



HAL
open science

BIBLME RecSys: Harnessing Bibliometric Measures for a Scholarly Paper Recommender System

Anaïs Ollagnier, Sébastien Fournier, Patrice Bellot

► **To cite this version:**

Anaïs Ollagnier, Sébastien Fournier, Patrice Bellot. BIBLME RecSys: Harnessing Bibliometric Measures for a Scholarly Paper Recommender System. BIR 2018 Workshop on Bibliometric-enhanced Information Retrieval, Mar 2018, Grenoble, France. <hal-01770588>

HAL Id: hal-01770588

<https://hal.science/hal-01770588v1>

Submitted on 19 Apr 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

BIBLME RecSys: Harnessing Bibliometric Measures for a Scholarly Paper Recommender System

Anais Ollagnier, Sébastien Fournier, and Patrice Bellot

Aix Marseille Univ, Toulon University, CNRS, LIS, Marseille, France
{anais.ollagnier,sebastien.fournier,patrice.bellot}@univ-amu.fr

Abstract. The iterative *continuum* of scientific production generates a need for filtering and specific crossing of ideas and papers. In this paper, we present *BIBLME RecSys* software which is dedicated to the analysis of bibliographical references extracted from scientific collections of papers. Our goal is to provide users with paper suggestions guided by the papers they are reading and by the references they contain. To do so, we propose a new approach based on a new bibliometric measure. We propose to determine the impact, the inner representativeness, of each bibliographical reference according to their occurrences in the paper the user is reading. By means of this approach, we suggest central references of the author’s paper. As a result, we obtain papers that are related to the paper selected by the user according to the influence of references on it. We evaluate the recommendation in the context of a digital library dedicated to humanities and social sciences.

Keywords: Recommender systems, Text mining, Digital libraries, Bibliographic information, Bibliometrics.

1 Introduction

There are 114 million of scholarly papers archived on the Web [9]. While certainly advantageous, researchers have unprecedented level of access which creates a problem of “information overload” [2]. To help generate relevant suggestions for researchers, scholarly paper recommenders emerged over the last decade to ease finding publications. In this research field, content-based filtering (CBF) algorithms are the predominant approaches [14]. In most CBF systems dedicated to textual applications (e.g., scholarly papers), item descriptions are represented by textual features such as plain words, phrases, or n-grams. However, these systems encounter difficulties caused by complications that originate from natural language ambiguity. In the context of digital libraries, references are a major source of links. Indeed, bibliographical references are an important part of the academic writing and allow to convey various information relating to the author’s research fields. In this paper, we propose a scholarly paper recommender system which suggests related papers to the user’s paper is reading. To do that, we were

interested in textual features through the identification of references extracted from the body of the text and the footnotes. By this way unlike traditional CBF approaches in which items are recommend to users by determining their similarities (inherent characteristics) with other items, we propose to determine an active user's interests from the content analysis of the article selected.

Leveraging information extracted from bibliographical reference analysis and from quantitative measurements, we propose a **centrality indicator** which allows to evaluating bibliographical references' importance of the user's paper is reading. Based on the findings that certain textual features, such as citation frequency and citation location might allow to predict references' importance [19], we construct this indicator from factors which use information on the authors' citation behavior. By means of this approach, citations are used in the same way as words were used in CBF algorithm in order to weighted each reference according to its level of centrality. By this way, our centrality indicator can reflect the strength of references' influence and so potential relevant readings. Reference analysis is integrated in the BILBO¹ [10] system, the software we have been developing for some years in the context of OpenEdition that is a large digital library dedicated to Humanities and Social Sciences (HSS).

2 Related Work

In the context of digital libraries and publishers' portals, recommender system have been introduced in order to provide relevant related literature. In 1998, Giles *et al.* introduced the first scholarly recommender system as part of the CiteSeer project [6]. Since several methods have been proposed and at least 216 papers relating to scholarly paper recommendation approaches were published [3]. Two types of algorithms are typically used in recommender systems: collaborative filtering (CF) and content-based filtering (CBF). CF algorithms recommend items to users based on the ratings of other users (with similar interest). In CBF algorithms, the user's interest is inferred from the items that the user interacted with. Items' representation consist in a content model in which features are typically word-based.

Several digital libraries have deployed recommender systems such as TechLens+ [12], BibTip [20] and CiteULike [5]. Each system uses different kind of data: TechLens+ uses citation data and usage data, BibTip rests upon the observation of user patterns and the statistical evaluation of the usage data, and CiteULike is based on users' bookmark items. These systems are essentially based on CBF algorithms. Most of the proposed approaches operate either by clustering similar items or by profiling users' behaviour (user-based collaborative filtering) or by combining the two (hybrid recommenders) [3]. In recent works, Asabere *et al.* [1] introduced a folksonomy-based paper recommendation algorithm which recommends papers issued by active participants, to other Group Profile Participants at the same conference based on preference similarity of their research interests.

¹ <https://github.com/OpenEdition/bilbo>

Philip *et al.* [17] proposed a CBF approach based on TF-IDF weighing scheme and cosine similarity measure.

Currently, it is not possible to identify the most effective recommendation approaches [3]. So according to our dataset, we have oriented our work on a CBF approach. The majority of these approaches are based on plain terms extracted from papers, n-grams or topics based on LDA. Certain researches have used non-textual features such as writing style, layout information, and XML tags. Concerning the use of citations, [6] have proposed the CC-IDF method in which citations are used in the same way as words-based features were used and weighted the citations with the standard TF-IDF measure. Inspired by this work, we propose a system in which the user's interests are inferred from citations extracted from the current paper that the user interact with in order to provide related papers.

3 Proposed Method

Our method starts with our former bibliographical references detection system dedicated to scholarly papers [15] which is integrated in the BILBO system. Based on this software the names of the authors, the titles, the year of publication and some meaningful elements of information are extracted from both the full texts and the reference sections. From bibliographical references annotated and extracted by our system, we build our scholarly recommender system entitled *BIBLME RecSys*. It consists of four steps:

- Step 1: Retrieve references both in the body of the text (*ref*) and in the reference section (P_{ref});
- Step 2: Check matches between each *ref* corresponding to the same paper P_{ref} ;
- Step 3: Compute for each P_{ref} a centrality indicator derived from the evaluation criteria based on objective quantitative measurements;
- Step 4: Rank each P_{ref} according to its centrality indicator and recommend papers with the high scores.

As we show in Fig. 1, each candidate paper to recommend (P_{ref_i}) is represented by a centrality indicator ($cIndic(P_{ref_i})$) which corresponds to quantitative measures used in Step 3. This indicator allows to highlight papers which are mainly employed by the authors in their paper. In this approach, a key innovative step is to model a paper of interest from factors based on authors' citation behaviours. The first factor corresponding to the $freqFactor(ref_i, P_{ref_i})$ function computes references ref_i which occur in the given paper. The second factor corresponding to the $granuFactor(ref_i, ref_j)$ function calculates for each P_{ref_i} how corresponding references in the text are used by the author in his paper. This factor has two levels of granularity (fine granularity/coarse granularity) which we detail in Section 3.2. From these factors, we determine central papers of the author's paper. By this way, we leverage the bibliographical references in order to shape a user's interests.

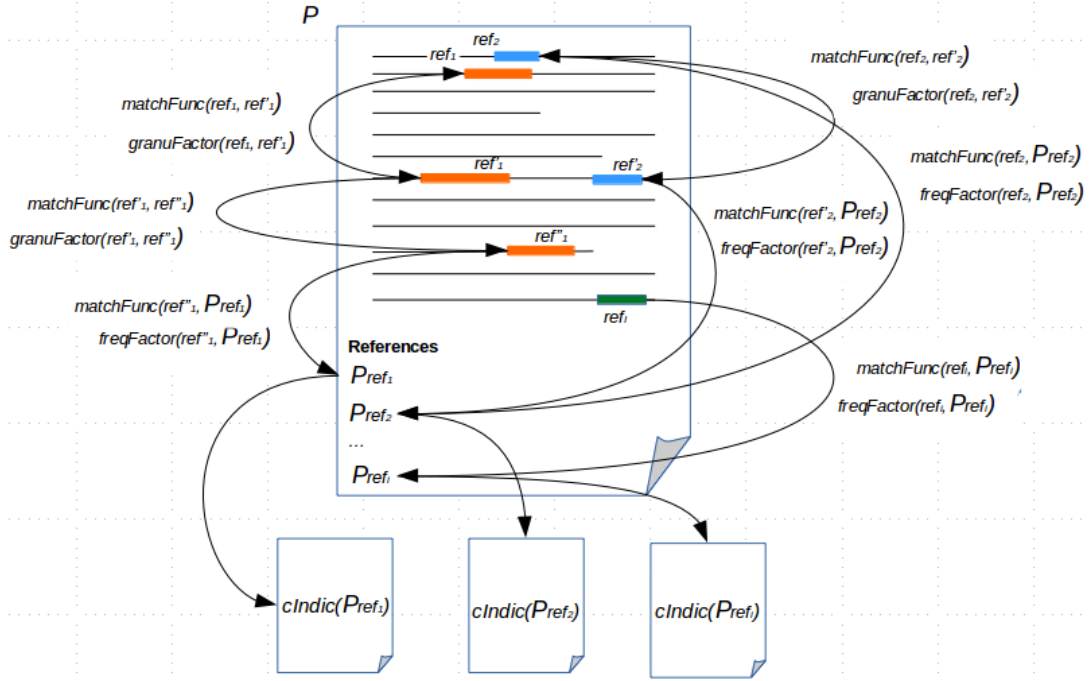


Fig. 1. Example of BIBLME RecSys operations

In Step 1, sets of references extracted both from the body of the text and from the reference section are constructed. Then in Step 2, we manage with these sets in order to determine which references correspond to the same paper from matching functions ($matchFunc(ref_i, ref_j)$). These functions based on a strict matching and a fuzzy matching allow to compare two bibliographical references extracted from the body of the text or from the reference section. Then for each reference whose matching functions are fulfilled, quantitative measures based on the frequency factor of usage and the granularity factors are computed. Lastly, candidate papers to recommend are ranked according to their centrality indicator which represents a linear combination of the quantitative measures. With this system, users can set the value assigned to each factor. According to the score assigned to each factor, users can obtain informations about author's citation behaviours. For example, if the coarse granularity factor is set at a high value, references which occur throughout the paper will be highlighted.

3.1 Matching Functions

The purpose of the matching functions ($matchFunc(ref_i, ref_j)$) in Fig. 1) is to gather the references corresponding to a given paper. These functions allow to compute the frequency factor and the granularity factors presented in section 3.2. Two functions are necessary: a strict matching function and a fuzzy matching

function. The strict matching function checks whether there is an exact matching between the content of the bibliographic fields found in a string and the content of the same fields found in another reference, even if the fields are not the same. The fuzzy matching function allows to estimate the matching degree between references whose content of the bibliographic fields is substantially different. Consider R as a set of references such as ref_i and $ref_j \in R$.

Definition 1. (The strict matching function) This function checks whether bibliographic fields of the ref_i reference ($i \in [1..n]$) and the ref_j reference ($j \in [1..m]$) match. ref_i is tokenized in (w_k^i) and ref_j is tokenized in (w_k^j) , with $k = \min(\text{len}(ref_i), \text{len}(ref_j))$.

$$matching_{strict}(ref_i, ref_j) = \begin{cases} 1 & \text{if } [w_1^i w_2^i .. w_k^i] = [w_1^j w_2^j .. w_k^j] \\ 0 & \end{cases} \quad (1)$$

Definition 2. (The fuzzy matching function) $sim(\theta, \delta)$ corresponds to the Levenshtein distance. θ refers to the vector $(w_1^i .. w_k^i)$ of tokens in ref_i and δ corresponds to the vector $(w_1^j .. w_k^j)$ of tokens extracted from ref_j , with $k = \min(\text{len}(ref_i), \text{len}(ref_j))$. ω specifies a similarity threshold.

$$matching_{similarity}(ref_i, ref_j) = \begin{cases} 1 & \text{if } sim(\theta, \delta) \geq \omega \\ 0 & \end{cases} \quad (2)$$

3.2 BIBLME RecSys Factors

BIBLME RecSys factors are based on the observations related to the construction of scientific discourse and especially authors' citation behaviours. They are computed using quantitative measures such as the frequency and the distribution granularity of each reference in the body of the text.

The Frequency Factor is computed from the number of references corresponding to the same paper even if their structures are different. The hypothesis emitted from this factor is: the impact of the reference will be considered more important if it is cited several times in the paper. Algorithm 1 describes the processing chain used to compute the frequency factor.

Consider R as the set of references ref_j ($j \in [1..m]$) extracted from the body of the text and R' as the set of references $Pref_i$ ($i \in [1..n]$) extracted from the reference section. We only perform these matches because we consider the reference section as a list in which occurs each reference extracted from the text. The *biblMeFactors* array stores the results in order to compute the centrality indicator. In this case, its size is equal to the size of R' . Each *biblMeFactors* index corresponds to the position index of references $Pref_i$ from R' . Then, for each reference in the body of the text ref_j from R , matching functions $matching_{strict}(Pref_i, ref_j)$ and $matching_{similarity}(Pref_i, ref_j)$ are applied with references from R' . If the conditions are verified, *biblMeFactors*[i] is incremented according to the parameter value assigned to α .

Algorithm 1 Computing the frequency factor

```

1: for i of 0 to (length of biblMeFactors) do
2:   biblMeFactors[i] ← 0
3: end for
4: for i of 0 to (length of R') do
5:   for j of 0 to (length of R) do
6:     if matchingstrict(refj, Prefi) = 1 or matchingsimilarity(refj, Prefi) = 1 then
7:       biblMeFactors[i] ← biblMeFactors[i] + α
8:     end if
9:   end for
10: end for

```

The Granularity Factors are computed by taking into account different levels of distribution granularity, namely, the fine granularity and the coarse granularity. The purpose of these factors is to distinguish how the references are used by the authors in their paper. To do that, an additional score is computed for each references extracted from the reference section if corresponding references in the body of the text fulfill distribution conditions. Based on the findings that references' importance increase proportionally with numbers of mentions and more detailed discussion of the cited document [19], we propose to construct the granularity factors according to the following assumptions:

- more citations of a reference occur throughout the paper more the author is influenced by this work;
- on the contrary, the concentration of citations within low textual density areas tends to strengthen author's arguments on specific aspects of his researches.

The fine granularity is computed from references in the body of the text referring to the same paper in the same paragraph and the number of words between each one of these references. Then, a score is assigned if the number of words between these references is less than the average of the distances between the references corresponding to the same paper. The fine granularity function is as follows:

$$Granularity_{fine}(ref_i, ref_j) = \begin{cases} 1 & \text{if } a_j - b_i < Avg_{Ref} \\ 0 & \end{cases} \quad (3)$$

Where *ref_i* and *ref_j* are references extracted from an ordered subset referring to the same document *d* in the same paragraph *P*. *a_j* is the *ref_j* start position in the paragraph *P* and *b_i* is the *ref_i* end position in the paragraph *P*. *Avg_{Ref}* is the average of all the averages of distances in words between references within a document *d* in a paragraph *P*.

The coarse granularity is measured from references in the body of the text corresponding to the same paper throughout a given paper. To do that, we count

the number of paragraphs which separate each of these references. Then, a score is assigned if the number of paragraphs between these references is less than the average of distances. The coarse granularity function can be calculated as follows:

$$Granularity_{coarse}(ref_i, ref_j) = \begin{cases} 1 & \text{if } index(Q) - index(P) < Avg'_{Ref} \\ 0 & \end{cases} \quad (4)$$

Where $index()$ is a function which gives the index of a given paragraph. P and Q two paragraphs and ref_i and ref_j are references extracted from an ordered subset referring to the same document d . Avg'_{Ref} is the average of all the averages of distances between paragraphs that separate two references to the same document d .

Algorithm 2 describes the processing chain used to compute the granularity factors.

Algorithm 2 Computing the granularity factors

```

1: for i of 0 to (length of R) do
2:   for j of 0 to (length of R) do
3:     if  $i \neq j$  then
4:       if  $matching_{strict}(ref_i, ref_j)$  or  $matching_{similarity}(ref_i, ref_j)$  then
5:         if  $granularity_{fine}(ref_i, ref_j) = 1$  then
6:            $biblMeFactors[i] \leftarrow biblMeFactors[i] + \beta$ 
7:         end if
8:         if  $granularity_{coarse}(ref_i, ref_j) = 1$  then
9:            $biblMeFactors[i] \leftarrow biblMeFactors[i] + \gamma$ 
10:        end if
11:       end if
12:     end if
13:   end for
14: end for

```

For each couple (ref_i, ref_j) of references from the set R , we apply matching functions $matching_{strict}(ref_i, ref_j)$ and $matching_{similarity}(ref_i, ref_j)$. If conditions are fulfilled, the measurement of granularity factors $granularity_{fine}(ref_i, ref_j)$ and $granularity_{coarse}(ref_i, ref_j)$ is performed. For each granularity function if their respective average of distances are verified, $biblMeFactors[i]$ is incremented according to the parameter value assigned to β or γ .

4 Context of the Experiments

To evaluate *BIBLME RecSys*, we compare the candidate papers proposed by the OpenEdition's search engine which is designed to search for documents on

OpenEdition portal. In the literature, we were unable to find open recommendation systems based on bibliometric measurements with which to compare to *BIBLME RecSys*'s suggestions. So, we used the OpenEdition search engine called *Search OpenEdition* which is based on Apache Lucene retrieval model. Candidate papers are extracted, for a given query, by a Boolean filter which identify papers containing the requested terms. Then, papers are ranked according to BM25 similarity. As for standard search engines, *Search OpenEdition* allows different querying modes which apply facets or filters. As part of our work, *Search OpenEdition* was queried without specifying search fields related to document characteristics and Boolean operators. However, the advanced search mode was used with the application of a filter which only query OpenEdition Journals platform² in order to deal with the same data that *BIBLME RecSys* uses. Concerning the queries submitted they are written with the name of the first author and the full title of given papers which correspond to the main fields available in the references extracted from the text.

The candidate papers to recommend is constructed from 12 papers³ extracted from various fields in HSS such as languages, anthropology, ethnology, communication, law and culture, health, economy and development, education, agriculture, and the environment. In order to estimate the impact of this approach and to avoid the possibility of bias product by the BILBO software, we manually annotated citations and references of these papers. In order to allow an evaluation of the candidate papers extracted from *BIBLME RecSys* and *Search OpenEdition*, a platform⁴ has been developed.

From this platform, users have access to the list of papers with their abstracts. For each paper two lists of five recommended papers are proposed for both systems. Only first five recommended papers were displayed for each system in order to avoid an evaluation too tedious. Users can choose the list containing the more relevant recommended papers and refine their evaluation by giving a rating from 0 to 5 (0 means that the recommended paper is off topic and 5 means that the recommended paper is in agreement with the topic of the current paper). The majority of the recommended papers have a clickable link to obtain the original version of the paper or an abstract. A "Suggestion" field is available in order to allow users to express an opinion or remarks on suggestions.

5 Experimental Results

From the platform presented previously, an analysis on the basis of the user feedback is performed. Through this study, we show the relevance of recommended papers according to the system selected by the users. The results have been recorded after one month. The OpenEdition employees were the main participants in this evaluation. These participants come from varying cultural and

² <http://journals.openedition.org/>

³ Due to the task complexity for the participants, we only have selected a little sample of papers.

⁴ <http://grapheval.openeditionlab.org/>

disciplinary backgrounds such as software engineering, sales and partnerships, finance, legal and public policy, marketing and communications, user services. Over this time frame, we counted 31 participants with an average of 2.3 items assessed per person.

Concerning the predefined settings for *BIBLME RecSys*, each centrality indicator is computed from the frequency factor and the granularity factors. The same parameter value is applied for each factor in order to obtain the sum of the coefficients equal to 1. Table 1 shows the obtained results for each proposed paper. The values correspond to the number of users which selected the suggestions provided by *BIBLME RecSys* or *Search OpenEdition*.

Table 1. Results Obtained by *BIBLME RecSys* and *Search OpenEdition*.

Proposed paper	BIBLME RecSys	Search OpenEdition
Jaubert - Correspondance as an Ethical Genre	7	5
Bisson - Sufism and Tradition	8	0
Nabti - Sufis in Parisian suburbs	5	0
Laborde - The hermit and the virtuoso	6	0
Esquenazi - From star system to people	4	4
Danou - On a Novella of Arthur Schnitzler (1862-1931), Flight into Darkness	2	1
Wrobel - Gothic, Reform and Panoptic	3	0
Amiraly - The impact of a pilot water metering project in an Indian city on users perception of the public water supply	4	0
Masdonati - The question of identity in the dual Swiss vocational training system: a contrasting picture	1	5
Delannoy - Karst: from palaeogeographic archives to environmental indicators	1	0
Duval - The deceased to the shackles	2	2
Angevin - The Magdalenian lithic industry from the openair site of la Corne-de-Rollay (Couleuvre, Allier): production standards and production lines variability	2	0

Result analyses. Users have selected 45 times *BIBLME RecSys*'s suggestions as the most relevant while *Search OpenEdition*'s suggestions have been selected 17 times. However, performances vary depending on the papers. Users have selected *BIBLME RecSys*'s suggestions as the most effective for 8 proposed papers while the others obtain almost similar performances. The user feedback

provided with the "Suggestion" field shows that the systems propose relevant suggestions. The main difference between the systems concerns the topic covered. Indeed, inspecting the recommended paper topics shows that *BIBLME RecSys* provides suggestions closest to the targeted paper topics. Unlike *Search OpenEdition* that tends to provide suggestions in terms of related themes. Let's take the example of the paper "Sufism and Tradition" which is focused on the influence of the intellectual René Guénon on European Sufi Islam. *BIBLME RecSys* provides suggestions whose the main topics are Sufism and/or René Guénon which are central topics to explaining content of the targeted paper. Unlike *Search OpenEdition* proposes suggestions based on the religious object topic. From this analysis, we can observe that *BIBLME RecSys*'s suggestions are more closely linked to the targeted paper topic. Conversely, those of *Search OpenEdition* are on more generic scientific aspects. Obviously, results are contrasted. An another example, the paper "From star system to people" which is focused on marketing strategies produced through the "star system". Given the candidate paper topics, *BIBLME RecSys* provides suggestions about the economic exploitation of notoriety and *Search OpenEdition* proposes suggestions based on the topic of the "celebrity culture". In this case, each system provides suggestions with at least one central topic related to the targeted paper. Despite this, these examples indicate how our approach can characterize user's interests by proposing papers based on important topics of the targeted paper.

Limitations. The user feedback has revealed different behaviours from each system. Indeed, our approach provides suggestions more closely related to the targeted paper topics than *Search OpenEdition*. The suggestions proposed by *BIBLME RecSys* are extracted directly from the text which explain close topic links. Sometimes it was difficult for users to evaluate recommended papers due to the topic complexity. Thus, the main limitations concern users' satisfaction. In this experiment, we focused on topic links between recommended papers and targeted papers. However, in the context of scholarly recommender system it is important to assess the scientific content of recommended papers. In this experiment, we can observe that users are able to evaluate the topic links although they are not specialists on proposed paper's topics. Papers not sufficiently evaluated, such as "Karst : from palaeogeographic archives to environmental indicators", reflect the needs of a specific expertise.

Finally, leveraging our centrality indicator, *BIBLME RecSys* is able to suggest central papers of the author's paper and therefore relevant readings according to the proposed paper. This is then reflected in the user choices which tend to select *BIBLME RecSys*'s suggestions. We believe that our approach is effective in characterizing candidate papers to recommend in order to obtain much higher recommendation relevance in the context of scholarly recommender system.

6 Conclusion

We have explored an approach based on the content analysis of the paper the user is reading. From bibliographical reference analysis and and objective quan-

titative measurements we proposed a centrality indicator. This indicator allows to evaluate bibliographical references' importance of the paper that the user interacted with. In this approach, we represent a candidate paper from the number of its mentions and how its citations are used by the authors. From these information, we determined central papers of the authors and therefore relevant readings. Our results showed that, in discovery of potential relevant readings, *BIBLME RecSys*'s suggestions are more closely linked to the targeted paper topic.

We believe that our approach can be applied more generally in the context of digital libraries. Harnessing bibliographical references in the full-text can be used wherever scientific domains. Moreover, this indicator is based on authors' citation behaviours specific to the targeted paper unlike recent works in bibliometrics [4]. The use of a such indicator can highlight, for a given paper, the influence of references on the author's paper. This indicator may be used for the scientific activity evaluation and scientific networks evaluation currently practiced in the bibliometric field. This indicator allows to harness information about central papers used by authors and potentially the most influential on their research.

To discover much more relevant papers, in future work, we plan to use graph data model for our data in order to exploit recommendation algorithms by graph analysis. By this modelling, we intent to involve external resources based on the same topics but also to consider the centrality indicator as edges. Moreover, we plan to determine relations between references and their proximity [21], and so propose references between two articles (citing-cited) based on their affiliation to a particular scientific domain, similar author collectives and their influence on each other. Our future work direction also aims to examine the degree of users' satisfaction regarding the OpenEdition Journals through library user surveys.

Acknowledgment

This research was supported by ANR program Investissements d'Avenir EquipEx DILOH (ANR-11-EQPX-0013).

References

1. Asabere, N. Y., Xia, F., Meng, Q., Li, F., Liu, H.: Scholarly paper recommendation based on social awareness and folksonomy. *International Journal of Parallel, Emergent and Distributed Systems*, vol. 30 no. 3, pp. 211-232, (2015)
2. Baeza-Yates, R. , Ribeiro-Neto, B.: *Modern information retrieval: The Concepts and Technology behind Search*. ACM Press, New York, (1999)
3. Beel, J., Gipp, B., Langer, S., Breiting, C.: paper recommender systems: a literature survey. *International Journal on Digital Libraries*, vol. 17 no. 4, pp. 305-338, (2016)
4. Belter, C.: "Bibliometric indicators: opportunities and limits", *Journal of the Medical Library Association*, vol. 103 no. 4, Medical Library Association, pp. 219, (2015)
5. Bogers, T., Van den Bosch, A.: Recommending scientific articles using citeulike. In: *the 2008 ACM conference on Recommender systems*, pp. 287-290, ACM, (2008)

6. Bollacker, K., Lawrence, S., Giles, L.: CiteSeer: An autonomous web agent for automatic retrieval and identification of interesting publications. In: Proc. of the 2nd Int. Conf. Autonomous agents, pp. 116–123, vol. 41 no. 1, ACM, (1998)
7. Haruna, K., Ismail, M. A., Damiasih, D., Sutopo, J., Herawan, T.: A collaborative approach for research paper recommender system. PloS one, vol. 12 no. 10, (2017)
8. Hristakeva, M., Kershaw, D., Rossetti, M., Knoth, P., Pettit, B., Vargas, S., Jack, K.: Building recommender systems for scholarly information. In: 1st International Workshop on Scholarly Web Mining (SWM), International Conference on Web Search and Data Mining (2017)
9. Khabsa, M., Giles, C. L.: The number of scholarly documents on the public web. PloS one, vol. 9 no. 5, pp. e93949, (2014)
10. Kim, Y.-M., Bellot, P., Faath, E., Dacos, M.: Automatic annotation of bibliographical references in digital humanities books, articles and blogs. In: Proc. of the 4th ACM workshop on Online books, complementary social media and crowdsourcing, pp. 41–48, ACM, (2011)
11. Knoth, P., Anastasiou, L., Charalampous, A., Cancellieri, M., Pearce, S., Pontika, N., Bayer, V. (2017). Towards effective research recommender systems for repositories. In: Proc. of Open Repositories, Open Repositories, (2017)
12. Mnnich, M., Spiering, M.: Adding value to the library catalog by implementing a recommendation system. D-Lib Magazine, 14.5/6, pp. 1082–9873, (2008)
13. Nascimento, C., Laender, A. H., da Silva, A. S., Goncalves, M. A.: A source independent framework for research paper recommendation. In : Proc. of the 11th annual international ACM/IEEE joint conference on Digital libraries, pp. 297–306, ACM, (2011)
14. Lops, P., De Gemmis, M., Semeraro, G.: Content-based recommender systems: State of the art and trends. Recommender systems handbook, pp. 73–105, Springer, (2011)
15. Ollagnier, A., Fournier, S., Bellot, P.: A Supervised Approach for Detecting Allusive Bibliographical References in Scholarly Publications. In : ACM WIMS Conference on Web Intelligence, Mining and Semantics (WIMS), Nmes, France, pp. 36–39, (2016)
16. Pazzani, M. J., Billsus, D.: Content-based recommendation systems. In: The adaptive web, pp. 325–341, Springer, Berlin, Heidelberg. (2007)
17. Philip, S., Shola, P., Ovyne, A.: Application of content-based approach in research paper recommendation system for a digital library. International Journal of Advanced Computer Science and Applications 5.10 (2014)
18. Su, X., Khoshgoftaar, T. M.: A survey of collaborative filtering techniques. Advances in artificial intelligence, pp. 4, (2009)
19. Tang, R., Safer., M. A.: Author-rated importance of cited references in biology and psychology publications. Journal of Documentation, vol. 62 no. 2, pp. 246–272, (2008)
20. Torres, R., Mcnee, S. M., Abel, M., Konstan, J. A., Riedl, J.: Enhancing digital libraries with TechLens. In : Proc. of the 2004 Joint ACM/IEEE Conference on. IEEE, pp. 228–236, (2004)
21. Tran, H. D., Cabanac, G., Hubert, G.: Expert suggestion for conference program committees. In : 11th International Conference on Research Challenges in Information Science (RCIS), pp. 221–232, IEEE, (2017)
22. Yang, C., Wei, B., Wu, J., Zhang, Y., Zhang, L.: CARES: a ranking-oriented CADAL recommender system. In : Proc. of the 9th ACM/IEEE-CS joint conference on Digital libraries, pp. 203–212, ACM, (2009)