Chapter 11

Recommendation for Social Networking in Academia

Tamara Heck

Abstract

Purpose – As researchers need partners to collaborate with, this study aims to provide author recommendation for academic researchers for potential collaboration, conference planning, and compilation of scientific working groups with the help of social information. Hereby the paper analyzes and compares different similarity metrics in information and computer science. AU :1

Methodology/approach – The study uses data from the multidiscipline information services Web of Science and Scopus as well as the social bookmarking service CiteULike to measure author similarity and recommend researchers to unique target researchers. The similarity approach is based on author co-citation, bibliographic coupling of authors and collaborative filtering methods. The developed clusters and graphs are then evaluated by these target researchers.

Findings – The analysis shows, for example, that different methods for social recommendation complement each other and that the researchers evaluated user- and tag-based data from a social bookmarking system positively.

Research limitations/implications – The present study, providing author recommendation for six target physicists, is supposed to be a starting point for further approaches on social academic author recommendation.

Social Information Research
Library and Information Science, 239–265

ISSN: ol;asi.21/doi:10.1002/asi.21495

Practical implications – The paper investigates in recommendation methods and similarity algorithm models as basis for an implementation of a social recommendation system for researchers in academics and knowledge-intensive organizations.

Originality/value of paper – The comparison of different similarity measurements and the user evaluation provide new insights into the construction of social data mining and the investigation of personalized recommendation.

Keywords: Social bookmarking; social recommendation; social networking; academic knowledge management

11.1. Scientific Collaboration

Collaborations with scientific colleagues are essential for most researchers, forming an important aspect of their career. One of the most visible acts of collaboration is coauthorship, where two or more researchers contribute to a publication (such as a journal article or book chapter). Other than coauthorship, there are several other situations that call for collaboration:

- Assembly of a (formal) working group in a large university department or company;
- Gathering of researchers in order to prepare a project proposal for a research grant;
- Establishment of a more informal Community of Practice (Wenger, 1998), in an institution or together with members of diverse institutions and companies;
- Search for contributors to a conference, a congress or a workshop;
- Search for contributors to a handbook or a specialised journal issue.

Studies show (e.g., Cronin, Shaw, & La Barre, 2003; Heck & Peters, 2010; Luukkonen, Persson, & Sivertsen, 1992) that there is a need for researchers, in academic circles as well as other knowledge-intensive organizations, to find qualified collaborators. In the ever-expanding World Wide Web it appears an easy task for a scientist to collect useful information about his or her potential partners. Social web sites in particular offer new (and more varied) information about scientists, potentially improving the researcher's decision-making. Nevertheless, it is still an extremely difficult task to find serious collaborators – not just any who happen to be available but the "right" ones, proven experts with solid reputations (Cruz, Motta, Santoro, & Elia, 2009). A researcher's reputation grows with the number of publications he contributes to peer-reviewed journals, and with the frequency at which those publications are cited by others (Cronin, 1984). If we wish to know whether two certain authors have a similar reputation and might thus be good collaborators, we can measure their similarity on the basis of author co-citation (ACC). This method, used in scientometrics (Leydesdorff, 2005; White & Griffith, 1981; White & Griffith, 1982), refers to a situation where two or more authors are co-cited by another researcher in one of his works. This may mean that authors co-cited in this way share similar interests and might therefore establish a fruitful partnership. Another scientometric method, bibliographic coupling (BC) (Kessler, 1963), suggests that two authors are similar to each other if they cite some of the same references in their works. Similarity in both cases means that these scientists might be interested in the same research areas and topics due to the overlap in the citations of their works or in their usage of the same references. Similarity measurement is then used as an indicator for high collaboration potential. Based on the social information available on a researcher – that is, the publications and authors he cites in his references as well as the authors who cite him and those who are cited in the same publication as himself – we can build a network of similar authors with several weak or strong connections to one another. The general assumption is that researchers whose interests and approaches are similar would automatically tend to collaborate with each other – however, many scientists actually look for the opposite, choosing dissimilar partners who might ideally complement their own skills. When speaking of the "right" candidates for collaboration, we must therefore keep in mind that this can mean people who share the same skills and contribute similar know-how to a research project, or people with complementary skills covering different approaches to the same goal.

The following approach tries to help researchers in finding the right people and to recommend scientist for potential collaboration. It is assumed that combined social information leads to better recommendations: To prove this assumption different methods and datasets are analyzed and compared to each other. Hereby social information about a researcher and his potential partners is collected not only in multidisciplinary information services such as Web of Science (WoS) and Scopus, based on ACC analysis and BC of authors (Li, Burnham, Lemley, & Britton, 2010; Meho, & Rogers, 2008; Meho, & Sugimoto, 2009). But social information is also collected in a social bookmarking service, using collaborative filtering (CF) methods to measure researcher similarity. The approach tries to answer two basic research questions which are at the core of how to further develop an expert recommendation system:

1. Can a relevant author network be proposed to a target scientist via social information in a social bookmarking service?

2. Are the results different from the results based on ACC and BC, and do they complement each other?

11.2. Social information and Social Networks

A critical aspect for choosing the right collaborator is the researcher's reliance on the information they must base their decision on. Using co-citation and reference data to build author networks takes into account the researchers' perspective: Who do they cite, by whom are they cited and who is co-cited alongside them? Both methods consider the social relations between researchers, that is, their relations to each other based on their published works, and are used to show scientific networks and the distribution of scientific knowledge. In this approach ACC and BC are not used to show relations in an overall scientific network within a single scientific discipline. The methods are used to measure similarity between a single target scientist and other researcher to recommend potential collaborators to this target scientist. The important question is: Which data should be collected, and which methods should be used to make good researcher recommendations? It seems insufficient to only consider common references or co-citations, since for many scientists there simply is not enough data available. This is particularly the case for young researchers who have only just started their academic career and have not built a scientific reputation for themselves yet. Blazek (2007) calls these "domain novice researchers," that is, academics who enter a new domain and wish to use a collection of academic documents. They face the cold-start problem: Citation analysis can hardly be applied to novice researchers as long as they have little or no references and citations to their name. Furthermore, there is a time lag when measuring citations and ACCs because an author's article will not be cited until several months after its publication, with differences in time span depending on the scientific discipline. This means that a researcher with a recently published article who might be a good collaboration partner simply will not be considered in ACC measurement.

To overcome the limitations of data scarcity, one can use further social information from the web in order to make better decisions about collaboration partners. On so-called social web sites, the users themselves contribute to the system's data, getting involved in the systems' data collection and even adding content themselves. On social networking services like Facebook and LinkedIn, social microblogging services like Twitter, and social bookmarking services like Del.icio.us, Bibsonomy, CiteULike, Connotea, and Mendeley, the users add personal information, short messages, bookmarks for web sites and academic literature, and tags. In general the amount of overall information grows with the amount of users in the system. Using this information has advantages vis-à-vis to that which is found in multidisciplinary information services like WoS and Scopus, which are often used for ACC and BC:

1. There is a greater variety of data available;
2. The users' perspective is taken into consideration.

The first aspect might have some shortages – these will be discussed in the paragraph about the limitations of data collection – but it is the second aspect that will turn out to be the more important one. ACC only takes into account the perspective of a third researcher citing two other authors who might be similar (and therefore potential collaboration partners). BC only considers the perspectives of two authors, marked by their choice of references. Social information from web services also considers the users' perspective, as it takes into account the content they themselves have contributed to the services. We thus have access to the perspectives of a large group of people: if, for instance, many users have added works by two certain authors to their bookmarking list in CiteULike, this would be a further hint that these authors share similar research interests – assuming, of course, that the users are interested in a specific topic and also bookmark the relevant literature. This social relation is similar to ACC: You might say the method of ACC analysis is assigned to a social bookmarking service, with the difference that we don't necessarily have scientific authors, but users of the social web. There are "pure" readers, that is, readers who read, but do not publish and hence do not cite. The difference lies in the new dataset: Not only do we now have the opinion of more people, but this new user perspective may also expand a researcher's known social network and uncover new relations between researchers. Here the focus lies mainly on young researchers once more. Senior scientists have already established their community network in which they connect to colleagues and collaboration partners. Junior researchers, on the other hand, first have to build up theirs. Helping them find the right scientific community is just one benefit of using social information. Seen from another angle, senior researchers will benefit from new ways of generating social information networks by finding new ways to identify potential collaborators.

Figure 11.1 shows the various social information-based relations between two researchers: On the right-hand side, there is data from information services that deal with scientific publications, references, and citations. On the left-hand side, we have information from a social bookmarking system for academic literature. Here users can bookmark an author's publications and assign tags – that is, keywords describing the source – to these bookmarks. Hence, there are direct relations connecting users, bookmarks and tags to each other. Additionally, since the authors are directly related to their publications, the works that are bookmarked in an information service

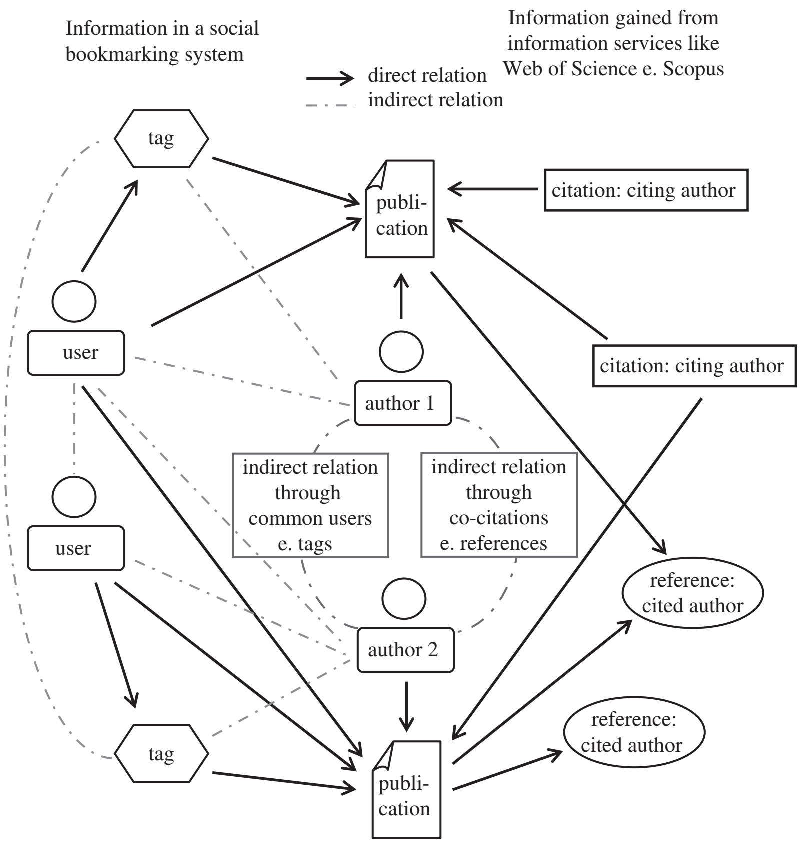

Figure 11.1: Direct and indirect relations for researchers based on social information in information services and bookmarking Systems.

establish an indirect connection between their respective authors and the tags assigned to their publications (as well as the users who bookmarked them). Ben Jabeur, Tamine, and Boughanem (2010) designate coauthorship and friendship as two other social relations between researchers. All these direct and indirect relations based on social information are used for social network analysis (Wasserman, Faust, & Iacobucci, 1994) and are further employed in the following approach to recommending potential collaboration partners to a target researcher.

11.3. Expert Recommendation

Recommendation (or recommender) systems ~~(RS)~~ have become an AU :2
important tool for overcoming information overload on the web and

advising people in selecting the right documents, products, or even people to satisfy their information needs. For a researcher in need of collaboration partners, a recommender system could point out relevant individuals on the basis of various characteristics. Nowadays, recommender systems use different methods and algorithms for different items, for example, products, movies, music, articles, etc., the goal being personalized recommendation (Berkovsky, Kuflik, & Ricci, 2007) of items unknown, yet relevant, to the target user. The first question is where to find the best resources for user *a* and how to rank them according to their relevance (Desrosiers & Karypis, 2011). Two different approaches are often used (from among other distinctive recommender methods and hybridizations): The content-based approach, which tries to identify similarities between items rated positively by user *a* on the basis of their content, and the CF approach, which not only considers the ratings of user *a*, but also those of other users (a.o. Goldberg, Nichols, Oki, & Terry, 1992; Herlocker, Konstan, Borchers, & Riedl, 1999; Parra & Brusilovsky, 2009; Resnick, Iacovou, Suchak, Bergstrom, & Riedl, 1994). One advantage of CF compared to the content-based method is that recommendations rely not only on the item's content, which may be insufficient for quality indication, but also on the evaluation of other users. When using CF to recommend potential collaborators to a target researcher, taking into account the users' perspective should yield new and more appropriate results.

Recommender systems work by assigning user ratings to the items, a method called user-item response (Desrosiers & Karypis, 2011): These ratings can be scalar (e.g., 1–5 stars), binary (like/dislike) or unary. The latter means that while a user has not rated an item, his purchase of or access to the item is interpreted as a positive response. This user-item response can be used for recommendations in social tagging systems, such as social bookmarking service (Marinho et al., 2011). Social tagging systems have a folksonomy structure with user-resource-tag relations, which forms the basis for CF. These systems provide for recommendations of not only items, but also of tags and users – this is the basis for academic author recommendation. On the basis of CF, potential collaboration partners are recommended to target academic researchers (Heck, Hanraths, & Stock, 2011; Heck, Peters, & Stock, 2011). One question that arises at this point is whether CF recommends different results in a social tagging system than the more established scientometric measurements of ACC and BC. In general, these measurements are not explicitly used for recommendation, but rather for author and scientific network analysis (Small, 1973).

Recommender systems can be constructed in many different ways, for example, by choosing the appropriate algorithm for personal recommendation in particular (Shepitsen, Gemmell, Mobasher, & Burke, 2008), defining user interactions and user models (Ramezani, Bergman, Thompson, Burke, & Mobasher, 2008), handling criteria like ~~RS~~ accuracy, efficiency, and stability (Desrosiers & Karypis, 2011) or focusing on ideal recommender system learning models (Rendle, Marinho, Nanopoulos, & Schmidt-Thieme, 2009). With the advent of bookmarking and collaboration services on the web, several algorithms and hybridizations have been developed (Hotho, Jäschke, Schmitz, & Stumme, 2006). They may differ in their specific combination of the relations between users, items, and tags as well as their weighting. Similarity fusion (Wang, de Vries, & Reinders, 2006), for example, combines user- and item-based filtering (which are subcategories of CF) and additionally uses ratings of similar items by similar users. Cacheda, Carneiro, Fernández, and Formoso (2011) provide an overview of the different algorithms comparing the performances of the methods, and further propose a new algorithm that takes into account the users' positive or negative ratings of the items. Bogers and van den Bosch (2008) compare three different CF algorithms, two of them item-based and one user-based; the latter outperformed the others. But the most evident problem seems to be the cold-start scenario, in which new items cannot be recommended at first (Ahn, 2008). Said, Wetzker, Umbrath, and Hennig (2009) also deal with this problem and investigate the performance of different algorithms within a certain time span. Their result: Adding tag similarity measures can improve the quality of item recommendation because tags offer more detailed information about items. Hotho et al. (2006) propose the FolkRank, a graph-based approach similar to the idea of the PageRank, which can be applied in a folksonomy structure such as that of a bookmarking service. Here users, tags, and resources are the nodes in the graph and the relations between them become the weighted edges, taking into account weight spreading in the manner of the PageRank.

There are several studies that investigate expert recommendation, mainly for commercial enterprises (Cai et al., 2011; Petry, Tedesco, Vieira, & Salgado, 2008; Reichling & Wulf, 2009). Petry et al. (2008), for instance, have developed the expert recommendation system ICARE, meant to recommend experts within an organization. In this system the primary spotlight is not directed onto an author's publications and citations, but rather on their organizational level, availability, and reputation, among other aspects. Following a field study and interviews with employees, Reichling and Wulf (2009) explore the options of a recommender system to support their knowledge management. In this system experts are defined via their collection of written documents, which have been analyzed automatically. The authors also used a post-integrated user profile with information about each individual's background and job description. The use of user profiles in bookmarking services could be helpful in terms of providing further information about a user's interests, thus improving the effectiveness of user recommendation. However, this approach might raise serious concerns about privacy and data security on the web.

In addition to user recommendation for commercial enterprises, several other approaches concentrate on Web 2.0 users and academics. Au Yeung, Noll, Gibbins, Meinel, and Shadbolt (2009), discussing the nonacademic bookmarking system Del.icio.us, define an expert user as someone who has deposited high-quality documents in their bookmark collection (many users who have high levels of expertise fulfill this criterion), and who tends to recognize useful documents well in advance of others (as seen in the timestamps on users' bookmarks). In contrast to the following approach, the "high-quality documents" in this experiment are the publications of the researcher to whom collaboration partners are meant to be recommended. Hence, it is vital for the purposes of recommendation that users bookmark at least one of the author's publications. Heck and Peters (2010) propose using social bookmarking systems for scientific literature, such as BibSonomy, CiteULike, and Connotea, to recommend researchers unknown to the user but who share the same interests and would thus be suitable partners for building a community of practice (Wenger, 1998). Users are recommended to each other when they have either bookmarks or tags in common. One precondition is that the researcher meant to be provided with relevant expert recommendations be active in the social bookmarking system, and store his relevant literature in his Internet library. Cabanac's (2010) approach is similar to this method but concentrates only on user similarity networks and relevant articles, not on the recommendation of unknown researchers. Cabanac (2010) uses the concepts of Ben Jabeur et al. (2010) to build a social network for recommending relevant literature. Additionally, social clues such as the connectivity of researchers and opportunities to meet in person, for example, at scientific conferences, are taken into account. It is assumed that these social clues lead to an improved performance of the recommendation system. Similarly Nocera and Ursino (2011) try to recommend similar users and resources, and thereby set their focus on "social folksonomy," that is, using information about user friendships and semantic information of tags.

Another important aspect for recommender systems is their evaluation. Recommender systems should not only prove accurate and efficient, but must also detect the users' respective needs in order to be of use to them (Herlocker, Konstan, Terveen, & Riedl, 2004). Several studies incorporate user evaluation in their investigation of the evaluation of model-based recommender systems (Krohn-Grimberghe, Nanopoulos, & Schmidt-Thieme, 2010). McNee, Kapoor, and Konstan (2006) show the pitfalls of recommender systems in order to foster user acceptance and promote further usage of recommender systems as knowledge management tools. The following example of a recommender system is also evaluated by the recipients of its collaboration recommendations.

11.4. An Example for Constructing Researcher Networks

The following approach attempts to recommend researchers to each other (Heck et al., 2011). Here it is assumed that combined social information – found in multidisciplinary information services such as WoS and Scopus, and in social bookmarking systems such as CiteULike – creates better recommendations.

11.4.1. Collaborative Filtering in CiteULike

CiteULike has become very popular (Linde & Stock, 2011, p. 268): Unlike bookmarking systems such as Del.icio.us, it focuses on the management of academic literature. The basis for social recommendation is the service's folksonomy structure. A folksonomy (Marinho et al., 2011; Peters, 2009) is defined as a tuple F: = (U, T, R, Y), where U, T and R are finite sets with the elements of "user name," "tag," and "resource" and *Y* is a ternary relation between them: $Y \subseteq U \times T \times R$ with the elements being called "tag actions" or "assignments." To use this information for recommending authors to each other, we expand the folksonomy to F_E: = (U, T, R, A, Y), where A is added as the finite set with the element "authors" and $Y \subseteq U \times T \times R \times A$ is their relation.

In our experimental comparison, we want to cluster scientific authors with similar research interests together. Results for author similarity based on ACC and BC are compared to results based on CF, using data from CiteULike. We are not interested in the networks and relations of the CiteULike users themselves but only in their bookmarks, that is, the bookmarked publications of our target scientist, and the tags assigned to those bookmarks. We define U_a (respectively U_b) for all users who have bookmarked at least one article by our target author *a* (respectively author *b*), R_a for all resources – in this case scientific articles – that have at least two tags in common with one bookmarked article by our target scientist *a* and T_a (respectively T_b) for all tags that are assigned to at least one bookmarked article by our target author *a* (respectively author *b*). We have two options for setting our database to author similarity measurement:

1. Searching for all users $u \in U$ who have at least one article by the target author *a* in their bookmark list: $U_a = \{u \in U | \exists\ r \in R, a \in A, (u, r, a) \in Y\}$.
2. Searching for all resources that have at least two tags in common with one bookmarked article by our target author *a*: $R_a = \{r \in R | t \in T_a, (r, t) \in Y\}$.

The disadvantage of the first method, for us, lies in the small number of users. Relying only on the users may not be enough to identify similarity (Lee & Brusilovky, 2010a). For this reason, we use the second method: Resources (here: scientific papers) can be deemed similar if they have been assigned some shared tags. From here, we assume that the authors of these documents are also similar. Tags point to topical relations, that is, authors connected via such relations regarding their research fields can be potential collaboration partners. Additionally, the more tags are shared by two documents, the more similar they are. In some cases, our target authors' articles were labelled with very general tags such as "nanotube" and "spectroscopy," so we decided to determine a minimum amount of unique tags that a document must have in common with a target author's document[1]

$$\mathrm{R}_a := \{\mathrm{r} \in \mathrm{R} | \mathrm{t} \in \mathrm{T}_a, (\mathrm{r}, \mathrm{t}) \in \mathrm{Y} \;\text{ with }\; |\mathrm{T}_a| 2\} \tag{1}$$

To measure similarity we use the cosine coefficient, one of the most common similarity measurements in Information Science besides Dice and Jaccard-Sneath (Ahlgren, Jarneving, & Rousseau, 2003; Ahn, 2008; Lee & Brusilovky, 2010b; Leydesdorff, 2008; Van Eck & Waltman, 2008). Our own experiences (Heck, 2011) as well as results from the literature (Rorvig, 1999) show that the cosine works extremely well.

In our dataset R_a, we measure author similarity in two different ways: (A) Based on shared tags $t \in \mathrm{T}$ assigned to the resources of authors a and b; (B) Based on shared users $u \in \mathrm{U}$ who have bookmarked the resources. The similarity between authors a and b is measured:

$$\text{A) } sim(a, b) := \frac{|\mathrm{T}_a \cap \mathrm{T}_b|}{\sqrt{|\mathrm{T}_a| * |\mathrm{T}_b|}} \qquad \text{B) } sim(a, b) := \frac{|\mathrm{U}_a \cap \mathrm{U}_b|}{\sqrt{|\mathrm{U}_a| * |\mathrm{U}_b|}} \tag{2}$$

Note that the latter method leads to different results than the proposed first method for database modelling does. If we were to apply the first method, we would identify all users who have at least one document by target author a in their bookmark list. With the second method we would be provided with a list of all users who have at least one document similar to one of the target author a's articles in their bookmark list, that is, it would be possible for users who have bookmarked an actual document by a might be left out. Since we want to apply one unique dataset for author similarity measurement, we do not merge both methods but instead measure tag-based and user-based similarity in the dataset described above. Nevertheless,

1. Bars denote the cardinality of the sets.

the first method was chosen where no tags were available (see ~~results~~ paragraph).

11.4.2. Author Co-Citation and Bibliographic Coupling for Recommendation

There are four relations between two authors with regard to their publications, references, and citations respectively: coauthorship, direct citation, BC of authors, and ACC. The first two relationships are not considered in this example, for here it is certain that one author knows the other: of course one knows who one's coauthors are and, we can assume, the author one has cited. Our goal is to recommend unknown scientists (BC) (Kessler, 1963) and co-citations (Leydesdorff, 2005; Marshakova, 1973; Schneider & Borlund, 2007a; Schneider & Borlund 2007b; Small, 1973; White & Griffith, 1981; White & Griffith, 1982) are undirected weighted linkages between two scientific papers, calculated via their fraction of shared references (BC) or co-citations. We then aggregate the data from the document level to the author level.

BC of authors means that two authors *a* and *b* are linked if they cite the same authors in their papers. We use WoS to mine data about BC, since this service allows searches for "related records," where relations are calculated via the number of references a certain document has in common with the source article (Cawkell, 2000; Stock, 1999). We now have the finite sets D, A, and Ref, featuring the elements "documents," "authors," and "references." Consider D and Ref as having similar elements, that is, the authors' articles. Our assumption is this: Two authors with one document each that share a high number of identical references are more similar than two authors with a large amount of shared references across many documents – the number of shared references per document being the vital quantity. For example, Let author *a* have six references in common with authors *b* and *c*. These six shared references are found in two unique documents by author *a* and author *b* respectively, but for author *c* they are distributed across six individual documents. In this scenario authors *a* and *b* are more similar than authors *a* and *c*, because the reference lists of *a* and *b*'s documents are more similar.

Our assumption leads to the following dataset model for BC, in which we take all related documents that share at least *n* references with any of the publications by target author *a*, where *n* may vary from case to case:

$$\mathrm{D}_{BC} := \{\mathrm{d} \in \mathrm{D} |\ |\mathrm{Ref}_{dj} \cap \mathrm{Ref}_{da}| \geq n, n \in \mathbb{N} \tag{3}$$

where Ref_{dj} designates the number of references in one document $d \in \mathrm{D}$ by author $j \in \mathrm{A}$ and Ref_{da} the number of references in one document $d \in \mathrm{D}$ by target author $a \in$ A. Authors of multiple documents in the dataset are summarised; this automatically generated list of the authors of the related documents in D_{BC} is cut off at $m \in \mathbb{N}$ unique authors ($m > 30$) because their publications and references for measuring BC must be analyzed manually in WoS. We measure the similarity of these related authors via the cosine coefficient and divide the amount of references shared by two authors by the product of the references of both authors (see Eq. (2)).

In ACC Leydesdorff, 2005; Schneider & Borlund, 2007a; Schneider & Borlund, 2007b), two authors *a* and *b* are linked if they are cited in the same documents. We cannot use WoS to mine ACC data because only the first author of any cited document is listed in the reference section of its bibliographic entries, whereas we require a complete list of all authors (Zhao & Strotmann, 2011). Therefore we will mine this data from Scopus, for here we find more than one author of the cited literature. We perform an inclusive all-ACC, in which two authors are considered co-cited when another author cites a paper they coauthored (Persson, 2001; Zhao & Strotmann, 2007). We have the finite sets D, A, and C with the elements "documents," "authors," and "citations," where the sets D and C share similar elements, that is, the authors' articles. The dataset contains all documents that cite at least one of the target author's articles in Scopus:

$$D_{ACC} := \{\mathrm{d} \in \mathrm{D} | \exists\, \mathrm{c} \in \mathrm{C}_{da}\} \tag{4}$$

where C_{da} is the set of cited documents by target author *a*. The list of potential similar authors is cut off at $m \in \mathbb{N}$ unique authors ($m > 30$), since their publications for ACC have to be analyzed manually in Scopus. For similarity measurement using the cosine coefficient we divide the number of documents co-citing two authors by the product of both authors' citations (see Eq. (2)). Regarding the results of research literature, the best performance in terms of representing research activities is achieved by both methods (BC and ACC) in combination (Boyack & Klavans, 2010; Gmur, 2003). Applying the proposed four mined datasets and similarity approaches, we are able to assemble four different sets of potential similar authors, which we call clusters. One cluster is based on BC in WoS, one is based on ACC in Scopus, one on CF of shared users in CiteULike, and the final cluster is based on CF of shared tags in CiteULike. We can now analyze those authors who, according to the cosine coefficient, are most similar to our target author, and evaluate the results. Also on the basis of the mined datasets we can measure the similarity between all authors of a cluster. These results are shown in visualizations, which we call graphs. Therefore, a visualized graph exists for each cluster and will likewise be evaluated.

11.5. Limitations of Data Collection

Various problems arise while filtering information in the three information services. We will briefly discuss these problems because recommendation results highly depend upon the source dataset. In Scopus, we detected differences in the metadata: One and the same article may appear in several different ways, that is, title and authors may be listed completely in one reference list but incompletely in the references of another article. In our case, several coauthors in the dataset went unmentioned and could not be considered for co-citation. The completeness of coauthorship varies considerably: In a random sample, with the co-citation dataset being adjusted via data from the Scopus web site, 5 of 14 authors have complete coverage, 3 have coverage between 70% and 90%, 5 between 55% and 70%, and 1 author is only covered to about 33%. Information services also face the problem of homonymy regarding author names. In CiteULike users also sometimes misspell author names, which mistakes were corrected for the purposes of our dataset. The ID n° for an author in Scopus is useful for identification, but it may fail when two or more authors with the same name are allocated to the same research field and change their workplace several times. In WoS there is no author ID, making it more difficult to distinguish individuals. Therefore, we check the filtered author's document list and (if necessary) correct it on the basis of the articles' subject area.

11.6. Experimental Project

Cooperating with six physicists[2], we built individual clusters for all of the six target academic authors (35–50 years old). These list authors supposed to be similar to the target authors. We limit the source for the dataset modelling to the authors' publications between 2006 and 2011 in order to make recommendations based on the physicists' actual research interests. To summarise: each scientist received the following four clusters: (1) based on ACC in Scopus, (2) based on BC in WoS, (3) based on CF of shared users in CiteULike (CULU) and (4) based on CF of shared tags in CiteULike (CULT). Using the cosine coefficient, we are also able to produce graphs for all clusters to show the similarity values between the authors (Figures 11.2–11.5). We used the Gephi[3] software for cluster visualization. The size of the nodes (= author names) depends on either the number of citations in

2. These physicists are researchers from the *Forschungszentrum Jülich*, Germany.
3. http://gephi.org/

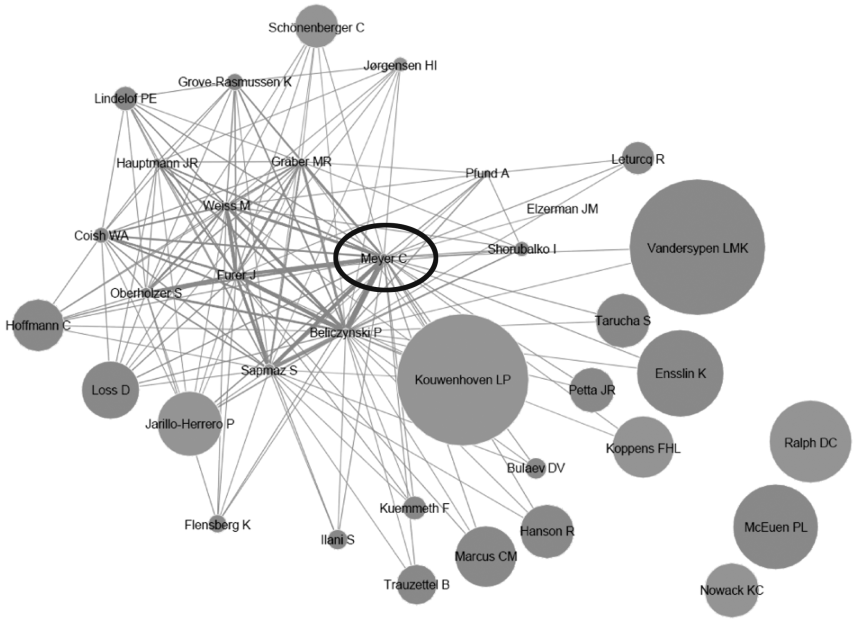

Figure 11.2: ACC Graph. Circle = Target Author 1. Cosine Threshold 0.06.

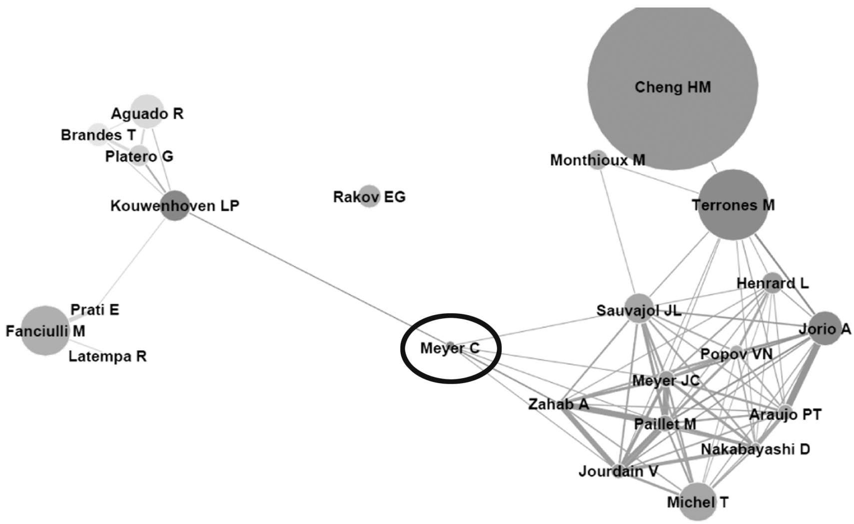

Figure 11.3: BC graph. Circle = Target Author 1. Cosine Threshold 0.06.

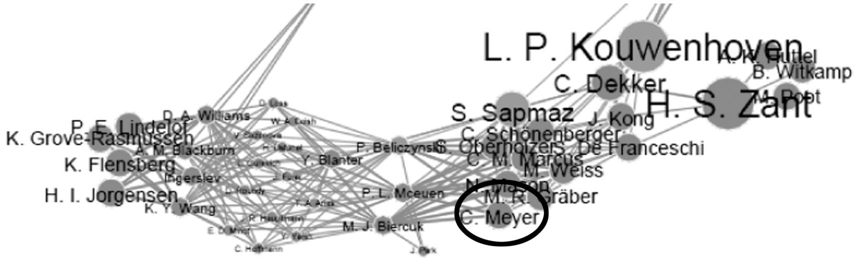

Figure 11.4: Extract of a CULU Graph. Circle = Target Author 1. Cosine Interval 0.99–~~0.4~~ AU :3

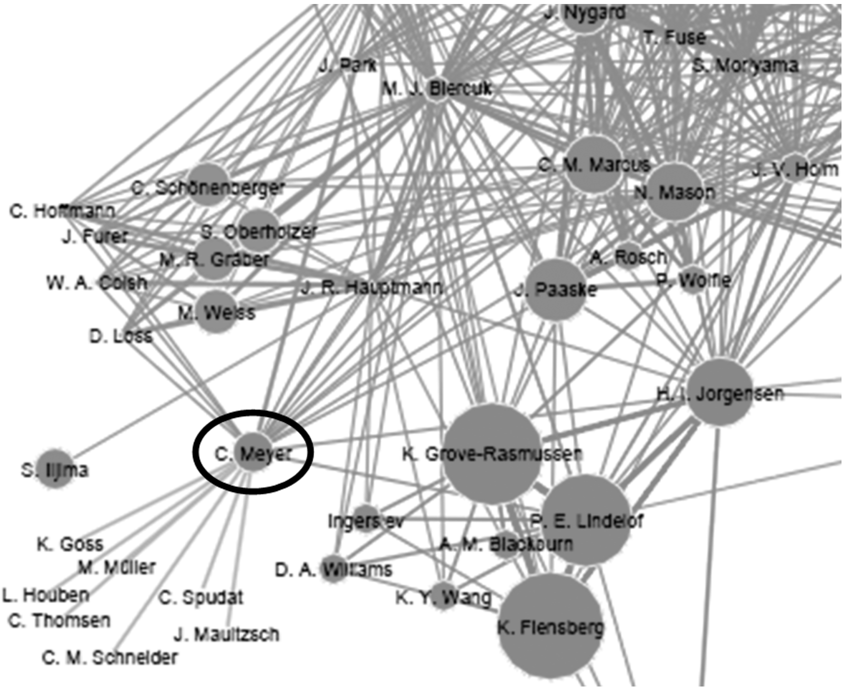

Figure 11.5: Extract of a CULT Graph. Circle = Target Author 1. Cosine Interval 0.99–0.45.

Scopus, references in WoS, or users or tags in CiteULike, and the edges are sized according to the cosine weight. Note that the CiteULike graphs are much larger, since we cut off the list of related authors in WoS and Scopus. To get a 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, as this would have distorted the results.

While modelling the datasets, we found that one of the six authors (author 6) didn't have any users who bookmarked his articles in CiteULike. Some articles were found but they had been adjusted to the system by the CiteULike operators themselves, so the CiteULike clusters couldn't be modelled for this scientist. One researcher's articles (author 5) were bookmarked, but not tagged. In this case we searched for all users who had bookmarked his articles (instead of all resources that have two tags in common with his articles). In all four clusters we ranked similar authors via the cosine coefficient. In general, it can be seen that the cosine coefficient for BC is very low compared to that for ACC as well as similarity measurements in CiteULike. This is because some authors have a lot of references, which minimizes similarity. Additionally, similarity is comparatively high for measurements in CiteULike because the number of users and assigned tags related to the target authors' publications was relatively low.

11.7. Evaluation

To prove our experimental results we let our target physicists evaluate both the clusters and the graphs. The evaluation is divided into three parts. Part 1 consists of a semi-structured interview featuring questions about the scientist's research behavior and his purchases of relevant literature as well as his working behavior – that is, does he work in teams, and if so, with whom does he cooperate? The answers paint a picture of the scientists' work and help us to estimate the evaluation results. In the second part the target author has to rank the proposed similar authors according to their relevance. The top 10 authors according to all four measurements are then listed in alphabetical order (coauthors being eliminated), and the interviewee must answer the following questions:

1. Do you know the recommended author?
2. If so, have you ever collaborated with him/her before?
3. Do you think the author's research is similar to yours?
4. How important are the known authors for your current research (rating from not important at all (1) to very important (10))?
5. With whom would you collaborate in a research project?
6. What are your reasons for collaboration or non-collaboration?
7. Are you missing any author who is important for your current research?

In Part 3 our author has to evaluate the cluster graphs (rating from 1 to 10) according to their distribution of authors and the generated groups. Here the questions are:

1. In your individual valuation, does the author distribution accurately reflect the reality of author collaboration in the research community?
2. Are the author communities clustered the right way?
3. Would this graph recommending similar authors help you, for example, to organize a workshop or to find collaboration partners?
4. Relating to question 3: How relevant are the shown graphs for you (rating from not relevant (1) to highly relevant (10))?

We will briefly summarize the most interesting answers for Part 1: As confirmed in our earlier studies (Heck & Peters, 2010), most of the physicists work in research teams (in groups generally no larger than five people). Regular meetings are important – although difficult, if international partners are involved. Novice researchers often meet new potential collaborators at meetings, for example, scientific conferences and workshops, or get introduced to them via senior colleagues. But it is more difficult for a novice researcher to find new relevant collaborators as they haven't established their social scientific network yet. The researchers' choice of possible collaborators highly depends on their research interests: There must be a high thematic overlap. On the other hand, an overlap that is too high could also be disadvantageous. Some authors who designated another author in a cluster as important stated (Part 2, question 6) that they wouldn't cooperate with him because he does exactly the same research, that is, they regard him as a competitor rather than a collaborator. The other reason given against collaboration was insufficient thematic overlap. Our interviewees regard collaborations with international institutes as desirable, and they tend to meet new colleagues at conferences and scientific workshops.

Part 2 of the evaluation is concerned with the similar author ranking. We analyzed all authors with at least a rating of 5 who were deemed important by an interviewee, as well as all important authors added by the interviewee which were not on any cluster's Top 10 list. In general, our target authors name up to 30 people they regard as important for their recent scientific work. Figure 11.6 shows the coverage of these important authors across the first 20 ranks, based on the cosine coefficient (consider that author 6 didn't have any publication bookmarked in CiteULike). For example, Target author 1 deems 25% of the 20 most similar ACC cluster to be important. In the BC cluster this number is 15%, in the user-based CiteULike (CULU) cluster it is 30% and in the tag-based CiteULike (CULT) cluster it is 25%. Compared to the other target authors there are great differences. The BC and ACC clusters can be said to provide the best

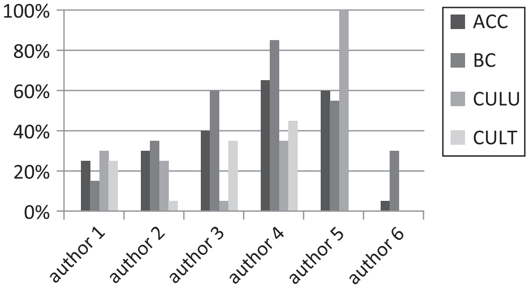

Figure 11.6: Coverage of important authors in the recommendation of the top 20 authors.

results except for authors 1 and 5. The CiteULike clusters fare slightly worse, but not in all cases: For author 1 the CULU provides the best coverage, and both CULU and CULT are better than BC. For author 5 the CULU shows full coverage, which means that all 20 authors top ranked by the cosine are deemed important by the target author. Here we must take into account that the bookmarked articles of author 5 had no tags assigned to them, and therefore had no CULT cluster. The great differences between some of the authors may also be a result of the interviewees' recent research activities: Some of the physicists said that their research interests had slightly changed. For this reason some similar authors who used to be important are no longer relevant. One disadvantage of our applied similarity measurement may be that it is based on past data, that is, on publications from the past five years. The authors deemed important by our interviewees are relevant for completed research projects. If our scope had been wide enough to include all important authors, past and present, the results for the clusters could have been improved.

Among the coverage data shown in Figure 11.6, it is interesting to take a look at the important authors who were only found in the CiteULike clusters: For example, 6 of the 29 important authors for target author 1 are only found in the CiteULike cluster, just like 5 of 19 important authors for target author 2. Table 11.1 shows the important authors for all target scientists. It is interesting to note that in all three services where different similarity measurements were applied – Scopus, WoS and CiteULike – the number of important authors identified is almost the same, but with little overlap. That means that in all of the datasets there are important authors missing that were present in at least one of the other datasets. Only 12 authors were found in all three datasets. This result indicates that the best way of finding important researchers for a target scientist is a combined approach.

Table 11.1: Distribution of important authors (for all target scientists) found in the three services, using author co-citation, bibliographic coupling and collaborative filtering respectively.

	ACC in Scopus	**BC in WoS**	**CF in CiteULike**
ACC in Scopus	**64**	27	24
BC in WoS	27	**67**	16
CF in CiteULike	24	16	**70**

In the third part of the evaluation, the interviewee had to evaluate the graphs. The average cluster relevance was:

- ACC: 5.08
- BC: 8.70
- ~~CF in CiteULike based on users (~~CULU~~)~~: 2.1
- ~~CF in CiteULike based on tags (~~CULT~~):~~ 5.25.

Note that only four authors had publications and tags in CiteULike that could be analyzed (authors 5 and 6 just rated the ACC and BC clusters). Two authors claimed that BC and CULT were very relevant and suggested combining these two in order to be yielded all important authors and relevant research communities. In BC and ACC, some interviewees missed important authors. Two of the interviewees stated that the authors in BC and ACC were too obvious to be similar and said they were interested in bigger graphs featuring more potential collaborators. A combined cluster could help them find research groups and cooperation partners and might help intensify working relationships among colleagues. Looking at the graphs, almost all target authors remembered important colleagues that they hadn't thought of at first but whom they found very helpful. They stated that bigger graphs like CULT showed more scientists who were unknown to them – to give a clear statement about these potentially similar researchers, the interviewees would have had to look at their publications. It may be assumed that if an unknown person is shown to be clearly connected to a known relevant research group, he or she probably does similar (and relevant) work. As the interviewees stated that the distribution of the researchers is shown correctly, it is likely, albeit not explicitly proven, that any unknown scientists would also be allocated correctly within the graph.

An important factor for all interviewees was a clear cluster arrangement. A possible problem with CiteULike clusters is their sparse dataset, that is, if only a small number of tags are assigned to one author's publications, or if only one user bookmarks them, the cluster cannot show clearly distinguishable communities. This was the case with authors 2 and 5. Author 2 gave less favorable ratings to the CiteULike graphs because they didn't show clear distributions and author groups. On the other hand, for author 1 the distribution (Figures 11.2–11.5) was very clear and the researcher regarded it as helpful. As a junior researcher she found the CULT and CULU graphs helpful for finding new researchers and getting an overview of her network community. Further categorizations of authors, for example, via tags or author keywords, might help to better classify scientists' work and avoid unclear distributions in a graph.

11.8. Conclusions

The amount of social information on the web has grown and continues to do so, constantly offering new possibilities for usage. Regarding the need of a researcher to collaborate with his or her colleagues, social information can be used to build networks of researchers and to recommend similar people to each other. Several approaches suggest methods and solutions for person-to-person recommendation for researchers, web users or employees. It is important for a service to recommend the right people in order to satisfy its users and be of advantage to them. For recommendations, the reputation of the potential partner is very important, hence citations and references must be considered. However, user-generated social information in a bookmarking service might also complement co-citation and BC measurements. In the model approach described above we analyzed academic author recommendation based on different author relations across three information services. The researchers confirmed that there is a need for author recommendation: Many physicists don't work by themselves, but in project teams. Cooperation with colleagues from the same research field is essential. This is where a recommender system could be of great help. The results and evaluations show that the best results are achieved by a combination of social information from different services. Similarity based on users and tags in an online bookmarking system may complement the methods of ACC and BC. Some target authors found more relevant similar authors in CiteULike than they did in Scopus or WoS, an assumption confirmed by the interviewees in the graph relevance ranking.

The challenge will be to combine the different similarity approaches. One method to do so 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 testing the best algorithm for similarity measurement are required. The relations between user-based and tag-based similarity in a bookmarking system should also be considered, for example, via a graph-based approach such as FolkRank (Hotho et al., 2006) or user expertise analysis (Au Yeung et al., 2009). Aspects such as accuracy and efficiency should be tested in an operating recommendation system. Apart from technical aspects, the target users – for example, researchers – should be involved in the system's evaluation. If they don't see its benefits and don't trust its recommendations, the service won't be of any use.

Acknowledgments

I would like to thank Isabella Peters and Wolfgang G. Stock for their contributions to the paper, Oliver Hanraths for helping with the datasets and the reviewers for their supportive comments. The project is financed by a grant of the Strategische Forschungsfonds of the Heinrich-Heine-University Düsseldorf.

References

Ahlgren, P., Jarneving, B., & 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. doi:10.1002/asi.10242

Ahn, H. J. (2008). A new similarity measure for collaborative filtering to alleviate the new user cold-starting problem. *Information Sciences*, *178*, 37–51. doi:10.1016/j.ins.2007.07.024

Au Yeung, C. M., Noll, M., Gibbins, N., Meinel, C., & Shadbolt, N. (2009). On measuring expertise in collaborative tagging systems. In *Proceedings of Web Science Conference: Society On-Line*. Retrieved from http://journal.webscience.org/109. Accessed on May 21, 2012.

Ben Jabeur, L., Tamine, L., & Boughanem, M. (2010). A social model for literature access: towards a weighted social network of authors. In *Proceedings of RIAO '10 International Conference on Adaptivity, Personalization and Fusion of Heterogeneous Information*, 32–39. Retrieved from http://dl.acm.org/citation.cfm?id=1937063. Accessed on May 21, 2012.

Berkovsky, S., Kuflik, T., & Ricci, F. (2007). Mediation of user models for enhanced personalization in recommender systems. *User Modeling and User-Adapted Interaction*, *18*(3), 245–286. doi:10.1007/s11257-007-9042-9

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. Retrieved from http://gradworks.umi.com/32/78/3278195.html. Accessed on May 21, 2012.

Bogers, T., & van den Bosch, A. (2008). Recommending scientific articles using CiteULike. In *Proceedings of the 2008 ACM Conference on Recommender Systems* (pp. 287–290). New York, NY: ACM. doi:10.1145/1454008.1454053

Boyack, K. W., & 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. doi:10.1002/asi.21419

Cabanac, G. (2010). Accuracy of inter-researcher similarity measures based on topical and social clues. *Scientometrics*, *87*(3), 597–620. doi:10.1007/s11192-011-0358-1

Cacheda, F., Carneiro, V., Fernández, D., & 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). doi:10.1145/1921591.1921593

Cai, X., Bain, M., Krzywicki, A., Wobcke, W., Kim, Y. S., Compton, P., & Mahidadia, A. (2011). Collaborative filtering for people to people recommendation in social networks. In J. Li (Ed.), *Lecture notes in computer science: Vol. 6464. Advances in artificial intelligence* (pp. 476–485). Berlin: Springer-Verlag. doi:10.1007/978-3-642-17432-2_48

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. doi:10.1177/016555150002600107

Cronin, B. (1984). *The citation process. The role and significance of citations in scientific communication*. London: Taylor Graham.

Cronin, B., Shaw, D., & La Barre, K. (2003). A cast of thousands: Coauthorship and subauthorship collaboration in the 20th century as manifested in the scholarly journal literature of psychology and philosophy. *Journal of the American Society for Information Science and Technology*, *54*(9), 855–871. doi:10.1002/asi.10278

Cruz, C. C. P., Motta, C. L. R., Santoro, F. M., & Elia, M. (2009). Applying reputation mechanisms in communities of practice. A case study. *Journal of Universal Computer Science*, *15*(9), 1886–1906. doi:10.3217/jucs-015-09-1886

Desrosiers, C., & Karypis, G. (2007). A comprehensive survey of neighborhood-based recommendation methods. In F. Ricci, L. Rokach, B. Shapira & P.B. Kantor (Eds.), *Recommender systems handbook* (pp. 197–144). New York, NY: Springer-Verlag.

Gmur, M. (2003). Co-citation analysis and the search for invisible colleges. A methodological evaluation. *Scientometrics*, *57*(1), 27–57. doi:10.1023/A:1023619503005

Goldberg, D., Nichols, D., Oki, B. M., & Terry, D. (1992). Using collaborative filtering to weave an information tapestry. *Communications of the ACM*, *35*(12), 61–70. doi:10.1145/138859.138867

Heck, T. (2011). A comparison of different user-similarity measures as basis for research and scientific cooperation. In *Proceedings of the Information Science and Social Media International Conference* (pp. 117–128). Retrieved from http://issuu.com/informationsvetenskap/docs/issome2011

Heck, T., Hanraths, O., & Stock, W. G. (2011). Expert recommendation for knowledge management in academia. In *Proceedings of the 74th ASIS&T Annual Meeting* (Vol. 48). doi:10.1002/meet.2011.14504801365

Heck, T., & Peters, I. (2010). Expert recommender systems: Establishing Communities of Practice based on social bookmarking systems. In *Proceedings of the 10th International Conference on Knowledge Management and Knowledge Technologies* (pp. 458–464). Retrieved from http://www.phil-fak.uni-duesseldorf.de/?id=12276. Accessed on May 21, 2012.

Heck, T., Peters, I., & Stock, W. G. (2011). Testing collaborative filtering against co-citation analysis and bibliographic coupling for academic author recommendation. In *Proceedings of the 3rd ACM RecSys'11 Workshop on Recommender Systems and the Social Web*. Retrieved from http://www.dcs.warwick.ac.uk/~ssanand/RSWeb11/index.htm. Accessed on May 21, 2012.

Herlocker, J. L., Konstan, J. A., Borchers, A., & Riedl, J. (1999). An algorithmic framework for performing collaborative filtering. In *Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 230–237). Berkeley, CA. doi:http://doi.acm.org/10.1145/312624.312682

Herlocker, J. L., Konstan, J. A., Terveen, L. G., & Riedl, J. T. (2004). Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems*, *22*(1), 5–53. doi:10.1145/963770.963772

Hotho, A., Jäschke, R., Schmitz, C., & Stumme, G. (2006). Information retrieval in folksonomies: Search and ranking. In Y. Sure & J. Domingue (Eds.), *Lecture notes in computer science: Vol. 4011. The semantic web: Research and applications* (pp. 411–426). Heidelberg: Springer-Verlag. doi:10.1007/11762256_31

Kessler, M. M. (1963). Bibliographic coupling between scientific papers. *American Documentation*, *14*, 10–25. doi:10.1002/zasi.5090140103

Krohn-Grimberghe, A., Nanopoulos, A., & 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*. Retrieved from http://ucersti.ieis.tue.nl/2010/program.html. Accessed on May 21, 2012.

Lee, D. H. & Brusilovky, P. (2010a). Social networks and interest similarity. The case of CiteULike. In *Proceedings of the 21st ACM Conference on Hypertext & Hypermedia* (pp. 151–155). New York, NY: ACM. doi:10.1145/1810617.1810643

Lee, D. H., & Brusilovky, P. (2010b). Using self-defined group activities for improving recommendations in collaborative tagging systems. In *Proceedings of the Fourth ACM Conference on Recommender Systems* (pp. 221–224). New York, NY: ACM. doi:10.1145/1864708.1864752

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. doi:10.1002/asi.v56:7

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. doi:10.1002/asi.v59:1

Li, J., Burnham, J. F., Lemley, T., & 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. doi:10.1080/15424065.2010.505518

Linde, F., & Stock, W. G. (2011). *Information Markets*. New York, NY: De Gruyter Saur.

Luukkonen, T., Persson, O., & Sivertsen, G. (1992). Understanding patterns of international scientific collaboration. *Science, Technology, & Human Values*, *17*(1), 101–126. doi:10.1177/016224399201700106

Marinho, L. B., Nanopoulos, A., Schmidt-Thieme, L., Jäschke, R., Hotho, A., Stumme, G., & Symeonidis, P. (2011). Social tagging recommenders systems. In F. Ricci, L. Rokach, B. Shapira & P. B. Kantor (Eds.), *Recommender systems handbook* (pp. 615–644). New York, NY: Springer-Verlag.

Marshakova, I. (1973). Система связей мкжду документами, построенная на основе ссылок: по данным Science Citation Index [System of document connections based on references (data of SCI)]. *Nauchno-Tekhnicheskaya Informatsiya, Seriya 2*(6), 3–8. Retrieved from http://garfield.library.upenn.edu/marshakova/marshakova-nauchtechn1973.pdf. Accessed on May 21, 2012.

McNee, S. M., Kapoor, N. & Konstan, J. A. (2006). Don't look stupid. Avoiding pitfalls when recommending research papers. In *Proceedings of the 20th Anniversary Conference on Computer Supported Cooperative Work* (pp. 171–180). doi:10.1145/1180875.1180903

Meho, L. I., & 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. doi:10.1002/asi.v59:11

Meho, L. I., & 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. doi:10.1002/asi.v60:12

Nocera, A., & Ursino, D. (2011). An Approach to Provide a User with recommendations of similar users and potentially interesting resources. *Knowledge-Based Systems*, *24*(8), 1277–1296. doi:10.1016/j.knosys.2011.06.003

Parra, D., & Brusilovsky, P. (2009). Collaborative filtering for social tagging systems. An experiment with CiteULike. In *Proceedings of the 3rd ACM Conference on Recommender Systems* (pp. 237–240). New York, NY: ACM. doi:10.1145/1639714.1639757

Peters, I. (2009). *Folksonomies. Indexing and retrieval in web 2.0*. Berlin: De Gruyter Saur.

Persson, O. (2001). All author citations versus first author citations. *Scientometrics*, *50*(2), 339–344. doi:10.1023/A:1010534009428

Petry, H., Tedesco, P., Vieira, V., & Salgado, A. C. (2008). ICARE. A context-sensitive expert recommendation system. *Proceedings of Workshop on Recommender Systems on the 18 th European Conference on Artificial Intelligence* (pp. 53–58). Retrieved from http://proserver3-iwas.uni-klu.ac.at/ECAI08-Recommender-Workshop/. Accessed on May 21, 2012.

Ramezani, M., Bergman, L., Thompson, R., Burke, R., & Mobasher, B. (2008). Selecting and applying recommendation technology. In *Proceedings of the International Workshop on Recommendation and Collaboration, in Conjunction with 2008 International ACM Conference on Intelligent User Interfaces*. Retrieved from http://maya.cs.depaul.edu/~mobasher/pubs.html. Accessed on May 22, 2012.

Reichling, T., & Wulf, V. (2009). Expert recommender systems in practice. Evaluating semi-automatic profile generation. In *Proceedings of the 27th International Conference on Human Factors in Computing Systems* (pp. 59–68). doi:10.1145/1518701.1518712

Rendle, S., Marinho, L. B., Nanopoulos, A. & 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* (pp. 727–736). New York, NY: ACM. doi:10.1145/1557019.1557100

Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., & Riedl, J. (1994). Grouplens: An open architecture for collaborative filtering of netnews. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work* (pp. 175–186). New York, NY: ACM. doi:10.1145/192844.192905

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. doi:10.1002/(SICI)1097-4571(1999)50:8<639::AID-ASI2>3.0.CO;2-C

Said, A., Wetzker, R., Umbrath, W., & Hennig, L. (2009). A hybrid PLSA approach for warmer cold start in folksonomy recommendation. In *Proceedings of the RecSys'09 Workshop on Recommender Systems & The Social Web* (pp. 87–90). Retrieved from http://ls13-www.cs.uni-dortmund.de/homepage/RSWEB/accepted papers.shtml. Accessed on May 21, 2012.

Schneider, J. W., & 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. doi:10.1002/asi.20643

Schneider, J. W., & 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. doi:10.1002/asi.20642

Shepitsen, A., Gemmell, J., Mobasher, B., & Burke, R. (2008). Personalized recommendation in social tagging systems using hierarchical clustering. In *Proceedings of the 2008 ACM Conference on Recommender Systems* (pp. 259–266). New York, NY: ACM. doi:10.1145/1454008.1454048

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. doi:10.1002/asi.4630240406

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. Retrieved from http://

www.phil-fak.uni-duesseldorf.de/infowiss/admin/public_dateien/files/1/1197986684stock120_h.htm. Accessed on May 21, 2012.

Van Eck, N. J., & 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. doi:10.1002/asi.20872

Wang, J., de Vries, A. P., & Reinders, M. J. T. (2006). Unifying userbased and itembased collaborative filtering approaches by similarity fusion. In *Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 501–508). New York, NY: ACM. doi:10.1145/1148170.1148257

Wasserman, S., Faust, K., & Iacobucci, D. (1994). *Social network analysis: Methods and applications (Structural analysis in the social science)*. Cambridge, UK: Cambridge University Press.

Wenger, E. (1998). *Communities of practice. Learning, meanings, and identity*. Cambridge, UK: Cambridge University Press.

White, H. D., & Griffith, B. C. (1981). Author cocitation: A literature measure of intellectual structure. *Journal of the American Society for Information Science*, *32*(3), 163–171. doi:10.1002/asi.4630320302

White, H. D., & Griffith, B. C. (1982). Authors as markers of intellectual space. Co-citation in studies of science, technology and society. *Journal of Documentation*, *38*(4), 255–272. doi:10.1108/eb026731

Zhao, D., & Strotmann, A. (2007). All-author vs. first author co-citation analysis of the information science field using scopus. In *Proceedings of the 70th Annual Meeting of the American Society for Information Science and Technology*, *44*(1), 1–12. doi:10.1002/meet.1450440262

Zhao, D., & 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), 654–676. doi:10.1002/asi.21495

AUTHOR QUERY FORM

Emerald	**Book: LIS-V005-3610623** **Chapter: 11**	**Please e-mail or fax your responses and any corrections to:** **E-mail:** **Fax:**

Dear Author,

During the preparation of your manuscript for typesetting, some questions may have arisen. These are listed below. Please check your typeset proof carefully and mark any corrections in the margin of the proof or compile them as a separate list.

Disk use
Sometimes we are unable to process the electronic file of your article and/or artwork. If this is the case, we have proceeded by:

☐ Scanning (parts of) your article ☐ Rekeying (parts of) your article

☐ Scanning the artwork

Bibliography
If discrepancies were noted between the literature list and the text references, the following may apply:

☐ **The references listed below were noted in the text but appear to be missing from your literature list. Please complete the list or remove the references from the text.**

☐ ***UNCITED REFERENCES:*** **This section comprises references that occur in the reference list but not in the body of the text. Please position each reference in the text or delete it. Any reference not dealt with will be retained in this section.**

Queries and/or remarks

Location in Article	Query / remark	Response
AU:1	Should the term "paper" be changed to "chapter" wherein being referred to the current chapter. Please suggest.	yes, chapter is the better term. The changes should be made on pages 1, 2 (2 times) and 22 (acknowledgm.)
AU:2	Should the terms "Recommendation (or recommender) systems (RS)", "CiteULike (CULU)" and "CiteULike (CULT)" be defined at their first occurrence and use their abbreviation thereafter? Please suggest.	as the abbr. RS is used only one time and would disturb readability, please delete the abbr. on page 6 and 7 Thanks for the remark. concerning CULU and CULT: chapter 11.6 summarizes the results at the beginning. Therefor the abbr. should be mentioned, not to be forgotten by the reader. So, I would not change this.
AU:3	We have enlarged the figure to gain high resolution. Please confirm	yes that's fine