

# Testing Collaborative Filtering against Co-Citation Analysis and Bibliographic Coupling for Academic Author Recommendation

Tamara Heck  
Heinrich-Heine-University  
Dept. of Information Science  
D-40225 Düsseldorf, Germany  
Tamara.Heck@hhu.de

Isabella Peters  
Heinrich-Heine-University  
Dept. of Information Science  
D-40225 Düsseldorf, Germany  
Isabella.Peters@hhu.de

Wolfgang G. Stock  
Heinrich-Heine-University  
Dept. of Information Science  
D-40225 Düsseldorf, Germany  
Stock@phil.hhu.de

## ABSTRACT

Recommendation systems have become an important tool to overcome information overload and help people to make the right choice of needed items, which can be e.g. documents, products, tags or even other people. Last attribute has aroused our interest: Scientists are in need of different collaboration partners, i.e. experts for a special topic similar to their research field, to work with. Co-citation and bibliographic coupling have become standard measurements in scientometrics for detecting author similarity, but it can be laborious to elevate these data accurately. As collaborative filtering (CF) has proved to show acceptable results in recommender systems, we investigate in the comparison of scientometric analysis methods and CF methods. We use data from the social bookmarking service CiteULike as well as from the multi-discipline information services Web of Science and Scopus to recommend authors as potential collaborators for a target scientist. The paper aims to answer how a relevant author cluster for a target scientist can be proposed with CF and how the results differ in comparison with co-citation and bibliographic coupling. In this paper we will show first result, complemented by an explicit user evaluation with the help of the target authors.

## Categories and Subject Descriptors

H.2.8 [Database Management]: Database Applications – *Scientific databases*. H.3.3 [Information Storage and Retrieval]: Information Search and Retrieval – *Information filtering*. H.3.5 [Information Storage and Retrieval]: Online Information Services – *Web-based services*.

## General Terms

Measurement, Experimentation, Human Factors, Management.

## Keywords

Collaborative Filtering, Recommendation, Evaluation, Social Bookmarking, Personalization, Similarity Measurement, Bibliographic Coupling, Author Co-Citation, Social Tagging.

## 1. INTRODUCTION

An important task for knowledge management in academic settings and in knowledge-intensive companies is to find the

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

*Conference '10*, Month 1–2, 2010, City, State, Country.  
Copyright 2010 ACM 1-58113-000-0/00/0010...\$10.00.

“right” people who can work together to solve successfully a scientific or technological problem. This can either be a partner having the same skills and providing similar know-how, or someone with complementary skills to form a collaborative team. In both cases the research interests must be similar. Amongst others this interest can be figured out with a person’s scientific publications. Exemplarily, we will list some situations in which expert recommendations are very useful:

- compilation of a (formal) working group in a large university department or company,
- compilation of researchers for preparing a project proposal for a research grant (inside and outside the department and company),
- forming a Community of Practice (CoP), independent from the affiliation with the institutions following only shared interests,
- accosting colleagues in preparation of a congress, a panel or a workshop,
- asking colleagues for contributions to a handbook or a specialized journal issue,
- finding appropriate co-authors.

It is very important for cooperation in science and technology that the reputation of the experts is proved [15]. A recommendation service must not suggest just anybody who is possibly relevant, but has to check up on the expert’s reputation. The reputation of a person in science and technology grows with her or his amount of publications in peer-reviewed journals and with the citations of those publications [14]. So we are going to use academic information services, which stores publication and citation data, as basis for our author recommendation. Multi-discipline information services which allow publication and citation counts are Web of Science (WoS) and Scopus [34, 40, 41]. Additionally our experimental expert recommendation applies also data from CiteULike, which is a social bookmarking service for academic literature [18, 22]. So we can not only consider the authors’ perspectives (by tracking their publications, references and citations via WoS and Scopus), but also the perspectives of the readers (by tracking their bookmarks and tags via CiteULike) to recommend relevant partners. Our research questions are: 1) Can we propose a relevant author cluster for a target scientist with CF applying CiteULike data? 2) Are these results different to the results based on co-citation and bibliographic coupling?

Recommender systems (RS) nowadays use different methods and algorithms to recommend items, for e.g. products, movies, music, articles, to a Web user. The aim is personalized recommendation [5], i.e. to get a list of items, which are unknown to the target user and which he might be interested in. One problem is to find the best resources for user *a* and to rank them according to their

relevance [16]. Two approaches are normally distinguished (among other distinctive recommender methods and hybridizations): The content-based approach, which tries to identify similarities between items based on their content and positively rated by user  $a$ , and the collaborative filtering approach (CF), which not only considers the ratings of user  $a$ , but also the ratings of other users [a.o. 16, 20, 25, 35, 37, 42, 48]. One advantage of CF compared to the content-based method is that recommendations rely on the evaluation of other users and not only on the item's content, which can be inappropriate for quality indication.

RS work with user ratings assigned to the items, also called user-item response [16]: They can be scalar (e.g. 1-5 stars), binary (like/dislike) or unary, i.e. a user doesn't rate an item, but his purchase or access of the item is assumed as a positive response. The latter user-item response can also be used for recommendations in social tagging systems (STS) as e.g. social bookmarking systems like BibSonomy, CiteULike and Connotea [38]. STS have a folksonomy structure with user-resource-tag relations, which is the basis for CF. In STS not only recommendations of items are possible, but also recommendations of tags and users, which is the basis for our academic author recommendations. We apply approaches of CF to recommend potential collaboration partners to academic researchers. Hereby we ask if CF in a STS recommends different results than the established scientometric measurements, author co-citation and bibliographic coupling of authors. In general these measurements are not explicitly used for recommendation, but rather for author and scientific network analysis [54].

## 2. RELATED WORK

RS can be constructed in many different ways, e.g. choosing the appropriate algorithm especially for personal recommendation [54], defining user interactions and user models [44], facing criteria like RS accuracy, efficiency and stability [16] and focusing on optimal RS learning models [47]. With the appearance of bookmarking and collaboration services in the Web, several algorithms and hybridizations have been developed [27]. They may differ in combination of the considered relations between users, items and tags and the used weights. Similarity fusion [59] for example combines user- and item-based filtering (subcategories of CF) and additionally adds ratings of similar items by similar users. Cacheda et al. give an overview of different algorithms and compare the performances of the methods, also proposing a new algorithm, which takes account of the users' positive or negative ratings of the items [11]. Bogers and van den Bosch compare three different collaborative filtering algorithms, two item-based and one user-based. The latter one outperformed the others throughout a time of 37 months [8]. But the most evident problem seems to be the cold-start, i.e. new items cannot be recommended at the beginning [2]. Said et al. are also concerned with the cold-start problem and the performance of different algorithms within a time span: Thereby adding tag similarity measures can improve the quality of item recommendation because tags offer more detailed information about items [50]. Hotho et al. propose the FolkRank [27], a graph based approach similar to the idea of the PageRank, which can be applied in a system with a folksonomy structure like a bookmarking service. Hereby users, tags and resources are the nodes in the graph and the relations between them become the weighted edges, taken into account weight-spreading like the PageRank does. In the current approach similarity based on users and tags within CiteULike is measured separately. Using the relations between them, like it is done in the FolkRank method, may lead to better recommendations. However this method may

not be applied to bibliographic coupling and author co-citation [see paragraph 3] without modifications.

Several papers investigate in expert recommendation mainly for business institutions [45, 46, 62]. Petry et al. developed the expert recommendation system ICARE, which should recommend experts in an organization. Therefore the focus doesn't lie on an author's publications and citations, but for example on his organizational level, his availability and his reputation [45]. Reichling and Wulf investigated in a recommender system for a European industrial association supporting their knowledge management, foregone a field study and interviews with the employees. Experts were defined according to their collection of written documents, which were automatically analyzed. Additionally a post-integrated user profile with information about their background and job is used [46]. Using user profiles in bookmarking services could also be helpful to provide further information about a user's interests and prove user recommendation, which could be an investigating new research approach. However this approach has serious problems with privacy and data security on the Web.

Apart from people recommendation for commercial companies [a.o. 12, 51] other approaches concentrate on Web 2.0 user and academics. Au Yeung et al., using the non-academic bookmarking system Del.icio.us, define an expert user as someone who has high-quality documents in his bookmark collection (many others users with high expertise have them in their collection) and who tends to identify useful documents before other users do it (according to the timestamp of a bookmark) [3]. In comparison their SPEAR algorithm is better for finding such experts than the HITS algorithm, which is used for link structure analysis. Compared to the current approach the "high-quality documents" in this experiment are the publications of our target author, i.e. a user who has bookmarked one of these publications is important for our user-based recommendation (see paragraph 3). A weighed approach like Yeung et al. did it when they weighted a user's bookmarks according to their quality could also be interesting to test. Blazek focuses on expert recommendation sets of articles for a "Domain Novice Researcher", i.e. for example new academics, who enter a new domain using a collection of academic documents [7]. A main aspect hereby is again the cold start problem: Citation analysis can hardly be applied for novice researchers, as long as there are no or only few references and citations. Therefore in the current approach only target authors where chosen, who have at least published five articles in the last five years. Blazek understands his expert recommendation mainly as a recommendation of relevant documents. Heck and Peters propose to use social bookmarking systems for scientific literature such as BibSonomy, CiteULike and Connotea to recommend researchers, who are unknown to the target researcher, but share the same interests and are therefore potential cooperation partners to build CoP [24]. Users are recommended when they have either common bookmarks or common tags, a method founded on the idea of CF. A condition is that the researcher, who should get relevant expert recommendations, must be active in the social bookmarking system and put his relevant literature to his internet library. In this project, beneath the additional comparison of CF against co-citation and bibliographic coupling, we avoid the problem of the "active researcher", i.e. we have a look at the users in CiteULike, who have bookmarked our target researcher's publications. Therefore the recommendation doesn't depend on the target scientist himself, which would be based on his bookmarks and assigned tags, but on the bookmarking users and their collaborative filtering. The approach of Cabanac is similar to [24], but he concentrates only on user similarity networks and

relevant articles, not on the recommendation of unknown researchers [10]. He uses the concepts of Ben Jabeur et al. to build a social network for recommending relevant literature [4]. The following entities can be used: Co-authorship, Authorship, Citation, Reference, Bookmarking, Tagging, Annotation and Friendship. Additionally Cabanac adds social clues like connectivity of researchers and meeting opportunities on scientific conferences. According to him these social clues lead to a better performance of the recommendation system. Both approaches [4, 10] aim to build a social network to show the researcher connectivity to each other. In this project co-authorship for example is not important, as we try to recommend unknown researchers or academics our target author has not in his mind. Zanardi and Capra, proposing a ‘‘Social Ranking’’, calculate similarity between users based on same tags and tag-pairs based on same bookmarks they both describe [63]. The tag similarity is compared with a user’s query tag; both user and tag similarity are then combined. The results show that user similarity improves accuracy whereas tag similarity improves coverage.

Another important aspect with RS is their evaluation. RS should not only prove accuracy and efficiency, but also usefulness for the users [26]. The users’ need must be detected to make the best recommendation. Beneath RS evaluation based on models [29], some papers investigate in user evaluation [39]. McNee et al. show recommender pitfalls to assure users acceptance and growing usage of recommenders as knowledge management tools. This is also one of our main aspects in this paper because we want to recommend potential collaboration partners to our target scientists. They have to prove the recommended people as useful for their scientific work.

### 3. MODELING RECOMMENDATION

#### 3.1 Similarity Algorithm

The most common similarity measures in Information Science are Cosine, Dice and Jaccard-Sneath [1, 2, 31, 33, 49, 57]. The last two are similar and come to similar results [17]. Additionally Hamers et al. proved that the similarity measures with the cosine coefficient are twice the number than the Jaccard coefficient showed referring to citation measurements [21]. According to van Eck and Waltman the most popular similarity measures are the association strength, the cosine, the inclusion index and the Jaccard index [58]. In our comparative experiment we make use of the cosine. Our own experiences [23] and results from the literature [49] show that cosine works well. But in later project steps we want to extent the similarity measures to Dice and Jaccard-Sneath as well.

#### 3.2 Collaborative Filtering Using Bookmarks and Tags in CiteULike

Social bookmarking systems like BibSonomy, Connotea and CiteULike have become very popular [36]: Unlike bookmarking systems like Del.icio.us they focus on academic literature management. Basis for social recommendation are their folksonomies. A folksonomy [38, 43] is defined as a tuple  $F := (U, T, R, Y)$ , where  $U$ ,  $T$  and  $R$  are finite sets with the elements usernames, tags and resources and  $Y$  is a ternary relation between them:  $Y \subseteq U \times T \times R$  with the elements called tag actions or assignments. The tripartite structure allows matching users, resources or tags which are similar to each other. CF uses data of the users in a system to measure similarity [16]. To get a 2-dimensional matrix for applying traditional CF, which is not possible in the ternary relation  $Y$ , one could split  $F$  in three 2-dimensional subsets: The docsonomy  $D_F := (T, R, Z)$  where  $Z \subseteq T \times R$ , the personomy  $P_{UT} := (U, T, X)$  where  $X \subseteq U \times T$ , and the user-

resource relation, which we call in our case personal bookmark list (PBL):  $PBL_{UR} := (U, R, W)$  where  $W \subseteq U \times R$ .

In our experimental comparison we want to cluster scientific authors who have similar research interests. Scientometric analyses are co-citation and bibliographic coupling, which we compare with data from CiteULike using CF. Therefore we are not interested in the CiteULike users themselves, but in their tags and bookmarks they connect with our target author, i.e. the bookmarked papers which our target author published. We set  $R_a$  for all bookmarked articles which our target author  $a$  published and  $T_a$  for all tags which are assigned to those articles. To set our database for author similarity measure we have two possible methods:

1. We search for all users  $u \in U$  who have at least one article of the target author  $a$  in their bookmark list:  $PBL_{URa} := (U, R_a, W)$  where  $W \subseteq U \times R_a$ .
2. We search for all documents, to which users assigned the same tags like to the target author’s  $a$  articles:  $D_{Fa} := (T_a, R, Z)$  where  $Z \subseteq T_a \times R$ .

The disadvantage in the first method, in our case, is the small number of users. It can be difficult to rely only on these users for identifying similarity [30]. Therefore we use the second method: Resources (here: scientific papers) can be supposed similar, if the same tags have been assigned to them. Our assumption is that also the authors of these documents are similar because users describe their papers with the same keywords. Tags show topical relations, and authors with thematically relations concerning their research field are potential collaboration partners. Additionally the more common tags two documents have, the more similar they are. In some cases very general tags like ‘‘nanotube’’ and ‘‘spectroscopy’’ were assigned to our target authors’ articles. So we decided to set a minimum of unique tags a document must have in common with a target author’s document:

$$D_{Fa} := (T_a, R, Z) \text{ where } Z \subseteq T_a \times R \text{ e. } \{r \in T_a \times R \text{ with } |T_a| \geq 2\} \quad (1)$$

On this database we measure author similarity in two different ways: (A) Based on common tags  $t$  assigned to the authors’ documents by users; (B) Based on common users  $u$ . We use the cosine coefficient as explained above:

$$a) \text{sim}(a, b) := \frac{T_a \cap T_b}{\sqrt{T_a * T_b}} \quad b) \text{sim}(a, b) := \frac{U_a \cap U_b}{\sqrt{U_a * U_b}} \quad (2)$$

Consider that the latter method leads to different results than applying the proposed first method for database modeling. If we would apply the first method, we would find all users who have at least one document of target author  $a$  in their bookmark list. With the second method, we get all users, who have at least one document in their bookmark list, which is similar to any of target author’s  $a$  articles, i.e. users who bookmarked a document of  $a$  may be left out. As we want to apply one unique dataset for author similarity measure, we do not merge both methods, but measure tag-based and user-based similarity in the dataset described above. Nevertheless where no tags were available, we chose the first method (see paragraph 5).

#### 3.3 Author Co-Citation and Bibliographic Coupling of Authors

There are four relations between two authors concerning their publications, references and citations: co-authorship, direct citation, bibliographic coupling of authors and author co-citation. The first two relationships are not appropriate for our problem, for here it is sure that one authors knows the other: of course, one knows his co-authors, and we can assume, that an author knows

who she has cited. Our goal is to recommend unknown scientists. Bibliographic coupling (BC) [28] and co-citations (CC) [55] are undirected weighted linkages of two papers, calculated through the fraction of shared references (BC) or co-citations (CC). We aggregate the data from the document level to the author level.

Bibliographic coupling of authors means that two authors  $a$  and  $b$  are linked if they cite the same authors in their references. We mine data about bibliographic coupling of authors by using WoS, for this information service allows searches for “related records”, where the relation is calculated by the number of references a certain document has in common with the source article [13, 56]. Our assumption is: Two authors who have two documents with a high number of same references are more similar than two authors who have a high number of same references in many documents, i.e. the number of same references per document is important. Consider authors  $a$ ,  $b$  and  $c$ :

$$\text{sim}(a, b) > \text{sim}(a, c) \quad (3)$$

if

$$\frac{\text{Ref}_a \cap \text{Ref}_b}{D_a \cup D_b} > \frac{\text{Ref}_a \cap \text{Ref}_c}{D_a \cup D_c} \quad (4)$$

where  $\text{Ref}$  is the set of references of an author and  $D$  the set of documents of an author  $\{d \in D \times \text{Ref}_a\}$ . For example: author  $a$  has 6 references in common with author  $b$  and  $c$ . These 6 common references are found in two unique documents of author  $a$ , respectively of author  $b$ , but in 6 unique documents of author  $c$ , i.e.:

$$\frac{6}{2+2} > \frac{6}{2+6} \quad (5)$$

Therefore it can be said that authors  $a$  and  $b$  are similar if there documents have similar reference lists. Our assumption leads to the following dataset model for BC, where we take all authors of related documents with at least  $n$  common references with any of the target author’s publications, where  $n$  may vary in different cases:

$$\text{BC} := (\text{Ref}_{d(a)}, D, S) \text{ where } S \subseteq \text{Ref}_{d(a)} \times D \text{ and } \{d \in D \mid |\text{Ref}_{d(a)}| \geq n, n \in \mathbb{N}\} \quad (6)$$

where  $\text{Ref}_{d(a)}$  is the number of references in one document  $d$  of target author  $a$ . Unique authors of the dataset are accomplished; the list of the generated authors of the related documents is cut at  $m \in \mathbb{N}$  unique authors ( $m > 30$ ) because their publications and references for BC have to be analyzed manually in WoS. For these related authors we measure similarity with the cosine (Eq. 2a), where  $T$  is substituted with  $H$  and  $H_a$  is the number of unique references of target author  $a$  and  $H_b$  the number of references of author  $b$ .

Author Co-Citation (ACC) [32, 52, 53, 60, 61] means that two authors  $a$  and  $b$  are linked if they are cited in the same documents. ACC is then measured with cosine (Eq. 2a), where  $T$  is substituted with  $J$  and  $J_a$  is the number of unique citing articles which cite target author  $a$  and  $J_b$  is the number of unique citing articles which cite author  $b$ . To mine the author-co-citation data it is not possible to work with WoS, for in the references section of a bibliographic entry there is only the first author of the cited documents and not, what is needed, a declaration of all authors [65]. Therefore we are going to mine those data from Scopus, for here we can find more than one author of the cited literature. We perform an inclusive all-author co-citation, i.e. two authors are considered co-cited when a paper they co-authored is cited [64]. The dataset is based on the documents which cite at least one of the target author’s articles in Scopus:

$$\text{ACC} := (D, C_a, Q) \text{ where } Q \subseteq D \times C_a \text{ with } |Q| > 0 \quad (7)$$

where  $C_a$  is the set of cited articles of target author  $a$ . The list of potential similar authors is cut at  $m \in \mathbb{N}$  unique authors ( $m > 30$ ) because their publications for ACC have to be analyzed manually in Scopus. With regards to the results of the research literature, both methods, BC and ACC in combination, perform best to represent research activities [6, 9, 19]. Applying the proposed four mined datasets and similarity approaches we can assemble four different sets of potential similar authors, which we call clusters. One cluster is based on BC in WoS, one cluster is based on ACC in Scopus, one cluster is based on common users in CiteULike and one cluster is based on common tags in CiteULike. We can now analyze the authors who are most similar to our target author according to the cosine coefficient and evaluate the results. Additionally based on the mined datasets we can also measure similarity between all authors of a cluster. These results are shown in visualizations, which we call graphs. Therefore for each cluster a visualized graph exists which will also be evaluated.

## 4. DATASET LIMITATIONS

While filtering the information in the three information services different problems arise, which we would point out briefly, because the recommendation results highly depend on the source dataset. In Scopus we detected differences in the metadata: An identical article may appear in different ways, i.e. for example title and authors may be complete in one reference list of an article, but incomplete in a reference list of another article. In our case, several co-authors in the dataset are missed and could not be considered for co-citation. The completeness of co-authorship highly varies: In a random sample, where the co-citation dataset is adjusted with data of the Scopus website, five of 14 authors have a complete coverage, three of them have coverage between 70 and 90 %, five between 55 and 70 % and one author only has coverage of about 33 %. In the information services there is the problem of homonymy concerning author names. Additionally in CiteULike users also misspell author names, which were rechecked for our dataset. The id-number for an author in Scopus is practical for identification, but it may also fail when two or more authors with the same name are allocated to the same research field and change their working place several times. In WoS we don’t have an author-id and it is more difficult to distinct a single person. Therefore we check the filtered author’s document list and if necessary correct it based on the articles’ subject area.

## 5. EXPERIMENTAL RESULTS

We cooperate with physicists of the *Forschungszentrum Jülich* and worked with 6 researchers so far. For any of the 6 target academic authors (35-50 years old) we build individual clusters with authors who are supposed to be similar to them. We limit source for the dataset modeling to the authors’ publications between 2006-2011 to make recommendations based on the actual research interest of the physicist. To summarize, any scientist got the following four clusters: 1. Based on author co-citation (COCI) in Scopus, 2. Based on bibliographic coupling (BICO) in WoS, 3. Based on common users in CiteULike (CULU) and 4. Based on common tags in CiteULike (CULT). Based on the cosine similarity we are also able to show graphs of all four clusters using the cosine coefficient for similarity measure between all authors (e.g. Fig. 1 and Fig. 2). We applied the software Gephi<sup>1</sup> for the cluster visualization. The nodes (=author names) are sized according to their connections, the edges are seized according to the cosine weight. Consider that the CiteULike graphs are much bigger because all related authors are taken into account. To get a

<sup>1</sup> <http://gephi.org/>

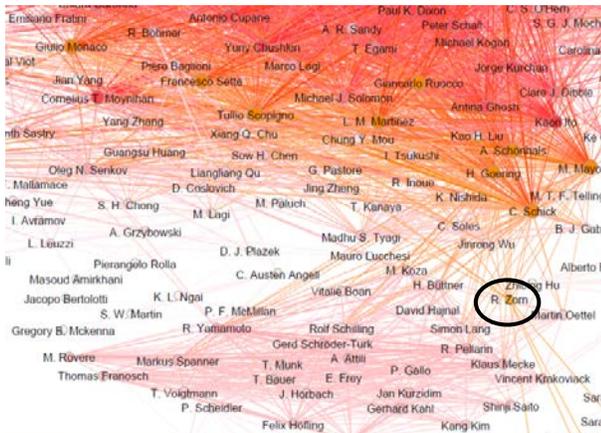


Figure 1. Extract of a CULT graph, circle = target author, cosine interval 0.99-0.49.

clear graph arrangement for a better evaluation, we set thresholds based on the cosine coefficient when needed. Additionally we left out author-pairs with a similarity of 1 if they had only one user or tag (in the CiteULike dataset) in common because this would distort the results.

While modeling the datasets we found out that one of the six authors didn't have any users, who bookmarked any of his articles in CiteULike. Some articles were found, but they were adjusted to the system by the CiteULike operators themselves, so the CiteULike clusters couldn't be modeled for this scientist. One researcher's articles were bookmarked, but not tagged. In this case, we used method 1 in 3.2 to model the dataset. In all four clusters we ranked the similar authors with the cosine. In general it can be seen that the cosine coefficient for BC is very low according to the one for ACC and similarity measures in CiteULike. This is because some authors have a lot of references, which minimize similarity. Additionally similarity is comparatively very high for measurements in CiteULike because the number of users and assigned tags related to the target authors' publications was relatively low.

## 6. EVALUATION

To prove our experimental results we let our 6 target physicists evaluate the clusters as well as the graphs. The evaluation is divided in three parts. Part one is arranged in a semi-structured interview with questions about the scientist's research behavior and the purchase of relevant literature as well as his working behavior, i.e. is he organized in teams and with whom does he cooperate? These questions should show a picture of the scientists work and help to estimate the following evaluation results. In the second part the target author has to rank the proposed similar authors according to their relevance. Therefore the ten top authors of all four measurements are listed in alphabetical order (co-authors eliminated). The interviewee should tell if he knew the proposed authors, how important these authors are for his research (rating from not important (1) to very important (10)), with whom he would cooperate and which important authors he misses.

In part three our author has to evaluate the cluster graphs (rating from 1 to 10) according to the distribution of the authors and the generated groups. Here the questions are: 1. Due to your individual valuation does the distribution of the authors reflect reality respective to the research community and the collaborations of them? 2. Are there any other important authors you didn't remember before? 3. Would this graph, i.e. the recommendation of similar authors, help you e.g. to organize a workshop or find collaboration partners?

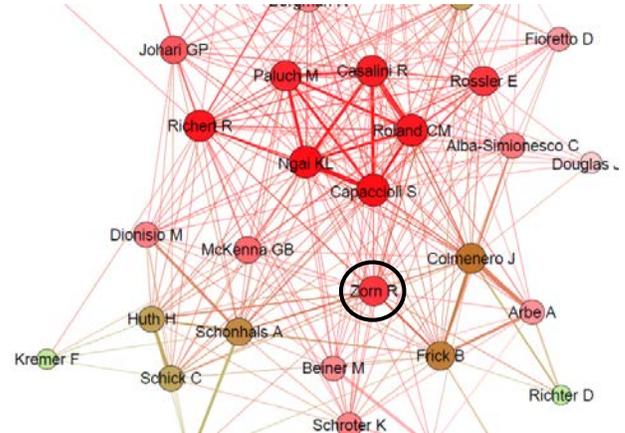


Figure 2. BICO graph, circle = target author, cosine threshold 0.2.

We will shortly summarize the interesting answers of part one: As confirmed in our earlier studies [24] most of the physicists work in research teams, i.e. they collaborate in small groups (in general not more than 5 people). The choice of people for possible collaboration highly depends on their research interest: There must be a high thematic overlap. On the other hand, if the overlap is too high, it could be disadvantageous. Some authors, who claimed a similar author in a cluster important, stated that they wouldn't cooperate with him because he exactly does the same research, i.e. he is important for their own work, but rather a competitor of them. Additionally another statement against collaboration was less thematic overlap. Successful collaborations with international institutes are aspired. In general our interviewees meet new colleagues at conferences and scientific workshops.

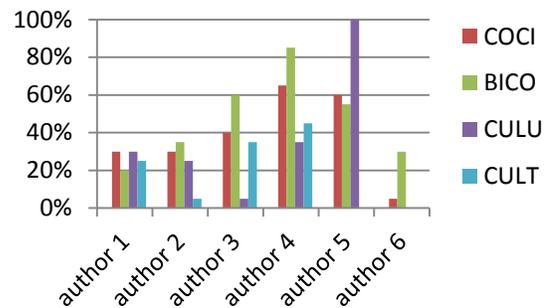


Figure 3. Coverage of important authors in the recommendation of the Top 20 authors.

Part two of the evaluation is concerned with the similar author ranking. We analyze all authors an interviewee claimed important with at least a rating of 5 and all important authors, which the researcher additionally added and which were not on the Top 10 list of any cluster. In general our target authors have up to 30 people they claim most important for their recent scientific work. Figure 3 shows the coverage of these important authors for the first 20 ranks based on the cosine (consider author 6 didn't have any publication bookmarked in CiteULike). For example concerning target author 1: 30 % of the 20 most similar authors of the co-citation cluster (COCI) are claimed important. In the bibliographic coupling cluster (BICO) it is 20 %, in the CiteULike cluster based on users (CULU) 30 % and in the CiteULike cluster based on tags 25 %. Compared to the other target authors there are great differences. The BICO and COCI cluster can be said to provide the best results except by author 1 and 5. Concerning the

CiteULike clusters they are slightly worse, but not in all cases: By author 1 the CULU provide the same coverage than COCI, both CULU and CULT are better than BICO. By author 5 (no CULT because no tags were assigned) CULU has full coverage, which means that all the 20 top authors ranked by the cosine are claimed important by the target author. Beneath the coverage results shown in figure 3 it is interesting to look at the important authors who were only found on the CiteULike clusters: E.g. 6 of the 29 important authors of target author 1 are only in the CiteULike cluster, the same applies to 5 of 19 important authors of target author 2.

The great differences may also depend on the interviewees' recent research activities: Some of the physicists said that they slightly changed their research interest. Hitherto similar authors who were important in the past aren't important nowadays. One problem with our applied similarity measure may be that it is based on past data, i.e. publications of the last five years. The authors the interviewees marked as important, are important for recent research. If we would have considered all authors who are or have been important, the results for the clusters would have been better.

In the third part of the evaluation the interviewee had to evaluate the graphs. The average cluster relevance (based on six target users) are 5.08 for COCI, 8.7 for BICO, 2.13 for CULU and 5.25 for CULT. Consider only four authors had publications and tags in CUL to be analyzed. For author 5, for whom we applied method 1 (see 3.2) in case of missing tags, no CiteULike graph could be modeled because only one user bookmarked his articles and we measured author similarity only on the numbers of authors this user had in his literature list. Two authors claimed BICO and CULT to be very relevant and proposed to combine these two to get all important authors and relevant research communities. In BICO and COCI some interviewees missed important authors. Two of the interviewees stated that the authors in BICO and COCI are too obvious to be similar and were interested in bigger graphs with new potential interesting colleagues. A combined cluster could help them to find researcher groups, partners for cooperation and it would be supportive to intensify relationships among colleagues. Looking at the graphs almost all target authors recollected important colleagues, who didn't come to their mind first, which they found very helpful. They stated that bigger graphs like CULT show more unknown and possible relevant people. However to give a clear statement about the similar researchers who were unknown by the target user, the interviewee would have had to look at these researchers' publications. Assumptions can be made that if an unknown person is clearly connected to a known relevant researcher group, this person would do similar relevant work. As the interviewees stated that the distribution of the researchers is shown correctly, it is likely, but not explicitly proved, that the unknown scientist are also allocated correctly within the graph.

An important factor for all interviewees is a clear cluster arrangement. A problem which may concern CUL clusters is the sparse dataset, i.e. if only few tags were assigned to one author's publications or only one user bookmarked them, the cluster cannot show high distinguishable communities. That was the case with author 2 and 5. Author 2 gave worse ratings to the CUL graphs because they didn't show clear distributions and author groups. Further categorizations of authors, e.g. via tags or author keywords, might help to classify scientists' work.

## 7. DISCUSSION

In our project we analyzed academic author recommendation based on different author relations in three information services. We combined two classical approaches (co-citation and

bibliographic coupling) with collaborative filtering methods. First results and the evaluation show that the combination of different methods leads to the best results. Similarity based on users and assigned tags of an online bookmarking system may complement co-citation and bibliographic coupling. By some target authors more important similar authors were found in CiteULike than in Scopus or WoS. The interviewees approved this assumption with the graph relevance ranking. They and other researchers in former studies confirm that there is a need for author recommendation: Many physicists don't work by oneself, but in project teams. The cooperation with colleagues of the same research field is essential. A recommender system could support them. Our paper shows a new approach to recommend relevant collaboration colleagues for scientific authors. The challenge will be to combine the different similarity approaches. One method is the simple summation of the cosine values. The cumulated cosine values provide better ranking results for some relevant researchers, but they are not satisfactory. Further investigations will be made in a weighted algorithm which considers the results of all four cluster. The relations between user- and tag-based similarity in a bookmarking system should also be considered and tested, e.g. with a graph based approach like FolkRank [27] or expertise analysis (SPEAR) [3]. Besides this the paper did not study important aspects of a running recommender system like accuracy and efficiency. Research has to be done on these fields. An issue which may as well be addressed is social network analysis and graph constructions.

## 8. ACKNOWLEDGMENTS

Tamara Heck is financed by a grant of the *Strategische Forschungsfonds* of the Heinrich-Heine-University Düsseldorf. We would like to thank Oliver Hanraths for data mining, Stefanie Hausteiner for valuable discussions and the physicists of *Forschungszentrum Jülich* for participating in the evaluation.

## 9. REFERENCES

- [1] Ahlgren, P., Jarneving, B. and Rousseau, R. 2003. Requirements for a cocitation similarity measure, with special reference to Pearson's correlation coefficient. *Journal of the American Society for Information Science and Technology*, 54(6), 550-560.
- [2] Ahn, H. J. 2008. A new similarity measure for collaborative filtering to alleviate the new user cold-starting problem. *Information Sciences*, 178, 37-51.
- [3] Au Yeung, C. M., Noll, M., Gibbins, N., Meinel, C. and Shadbolt, N. 2009. On measuring expertise in collaborative tagging systems. *Web Science Conference: Society On-Line*, 18th-20th March 2009, Athens, Greece.
- [4] Ben Jabeur, L., Tamine, L. and Boughanem, M. 2010. A social model for literature access: towards a weighted social network of authors. *Proceedings of RIAO '10 International Conference on Adaptivity, Personalization and Fusion of Heterogeneous Information*. Paris, France, 32-39.
- [5] Berkovsky, S., Kuflik, T. and Ricci F. 2007. Mediation of user models for enhanced personalization in recommender systems. *User Model User-Adap Inter*, 18, 245-286.
- [6] Bichteler, J. and Eaton, E. A. 1980. The combined use of bibliographic coupling and cocitation for document-retrieval. *Journal of the American Society for Information Science*, 31(4), 278-282.
- [7] Blazek, R. 2007. *Author-Statement Citation Analysis Applied as a Recommender System to Support Non-Domain-Expert Academic Research*. Doctoral Dissertation. Fort Lauderdale, FL: Nova Southeastern University.

- [8] Bogers, T. and van den Bosch, A. 2008. Recommending scientific articles using CiteULike. Proceedings of the 2008 ACM Conference on Recommender Systems. New York, NY, 287-290.
- [9] Boyack, K. W. and Klavans, R. 2010. Co-citation analysis, bibliographic coupling, and direct citation. Which citation approach represents the research front most accurately? *Journal of the American Society for Information Science and Technology*, 61(12), 2389-2404.
- [10] Cabanac, G. 2010. Accuracy of inter-researcher similarity measures based on topical and social clues. *Scientometrics*, 87(3), 597-620.
- [11] CACHED, F., Carneiro, V., Fernández, D. and Formoso, V. 2011. Comparison of collaborative filtering algorithms: Limitations of current techniques and proposals for scalable, high-performance recommender systems. *ACM Transactions on the Web*, 5(1), article 2.
- [12] Cai, X., Bain, M., Krzywicki, A., Wobcke, W., Kim, Y. S., Compton, P. and Mahidadia, A. 2011. Collaborative filtering for people to people recommendation in social networks. *Lecture Notes in Computer Science*, 6464, 476-485.
- [13] Cawkell, T. 2000. Methods of information retrieval using *Web of Science*. Pulmonary hypertension as a subject example. *Journal of Information Science*, 26(1), 66-70.
- [14] Cronin, B. 1984. *The Citation Process. The Role and Significance of Citations in Scientific Communication*. London, UK: Taylor Graham.
- [15] Cruz, C. C. P., Motta, C. L. R., Santoro, F. M. and Elia, M. 2009. Applying reputation mechanisms in communities of practice. A case study. *Journal of Universal Computer Science*, 15(9), 1886-1906.
- [16] Desrosiers, C. and Karypis, G. 2011. A comprehensive survey of neighborhood-based recommendation methods (pp. 197-144). In Ricci, F., Rokach, L., Shapira, B. and Kantor, P.B (Eds.), *Recommender Systems Handbook*. Springer, NY.
- [17] Egghe, L. 2010. Good properties of similarity measures and their complementarity. *Journal of the American Society for Information Science and Technology*, 61(10), 2151-2160.
- [18] Emamy, K. and Cameron, R. 2007. CiteULike. A researcher's bookmarking service. *Ariadne*, 51.
- [19] Gmur, M. 2003. Co-citation analysis and the search for invisible colleges. A methodological evaluation. *Scientometrics*, 57(1), 27-57.
- [20] Goldberg, D., Nichols, D., Oki, B. M. and Terry, D. 1992. Using collaborative filtering to weave an information tapestry. *Communications of the ACM*, 35(12), 61-70.
- [21] Hamers, L., Hemeryck, Y., Herweyers, G. and Janssen, M. 1989. Similarity measures in scientometric research: The Jaccard Index versus Salton's cosine formula. *Information Processing & Management*, 25(3), 315-318.
- [22] Haustein, S. and Siebenlist, T. 2011. Applying social bookmarking data to evaluate journal usage. *Journal of Informetrics*, 5, 446-457.
- [23] Heck, T. (2011). A comparison of different user-similarity measures as basis for research and scientific cooperation. *Information Science and Social Media International Conference August 24-26, Åbo/Turku, Finland*.
- [24] Heck, T. and Peters, I. 2010. Expert recommender systems: Establishing Communities of Practice based on social bookmarking systems. In *Proceedings of I-Know 2010, 10<sup>th</sup> International Conference on Knowledge Management and Knowledge Technologies*, 458-464.
- [25] Herlocker, J. L., Konstan, J. A., Borchers, A. and Riedl, J. 1999. An algorithmic framework for performing collaborative filtering. *Proceedings of the 22<sup>nd</sup> Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. New York, 230-237.
- [26] Herlocker, J. L., Konstan, J. A., Terveen L. G. and Riedl, J. T. 2004. Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems*, 22(1), 5-53.
- [27] Hotho, A., Jäschke, R., Schmitz, C. and Stumme, G. 2006. Information retrieval in folksonomies: Search and ranking (pp. 411-426). In Sure, Y., Domingue, J. (Eds.), *The Semantic Web: Research and Applications*, *Lecture Notes in Computer Science 4011*, Springer, Heidelberg.
- [28] Kessler, M. M. 1963. Bibliographic coupling between scientific papers. *American Documentation*, 14, 10-25.
- [29] Krohn-Grimberghe, A., Nanopoulos, A. and Schmidt-Thieme, L. 2010. A novel multidimensional framework for evaluating recommender systems. In *Proceedings of the ACM RecSys 2010 Workshop on User-Centric Evaluation of Recommender Systems and Their Interfaces (UCERSTI)*. New York, NY, ACM.
- [30] Lee, D. H. and Brusilovsky, P. 2010a. Social networks and interest similarity. The case of CiteULike. In *Proceedings of the 21<sup>st</sup> ACM Conference on Hypertext & Hypermedia*, Toronto, Canada, 151-155.
- [31] Lee, D. H. and Brusilovsky, P. 2010b. Using self-defined group activities for improving recommendations in collaborative tagging systems. In *Proceedings of the Fourth ACM Conference on Recommender Systems*. NY, 221-224.
- [32] Leydesdorff, L. 2005. Similarity measures, author cocitation analysis, and information theory. *Journal of the American Society for Information Science and Technology*, 56(7), 769-772.
- [33] Leydesdorff, L. 2008. On the normalization and visualization of author co-citation data. Salton's cosine versus the Jaccard index. *Journal of the American Society for Information Science and Technology*, 59(1), 77-85.
- [34] Li, J., Burnham, J. F., Lemley, T. and Britton, R. M. 2010. Citation analysis. Comparison of Web of Science, Scopus, SciFinder, and Google Scholar. *Journal of Electronic Resources in Medical Libraries*, 7(3), 196-217.
- [35] Liang, H., Xu, Y., Li, Y. and Nayak, R. 2008. Collaborative filtering recommender systems using tag information. *ACM International Conference on Web Intelligence and Intelligent Agent Technology*. New York, NY, 59-62.
- [36] Linde, F. and Stock, W.G. 2011. *Information Markets*. Berlin, Germany, New York, NY: De Gruyter Saur.
- [37] Luo, H., Niu, C., Shen, R. and Ullrich, C. 2008. A collaborative filtering framework based on both local user similarity and global user similarity. *Machine Learning*, 72(3), 231-245.
- [38] Marinho, L. B., Nanopoulos, A., Schmidt-Thieme, L., Jäschke, R., Hotho, A., Stumme, G. and Symeonidis, P. 2011. Social tagging recommenders systems (pp. 615-644). In Ricci, F., Rokach, L., Shapira, B. and Kantor, P.B (Eds.), *Recommender Systems Handbook*. Springer, NY.

- [39] McNee, S. M., Kapoor, N. and Konstan, J.A. 2006. Don't look stupid. Avoiding pitfalls when recommending research papers. In Proc. of the 20th anniversary Conference on Computer Supported Cooperative Work. New York, NY, ACM, 171-180.
- [40] Meho, L. I. and Rogers, Y. 2008. Citation counting, citation ranking, and h-index of human-computer interaction researchers. A comparison of Scopus and Web of Science. *Journal of the American Society for Information Science and Technology*, 59(11), 1711-1726.
- [41] Meho, L. I. and Sugimoto, C. R. 2009. Assessing the scholarly impact of information studies. A tale of two citation databases – Scopus and Web of Science. *Journal of the American Society for Information Science and Technology*, 60(12), 2499-2508.
- [42] Parra, D. and Brusilovsky, P. 2009. Collaborative filtering for social tagging systems. An Experiment with CiteULike. In Proc. of the Third ACM Conference on Recommender Systems. New York, NY, ACM, 237-240.
- [43] Peters, I. 2009. Folksonomies. Indexing and Retrieval in Web 2.0. Berlin, Germany: De Gruyter Saur.
- [44] Ramezani, M., Bergman, L., Thompson, R., Burke, R. and Mobasher, B. 2008. Selecting and applying recommendation technology. In Proc. of International Workshop on Recommendation and Collaboration, in Conjunction with 2008 International ACM Conference on Intelligent User Interfaces. Canaria, Canary Islands, Spain.
- [45] Petry, H., Tedesco, P., Vieira, V. and Salgado, A. C. 2008. ICARE. A context-sensitive expert recommendation system. In The 18<sup>th</sup> European Conference on Artificial Intelligence. Workshop on Recommender Systems. Patras, Greece, 53-58.
- [46] Reichling, T. and Wulf, V. 2009. Expert recommender systems in practice. Evaluating semi-automatic profile generation. CHI '09. In Proc. of the 27<sup>th</sup> International Conference on Human Factors in Computing Systems. New York, NY, 59-68.
- [47] Rendle, S., Marinho, L. B., Nanopoulos, A. and Schmidt-Thieme, L. 2009. Learning optimal ranking with tensor factorization for tag recommendation. In Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York, NY, 727-736.
- [48] Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P. and Riedl, J. 1994. Grouplens: An open architecture for collaborative filtering of netnews. In Proc. of CSCW'94, ACM Conference on Computer Supported Cooperative Work. New York, NY, ACM, 175-186.
- [49] Rorvig, M. 1999. Images of similarity: A visual exploration of optimal similarity metrics and scaling properties of TREC topic-document sets. *Journal of the American Society for Information Science and Technology*, 50(8), 639-651.
- [50] Said, A., Wetzker, R., Umbrath, W. and Hennig, L. 2009. A hybrid PLSA approach for warmer cold start in folksonomy recommendation. In Proc. of the International Conference on Recommender Systems. New York, NY, 87-90.
- [51] Sarwar, B., Karypis, G., Konstan, J. and Riedl, J. 2000. Analysis of recommendation algorithms for ecommerce. In Proceedings of the 2<sup>nd</sup> ACM Conference on Electronic Commerce. New York, NY, 158-167.
- [52] Schneider, J.W. and Borlund, P. 2007a. Matrix Comparison, Part 1: Motivation and important issues for measuring the resemblance between proximity measures or ordination results. *Journal of the American Society for Information Science and Technology*, 58(11), 1586-1595.
- [53] Schneider, J. W. and Borlund, P. 2007b. Matrix Comparison, Part 2: Measuring the resemblance between proximity measures or ordination results by use of the Mantel and Procrustes statistics. *Journal of the American Society for Information Science and Technology*, 58(11), 1596-1609.
- [54] Shepitsen, A., Gemmell, J., Mobasher, B. and Burke, R. 2008. Personalized recommendation in social tagging systems using hierarchical clustering. In Proc. of the 2008 ACM Conference on Recommender Systems. NY, 259-266.
- [55] Small, H. 1973. Cocitation in scientific literature. New measure of relationship between 2 documents. *Journal of the American Society for Information Science*, 24(4), 265-269.
- [56] Stock, W. G. 1999. Web of Science. Ein Netz wissenschaftlicher Informationen – gesponnen aus Fußnoten [Web of Science. A web of scientific information – cocooned from footnotes]. *Password*, no. 7+8, 21-25.
- [57] Van Eck, N. J. and Waltman, L. 2008. Appropriate similarity measures for author co-citation analysis. *Journal of the American Society for Information Science and Technology*, 59(10), 1653-1661.
- [58] Van Eck, N. J. and Waltman, L. 2009. How to normalize cooccurrence data? An analysis of some well-known similarity measures. *Journal of the American Society for Information Science and Technology*, 60(8), 1635-1651.
- [59] Wang, J., de Vries, A. P. and Reinders, M. J. T. 2006. Unifying userbased and itembased collaborative filtering approaches by similarity fusion. In Proc. of the 29<sup>th</sup> Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. NY, 501-508.
- [60] White, H. D., & Griffith (1981). Author cocitation. A literature measure of intellectual structure. *Journal of the American Society for Information Science*, 32(3), 163-171.
- [61] White, H. D. and Griffith, B. 1982. Authors as markers of intellectual space. Co-citation in studies of science, technology and society. *Journal of Documentation*, 38(4), 255-272.
- [62] Yukawa, T., Kasahara, K., Kita, T. & Kato, T. (2001). An expert recommendation system using concept-based relevance discernment. In Proceedings of ICTAI '01, 13th IEEE International Conference on Tools with Artificial Intelligence (pp. 257-264. Dallas, Texas, 257-264.
- [63] Zanardi, V. and Capra, L. 2008. Social ranking: Uncovering relevant content using tag-based recommender systems. Proceedings of the 2008 ACM Conference on Recommender Systems. New York, NY, 51-58.
- [64] Zhao, D. and Strotmann, A. 2007. All-author vs. first author co-citation analysis of the Information Science field using Scopus. Proceedings of the 70<sup>th</sup> Annual Meeting of the American Society for Information Science and Technology, 44(1). Milwaukee, Wisconsin, 1-12.
- [65] Zhao, D. and Strotmann, A. 2011. Counting first, last, or all authors in citation analysis. Collaborative stem cell research field. *Journal of the American Society for Information Science and Technology*, 62(4), 654-67.