Science Journal of Education
2017; 5(2): 60-65
http://www.sciencepublishinggroup.com/j/sjedu
doi: 10.11648/j.sjedu.20170502.14
ISSN: 2329-0900 (Print); ISSN: 2329-0897 (Online)

Searching Similar Books Based on Student's Preference for Personalized Education

Mingxi Zhang, Tianxing Liu, Xiaohong Wang, Liujie Sun

College of Communication and Art Design, University of Shanghai for Science and Technology, Shanghai, China

Email address:

WAXL7461@aliyun.com (Mingxi Zhang), mliutianxing@126.com (Tianxing Liu), wang_keyan@163.com (Xiaohong Wang), liujiesunx@163.com (Liujie Sun)

To cite this article:

Mingxi Zhang, Tianxing Liu, Xiaohong Wang, Liujie Sun. Searching Similar Books Based on Student's Preference for Personalized Education. *Science Journal of Education.* Vol. 5, No. 2, 2017, pp. 60-65. doi: 10.11648/j.sjedu.20170502.14

Received: March 1, 2017; **Accepted**: March 9, 2017; **Published**: March 25, 2017

Abstract: Personalized education aims to give students a personalized learning schedule according to students' backgrounds and preferences, and the required learning resources for learning are personalized. On-line bookstore allows students to collect learning recourses on-line through Internet, but the problem of information overload plagues students since it is difficult to find the suitable books with the data becoming diverse and massive. Similarity search aims to find the similar objects to a given query, which can be regarded as a promising solution to the problem of information overload. However, the existing similarity search approaches limit the query into only one object, the students cannot express their preferences personally. In this paper, we proposed a personalized similarity search framework, towards finding the similar books based on student's preference for personalized education. We build the student-book network based on the students' ratings for books, and use SimRank to measure the similarities between books according to the student-book network. For satisfying student's personalized query preference, we allow student to express query with multi-books. A personalized similarity measure is proposed for measuring the similarity between query and candidate book by combining the similarities between books. Experiments on Amazon dataset demonstrate that, when the number of input books are not limited into one, the returned rankings are more consistent with students' query intentions.

Keywords: Student's Preference, Personalized Similarity Search, Personalized Education

1. Introduction

Personalized education aims to give students a personalized learning schedule according to students' backgrounds and preferences [1, 2]. Every student has his own preference while collecting learning resources, so the required learning resources for learning are personalized, even though they are come from a same major. For example, a student may want to get a book list in which the returned books are relevant to the topic "data mining", and another student may prefer the book list relevant to the topic "software engineering". There are many students who are searching the learning resources related to their personalized education processes, and it is difficult to find so many learning resources for satisfying the personalized preferences of students.

The rapid development of the Internet makes the learning resource collection of students more convenient as billions of learning resources are available online. On-line bookstore allows students to collect learning recourses on-line through Internet at home or some other places, which transcends the barriers of geography and makes the study process easier. Through on-line bookstore, students can get kinds of learning resources, including books, audio and video resources. Today, on-line bookstores, such as Amazon (https://www.amazon.com/) and China-pub (http://www.chinapub.com/), have attracted millions of students and helped provide them a large amount of valuable learning recourses. With the data of on-line bookstores becoming diverse and massive, the problem of information overload plagues us every day since it is difficult to find suitable books for learning.

Similarity search can be regarded as a promising way for efficiently solving the problem of information overload since it can effectively find the similar objects to a given object from large dataset. Similarity measures are the core task of similarity search problem, which can be divided into two broad categories: 1) content-based similarity measures treat each object as a bag of items or as a vector of word weights

[3–7]; and 2) structural-based similarity measures, consider object-to-object relationships expressed in terms of links [8–14]. Compared to the content-based similarity measures, the link-based similarity measures produce systematically better correlation with human judgements [15]. When applying link-based similarity measures in on-line bookstore, the students can find the similar books by providing a book as query, which would simplify the learning resource collection over largescale on-line resources. However, the existing similarity search framework limit the input query into one object, the student can choose only one book as query. The query intentions of students cannot be thoroughly expressed since it is difficult for students to choose a suitable query object related to their personalized preferences.

In this paper, we study the similarity search problem in on-line bookstore and propose a personalized similarity search framework, towards finding the similar books based on student's preferences for personalized education. For satisfying student's preferences, we allow student express the query with multi-books. Based on students' rating for books, we build the student-book network and compute the similarities between books over the student-book network. We define a personalized similarity measure for measuring the similarity between query and candidate book by combining the similarities between books. Experiments on Amazon dataset demonstrate that, when the number of input books are not limited into one, the returned rankings are more consistent with students' query intentions.

The rest of this paper is organized as follows. Section 2 defines the student-book network and discusses the similarity measure between books. Section 3 gives the personalized similarity measure and describes the personalized similarity search framework. Experimental studies are reported in section 4. Section 2 discusses the related work. Section 6 concludes this paper and discusses the future work.

2. Similarity Between Books

For our further discussions on personalized similarity, in this section, we first give the definition of student-book network, and then discuss the similarity measure between books based on the student-book network.

2.1. Student-Book Network

In the data of on-line book stores system, there are many objects of different types, including books, categories and attributes of these books, and the relationship between these objects are diverse and complex. The relationships between the books of these types are diverse and complex. Among these objects of different types, the student and books as well as the "rating" relationship between them are more informative for measuring similarities, since the task of our research is mainly to find the similar books to student preference.

The "rating" relationship which means the students have rated the books. Based on the "rating" relationship, we next give the definition of student-book network. Formally, the student-book network is defined as:

Definition 1 (Student-book network): A student-book network is defined as a bipartite graph $\mathrm{G} = (V_S \cup V_B, E)$, where V_S and V_B are the set of nodes of students and books type respectively, and E is the set of links of "rating" type between students and books, i.e., $\forall(\mathrm{u},\mathrm{v}) \in \mathrm{E}\colon \mathrm{u} \in V_S, \mathrm{v} \in V_B$.

Usually, a student prefers a book if he/her rated for the books with high score. So the nodes of student and book types as well as the "rating" relationship between them are informative for measuring similarities between books, which is the base to find the similar books to the preferences of students.

2.2. Similarity Between Books

There many existing link-based similarity measures in recent work, including SimRank [8], SimFusion [9], P-Rank [10], PathSim [12] and NetSim [14]. Among existing link-based similarity measures, SimRank can be considered as a promising solution to measure the similarities between books in student-book network. The intuition behind SimRank is that "two nodes are similar if they are referenced by similar nodes", which conforms to our basic understandings. When compared to the 1-hop similarity measures [16–18], SimRank considers not only direct connections among nodes but also indirect connections, which can find more valuable underlying relationships.

In a given network, the SimRank similarity between objects $a, b \in V$ is denoted by $s(a,b) \in [0,1]$, which is defined as $s(a,b) = 1$ if $a = b$, otherwise:

$$s(a,b) = \frac{C}{|I(a)||I(b)|}\sum_{x\in I(a)}\sum_{y\in I(b)} s(x,y) \qquad (1)$$

where $C \in (0,1)$ is the decay factor. For preventing division by zero in (1), $\mathrm{s(a,b)}$ is defined as zero when $I(a) = \emptyset$ or $I(b) = \emptyset$.

The SimRank similarities can be computed iteratively. At iteration l, the similarity between a and b is denoted by $R_l(a,b)$. The iterative computation is started with $R_0(*,*)$, which is initialized as: $R_0(a,b) = 1$ if $\mathrm{a} = \mathrm{b}$, and $R_0(a,b) = 0$ for otherwise. And when $\mathrm{l} = 1,2,\ldots$, $R_l(a,b)$ is defined as $R_l(a,b) = 1$ if $\mathrm{a} = \mathrm{b}$, otherwise:

$$R_l(a,b) = \frac{C}{|I(a)||I(b)|}\sum_{x\in I(a)}\sum_{y\in I(b)} R_{l-1}(x,y) \qquad (2)$$

The time cost for computing the similarities of all node pairs at the $\mathrm{l}-\mathrm{th}$ iteration is $\mathrm{O}(\mathrm{l}d^2n^2)$, and the space cost is $\mathrm{O}(n^2)$, where d is the average degree and n is the node number of a given graph. The iterative SimRank computation converges very fast, and there is little change in the returned rankings after five iterations [8].

When applying SimRank to student-book network, the intuition under the similarity can be described as "two books are similar if they are rated by similar students, and two students are similar if they rated similar books". During similarity computation, the similarity between books is computed by accumulating only the similarities between students, and the similarity between students is computed by accumulating only the similarities between books. Thus, the similarity between books is computed as: $R_0(b_1,b_2) = 1$ if $b_1 = b_2$, and $R_0(b_1,b_2) = 0$ for otherwise; and when $\mathrm{l} \neq 0$,

$R_l(b_1, b_2)$ is defined as $R_l(b_1, b_2) = 1$ if $b_1 = b_2$, otherwise: for $b_1, b_2 \in V_B$:

$$R_l(b_1, b_2) = \frac{C}{|I(b_1)||I(b_2)|} \sum_{x \in I_S(b_1)} \sum_{y \in I_S(b_2)} R_{l-1}(x, y) \quad (3)$$

and for $s_1, s_2 \in V_S$: $R_0(s_1, s_2) = 1$ if $s_1 = s_2$, and $R_0(s_1, s_2) = 0$ for otherwise; and when $l \neq 0$, $R_l(s_1, s_2)$ is defined as $R_l(s_1, s_2) = 1$ if $s_1 = s_2$, otherwise:

$$R_l(s_1, s_2) = \frac{C}{|I(s_1)||I(s_2)|} \sum_{x \in I_C(s_1)} \sum_{y \in I_C(s_2)} R_{l-1}(x, y) \quad (4)$$

where $I_S(b_1)$ is the $b_1's$ in-neighbor sets of student type, and $I_B(s_1)$ is the $s_1's$ in-neighbor sets of book type.

The disadvantage of SimRank is the computational cost. With the student-book network becoming large, the computation of SimRank would be expensive in terms of time and space cost. Fortunately, there are extensive optimization techniques on SimRank computation in previous work [19–24], which significantly reduced the computation cost. For example, in our previous research [24], the reduction of the time and space cost of the iterative SimRank computation is on average 99.83%, accuracy loss is on average 0.02% NDCG, which can be used to optimize the similarity computation in student-book network. The similarities can be computed in the off-line stage, which would not affect the response time of query processing.

3. Personalized Similarity Search

3.1. Personalized Similarity Measure

For supporting student preferences, we allow students express their queries with multi-books. Formally, the student preference is defined as:

Definition 2 (Student preference): The preference of a student is represented by vector $P(P_1, P_2, \ldots, P_N)$, where the entry P_i of vector P is either 0 or 1, and N is the number of books. $P_i = 1$ represents the current student prefers book i when inputting query and $P_i = 0$ represents the current student does not prefer book i.

The student preference is taken as query. When the query is not limited into one book, the definition of similarity between query and book would become more complex, since the query and book is not belong to the same type. For modeling the similarity between query and books, we define the similarity between query and book by combining the similarities between candidate books and the preferred books. The similarity between query P and book x is called personalized similarity, defined as:

$$PS(P, x) = \frac{\sum_{i=1}^{N} R_l(i, x) P_i}{\sum_{i=1}^{N} P_i} \quad (5)$$

Based on the personalized similarity, the students can express their preferences on different topics by choosing different books on different topics. For example, a student can choose some preferred books on "similarity computation" and "recommendation systems" as query, and the system returns the similar books to this query, which would be more personalized than the result when providing only one book as query.

3.2. Framework of Personalized Similarity Search

The framework of personalized similarity search is shown in Fig. 1. The process of the off-line and on-line stages are respectively shown in the below and above of the dotted line. In the off-line stage, the raw data is cleaned, including unnecessary links and noise data, and the "rating" relationship between students and books are chosen for building student-book network. The similarities between books are computed based on the student-book network, which are stored in a similarity matrix. In the on-line stage, the student input some preferred books as query, and system takes these books and transform them into the vector of student preference. The similarities between query and books are computed by combining the similarities between books, and then the candidate books are sorted according to the similarities. Finally, the system returns the top-k more similar books.

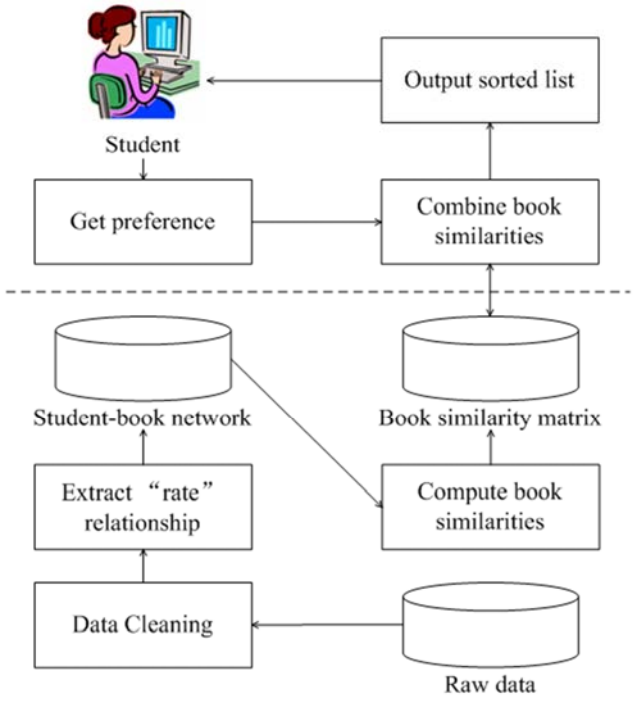

Figure 1. *The framework of personalized similarity search.*

4. Experimental Result

4.1. Setup

In this section, we compare our proposed personalized similarity measure (PSR) with the SimRank similarity measure (SR). The similarity computation in off-line stage are speeded up via partial sums function [20]. According to the literature, the decay factor are set as 0.8. Our experiments were conducted on a 2.30 GHz Intel(R) Xeon(R) CPU with 12 GB RAM, running Windows 8. All algorithms were implemented in C++ and compiled by using VS 2010.

We use Amazon dataset [25] to evaluate our approach. There are 355,601 products with 2,359,584 co-purchased relationships, 36,591 categories, and 42,890 terms appearing more than once. From which, we choose 5,521 users as students and 2,810 books with 18,901 links of "rating" relationship and 18,901 links of "be rated by" relationship.

We use Normalized Discounted Cumulative Gain (NDCG) [26] to evaluate the effectiveness of similarity ranking lists. The NDCG at position k is defined as $\mathrm{NDCG@}k = \frac{DCG@k}{IDCG@k}$ where DCG@k is defined as $\mathrm{DCG@k} = \mathrm{r(v}, v_i)$ if $\mathrm{i} < 2$, and $\mathrm{DCG@k} = \mathrm{DCG@}i + \sum_{i=2}^{k} \frac{r(v,v_i)}{log_2 i}$ for otherwise, where i denotes rank of v_i in the returned list, and $\mathrm{r(v}, v_i)$ is set as: 2 (highly relevant), 1 (marginally relevant), and 0 (irrelevant). And the similarity levels are labeled in a double-blind fashion.

4.2. Performance

Fig. 2 shows the NDCG values of both PSR and SR on varying k. For each algorithm, we use 20 queries to test the effectiveness. At each query, we indicate the expected topic of the returned books. Specifically, for SR, the student is allowed to choose only one book as query very time; and for PSR, the students is allowed to choose multi-books as query every time. We find that the NDCG increases with k increasing and finally becomes stable, this is because the rankings for different queries become relatively stable as k increases. We also find that, the NDCG values at different k of PSR are evidently higher than SR. Generally, when the number of input books is not limited into one, the returned rankings are more consistent with students' query intentions. Fig. 3 shows the NDCG values of PSR on varying query size n. We choose 10 queries at different sizes to test the influence of query sizes. Specifically, for each query, we limit the number of input books into $1,2,\ldots,10$, respectively, and recorded the NDCG value for each query. From this figure, we find that, the size of the input query can really affect the effectiveness of the returned rankings, and in the range of $n = 2$ to 6, the NDCG values are relatively higher.

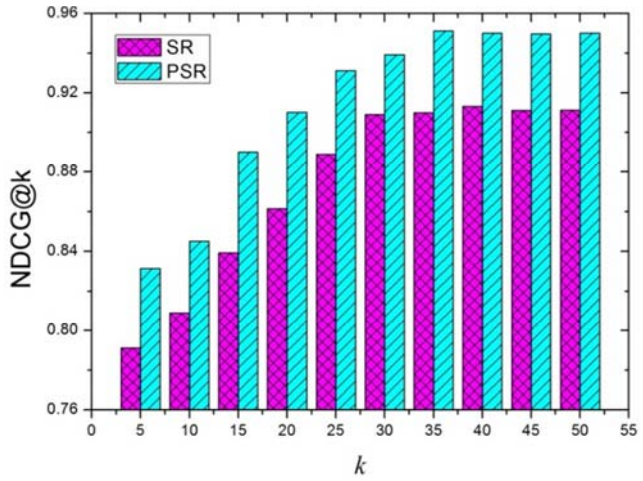

Figure 2. *NDCG values on varying rank k.*

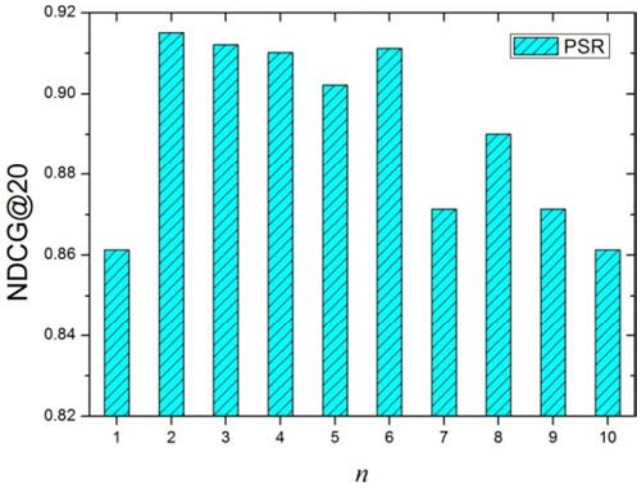

Figure 3. *NDCG values on varying query sieze n.*

5. Related Work

There are extensive link-based similarity measures that can be used for measuring similarities between books. With respect to the focus of this paper, next we introduce some similarity measures that are most relevant to the current work.

SimRank [8] is a classical similarity measure proposed by Glen Jeh and Jennifer Widom, which defines the similarities between objects based on the intuition that "two nodes are similar if they are referenced by similar nodes". SimFusion [9] is one of the influent similarity measures for computing link-based similarities in heterogeneous network, which aims to combine relationships from multiple heterogeneous data sources. SimFusion computes the similarities iteratively over a unified relationship matrix (URM). Compared to SimRank, SimFusion utilizes the relationship for distinguishing link importance, but there only one type links in student-book network defined in our research, which makes SimRank more efficient and suitable for measuring similarities between books.

P-Rank [10] enriches SimRank by considering both in- and out-links for solving improving the "limited information" of similarity computation and improving the effectiveness. The intuition behind P-Rank is that "two objects are similar if they are referenced by similar objects or they reference similar objects". C-Rank [12] ignores the direction of links when computing similarities, the meetings of both backward and forward directions are exploited for similarity computation in scientific literature databases. Both P-Rank and C-Rank can find more similar objects by considering the meetings of different directions, however, the student-book network is defined an undirected graph.

PathSim [13] assesses similarities in heterogeneous network by utilizing a meta path provided by users, which captures the similarity semantics among peer objects in networks. This measure allows users to measure similarities from different perspectives. HeteSim [14] adopts the spiritual of meta path, which can find similar objects from network to a query object of any type. Both of PathSim and HeteSim require users provide meta paths, which is difficult for the

users to choose a suitable meta path especially when the network schema becoming diverse. NetSim [15] measures the similarities between objects based on the similarities between attributes, the intuition of NetSim is that "similar centers are linked with similar attributes". However, this measure suitable only the network of x-star network schema.

There are also some similarity measures that utilize the 1-hop neighborhood for similarity computation. Co-citation [16] measures the similarity between two papers in citation network based on the common papers which cite both of them. Formally, the similarity between papers is defined as the number of papers which cites them. And Bibliographic Coupling [17] defines similarity as the number of papers cited by them. Jaccard similarity coefficient [18] defines the similarity measures between two objects as the ratio of the common neighbors of their neighbors. These approaches use 1-hop neighbors for defining similarities. When compared to SimRnak, the indirect connections are not considered, which would ignores some similar results when find similar objects.

For fast similarity computation, a lot of optimization techniques are proposed. BlockSimRank [19] reduces the computation cost of SimRank by partitioning the graph into several blocks according to the block structure of graph data. By which, the similarity for each node-pair can be efficiently obtain from these blocks. D. Lizorkin and P. Velikhov [20] optimized SimRank based on partial sums, essential node pairs and threshold-sieved similarity. W. Zheng and L. Zou [21] proposed an efficient algorithm for finding the most similar object pairs in large networks. W. Yu and X. Lin [22] developed an incremental SimRank computation algorithm for fast similarity computation in dynamic networks. W. Yu and J. A. McCann [23] modified SimRank to compute the similarities for partial object pairs, which is important when only the similarities of partial object pairs are required in some applications. M. Zhang and H. Hu [24] proposed WebSim that reduces the computation cost of similarity search by limiting the iteration number into two, and uses a partial index to reduce the execution time of on-line query processing. These approaches can be easily taken into student-book network for speeding up the similarity computation between books.

6. Conclusion and Future Work

This paper introduced a personalized similarity search framework, which aims to find the similar books to student's preferences for personalized education. In contrast to traditional similarity search framework, our proposed approach allows students express queries by any number books according to their preferences. We integrate the student preference into similarity computation, and define the personalized similarity measure by splitting into the similarities between candidate books and the preferred books. Through the experiments on real datasets we conclude that, when the number of input books are not limited into one, the returned rankings are more consistent with students' query intentions.

There are numbers of directions in our future work. First, we would like to study the efficiency problem of on-line query processing, since the time cost of query processing would be significantly increased when the student-book network grows large. Second, we want to integrate the prerequisite relation corresponds to different books into similarity search to search more suitable books for personalized education. We can get the prerequisite relationship from the course schedule of some universities or learned from the purchasing behavior from the on-line bookstore. Third, we plan to apply our proposed personalized similarity search framework to other real datasets in some real applications, including literature search [26, 27] and web search [28, 29]. Our approach can be applied to any datasets of bipartite network schema besides the student-book network, such as product co-purchasing network [30, 31] and bibliographic network [32].

Acknowledgment

This work was supported by Natural Science Foundation of Shanghai grant 16ZR14228; Training Project of University of Shanghai for Science and Technology grant 16HJPY-QN04; Innovation Program of Shanghai Municipal Education Commission grants 15ZZ073 and 15ZZ074.

References

[1] C. Tekin, J. Braun, and M. van der Schaar, "etutor: Online learning for personalized education," in 2015 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2015, South Brisbane, Queensland, Australia, April 19-24, 2015, pp. 5545–5549.

[2] T. Morrow, S. S. Sarvestani, and A. R. Hurson, "Algorithmic decision support for personalized education (invited paper)," in 17th IEEE International Conference on Information Reuse and Integration, IRI 2016, Pittsburgh, PA, USA, July 28-30, 2016, 2016, pp. 188–197.

[3] M. C. Lee, "A novel sentence similarity measure for semantic-based expert systems," Expert Syst. Appl., vol. 38, no. 5, pp. 6392–6399, 2011.

[4] C. Liu, "Discriminant analysis and similarity measure," Pattern Recognition, vol. 47, no. 1, pp. 359–367, 2014.

[5] C. D. Boom, S. V. Canneyt, S. Bohez, T. Demeester, and B. Dhoedt, "Learning semantic similarity for very short texts," in IEEE International Conference on Data Mining Workshop, ICDMW 2015, Atlantic City, NJ, USA, November 14-17, 2015, pp. 1229–1234.

[6] H. Hu, G. Li, Z. Bao, J. Feng, Y. Wu, Z. Gong, and Y. Xu, "Top-k spatio-textual similarity join," in 32nd IEEE International Conference on Data Engineering, ICDE 2016, Helsinki, Finland, May 16-20, 2016, pp. 1576–1577.

[7] F. Chen, C. Lu, H. Wu, and M. Li, "A semantic similarity measure integrating multiple conceptual relationships for web service discovery," Expert Syst. Appl., vol. 67, pp. 19–31, 2017.

[8] G. Jeh and J. Widom, "Simrank: a measure of structuralcontext similarity," in Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, July 23-26, 2002, Edmonton, Alberta, Canada, 2002, pp. 538–543.

[9] W. Xi, E. A. Fox, W. Fan, B. Zhang, Z. Chen, J. Yan, and D. Zhuang, "Simfusion: measuring similarity using unified relationship matrix," in SIGIR 2005: Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Salvador, Brazil, August 15-19, 2005, pp. 130–137.

[10] P. Zhao, J. Han, and Y. Sun, "P-rank: a comprehensive structural similarity measure over information networks," in Proceedings of the 18th ACM Conference on Information and Knowledge Management, CIKM 2009, Hong Kong, China, November 2-6, 2009, pp. 553–562.

[11] S. Yoon, S. Kim, and S. Park, "C-rank: A link-based similarity measure for scientific literature databases," Inf. Sci., vol. 326, pp. 25–40, 2016.

[12] Y. Sun, J. Han, X. Yan, P. S. Yu, and T. Wu, "Pathsim: Meta path-based top-k similarity search in heterogeneous information networks," PVLDB, vol. 4, no. 11, pp. 992– 1003, 2011.

[13] C. Shi, X. Kong, Y. Huang, P. S. Yu, and B. Wu, "Hetesim: A general framework for relevance measure in heterogeneous networks," IEEE Trans. Knowl. Data Eng., vol. 26, no. 10, pp. 2479–2492, 2014.

[14] M. Zhang, H. Hu, Z. He, and W. Wang, "Top-k similarity search in heterogeneous information networks with x-star network schema," Expert Syst. Appl., vol. 42, no. 2, pp. 699–712, 2015.

[15] A. G. Maguitman, F. Menczer, F. Erdinc, H. Roinestad, and A. Vespignani, "Algorithmic computation and approximation of semantic similarity," World Wide Web, vol. 9, no. 4, pp. 431–456, 2006.

[16] H. G. Small, "Co-citation in the scientific literature: A new measure of the relationship betweenf two documents," Journal of the American Society for Information Science, vol. 24, no. 4, pp. 265–269, 1973.

[17] M. M. Kessler, "Bibliographic coupling between scientific papers," American Documentation, vol. 14, pp. 10–25, 1963.

[18] J. Santisteban and J. Tejada-Cárcamo, "Unilateral jaccard similarity coefficient," in Proceedings of the First International Workshop on Graph Search and Beyond, GSB 2015, co-located with The 38th Annual SIGIR Conference (SIGIR'15), Santiago, Chile, August 13th, 2015, pp. 23–27.

[19] P. Li, Y. Cai, H. Liu, J. He, and X. Du, "Exploiting the block structure of link graph for efficient similarity computation," in Advances in Knowledge Discovery and Data Mining, 13th Pacific-Asia Conference, PAKDD 2009, Bangkok, Thailand, April 27-30, 2009, Proceedings, 2009, pp. 389–400.

[20] D. Lizorkin, P. Velikhov, M. N. Grinev, and D. Turdakov, "ccuracy estimate and optimization techniques for simrank computation," VLDB J., vol. 19, no. 1, pp. 45–66, 2010.

[21] W. Zheng, L. Zou, Y. Feng, L. Chen, and D. Zhao, "Efficient simrank-based similarity join over large graphs," PVLDB, vol. 6, no. 7, pp. 493–504, 2013.

[22] W. Yu, X. Lin, and W. Zhang, "Fast incremental simrank on link-evolving graphs," in IEEE 30th International Conference on Data Engineering, Chicago, ICDE 2014, IL, USA, March 31 - April 4, 2014, pp. 304–315.

[23] W. Yu and J. A. McCann, "Efficient partial-pairs simrank search for large networks," PVLDB, vol. 8, no. 5, pp. 569-580, 2015.

[24] M. Zhang, H. Hu, Z. He, L. Gao, and L. Sun, "Efficient link-based similarity search in web networks," Expert Syst. Appl., vol. 42, no. 22, pp. 8868–8880, 2015.

[25] J. Leskovec, L. A. Adamic, and B. A. Huberman, "The dynamics of viral marketing," TWEB, vol. 1, no. 1, 2007.

[26] K. Järvelin and J. Kekäläinen, "Cumulated gain-based evaluation of ir techniques," ACM Trans. Inf. Syst., vol. 20, no. 4, pp. 422-446, 2002.

[27] Y. Ji, H. Ying, J. Tran, P. Dews, R. M. Massanari. Integrating unified medical language system and association mining techniques into relevance feedback for biomedical literature search. BMC Bioinformatics, vol. 17, no. S-9, pp.264, 2016.

[28] N. Yi, N. Standaert, B. Nemery, K. Dierickx. Research integrity in China: precautions when searching the Chinese literature. Scientometrics, vol. 110, no. 2, pp. 1011-1016, 2017.

[29] S. Plansangket, J. Q. Gan. Re-ranking Google search returned web documents using document classification scores. Artif. Intell. Research, vol. 6, no. 1, pp. 59-68, 2017.

[30] R. Song, K. Umemoto, J. Y. Nie, X. Xie, K. Tanaka, Y. Rui. UniClip: Leveraging Web Search for Universal Clipping of Articles on Mobile. Data Science and Engineering, vol. 1, no. 2, pp. 101-113, 2016.

[31] T. Yamazaki, N. Shimizu, H. Kobayashi, S. Yamauchi. Weighted Micro-Clustering: Application to Community Detection in Large-Scale Co-Purchasing Networks with User Attributes. WWW (Companion Volume), pp. 131-132, 2016.

[32] M. Jebabli, H. Cherifi, C. Cherifi, A. Hamouda. Overlapping Community Detection Versus Ground-Truth in AMAZON Co-Purchasing Network. The 11th International Conference on Signal-Image Technology & Internet-Based Systems, SITIS 2015, Bangkok, Thailand, November, 2015, pp. 23-27.

[33] H. Hazimeh, I. Youness, J. Makki, H. Noureddine, J. Tscherrig, E. Mugellini, O. A. Khaled. Leveraging Co-authorship and Biographical Information for Author Ambiguity Resolution in DBLP. The 30th {IEEE} International Conference on Advanced Information Networking and Applications, AINA 2016, Crans-Montana, Switzerland, 23-25 March, 2016, pp. 1080-1084.

[34] Tinghuai Ma, Huan Rong, Changhong Ying, Yuan Tian, Abdullah Al-Dhelaan, Mznah Al-Rodhaan: Detect structural-connected communities based on BSCHEF in C-DBLP. Concurrency and Computation: Practice and Experience, vol. 28, no. 2, pp. 311-330, 2016.