Telecommunications, under grant No. 505022019, and NO. 500422415, and Research Innovation Fund for College Students of Beijing University of Posts and Telecommunications.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

1. Lü L, Medo M, Yeung CH, Zhang YC, Zhang ZK, Zhou T. Recommender systems. *Phys Rep* (2012) 5191:1–49. doi:10.1016/j.physrep.2012.02.006

2. Resnick P, Iacovou N, Suchak M, Bergstrom P, Riedl JG. An open architecture for collaborative filtering of netnews. In: Proceedings of the 1994 ACM conference on Computer supported cooperative work; Chapel Hill North Carolina USA (1994). p. 175–86.

3. Sarwar B, Karypis G, Konstan J, Riedl J. Item-based collaborative filtering recommendation algorithms. In: Proceedings of the 10th international conference on World Wide Web; Hong Kong Hong Kong (2001). p. 285–95.

4. Breese JS, Heckerman D, Kadie C. Empirical analysis of predictive algorithms for collaborative filtering (2013). Available from: https://arxiv.org/ftp/arxiv/papers/1301/1301.7363.pdf.1301.7363

5. Goldberg D, Nichols D, Oki BM, Terry D. Using collaborative filtering to weave an information tapestry. *Commun ACM* (1992) 3512:61–70. doi:10.1145/138859.138867

6. Ungar LH, Foster DP. Clustering methods for collaborative filtering. *AAAI Workshop recommendation Syst* (1998) 1:114–29.

7. Ungar L, Foster DP. A formal statistical approach to collaborative filtering. CONALD'98 (1998). Available from: https://www.cis.upenn.edu/~ungar/Datamining/Publications/CONALD.pdf.

8. Azar Y, Fiat A, Karlin A, McSherry F, Saia J Spectral analysis of data. In: Proceedings of the thirty-third annual ACM symposium on Theory of computing: San Jose California USA (2001). 619–26.

9. Koren Y, Bell R, Volinsky C. Matrix factorization techniques for recommender systems. *Computer* (2009) 42(8):30–7. doi:10.1109/mc.2009.263

10. Blei DM, Ng AY, Jordan MI. Latent Dirichlet allocation. *J machine Learn Res* (2003) 3(Jan):993–1022.

11. Keshavan R, Montanari A, Oh S. Matrix completion from a few entries. *IEEE Trans Inf Theor* (2010) 56(6):2980–98. doi:10.1109/tit.2010.2046205

12. Albert R, Barabási AL. Statistical mechanics of complex networks. *Rev Mod Phys* (2002) 74(1):47–97. doi:10.1103/revmodphys.74.47

13. Castellano C, Fortunato S, Loreto V. Statistical physics of social dynamics. *Rev Mod Phys* (2009) 81(2):591–646. doi:10.1103/revmodphys.81.591

14. Zhang YC, Blattner M, Yu YK. Heat conduction process on community networks as a recommendation model. *Phys Rev Lett* (2007) 99(15):154301. doi:10.1103/physrevlett.99.154301

15. Zhou T, Ren J, Medo M, Zhang YC. Bipartite network projection and personal recommendation. *Phys Rev E* (2007) 76(4):046115. doi:10.1103/physreve.76.046115

16. Zhang YC, Medo M, Ren J, Zhou T, Li T, Yang F. Recommendation model based on opinion diffusion. *Europhys Lett* (2007) 80(6):68003. doi:10.1209/0295-5075/80/68003

17. Liu JH, Zhang ZK, Chen L, Liu C, Yang C, Wang X. Gravity effects on information filtering and network evolving. *PloS one* (2014) 9(3):e91070. doi:10.1371/journal.pone.0091070

18. Zeng W, Zeng A, Liu H, Shang MS, Zhou T. Uncovering the information core in recommender systems. *Sci Rep* (2014) 4(1):6140–8. doi:10.1038/srep06140

19. Zhang QM, Zeng A, Shang MS. Extracting the information backbone in online system. *PloS one* (2013) 8(5):e62624. doi:10.1371/journal.pone.0062624

20. Zhou T, Kuscsik Z, Liu JG, Medo M, Wakeling JR, Zhang YC. Solving the apparent diversity-accuracy dilemma of recommender systems. *Proc Natl Acad Sci U S A* (2010) 107(10):4511–5. doi:10.1073/pnas.1000488107

21. Fiasconaro A, Tumminello M, Nicosia V, Latora V, Mantegna RN. Hybrid recommendation methods in complex networks. *Phys Rev E* (2015) 92(1):012811. doi:10.1103/physreve.92.012811

22. Stojmirović A, Yu YK. Information flow in interaction networks. *J Comput Biol* (2007) 14(8):1115–43. doi:10.1089/cmb.2007.0069

23. Liu JG, Zhou T, Guo Q. Information filtering via biased heat conduction. *Phys Rev E* (2011) 84(3):037101. doi:10.1103/physreve.84.037101

24. Shang MS, Jin CH, Zhou T, Zhang YC. Collaborative filtering based on multi-channel diffusion. *Physica A: Stat Mech its Appl* (2009) 388(23):4867–71. doi:10.1016/j.physa.2009.08.011

25. Zhou T, Jiang LL, Su RQ, Zhang YC. Effect of initial configuration on network-based recommendation. *Europhys Lett* (2008) 81(5):58004. doi:10.1209/0295-5075/81/58004

26. Lü L, Liu W. Information filtering via preferential diffusion. *Phys Rev E* (2011) 83(6):066119. doi:10.1103/physreve.83.066119

27. Liu JG, Shi K, Guo Q. Solving the accuracy-diversity dilemma via directed random walks. *Phys Rev E* (2012) 85(1):016118. doi:10.1103/physreve.85.016118

28. Zhang ZK, Liu C. A hypergraph model of social tagging networks. *J Stat Mech* (2010) 10:P10005. doi:10.1088/1742-5468/2010/10/p10005

29. Zhang ZK, Zhou T, Zhang YC. Tag-aware recommender systems: A state-of-the-art survey. *J Comput Sci Technol* (2011) 26(5):767–77. doi:10.1007/s11390-011-0176-1

30. Yu F, Koltun V. Multi-scale context aggregation by dilated convolutions. Available from: https://arxiv.org/abs/1511.07122.

31. Ren X, Lü L, Liu R, Zhang J. Avoiding congestion in recommender systems. *New J Phys* (2014) 16(6):063057. doi:10.1088/1367-2630/16/6/063057

32. Deng X, Wu L, Ren X, Jia C, Zhong Y, Lü L. Inferring users' preferences through leveraging their social relationships. In: Proceedings of the IECON 2017-43rd Annual Conference of the IEEE Industrial Electronics Society; Beijing, China (October 2017). p. 5830–6.

33. Davis J, Goadrich M. The relationship between Precision-Recall and ROC curves. In: Proceedings of the 23rd international conference on Machine learning; Pittsburgh Pennsylvania USA (June 2006). p. 233–40.

34. Zhang Y-C, Marcel B, Yi-Kuo Y. Heat conduction process on community networks as a recommendation model. *Phys Rev Lett* (2007) 99(15):154301. doi:10.1103/physrevlett.99.154301

35. Ren ZM, Kong Y, Shang MS, Zhang YC. A generalized model via random walks for information filtering. *Phys Lett A* (2016) 380(34):2608–14. doi:10.1016/j.physleta.2016.06.009

36. Chea E, Livesay DR. How accurate and statistically robust are catalytic site predictions based on closeness centrality? *Bmc Bioinformatics* (2007) 8(1):153–14. doi:10.1186/1471-2105-8-153

37. Gabriel P, Francis Narin. Citation influence for journal aggregates of scientific publications: Theory, with application to the literature of physics. *Inf Process Management* (1976) 12(5):297–312. doi:10.1016/0306-4573(76)90048-0

38. Katz L. A new status index derived from sociometric analysis. *Psychometrika* (1953) 18:39–43. doi:10.1007/bf02289026

39. Brin S, Page L. The anatomy of a large-scale hypertextual web search engine. *Computer networks ISDN Syst* (1998) 30(1-7):107–17. doi:10.1016/s0169-7552(98)00110-x