Received September 20, 2019, accepted September 30, 2019, date of publication October 8, 2019, date of current version October 23, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2946262

Unbalanced Multistage Heat Conduction and Mass Diffusion Algorithm in an Educational Digital Library

PENGFEI YIN[1,2,3], GUOJUN WANG[2], (Member, IEEE), MD ZAKIRUL ALAM BHUIYAN[4], (Senior Member, IEEE), MEIJING SHAN[5], AND FANG QI[1]

[1]School of Computer Science and Engineering, Central South University, Changsha 410083, China
[2]School of Computer Science, Guangzhou University, Guangzhou 510006, China
[3]College of Information Science and Engineering, Jishou University, Jishou 416000, China
[4]Department of Computer and Information Sciences, Fordham University, New York, NY 10458, USA
[5]Institute of Information Science and Technology, East China University of Political Science and Law, Shanghai 201620, China

Corresponding author: Fang Qi (csqifang@csu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61632009, in part by the Guangdong Provincial Natural Science Foundation under Grant 2017A030308006, in part by the National Social Science Fund of China under Grant 16BFX085, in part by the Science and Technology Project of Changsha under Grant kq1701089, and in part by the High-Level Talents Program of Higher Education in Guangdong Province under Grant 2016ZJ01.

ABSTRACT Discovering valuable and diverse resources in an educational digital library can be difficult with the existing large number of content collections. Researchers in information retrieval and related domains have moved from considering only keyword-based matching to modeling the underlying behavior pattern of users. To achieve the purpose of both precision and diversity while providing online educational resource services, in this paper, we propose a weighted network-based information filtering framework that models user usage as a bipartite user-resource network; users and resources are treated as nodes in this network, each edge from a user to a resource means usage, and the weight represents the accumulation of multiple usage scenarios. Under this framework, we propose two individual algorithms, the unbalanced heat conduction algorithm and the unbalanced mass diffusion algorithm, and one hybrid multistage heat conduction and mass diffusion algorithm. There are two stages in these algorithms. In stage 1, an initial energy is assigned to each resource that has been visited by the target and passes to users according to a specific strategy. In stage 2, the energy is similarly transferred from users to resources, and a sorted resource list ranked by energy is presented to the target user. Experiments on a real-world dataset of one year of academic search logs showed improved performance from multiple indicators compared to existing algorithms.

INDEX TERMS Heat conduction, mass diffusion, network-based search, information filtering.

I. INTRODUCTION

The The rapid development of the Internet has created great changes in the field of education. Educational resources are gradually transitioning from paper to electronic. In the past, when electronic resources were limited, catalog technology solved the problem of people looking for resources [1]. With the explosive development of educational resources, traditional catalog technology has been unable to meet the demand, and search engines have emerged. For researchers, a large number of academic search engines, such as Google Scholar [2], PubMed [3], and arXiv.org [4], have increased searching convenience. The educational digital library (edu-DL) plays a key role in retrieving resources from educational platforms, and their functions are mostly the same [5]. People can search by author's name, organization, journal name, etc. The search results can also be sorted based on various indicators, such as citation, date, relevance, and so on. Figure 1 shows the search result of Web of Science, which we searched by the keyword – mass diffusion.They have collected a large quantity of research information that is widely distributed across different databases, systems and information portals.

The associate editor coordinating the review of this manuscript and approving it for publication was Yuyu Yin.

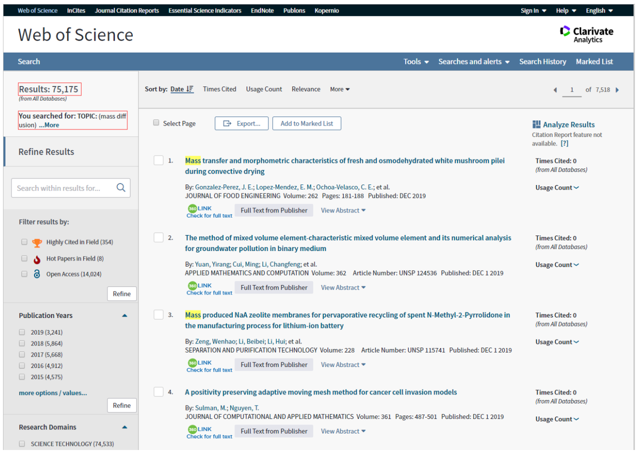

FIGURE 1. The search result in web of science with keyword "mass diffusion".

Supporting services include document retrieval, reading, and downloading. However, the accumulation of considerable resources and activities leads to the problem of information overload–finding high-quality resources is similar to finding a needle in a haystack. Due to the openness of most educational systems, users can navigate anonymously. Less identifiable users and rare feedback make it extremely difficult to distinguish valuable resources. The user's intent is ambiguous in most cases, which makes keyword-based retrieval unsatisfactory. At the same time as providing services, the quality of service must also be guaranteed [6], [7].

Interactive information retrieval (IIR) considers the past interactions of the search user and has improved user experiences. In such a system, functions such as search strategies [8], [9], term suggestions [10], and personalization are studied. However, IIR ignores the interactions of other users and inevitably loses considerable valuable information. This usage information can provide useful decision assistance with the quality of resources. Usage-based search is becoming increasingly important in the current era of personalization and collective intelligence. As two means of information filtering, search engines and recommender systems are common, they provide users with separate or combined services [11], [12]. A keyword-based search can be treated as one kind of content-based filtering in the recommender system. Similarly, we can learn from the idea of collaboration in the recommender system, which emphasizes collective intelligence. Under this consideration, in our paper, we map the usage of all users into a user-resource bipartite network, and we propose an improved heat conduction (HC) and mass diffusion (MD) algorithm for energy transmission in the network. Existing work has shown acceptable performance on accuracy and diversity of the heat conduction algorithm and the mass diffusion algorithm, respectively [13].

Both the HC algorithm and the MD algorithm belong to network-based methods. In these two models, the user's past interactions are mapped into a bipartite user-resource network. Users and resources are nodes in the network, and the edge between the user and resource represents usage. There is a two-stage energy transmission for these algorithms. In stage 1, each resource that has been visited by the target user is assigned an initial energy; then, the energy is passed to users through the connected edges. The transmission process in stage 2, which is from users to resources, is similar. Finally, the resources are ranked according to the energy they own and provided to the target user. The network structure of these two methods is the same, and the difference lies in how energy is transferred. In a certain transmission process, energy transfer depends on the degree of either the user or the resource.

The activities of users in edu-DL include two kinds: **explicit** and **implicit**. **Explicit** activities are the behaviors that users show in their attitude toward documents very clearly, such as ratings, comments, and feedback. However, **implicit** activities are often not visible to other users in the system, such as searches, views, and downloads. These activities can help build different models to find similarities between users and help users discover high-quality resources that other users have visited. In the traditional HC algorithm and MD algorithm, explicit activities are modeled as the edges between users and resources in the user-resource bipartite network. Explicit activities are sparse due to the openness and anonymity of current edu-DL. However, implicit activities can better reflect the real attitudes of users because they can give false ratings for their purposes.

For these considerations, in our paper, we propose integrating explicit and implicit activities to build the hybrid HC and MD model. In the traditional model, there are cases where some resources are not covered due to the limited energy transmission process. In our proposed model, we overcome this issue to use multistage energy transmission. Moreover, energy transmission is balanced in the traditional model, while in the real world, we know that different people and resources are different in importance because influential people can spread resources more widely and quickly. Considering this problem, we propose using unbalanced transmission, which increases the effect of influential users and reduces that of the popular resources.

Our contributions in this paper are threefold:

1) We propose a weighted bipartite network-based framework for solving the resource ranking problem in edu-DL. Both explicit and implicit activities of users are used to construct our network model.
2) Under this framework, we propose two individual algorithms, unbalanced HC (UHC) and unbalanced MD (UMD), and one hybrid algorithm, unbalanced multistage HC and MD (UHM), to balance the precision and diversity while performing information filtering in edu-DL. Our proposed algorithms consider the user's influence and resource's popularity compared with classic algorithms. Moreover, the energy transmission in our algorithms is unbalanced, which fully utilizes the interaction data of users.
3) Considering the low coverage problem of classic algorithms, we propose the multistage energy transmission

process, which has the disadvantage of the classic two-stage energy transmission process.

The remainder of this paper is organized as follows. We introduced related works in Section II. Section III presents the classic HC and MD algorithms. The proposed new algorithms are shown in Section IV. Section V shows a clear analysis of the experiment on a real-world dataset for both individual and hybrid algorithms. The conclusion of our paper is given in Section VI.

II. RELATED WORK

The explosive increase in information has promoted the transformation of information filtering technology, and the network-based method has achieved satisfactory results and has been greatly favored by academics and industry. One of the more well-known technologies is Google's PageRank algorithm [14], which treats web pages as nodes, creates a directed graph based on the relationships between them, and sorts them by the importance. The various resources on the Internet are connected by complex relationships to form a complex network [15]. The development of network structure has led researchers to apply it to information filtering, such as social networks [16], bipartite networks [17], and tripartite networks [18]. Among these models, the bipartite network has received extensive attention.

The bipartite network projection can be more informative than the one-mode projection [17]. Similar to the probability-based model [19], the random walk methods on the bipartite network have been widely studied in information filtering, especially in recommender systems [20]. With the weighted projection of the user-object network, a network-based inference (NBI) model was proposed and demonstrated higher accuracy than the classic algorithms [21]. Learning from the principle of the heat conduction algorithm [22] and combined with the proposed mass diffusion algorithm in [17], the hybrid heat conduction and mass diffusion (HHM) algorithm solved the diversity-accuracy dilemma [13]. Researchers have also proposed a series of improvements for the heat conduction algorithm, such as biased heat conduction (BHC) [23], weighted heat conduction (WHC) [24], local heat conduction (LHC) [25], heterogeneous heat conduction (HHC) [26], and top-n-stability heat conduction (TNS-HC) [27]. Social influence has also been applied on the user-object bipartite network [28], [29]. By taking only a group of core users into consideration, the K-Nearest neighbor mass diffusion (KNNMD) algorithm can help improve recommendation efficiency in a system with a large number of users [30]. The problem of superfluous diffusion in the mass diffusion model can be solved with symmetrical punishment [31]. From the dataset aspect, the performance of HHM was improved by adopting partial recent behaviors [32] or negative ratings [33]. To increase diversity and novelty, balanced diffusion (BD) [34], preferential diffusion (PD) [35], heat bidirectional transfer (HBT) [36], preferential bidirectional mass diffusion (PBMD) [37] and a

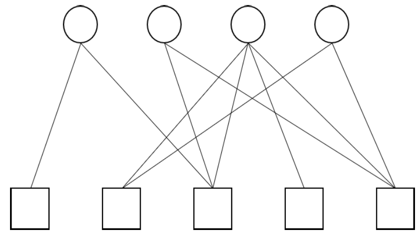

FIGURE 2. Bipartite network of user-resource. Users are shown in circles, and resources are shown in squares.

model through eliminating redundant diffusion and compensating balance (ERD-CB) [38] have been proposed. As the key components of the information network, time and trust information were used in the diffusion-based model, such as the time-aware HHM (THHM) model [39] and the trust-based diffusion model [40], [41].

Here, we propose an improved hybrid heat conduction and mass diffusion algorithm for solving the precision, diversity, and coverage problem while searching in educational digital libraries. The difference between our model and existing methods is mainly in three aspects. First, we map both the explicit and implicit usage of resources into a weighted bipartite network, while in most literature, the explicit ratings are used to construct the network without weight. Second, we enhance the diffusion ability of high-impact users, reduce the diffusion ability of popular resources, and unbalance energy transmission. Finally, the multistage energy transmission process is executed in our model to overcome the low coverage problem of the existing method. The energy of some resources will always be 0 with classic two-stage methods and will not be covered.

III. PRELIMINARY

As the intersections between disciplines become more frequent, the theoretical approaches across different disciplines have received much attention. The heat conduction and mass diffusion algorithms are borrowed from physics theory. Among them, the heat conduction algorithm has better diversity, while the mass diffusion algorithm has higher accuracy. In computer science, they are transformed into a bipartite network; users and resources are treated as nodes, the attributes of users and resources are ignored, and only the interactions between users and resources are considered.

The bipartite network model is defined by graph $G = < U, R, E >$, which is shown in Figure 2, where $U = \{u_1, u_2, \ldots, u_n\}$ is the user set, $R = \{r_1, r_2, \ldots, r_m\}$ is the resource set, and $E = \{e_{ij}, i \in U; j \in R\}$ is the edge set. If there are interactions between u_i and r_j, then $e_{ij} = 1$, otherwise $e_{ij} = 0$. When the algorithm is executed, each resource that has been visited by the target user is given an initial energy, and these resources can transfer part of their energy to other resources. Assume that the initial energy of r_β is e_β, and the transfer weight between r_α and r_β is $w_{\alpha\beta}$, which is the ratio passed from r_β to r_α. After execution, the total energy obtained by r_α is calculated by (1). Finally, the ranking

TABLE 1. Notations.

Notation	Description
R	Resource space
U	User space
N_j	Activity/Popularity level of u_j/r_j, which is the sum of the weight of all edges connected with u_j/r_j.
u_i	User u_i
r_j	Resource r_j
e_k	Energy owned by r_k or u_k
d_m	Degree of r_m or u_m, which is defined as the number of connected edges
$w^A_{\alpha\beta}$	Transfer weight from resource α to resource β under algorithm A
g_n	Gained energy of r_n or u_n
a_{ij}	Weight of edge <i, j>, which connects u_i and r_j, and $a_{ij} = a_{ji}$. In the HC and MD algorithm, if there is an edge between u_i and r_j, then $a_{ij} = 1$; otherwise, $a_{ij} = 0$. In the UHM algorithm, if there is an edge between u_i and r_j, then $a_{ij} \geq 1$; otherwise, $a_{ij} = 0$.

list sorted by energy is provided to the target user. A summary of the notations used in this paper is shown in Table 1.

$$g_\alpha = \sum_{\beta=1}^{|R|} w_{\alpha\beta} e_\beta \tag{1}$$

A. HEAT CONDUCTION (HC)

The premise of the heat conduction algorithm is to assume that each resource that interacts with the user has a certain initial energy and can be transmitted through the user-resource bipartite network. The process of transmission mainly consists of the following two stages:

Stage 1: Resource → user. The initial energy of each resource that interacts with the target user is set to 1, and the energy of the other nodes is set to 0. Each resource passes energy to the connected users according to the degree of the users. Finally, the energy obtained by each user is the mean of the energy passed by all connected resources, which is shown in (2).

$$g_i = \sum_{j=1}^{|R|} \frac{a_{ij} e_j}{d_i} \tag{2}$$

where a_{ij} is the weight of edge <i, j>, which connects u_i and r_j, e_j is the energy owned by r_j, and d_i is the degree of u_i.

Stage 2: User → resource. The initial energy of the users is calculated from **Stage 1**, and the energy of all resources is set to 0. Each user passes energy to the connected resources according to the degree of resources. Finally, the energy obtained by each resource is the mean of the energy passed by all connected users, which is shown in (3).

$$g_j = \sum_{i=1}^{|U|} \frac{a_{ij} e_i}{d_j} \tag{3}$$

where a_{ij} is the weight of edge <i, j>, which connects u_i and r_j, e_i is the energy owned by u_i, and d_j is the degree of r_j.

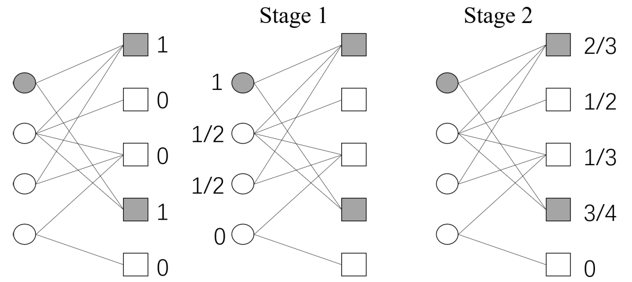

FIGURE 3. The HC algorithm at work on user-resource bipartite network.

From Equation 1, 2 and 3, we derive the transfer weight of the heat conduction algorithm, which is shown in (4).

$$w^H_{\alpha\beta} = \frac{1}{d_\alpha} \sum_{i=1}^{|U|} \frac{a_{i\alpha} a_{i\beta}}{d_i} \tag{4}$$

The derivation process is as follows.

$$\begin{aligned} g_\alpha &= \sum_{i=1}^{|U|} \frac{a_{i\alpha} e_i}{d_\alpha} \\ &= \sum_{i=1}^{|U|} \frac{a_{i\alpha}}{d_\alpha} \sum_{j=1}^{|R|} \frac{a_{ij} e_j}{d_i} \\ &= \sum_{i=1}^{|U|} \frac{a_{i\alpha}}{d_\alpha} \sum_{\beta=1}^{|R|} \frac{a_{i\beta} e_\beta}{d_i} \\ &= \sum_{\beta=1}^{|R|} \left(\frac{1}{d_\alpha} \sum_{i=1}^{|U|} \frac{a_{i\alpha} a_{i\beta}}{d_i} \right) e_\beta \end{aligned} \tag{5}$$

Compare (5) with (1), then (4) is obvious. The following is the example of the execution process of heat conduction algorithm.

Example 1 ***Heat Conduction****: Consider the small interactions log depicted in Figure 3, where users are shown as circles, and resources are shown as squares. The edges represent the usage of users on resources. The target user is marked as a shaded circle; assume two resources, marked as shaded squares, have been visited by the target user, and the initial energy owned by each resource is* 1*. The degree of user/resource is defined as the number of nodes that are connected to that user/resource. In stage 1, the energy is transferred from resources to users according to the degree of the users. For the target user, the energy gained after transmission is* $(1 + 1)/2 = 1$*. In stage 2, the energy is transferred back to the resources according to the degree of resources. Finally, we obtain the amount of energy that all resources own. We number the resources from top to bottom as* $1 \sim 5$*, then the search result for target user under HC algorithm is* $\{r_2, r_3\}$.

B. MASS DIFFUSION (MD)

The energy transfer process of the mass diffusion algorithm is similar to that of the random walk. It is assumed that the resources that interact with the user have a certain initial energy and can transfer energy to other resources. The execution process of the algorithm also includes two stages:

Stage 1: Resource → user. The process of energy transmission is similar to the heat conduction algorithm, except that each resource transfers energy to users according to the degree of the resource, which is shown in (6).

$$g_i = \sum_{j=1}^{|R|} \frac{a_{ij} e_j}{d_j} \tag{6}$$

where a_{ij} is the weight of edge <i, j>, which connects u_i and r_j, e_j is the energy owned by r_j, and d_j is the degree of r_j.

Stage 2: User → resource. At this stage, each user also passes energy to the resources according to the degree of the user, which is shown in (7).

$$g_j = \sum_{i=1}^{|U|} \frac{a_{ij} e_i}{d_i} \tag{7}$$

where a_{ij} is the weight of edge <i, j>, which connects u_i and r_j, e_i is the energy owned by u_i, and d_i is the degree of u_i.

The derived transfer weight of the mass diffusion algorithm is shown in (8).

$$w_{\alpha\beta}^{M} = \frac{1}{d_\beta} \sum_{i=1}^{|U|} \frac{a_{i\alpha} a_{i\beta}}{d_i} \tag{8}$$

The following is the derivation process.

$$\begin{aligned} g_\alpha &= \sum_{i=1}^{|U|} \frac{a_{i\alpha} e_i}{d_i} \\ &= \sum_{i=1}^{|U|} \frac{a_{i\alpha}}{d_i} \sum_{j=1}^{|R|} \frac{a_{ij} e_j}{d_j} \\ &= \sum_{i=1}^{|U|} \frac{a_{i\alpha}}{d_i} \sum_{\beta=1}^{|R|} \frac{a_{i\beta} e_\beta}{d_\beta} \\ &= \sum_{\beta=1}^{|R|} \left(\frac{1}{d_\beta} \sum_{i=1}^{|U|} \frac{a_{i\alpha} a_{i\beta}}{d_i} \right) e_\beta \end{aligned} \tag{9}$$

Compare (9) with (1), we obtain (8). The following is the example of the execution process of mass diffusion algorithm.

Example 2 ***Mass Diffusion****: The lower part of Figure 4 illustrates the energy transmission process of the mass diffusion algorithm. For the resources visited by the target user, we set an initial energy of 1, and the other resources and the user's energy are initialized to 0. Then the energy of the resources is passed through two stages. In stage 1, the energy is transferred from resources to users according to the degree of the resources. After transmission, the energy gained by the target user is* $(1/3 + 1/2) = 5/6$*. Then, users transfer their energy to resources according to the degree of the users in stage 2. Finally, we obtain the resource list ranked by the energy that they own. We know that the search result for target user under MD algorithm is* $\{r_3, r_2\}$.

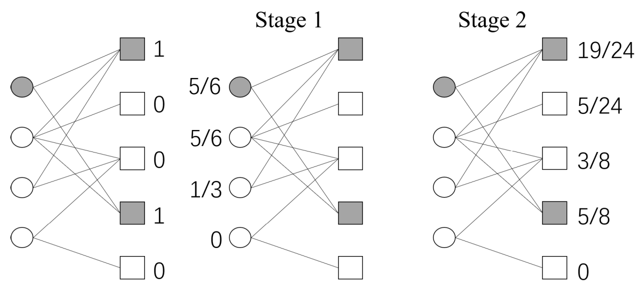

FIGURE 4. The MD algorithm at work on user-resource bipartite network.

IV. THE PROPOSED ALGORITHMS

The characteristics of the heat conduction algorithm and mass diffusion algorithm can be seen in Figure 3 and 4. In the heat transfer algorithm, the second resource is eventually placed in front of the third resource, and it is the opposite in the mass diffusion algorithm. From the degree of resources, we can see that the heat transfer algorithm tends to provide novel resources, which is very diverse and novel, while the mass diffusion algorithm tends to provide popular resources, which ensures accuracy. However, in these two algorithms, the energy transmission process of stage 1 treats the first resource and the fourth resource equally, which is unfair, because the distinguishing ability between the popular resource and the unpopular resource is different, unpopular resources can better determine the user's personality. Moreover, it is unreasonable to treat all users equally in the process of energy transmission in stage 2 because the influence of users is different. Moreover, after the execution of the two algorithms, the energy of the fifth resource is 0 and cannot be provided to the target user. Here, we propose an unbalanced multistage energy transmission strategy to solve these problems and to balance accuracy and diversity through the hybrid of these two algorithms.

Assume that the initial energy owned by resources is represented by the vector $\mathbf{e}$, then the energy is transferred through (10):

$$\tilde{\mathbf{e}} = \left(\mathbf{W}^{\mathbf{UHM}}\right)^{\mathbf{m}} \mathbf{e} \tag{10}$$

where $\tilde{\mathbf{e}}$ is the energy vector after transmission, $\mathbf{W}^{\mathbf{UHM}}$ is the transfer weight, and $\mathbf{m} \in \mathbb{N}_+$.

If the user's behavior in the system is more active, we should improve its energy transmission ability. Here, we achieve this by stretching the energy the user owns, which is shown in (11).

$$e_j{}' = log(1 + N_j) e_j \tag{11}$$

where N_j is the activity level of u_j, $N_j \in \mathbb{N}_+$.

Similarly, we also compress the energy of popular resources to reduce their ability to transfer.

$$e_j{}'' = \frac{1}{log(1 + N_j)} e_j \tag{12}$$

where N_j is the popularity level of r_j, $N_j \in \mathbb{N}_+$.

A. UNBALANCED HEAT CONDUCTION (UHC)

In the classic heat conduction algorithm, energy is transferred in a balanced manner. In the edu-DL, we consider that different methods of usage should create different methods of energy transmission. Therefore, we assign a certain weight to the bipartite network and associate it with energy transmission. The execution process of the unbalanced heat conduction algorithm is as follows:

Stage 1: Resource → user. The initial energy of each resource that interacts with the target user is set to 1, and the energy of other nodes is set to 0. Each resource passes energy to connected users according to the weight of edges. Finally, the energy obtained by u_i is shown in (13).

$$g_i = \sum_{j=1}^{|R|} \frac{a_{ij}}{N_i} e_j'' = \sum_{j=1}^{|R|} \frac{a_{ij} e_j}{N_i log(1+N_j)} \tag{13}$$

where e_j'' is defined in (12), a_{ij} is the weight of edge <i, j>, which connects u_i and r_j, N_j is the activity level of u_j, and $N_j \in \mathbb{N}_+$.

Stage 2: User → resource. At this stage, the energy is transferred back to resources. The gained energy of r_j is shown in (14).

$$g_j = \sum_{i=1}^{|U|} \frac{a_{ji}}{N_j} e_i' = \sum_{i=1}^{|U|} \frac{a_{ij} e_i log(1+N_i)}{N_j} \tag{14}$$

where e_i' is defined in (11), a_{ij} is the weight of edge <i, j>, which connects u_i, r_j, N_j is the activity level of u_j, and $N_j \in \mathbb{N}_+$.

The derived transfer weight is shown in (15).

$$w_{\alpha\beta}^{UH} = \frac{1}{N_\alpha log(1+N_\beta)} \sum_{i=1}^{|U|} \frac{a_{i\alpha} a_{i\beta} log(1+N_i)}{N_i} \tag{15}$$

The following is the derivation process.

$$\begin{aligned} g_\alpha &= \sum_{i=1}^{|U|} \frac{a_{i\alpha} e_i log(1+N_i)}{N_\alpha} \\ &= \sum_{i=1}^{|U|} \frac{a_{i\alpha} log(1+N_i)}{N_\alpha} \sum_{j=1}^{|R|} \frac{a_{ij} e_j}{N_i log(1+N_j)} \\ &= \sum_{i=1}^{|U|} \frac{a_{i\alpha} log(1+N_i)}{N_\alpha} \sum_{\beta=1}^{|R|} \frac{a_{i\beta} e_\beta}{N_i log(1+N_\beta)} \\ &= \sum_{\beta=1}^{|R|} \left(\frac{1}{N_\alpha log(1+N_\beta)} \sum_{i=1}^{|U|} \frac{a_{i\alpha} a_{i\beta} log(1+N_i)}{N_i} \right) e_\beta \end{aligned} \tag{16}$$

We know that (15) is obvious when (16) is compared with (1). For simplicity, we define

$$I_i = \frac{N_i}{log(1+N_i)} \tag{17}$$

$$P_{\alpha\beta} = N_\alpha log(1+N_\beta) \tag{18}$$

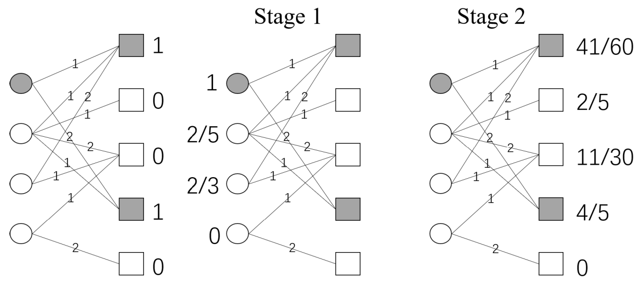

FIGURE 5. The UHC algorithm at work on user-resource bipartite network.

Finally, we obtain the following transfer weight:

$$\mathbf{W}_{\alpha\beta}^{\mathbf{UH}} = \frac{1}{P_{\alpha\beta}} \sum_{i=1}^{|U|} \frac{a_{i\alpha} a_{i\beta}}{I_i} \tag{19}$$

The energy transmission process of UHC algorithm is shown in Figure 5.

B. UNBALANCED MASS DIFFUSION (UMD)

The energy transmission process of this algorithm is similar to the MD algorithm. The difference is that the energy transmission is unbalanced. The execution process is as follows.

Stage 1: Resource → user. The initial energy of each resource that interacts with the target user is set to 1, and the energy of the other nodes is set to 0. Each resource passes energy to the connected users according to the popularity level of the resource and weight of the connected edge. Gained energy of u_i is calculated by (20).

$$g_i = \sum_{j=1}^{|R|} \frac{a_{ij}}{N_j} e_j'' = \sum_{j=1}^{|R|} \frac{a_{ij} e_j}{N_j log(1+N_j)} \tag{20}$$

where e_j'' is defined in (12), a_{ij} is the weight of edge <i, j>, which connects u_i and r_j, N_j is the activity level of u_j, and $N_j \in \mathbb{N}_+$.

Stage 2: User → resource. At this stage, the energy is transferred back to the resources according to the activity level of the users and the weight of the connected edge. Gained energy of r_j is calculated by (21).

$$g_j = \sum_{i=1}^{|U|} \frac{a_{ij}}{N_i} e_i' = \sum_{i=1}^{|U|} \frac{a_{ij} e_i log(1+N_i)}{N_i} \tag{21}$$

where e_i' is defined in (11), a_{ij} is the weight of edge <i, j>, which connects u_i and r_j, and N_i is the activity level of u_i, and $N_i \in \mathbb{N}_+$.

The derived transfer weight is shown in (22).

$$w_{\alpha\beta}^{UM} = \frac{1}{N_\beta log(1+N_\beta)} \sum_{i=1}^{|U|} \frac{a_{i\alpha} a_{i\beta} log(1+N_i)}{N_i} \tag{22}$$

The following is the derivation process.

$$g_\alpha = \sum_{i=1}^{|U|} \frac{a_{i\alpha} e_i log(1+N_i)}{N_i}$$

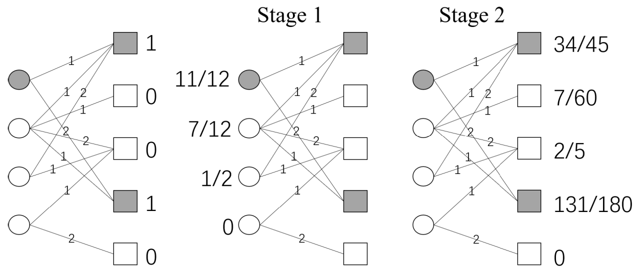

FIGURE 6. The UMD algorithm at work on user-resource bipartite network.

$$= \sum_{i=1}^{|U|} \frac{a_{i\alpha} log(1+N_i)}{N_i} \sum_{j=1}^{|R|} \frac{a_{ij} e_j}{N_j log(1+N_j)}$$
$$= \sum_{i=1}^{|U|} \frac{a_{i\alpha} log(1+N_i)}{N_i} \sum_{\beta=1}^{|R|} \frac{a_{i\beta} e_\beta}{N_i log(1+N_\beta)}$$
$$= \sum_{\beta=1}^{|R|} \left(\frac{1}{N_\beta log(1+N_\beta)} \sum_{i=1}^{|U|} \frac{a_{i\alpha} a_{i\beta} log(1+N_i)}{N_i} \right) e_\beta \tag{23}$$

Equation 22 is obvious when (23) is compared with (1). For simplicity, we define

$$P_\beta = N_\beta log(1+N_\beta) \tag{24}$$

From (17), (22) and (24), we obtain the final transfer weight as follows.

$$\mathbf{W}_{\alpha\beta}^{\mathbf{UM}} = \frac{1}{P_\beta} \sum_{i=1}^{|U|} \frac{a_{i\alpha} a_{i\beta}}{I_i} \tag{25}$$

The energy transmission process of UMD algorithm is shown in Figure 6.

C. UNBALANCED MULTISTAGE HHM (UHM)

The heat conduction algorithm has good diversity but low accuracy, while the mass diffusion algorithm has good accuracy but low diversity. In response to this problem, we propose an unbalanced multistage hybrid algorithm that not only balances accuracy and diversity but also further improves coverage. The most basic hybrid algorithm uses linear weighting (also called parallelized hybridization). However, it needs to execute both individual algorithms, which has a higher computational cost. We propose using pipelined hybridization. Assuming that the initial energy of r_j is e_j, the transmission ratio by the mass diffusion algorithm is m, and transmission ratio by the heat conduction algorithm is h, then the energy obtained by the connected u_i is calculated by (26).

$$g_i = m^\lambda h^{1-\lambda} e_j \tag{26}$$

where $\lambda \in [0, 1]$ is the weight parameter.

The execution process of our proposed UHM algorithm is as follows.

Stage 1: Resource $\rightarrow$ user. The initial energy of each resource that interacts with the target user is set to 1, and the energy of other nodes is set to 0. Each resource passes energy to connected users through both the UH and UM algorithms. Gained energy of u_i is calculated by (27).

$$g_i = \sum_{j=1}^{|R|} e_j'' \left(\frac{a_{ij}}{N_i}\right)^\lambda \left(\frac{a_{ij}}{N_j}\right)^{1-\lambda}$$
$$= \sum_{j=1}^{|R|} \frac{e_j}{log(1+N_j)} \left(\frac{a_{ij}}{N_i}\right)^\lambda \left(\frac{a_{ij}}{N_j}\right)^{1-\lambda}$$
$$= \sum_{j=1}^{|R|} \frac{e_j}{log(1+N_j)} \frac{a_{ij}}{N_i^\lambda N_j^{1-\lambda}} \tag{27}$$

where e_j'' is defined in (12), $\lambda \in [0, 1]$ is the weight parameter, a_{ij} is the weight of edge <i, j>, which connects u_i and r_j, and N_i is the activity level of u_i, $N_i \in \mathbb{N}_+$.

Stage 2: User $\rightarrow$ resource. At this stage, the energy is transferred back to resources according to the hybrid UH and UM algorithm. The gained energy of r_j is calculated by Equation 28.

$$g_i = \sum_{j=1}^{|R|} e_j' \left(\frac{a_{ij}}{N_j}\right)^\lambda \left(\frac{a_{ij}}{N_i}\right)^{1-\lambda}$$
$$= \sum_{j=1}^{|R|} \frac{e_j}{log(1+N_j)} \left(\frac{a_{ij}}{N_j}\right)^\lambda \left(\frac{a_{ij}}{N_i}\right)^{1-\lambda}$$
$$= \sum_{j=1}^{|R|} \frac{e_j}{log(1+N_j)} \frac{a_{ij}}{N_j^\lambda N_i^{1-\lambda}} \tag{28}$$

where e_j' is defined in (11), $\lambda \in [0, 1]$ is the weight parameter, a_{ij} is the weight of edge <i, j>, which connects u_i and r_j, and N_i is the activity level of u_i, $N_i \in \mathbb{N}_+$.

The derived transfer weight is shown in (29).

$$w_{\alpha\beta}^{UHM} = \frac{1}{log(1+N_\beta) N_\alpha^\lambda N_\beta^{1-\lambda}} \sum_{i=1}^{|U|} \frac{a_{i\alpha} a_{i\beta} log(1+N_i)}{N_i} \tag{29}$$

The following is the derivation process.

$$g_\alpha = \sum_{i=1}^{|U|} e_i log(1+N_i) \frac{a_{i\alpha}}{N_\alpha^\lambda N_i^{1-\lambda}}$$
$$= \sum_{i=1}^{|U|} log(1+N_i) \frac{a_{i\alpha}}{N_\alpha^\lambda N_i^{1-\lambda}} \sum_{j=1}^{|R|} \frac{e_j}{log(1+N_j)} \frac{a_i j}{N_i^\lambda N_j^{1-\lambda}} \tag{30}$$
$$= \sum_{i=1}^{|U|} log(1+N_i) \frac{a_{i\alpha}}{N_\alpha^\lambda N_i^{1-\lambda}} \sum_{\beta=1}^{|R|} \frac{e_\beta}{log(1+N_\beta)} \frac{a_i \beta}{N_i^\lambda N_\beta^{1-\lambda}} \tag{31}$$
$$= \sum_{\beta=1}^{|R|} \left(\frac{1}{log(1+N_\beta) N_\alpha^\lambda N_\beta^{1-\lambda}} \sum_{i=1}^{|U|} \frac{a_{i\alpha} a_{i\beta} log(1+N_i)}{N_i} \right) e_\beta \tag{32}$$

TABLE 2. Sample of a session search for the user with ID 42357.

Date	Action	Action_length (s)
2014/5/11 23:30	goto_login	1
2014/5/11 23:30	goto_favorites	35
2014/5/11 23:31	goto_home	42
2014/5/11 23:32	search	18
2014/5/11 23:32	searchterm_2	18
2014/5/11 23:32	resultlistids	18
2014/5/11 23:32	view_record	26
2014/5/11 23:32	docid	26
2014/5/11 23:42	search	3
2014/5/11 23:42	searchterm_2	3
2014/5/11 23:42	resultlistids	3
2014/5/11 23:42	view_record	805
2014/5/11 23:42	docid	805
2014/5/11 23:55	goto_home	14
2014/5/11 23:55	goto_favorites	0

Equation 29 is obvious when (30) is compared with (1). From (18) and (24), the following equation is obvious.

$$P_{\alpha\beta}^{\lambda}P_{\beta}^{1-\lambda} = log(1+N_{\beta})N_{\alpha}^{\lambda}N_{\beta}^{1-\lambda} \tag{33}$$

From (17), (29) and (33), we obtain the final transfer weight as follows.

$$\mathbf{W}_{\alpha\beta}^{\mathbf{UHM}} = \frac{1}{P_{\alpha\beta}^{\lambda}P_{\beta}^{1-\lambda}} \sum_{i=1}^{|U|} \frac{a_{i\alpha}a_{i\beta}}{I_i} \tag{34}$$

Considering the complexity of the calculation, the example of energy transmission process of the UHM algorithm will not be shown here.

V. EMPIRICAL STUDY

A. DATASET

The **Sowiport User Search Sessions Dataset (SUSS)** was used to test our proposed algorithms [42]. Sowiport is an academic search engine that includes 22 different databases whose content is in English and German [43], [44]. This dataset contains users' search logs between April 2, 2014 and April 2, 2015. The usage behaviors (e.g., typing a query, visiting a document, selecting a facet, etc.) in the system are mapped to a list of 58 actions (35.79% are document-related actions). For each action, a session id, the date stamp and additional information (e.g., queries, document ids, and result lists) are stored. The dataset contains 558,008 individual search sessions (0.69% were performed by registered users) and a total of 7,982,427 logs entries. Table 2 shows a piece of specific session information within the dataset. In this session, the user logged into the system and visited his favorites and home pages; then, he/she performed a search (searchterm_2). After obtaining the result list, he/she viewed a document. After performing an additional search, he/she returned to the home page and visited his/her favorites page.

For sessions generated by unregistered users, we treat individual sessions as an independent user. Document-related actions are preprocessed as the weight between users and documents.

B. EVALUATION METRICS

Appropriate metrics can measure whether the algorithm achieved the desired results. In our experiments, we used 70% of the data for training and the remaining 30% for testing, and the length of the ranking list was set to 10. Common indicators such as precision, recall, diversity, coverage, popularity, and ranking score are used here.

The goal of the information filtering task is to preserve the most relevant resources to the user. Precision and recall are the two most popular metrics, both of which are calculated by the hit ratio. Precision is the ratio of the number of hits in the result list.

$$Precision = \frac{1}{T}\sum_{i=1}^{n}\frac{hits}{L} \tag{35}$$

where T is the length of the testing set, $hits$ is the number of hits, and L is the length of the ranking list.

Relatively speaking, the recall is calculated as the proportion of the number of hits in the theoretical maximum number of hits.

$$Recall = \frac{1}{T}\sum_{i=1}^{n}\frac{hits}{H_i} \tag{36}$$

where H_i is the number of resources that u_i interacted with.

Two kinds of diversity are used in most literature: extrinsic diversity and intrinsic diversity [45]. The former comes from the uncertainty of information needed in a given a query, and the latter is mainly designed to avoid redundancy. Here, we focus on intrinsic diversity because we hope to optimize novelty in our results. The calculation of diversity is as follows.

$$Diversity = \frac{\sum_{i=1}^{|U|}\sum_{j=1}^{|U|}H_{ij}}{|Q|} \tag{37}$$

where H_{ij} is the Hamming Distance between u_i and u_j, which is calculated by (38).

$$H_{ij} = 1 - \frac{Q_{ij}}{L} \tag{38}$$

where Q_{ij} is the number of common resources in the ranking list of both u_i and u_j.

In an information filtering system, the higher the position of resources, the greater the probability of being visited. The ranking score not only measures the number of hits but also considers the relative position of the resources [46]. It is calculated by (39).

$$Ranking - score = \sum_{i=1}^{|U|}\frac{score_i}{|U|} \tag{39}$$

where $score_i$ is the ranking-score of u_i, which is calculated by (40)

$$score_u = \sum_{i=1}^{hits}\frac{1}{2^{\frac{P_i-1}{10}}} \tag{40}$$

where P_i is the position of r_i in the ranking list.

TABLE 3. Performance of four individual algorithms according to metrics including precision, recall, diversity, coverage, ranking-score, and popularity.

Method	Precision	Recall	Diversity	Coverage	Ranking-score	Popularity
HC	0.150485008263	0.141876357147	0.674355422284	0.573056340569	0.398080349452	0.742421643159
MD	0.152993782954	0.14434224532	0.625178293626	0.553032770128	0.428570883667	0.766279750139
UHC	0.148927362871	0.141506480496	0.678940570395	0.574048463712	0.383607827319	0.741355676418
UMD	0.155616589281	0.152150773277	0.612031372036	0.550312669112	0.412919526457	0.795957983147

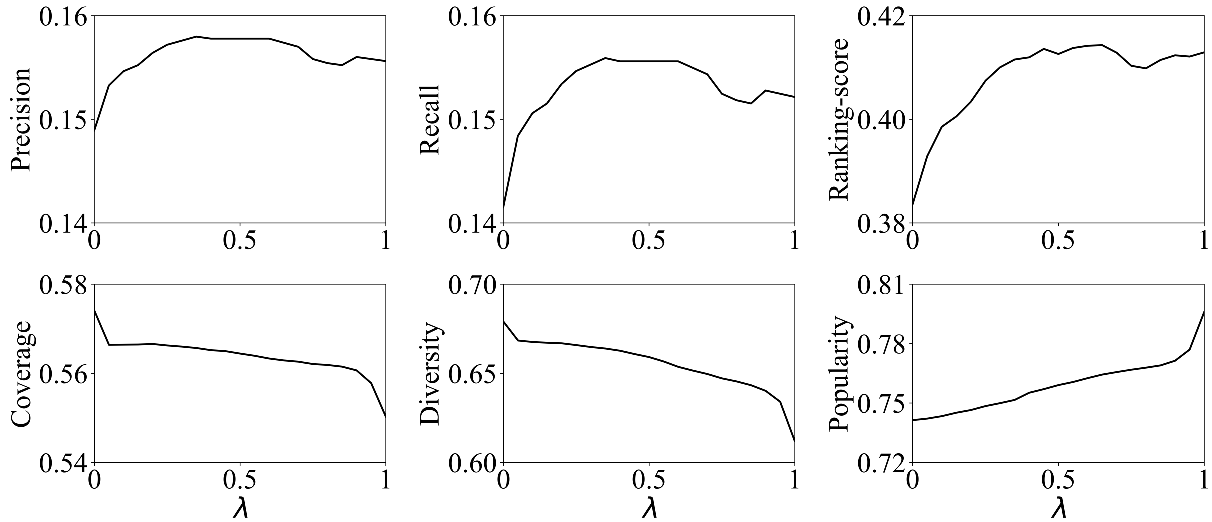

FIGURE 7. Performance of the UHM algorithm with different λ. λ ∈ [0, 1], which means varying from the pure UHC and pure UMD algorithm.

Coverage and popularity are also two important indicators in information filtering systems. Coverage measures how many of the resources are provided to users at least once, while popularity indicates the ability of the algorithm to provide unpopular resources. The better the two indicators, the better the ability of the algorithm to explore long tails. Their calculation method is as follows.

$$Coverage = \frac{|\bigcup_{i \in U} Rank_i|}{|R|} \tag{41}$$

where $Rank_i$ is the ranking list for u_i.

$$Popularity = \frac{\sum_{i \in Rank} log(1 + N_i)}{|U|} \tag{42}$$

where N_i is defined in Table 1.

C. INDIVIDUAL METHODS

We evaluated the individual HC, MD, UHC, and UMD algorithms from six indicators on the SUSS dataset, and the results of the evaluation are the average of ten experiments. The experimental performance of these four algorithms is shown in Table 3. From the perspective of accuracy, the UMD algorithm performs best, followed by the MD algorithm. The same is true for recall and ranking-score. Additionally, we can see that the accuracy of the UMD algorithm is much higher than that of the UHC algorithm, which can also be seen in the comparison of the MD and HC algorithms. The UHC algorithm and the HC algorithm are excellent in diversity indicators, and the UHC algorithm is superior. Starting from the two indicators of coverage and popularity, we can also conclude that the UHC and HC algorithms are better than the UMD and MD algorithms. Because the UHC and HC algorithms are more biased toward unpopular resources, resources are more diverse and covered. It can be seen from the above analysis that the proposed UMD algorithm and UHC algorithm are improved in terms of accuracy and diversity, respectively. Compared with the HC algorithm, the accuracy of the UHC algorithm decreases, and the diversity of the UMD algorithm also decreases compared with the MD algorithm. This shortcoming provides an idea for the latter hybrid algorithm.

D. HYBRID METHODS

From the previous evaluation results of independent algorithms on the dataset, it can be seen that different algorithms have performance differences in various indicators, which is determined by the theoretical framework of the algorithm. For example, the heat transfer algorithm has good diversity because it is more inclined to the unpopular resources in the energy transmission process, and the good accuracy of the mass diffusion algorithm is determined by its preference for popular resources. In the real world, we often need to consider multiple aspects of the indicators to achieve an overall optimal performance. Therefore, it is a feasible solution to combine the various algorithms with their strengths.

TABLE 4. Performance comparison between the UHM algorithm and HHM algorithm while $m = 2$, $\lambda = 0.44$.

Method	Precision	Recall	Diversity	Coverage	Ranking-score	Popularity
HHM	0.132988392681	0.116148018283	0.734118016335	0.626149960916	0.30594532504	0.758068223396
UHM	0.136136140075	0.121157097239	0.71618782295	0.617611689014	0.32779530303	0.765238040276

Based on the HC and MD algorithms, we find the problem of user discrimination and resource discrimination and propose an unbalanced strategy. The newly proposed UHC algorithm and UMD algorithm significantly improve in terms of diversity and accuracy. To meet the actual needs and balance in various indicators, we propose a UHM algorithm to mix the UHC and UMD algorithms. Considering that the energy propagation of the UHC algorithm and UMD algorithm have a similar form, our UHM algorithm combines the two by a parameter λ and has a similar propagation mode. The performance of the improved hybrid algorithm under different parameter values of λ is shown in Figure 7.

From this figure, we can see that when λ gradually increases, the precision, recall and ranking-score gradually improve, but the coverage, diversity and popularity indicators decrease (it should be noted here that the increase in popularity value indicates that the algorithm is more inclined to provide popular resources. From the perspective of personalization, we think this is a reduction in performance) because the hybrid algorithm considers the strengths of the UMD algorithm. Of course, the drawback is also manifested in the reduction in diversity. As λ continues to increase, indicators such as precision begin to show a downward trend because the change in parameters makes the hybrid algorithm too dependent on the UHC or UMD algorithm and misses the optimal balance.

Next, we fixed the parameter λ to evaluate the effect of the parameter m. Here, the value of our λ is chosen as the value of the best accuracy. That is, $\lambda = 0.44$. As we said before, algorithms such as the HC, MD, and HHM belong to two-stage energy transmission. The disadvantage of this method is its low coverage. Therefore, we propose multistage energy transmission. Here, we evaluate the performance of the HHM algorithm and the UHM algorithm with different m. The results are shown in Table 4. From the table, we can see that the four-stage ($m = 2$) energy transmission improved in indicators of precision, recall, and ranking-score compared to HHM. The efficiency of the multistage method has been guaranteed.

VI. CONCLUSION

The rapid development of the Internet has led to changes in the educational industry. In the past, people carried many teaching materials to the classroom and took notes, which has gradually transitioned to electronic formats. The increase in online educational resources has fueled the explosive growth of user interaction data. How to use the interaction data between users and educational resources to provide personalized information filtering services for users becomes increasingly important. The traditional keyword-based search model can play an important role when users have a clear need, but when the user's intention is unclear, this search method shows its drawbacks. It is difficult for users to discover diverse and innovative educational resources under this classic mode.

To solve the problem of resource search in edu-DL, this paper proposed an unbalanced multistage heat conduction and mass diffusion algorithm. Based on existing research, the algorithm mainly considers user differences and resource differences. Through a large number of experiments on the open academic resource dataset, we evaluated the performance of the proposed algorithms from six indicators. The experimental results show that the proposed UHM algorithm is superior to the HHM algorithm from both individual and hybrid aspects, which indicates that our algorithm is effective.

Of course, there are still many factors that we have not considered, which shows that there is still considerable research space from the classic algorithm of the HC and MD. For example, how to include the characteristics of users and resources, not just the interaction between users and resources, and how to combine the bipartite network model with social networks. In future research, we will continue to conduct various investigations and research and strive to continuously improve and innovate our UHM model.

REFERENCES

[1] A. Callery and D. T. Proulx, "Yahoo! Cataloging the Web," *J. Internet Cataloging*, vol. 1, no. 1, pp. 57–64, 1997.

[2] P. Jacsó, "Google scholar: The pros and the cons," *Online Inf. Rev.*, vol. 29, no. 2, pp. 208–214, Apr. 2005.

[3] N. Fiorini, K. Canese, G. Starchenko, E. Kireev, W. Kim, V. Miller, M. Osipov, M. Kholodov, R. Ismagilov, and S. Mohan, "Best match: New relevance search for pubmed," *PLoS Biol.*, vol. 16, no. 8, Aug. 2018, Art. no. e2005343.

[4] C. Sutton and L. Gong, "Popularity of arXiv.org within computer science," 2017, *arXiv:1710.05225*. [Online]. Available: https://arxiv.org/abs/1710.05225

[5] A. Martín-Martín, E. Orduna-Malea, M. Thelwall, and E. D. López-Cózar, "Google scholar, Web of science, and scopus: A systematic comparison of citations in 252 subject categories," *J. Informetrics*, vol. 12, no. 4, pp. 1160–1177, Nov. 2018.

[6] H. Gao, K. Zhang, J. Yang, F. Wu, and H. Liu, "Applying improved particle swarm optimization for dynamic service composition focusing on quality of service evaluations under hybrid networks," *Int. J. Distrib. Sensor Netw.*, vol. 14, no. 2, Feb. 2018, Art. no. 1550147718761583.

[7] Y. Yin, L. Chen, Y. Xu, J. Wan, H. Zhang, and Z. Mai, "QoS prediction for service recommendation with deep feature learning in edge computing environment," *Mobile Netw. Appl.*, pp. 1–11, Apr. 2019. doi: 10.1007/s11036-019-01241-7.

[8] Z. Carevic, M. Lusky, W. van Hoek, and P. Mayr, "Investigating exploratory search activities based on the stratagem level in digital libraries," *Int. J. Digit. Libraries*, vol. 19, nos. 2–3, pp. 231–251, Sep. 2018.

[9] Z. Carevic and P. Mayr, "Survey on high-level search activities based on the stratagem level in digital libraries," in *Proc. Int. Conf. Theory Pract. Digit. Libraries*. Cham, Switzerland: Springer, 2016, pp. 54–66.

[10] D. Hienert and P. Mutschke, "A usefulness-based approach for measuring the local and global effect of IIR services," in *Proc. ACM Conf. Hum. Inf. Interact. Retr.*, Mar. 2016, pp. 153–162.
[11] H. Gao, W. Huang, Y. Duan, X. Yang, and Q. Zou, "Research on cost-driven services composition in an uncertain environment," *J. Internet Technol.*, vol. 20, no. 3, pp. 755–769, 2019.
[12] Y. Yin, Y. Xu, W. Xu, M. Gao, L. Yu, and Y. Pei, "Collaborative service selection via ensemble learning in mixed mobile network environments," *Entropy*, vol. 19, no. 7, p. 358, 2017.
[13] T. Zhou, Z. Kuscsik, J.-G. Liu, M. Medo, J. R. Wakeling, and Y.-C. Zhang, "Solving the apparent diversity-accuracy dilemma of recommender systems," *Proc. Nat. Acad. Sci. USA*, vol. 107, no. 10, pp. 4511–4515, 2008.
[14] L. Page, S. Brin, R. Motwani, and T. Winograd, "The pagerank citation ranking: Bringing order to the Web," Dept. Comput. Sci., Stanford Univ., Stanford, CA, USA, Tech. Rep., 1999. [Online]. Available: http://ilpubs.stanford.edu:8090/422/
[15] R. Albert and A.-L. Barabási, "Statistical mechanics of complex networks," *Rev. Mod. Phys.*, vol. 74, no. 1, p. 47, 2002.
[16] S. P. Borgatti, A. Mehra, D. J. Brass, and G. Labianca, "Network analysis in the social sciences," *Science*, vol. 323, no. 5916, pp. 892–895, 2009.
[17] T. Zhou, J. Ren, M. Medo, and Y.-C. Zhang, "Bipartite network projection and personal recommendation," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 76, no. 4, Oct. 2007, Art. no. 046115.
[18] R. Lambiotte and M. Ausloos, "Collaborative tagging as a tripartite network," in *Proc. Int. Conf. Comput. Sci.* Cham, Switzerland: Springer, 2006, pp. 1114–1117.
[19] H. Gao, W. Huang, and X. Yang, "Applying probabilistic model checking to path planning in an intelligent transportation system using mobility trajectories and their statistical data," *Intell. Automat. Soft Comput.*, vol. 25, no. 3, pp. 547–559, Jan. 2019.
[20] Y. Yin, F. Yu, Y. Xu, L. Yu, and J. Mu, "Network location-aware service recommendation with random walk in cyber-physical systems," *Sensors*, vol. 17, no. 9, p. 2059, 2017.
[21] T. Zhou, R.-Q. Su, R.-R. Liu, L.-L. Jiang, B.-H. Wang, and Y.-C. Zhang, "Accurate and diverse recommendations via eliminating redundant correlations," *New J. Phys.*, vol. 11, no. 12, 2009, Art. no. 123008.
[22] Y.-C. Zhang, M. Blattner, and Y.-K. Yu, "Heat conduction process on community networks as a recommendation model," *Phys. Rev. Lett.*, vol. 99, no. 15, Oct. 2007, Art. no. 154301.
[23] J.-G. Liu, T. Zhou, and Q. Guo, "Information filtering via biased heat conduction," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 84, no. 3, 2011, Art. no. 037101.
[24] J.-G. Liu, Q. Guo, and Y.-C. Zhang, "Information filtering via weighted heat conduction algorithm," *Phys. A, Stat. Mech. Appl.*, vol. 390, no. 12, pp. 2414–2420, Jun. 2011.
[25] Q. Guo, R. Leng, K. Shi, and J. Liu, "Heat conduction information filtering via local information of bipartite networks," *Eur. Phys. J. B*, vol. 85, no. 8, p. 286, Aug. 2012.
[26] G. Chen, T. Qiu, L. Zhong, X. Zhang, and A. Ye, "A personalized recommendation algorithm via heterogeneous heat conduction," in *Proc. IEEE 3rd Int. Conf. Inf. Sci. Technol. (ICIST)*, Mar. 2018, pp. 602–606.
[27] L. Hou, K. Liu, J. Liu, and R. Zhang, "Solving the stability–accuracy–diversity dilemma of recommender systems," *Phys. A, Stat. Mech. Appl.*, vol. 468, pp. 415–424, Feb. 2017.
[28] J.-G. Liu, Z. Hu, and Q. Guo, "Effect of the social influence on topological properties of user-object bipartite networks," *Eur. Phys. J. B*, vol. 86, no. 11, p. 478, Nov. 2013.
[29] J. Lian, Y. Liu, Z.-J. Zhang, and C.-N. Gui, "Personalized recommendation via an improved nbi algorithm and user influence model in a microblog network," *Phys. A, Stat. Mech. Appl.*, vol. 392, no. 19, pp. 4594–4605, Oct. 2013.
[30] W. Zeng, A. Zeng, H. Liu, M.-S. Shang, and T. Zhou, "Uncovering the information core in recommender systems," *Sci. Rep.*, vol. 4, Aug. 2014, Art. no. 6140.
[31] X. Zhu, H. Tian, and T. Zhang, "Symmetrical information filtering via punishing superfluous diffusion," *Phys. A, Stat. Mech. Appl.*, vol. 508, pp. 1–9, Oct. 2018.
[32] W.-J. Song, Q. Guo, and J.-G. Liu, "Improved hybrid information filtering based on limited time window," *Phys. A, Stat. Mech. Appl.*, vol. 416, pp. 192–197, Dec. 2014.
[33] L. Hu, L. Ren, and W. Lin, "A reconsideration of negative ratings for network-based recommendation," *Phys. A, Stat. Mech. Its Appl.*, vol. 490, pp. 690–701, Jan. 2018.
[34] D.-C. Nie, Y.-H. An, Q. Dong, Y. Fu, and T. Zhou, "Information filtering via balanced diffusion on bipartite networks," *Phys. A, Stat. Mech. Appl.*, vol. 421, pp. 44–53, Mar. 2015.
[35] L. Lü and W. Liu, "Information filtering via preferential diffusion," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 83, no. 6, 2011, Art. no. 066119.
[36] W. Ma, X. Feng, S. Wang, and M. Gong, "Personalized recommendation based on heat bidirectional transfer," *Phys. A, Stat. Mech. Appl.*, vol. 444, pp. 713–721, Feb. 2016.
[37] G. Chen, T. Gao, X. Zhu, H. Tian, and Z. Yang, "Personalized recommendation based on preferential bidirectional mass diffusion," *Phys. A, Stat. Mech. Its Appl.*, vol. 469, pp. 397–404, Mar. 2017.
[38] X. Liu, X. Su, J. Ma, Y. Zhu, X. Zhu, and H. Tian, "Information filtering based on eliminating redundant diffusion and compensating balance," *Int. J. Mod. Phys. B*, vol. 33, no. 13, May 2019, Art. no. 1950129.
[39] A. Vidmer and M. Medo, "The essential role of time in network-based recommendation," *EPL*, vol. 116, no. 3, 2016, Art. no. 30007.
[40] X. Wang, Y. Liu, G. Zhang, F. Xiong, and J. Lu, "Diffusion-based recommendation with trust relations on tripartite graphs," *J. Stat. Mech., Theory Exp.*, vol. 2017, no. 8, Aug. 2017, Art. no. 083405.
[41] L.-J. Chen and J. Gao, "A trust-based recommendation method using network diffusion processes," *Phys. A, Stat. Mech. Its Appl.*, vol. 506, pp. 679–691, Sep. 2018.
[42] P. Mayr, "Sowiport user search sessions data set (SUSS)," GESIS Datorium, Tech. Rep., 2016. [Online]. Available: https://data.gesis.org/sharing/#!Detail/10.7802/1380
[43] P. Mayr and A. Kacem, "A complete year of user retrieval sessions in a social sciences academic search engine," in *Proc. Int. Conf. Theory Pract. Digit. Libraries*. Cham, Switzerland: Springer, 2017, pp. 560–565.
[44] D. Hienert, F. Sawitzki, and P. Mayr, "Digital library research in action–supporting information retrieval in sowiport," *D-Lib Mag.*, vol. 21, nos. 3–4, pp. 1–8, 2015.
[45] F. Radlinski, P. N. Bennett, B. Carterette, and T. Joachims, "Redundancy, diversity and interdependent document relevance," *ACM SIGIR Forum*, vol. 43, no. 2, pp. 46–52, Dec. 2009.
[46] J. S. Breese, D. Heckerman, and C. Kadie, "Empirical analysis of predictive algorithms for collaborative filtering," in *Proc. 14th Conf. Uncertainty Artif. Intell.* San Mateo, CA, USA: Morgan Kaufmann, 1998, pp. 43–52.

PENGFEI YIN received the B.Sc. degree in computer science and technology from Jishou University, China, and the M.Sc. degree in instructional technology from East China Normal University, China. He is currently pursuing the Ph.D. degree with the School of Computer Science and Engineering, Central South University, Changsha, China. He is also doing part-time Research and Development at both Guangzhou University, China, and Jishou University, China. His research interests include information retrieval and recommender systems.

GUOJUN WANG received the B.Sc. degree in geophysics, the M.Sc. degree in computer science, and the Ph.D. degree in computer science from Central South University, Changsha, China, in 1992, 1996, and 2002, respectively, where he was a Professor. He was an Adjunct Professor with Temple University, USA, a Visiting Scholar with Florida Atlantic University, USA, a Visiting Researcher with The University of Aizu, Japan, and a Research Fellow with The Hong Kong Polytechnic University. He is currently a Professor with the School of Computer Science, Guangzhou University. His research interests include artificial intelligence, network and information security, the Internet of Things, and cloud computing. He is also a Distinguished Member of CCF and a member of ACM and IEICE.

MD ZAKIRUL ALAM BHUIYAN was an Assistant Professor with Temple University and a Postdoctoral Research Fellow with the School of Information Science and Engineering, Central South University. He is currently an Assistant Professor with the Department of Computer and Information Sciences, Fordham University, NY, USA. His research interests include dependability, cybersecurity, big data, and cyberphysical systems. He is also a member of ACM.

MEIJING SHAN received the Ph.D. degree from East China Normal University, China, in 2009. She is currently an Assistant Professor with the East China University of Political Science and Law. She has published the results of her research in more than 20 articles in international journals, conference proceedings, and book chapters. Her research interests include cybercrime and computer forensics.

FANG QI received the B.Sc. degree from the National University of Defense Technology and the M.Sc. and Ph.D. degrees from Central South University, where she is currently an Associate Professor with the School of Computer Science and Engineering. Her current research interests include network and information security and the Internet of Things.