Preprint to appear in Scientometrics
DOI 10.1007/s11192-011-0358-1

Accuracy of inter-researcher similarity measures based on topical and social clues

Guillaume Cabanac

Received: November 14, 2010 / Published online: March 1, 2011 /

Abstract Scientific literature recommender systems (SLRSs) provide papers to researchers according to their scientific interests. Systems rely on inter-researcher similarity measures that are usually computed according to publication contents (i.e., by extracting paper topics and citations). We highlight two major issues related to this design. The required full-text access and processing are expensive and hardly feasible. Moreover, clues about meetings, encounters, and informal exchanges between researchers (which are related to a social dimension) were not exploited to date. In order to tackle these issues, we propose an original SLRS based on a threefold contribution. First, we argue the case for defining inter-researcher similarity measures building on *publicly available metadata*. Second, we define topical and social measures that we combine together to issue *socio-topical recommendations*. Third, we conduct an evaluation with 71 volunteer researchers to check researchers' perception against socio-topical similarities. Experimental results show a significant 11.21% *accuracy improvement* of socio-topical recommendations compared to baseline topical recommendations.

Keywords Similarity among Researchers · Topical Clues · Social Clues · Literature Review · Recommendation · Experiment · Human Perception · Measurement

1 Introduction

Researchers continually investigate the scientific literature about their field, as well as frontiers with related domains. This complex, steady, and thorough activity enables them to constitute and then incrementally update their knowledge of state-of-the-art contributions. This is usually achieved by reading journal issues and conference proceedings, as well as chatting with other researchers at work or during conferences. Alternatively, researchers may benefit from modern computer capabilities offered in a scientific literature recommender

G. Cabanac
University of Toulouse
Computer Science Department – IRIT UMR 5505 CNRS
118 route de Narbonne, F-31062 Toulouse cedex 9
E-mail: guillaume.cabanac@irit.fr

system (SLRS). Such systems allow the retrieval of experts and articles for a given scientific subject. They also support the retrieval of the researchers most matching the user's research interests. Designers of SLRSs have been conceiving rich models for representing researchers; these are instantiated by downloading and processing publication contents (i.e., full-text documents), by extracting citations, by asking researchers to rate the papers they read, and so on. Finally, inter-researcher similarity measures are calculated on these data to identify the scientists sharing most interests with the user of the SLRS.

Although mature and appealing in theory, such SLRSs relying on rich models fail to get implemented in publicly available services for one major reason. They rely on expensive data (requiring costly subscriptions to several publishers such as Elsevier, Springer, or Wiley) that are tough processing (especially regarding the downloading and parsing of loosely structured documents based on various templates). As a result, to the best of our knowledge, there is no publicly available service issuing scientific recommendations in reply to researchers' needs concerning literature review.[1] However, it would be of great help to researchers if they could automatically get reading recommendations according to their own research activity.

Apart from ultimately offering access to (fee-paying) publications, all scientific literature digital libraries[2] freely release publication metadata—forming their library catalog. Publicly available metadata associated with any publication are generally comprised of author names, title, publication year, venue (conference of journal name), keywords, and abstract (i.e., a 150–300 words summary). In this paper, we intend to harness this *publicly available metadata* by modeling researchers and defining inter-researcher similarity measures. Besides considering a common *topical* dimension by analyzing publication titles, we define an original *social* dimension allowing the modeling of researchers' time-evolving social environment. Despite the poor amount of information used (compare metadata versus full-text), we claim and demonstrate that it is sufficient to issue socio-topical recommendations that are accurate and useful according to researchers themselves.

This paper is organized as follows. In Section 2, the computer-aided literature review activity that researchers steadily conduct is introduced. We also point out the associated current issues. Our twofold contribution is presented in Section 3. We propose 1) to harness *publicly available metadata* from which 2) original *social clues* about researchers' environment are extracted. These are considered as additional evidence to topical clues for measuring the similarity between researchers in order to issue recommendations. In Section 4, we define an evaluation protocol for measuring the accuracy of recommendations according to researchers' perception. It allows us to assess the effectiveness of topical recommendations, and then to check topical recommendations against socio-topical recommendations. The difference between these two approaches represents the degree of improvement provided by integrating social clues into the recommendation system. Then, we describe the experiment that we conducted with 71 volunteer independent researchers. We report analyses of the accuracy of socio-topical recommendations that we propose in this paper. Finally, we conclude the paper and give insights into future work in Section 5.

[1] It may be argued that Google Scholar (`http://scholar.google.com`) and derivatives, such as ArnetMiner (Tang et al 2008) (`http://arnetminer.org`) meet this need. These search engines surely are helpful for finding document related to a *query* (e.g., bibliometrics). However, they do not succeed in taking a researcher's name as input for recommending him/her papers or other researcher names that would be relevant for his/her overall scientific *activity* (as we intend to do in this paper).

[2] Subject to charges like the ACM Portal (`http://portal.acm.org`) and SpringerLink (`http://springerlink.com`) or free like CiteSeer^X (`http://citeseerx.ist.psu.edu`), DBLP (`http://www.informatik.uni-trier.de/~ley/db`) or arXiv (`http://arxiv.org`).

2 Computer-aided literature review with SLRSs

We are interested in the literature review activity, as commonly achieved by any researcher. This demanding task is nevertheless mandatory for researchers who author papers or serve as reviewers for conferences or journals. The thorough knowledge of their field allows them to tackle current, unsolved issues, to underline the original contributions of their own research, and to track down ideas already published. We highlight below three issues related to usual literature review.

1. Researchers must identify the appropriate (digital) libraries allowing them to access interesting publication contents (either published in conference proceedings or journals). These repositories may be non-specialized like Google Scholar, or domain-specific like PubMed Central[3] for biomedical literature or DBLP (Ley 2002) for Computer Science.
2. In the current 'publish or perish' atmosphere, researchers are pressured to publish more and quickly (Garfield 1996). This global production increase is reported by several studies (e.g., Elmacioglu and Lee 2005). As a result, researchers conducting literature review must seek relevant publications in an ever growing number of venues (both conference proceedings and journals), which makes this task increasingly time consuming and tedious.
3. Researchers expect to effectively browse digital libraries, which requires training and specific skills for using the various tools and platforms giving access to publications.

Complementary approaches tackle these issues for improving researchers' experience with scientific literature digital libraries. We organized them into the four following categories.

- Approaches in the first category aim to rank most prominent, influential, prestigious researchers of scientific fields (Hirsch 2005, 2010; Mimno and McCallum 2007; Tang et al 2008; Yan and Ding 2009; Yang et al 2010).
- Approaches in the second 'pull' category provide researchers with a search engine for retrieving papers matching a given query (Zhou et al 2007; Tang et al 2008; Ben Jabeur et al 2010).
- Approaches in the third 'push' category issue reading recommendations with SLRSs (as explained in greater detail afterwards) or classify incoming messages (e.g., call for papers) and documents (Hurtado Martín et al 2009; Tsatsaronis et al 2009).
- Approaches in the fourth category allow the visualization of research fields according to the related subjects (Glenisson et al 2005a,b).

These references constitute by no means an exhaustive survey but they give some indication of the many research directions currently undertaken for scientific literature digital libraries. In Section 2.1, we focus on the subject of this paper (SLRSs), as represented in the second 'pull' category. We study SLRSs alleviating researchers from part of the literature review task by recommending them researcher names who are relevant to their interests. Then, we highlight in Section 2.2 some major issues that we intend to address in this paper.

2.1 Scientific literature recommender systems

Recommender systems have been applied to many items (Resnick and Varian 1997), such as products, music, movies, and people. They rely on three key techniques (Montaner et al

[3] `http://www.ncbi.nlm.nih.gov/pubmed`

2003; Adomavicius and Tuzhilin 2005), namely collaborative filtering, content-based filtering, and hybrid filtering. In the next sections, we present how these approaches have been applied for recommending scientific resources in order to foster knowledge transfer among research communities. The interested reader may refer to (McNee et al 2006) for greater technical details.

2.1.1 Collaborative filtering

SLRSs based on collaborative filtering (Goldberg et al 1992; Herlocker et al 2004) compute recommendations according to judgments (e.g., marks, ratings, votes) that researchers attach to various resources. Alternatively, relevance judgments may be inferred by extracting references from publication contents (e.g., Powley and Dale 2007), where any citation from a paper p_1 to another paper p_2 is considered as a supporting vote given by the authors of p_1. Note, however, that textual contents are not further analyzed.

The ultimate aim of a collaborative filtering SLRS is to exploit patterns of interest identified among a research community for issuing recommendations. In a practical way, two researchers r_1 and r_2 sharing the same interests (i.e., liking a same set of papers and disliking another one) are first matched. Recommendations are then issued accordingly: r_1 will be offered the resources that r_2 liked (provided that r_1 has not encountered them yet) and vice versa. Several SLRSs based on collaborative filtering were implemented; they recommend scientific resources according to extracted citations (McNee et al 2002; Gori and Pucci 2006), laboratory coworkers' browsing activity (Agarwal et al 2005), researchers' opinions (Cazella and Campos Alvares 2005), or social bookmarks attached to publications (Bogers and van den Bosch 2008).

2.1.2 Content-based filtering

SLRSs relying on content-based filtering (Belkin and Croft 1992) compute recommendations according to publication contents (i.e., full-text). This implies the implementation of an indexing process (Manning et al 2008, chap. 2), which is a core Information Retrieval (IR) process. Any given publication is generally processed as follows. First, full-text is tokenized into terms. Second, meaningless terms (e.g., 'a', 'of') are dismissed. Third, terms are normalized by truncation or stemming (e.g., 'computers' and 'computing' are stemmed as 'comput'). Fourth, remaining terms are weighted according to their frequency in the text, as well as according to their rarity in the collection of publications.

Then a matching function is defined to compute the similarity between two given publications. It depends on the underlying IR model, the most common matching function being cosine for the Vector Space Model (Salton et al 1975). Inter-document matching functions may be generalized to inter-researcher functions by modeling a researcher as a mega-document (Klas and Fuhr 2000) concatenating all his/her publication contents.

2.1.3 Hybrid filtering

SLRSs based on hybrid filtering (Balabanović and Shoham 1997) compute recommendations by allying collaborative filtering and content-based filtering. They thus consider researchers' judgments and publication contents at the same time. For instance, this approach is implemented in the COBRAS system (Karoui et al 2006) harnessing researchers' judgments, as well as their BibTeX (Mittelbach and Goossens 2005, chap. 13) bibliographic

entries (paper title, keywords, and so on). Several other SLRSs based on hybrid filtering were implemented (see Naak et al 2009; Porcel et al 2009).

2.2 Issues of current recommendation approaches

Collaborative filtering and content-based filtering techniques for SLRSs have several limitations that we discuss in the following sections.

2.2.1 Issues of collaborative filtering

Data availability. Collaborative filtering relies on private data (e.g., ratings, bookmarks, annotations) that are sparse and difficult to acquire since they are stored in various formats on heterogeneous platforms. Moreover, citation extraction from full-text (when available) may perform with a variable error rate.

Data quality. Many approaches use citation graphs for computing papers and researchers' reputation or 'prestige' (Hirsch 2005, 2010; Mimno and McCallum 2007; Yan and Ding 2009; Ben Jabeur et al 2010; Yang et al 2010). It is generally accepted that the more a paper is referred to, the more interesting it is. A citation is viewed as a vote given from the papers' authors to the referred paper (and its authors) without further analysis of:

- Citation location. The degree of relatedness of a citation to the author's work varies according to the location of the citation in his/her paper. A citation in the 'Introduction' section compared to a citation in the 'Contribution' section are neither equally general/specific nor equally related to the author's work.
- Citation polarity. Researchers use negative (positive) citations to criticize (praise) existing works. Obviously, negative citations should not count the same way as positive citations do.
- Citation target and ethics. Researchers use self-citations to refer to previous works; these should not benefit to the authors' incoming citation count. Moreover, some researchers may add 'deference' citations to advertise some acquaintances' works, even if not directly related to the subject of their paper. Such a phenomenon is dishonest, since it unfairly contributes to the increase of indicators such as researchers' *h*-index (Hirsch 2005, 2010) and journal Impact Factors (Garfield 1955, 2006).

Although conveying different authorial intentions, current recommender systems do process citations with no analysis of polarity and target.

Cold start. Junior researchers cannot get recommendations since there are no citations to extract from their (not existing yet) papers. As a workaround, they may query the SLRS with their advisor's or mentor's identity.

Nepotism. Recommending papers according to incoming citations makes the most cited papers attract even more citations, which is a vicious cycle. On the contrary, it may be worth focusing on lesser known works, as they may be contributing innovative ideas to the field. Currently, this kind of works remains in the shadow of the bright much cited and then much recommended papers.

Summing up collaborative filtering limitations, proposed prestige-based approaches neither foster the emergence of new researchers to the field nor they support new ideas, as there is a tendency to favor already highly visible ideas from famous scientists instead of topically related contributions from less famous researchers.

2.2.2 Issues of content-based filtering

Syntactic issues. Term semantics are not considered. For instance, 'taxonomy' and 'taxinomy' are processed as two different terms though referring to the same concept.

Compound nouns. Expressions are split into terms (e.g., 'information retrieval' is indexed by 'information' and 'retrieval') that separately fail to capture the original meaning.

Multilingual issues. Documents authored in different languages are not comparable (e.g., 'retrieval' in English is 'interrogation' in French).

Besides the aforementioned limitations, collaborative filtering and content-based filtering techniques only consider researchers' *scientific production* and ignore their social relationships. In our view, researchers shall not be modeled according to their production only. This results in a partial representation lacking many aspects of their daily social interactions, such as coauthoring, participating in conferences, transferring knowledge to colleagues, and so on. Intending to counterbalance this issue, we introduce the *social environment* concept and formalize it in the next section.

3 Socio-topical recommender system based on publicly available metadata

In Section 3.1, we give an overview of the proposed socio-topical recommender system supporting computer-aided literature review. This is based on the proposed social (Section 3.2) and topical (Section 3.3) similarity measures combined together, as explained in Section 3.4.

3.1 Overview of the proposed socio-topical recommender system

We highlighted in the previous section that state-of-the-art SLRSs require publication contents to be retrieved and analyzed (i.e., terms and citations extraction). These approaches rely on rich conceptual models for representing authors and the resources they are interacting with (i.e., other authors and publications). For instance and as illustrated in Figure 1, Ben Jabeur et al (2010) consider a wide range of information, such as citation links between papers, researchers' bookmarks associated with the papers they are interested in, annotations they scribble on papers, friendship between researchers, and so on.

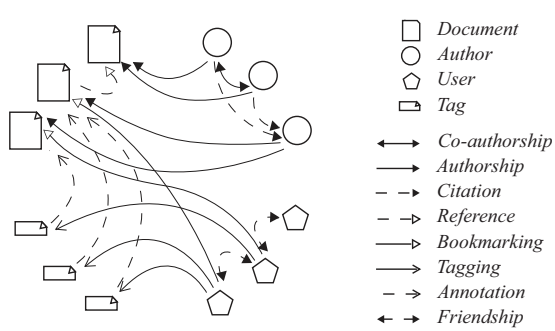

Fig. 1 Model of 'The social information network for literature access' as proposed in (Ben Jabeur et al 2010)

Although such rich and detailed modeling capabilities seem appealing in theory they are pitfalls in practice, mostly because the data involved are hardly available. Bookmarks, annotations, and social networking data are private indeed. Regarding full-text papers, researchers generally have to transfer their copyright to publishers (e.g., ACM, Springer, Wiley). Most publishers then grant their subscribers access to papers provided that they pay a membership fee or pay-per-use fee (Zamparelli 1998). This economic model is a hindrance to content-based SLRSs. First, the user must have subscribed to all the publishers in his domain (and afford to be charged for it). Second, the system must connect to publishers' digital libraries, download papers, and process them (dealing with multiple paper templates and reference styles). This expensive and technically difficult approach is hardly feasible.

We intend to tackle these data-related constraints by avoiding access (and payment) to papers full-text. Our contribution is a SLRS building on free and publicly available data, as the input of hybrid filtering inter-researcher similarity measures. These only process the metadata associated with any publication: title, authors, year, and venue name. Note that abstract and full-text are dismissed. This design makes *paper* recommendation barely impossible because the only title is poorly representative of a whole paper. Nevertheless, we believe that aggregating all the titles of papers written by a researcher is an accurate representation of the researcher's scientific interests. Consequently, instead of recommending papers to the user as currently done by most SLRSs (McNee et al 2002, 2006; Agarwal et al 2005; Cazella and Campos Alvares 2005; Gori and Pucci 2006; Karoui et al 2006; Bogers and van den Bosch 2008; Naak et al 2009; Porcel et al 2009), we propose to recommend *researchers' names* and then allow the user to retrieve their publications. Our proposed SLRS *recommends a list of researchers* related to any given researcher r (the user of the system), according to the following two requirements.

R1. The recommended researchers and r shall *share long-term interests* (i.e., they shall have been working on similar or identical subjects). A researcher is all the more relevant as he/she wrote many papers matching r's interests.
R2. The recommended researchers shall be *accessible*. We favor researchers to whom r may be introduced thanks to acquaintances (e.g., during a conference gala dinner). Those shall be closely reachable in r's social environment.

According to these requirements, any recommended researcher will be relevant to the user's scientific subjects while being closely related to the user's social network and research communities. Such recommendations support the identification of potential partners for a project, for instance. We believe that these recommendations may lead to serendipitous findings (i.e., interesting recommendations, though unexpected).

3.2 Inter-researcher social similarity measures

A major originality of our approach lies in the extraction and exploitation of researchers' social clues for the recommendation process. This is the collaborative filtering part of our SLRS. In order to compensate for the issues of the topical dimension presented in Section 2.2, we propose to enhance this latter with a social dimension, as covered by the three similarity measures introduced in the next sections. They are defined according to the undirected *graph of coauthors* (see Figure 2a, where a node stands for a researcher and an edge connects two researchers who coauthored a publication), and the directed *graph of venues* (see Figure 2b, where a node is either a researcher or a conference edition, and an edge connects a researcher to a conference venue where he had papers accepted). In the remainder of

the paper, note that $R = \{r_1, \ldots, r_n\}$ is the set of all researchers known in scientific literature digital libraries.

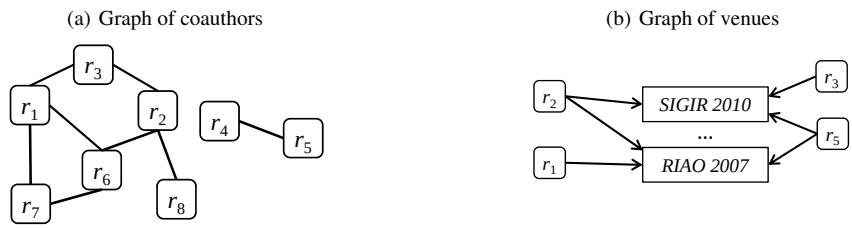

Fig. 2 Modeling researchers' social interactions: coauthors (author ↔ author) and venues (author ↔ venue)

3.2.1 Proximity in the graph of coauthors

Milgram (1967) defined the *degree of separation* of two people as the minimum number of acquaintances (i.e., intermediaries) required to connect them through their social network. In Graph Theory, this corresponds to the length of the shortest path connecting two nodes together (Easley and Kleinberg 2010, chap. 2). Empirical experiments involving subjects whose task was to reach a given unknown person by mail through relatives and acquaintances showed that the median degree of separation between two persons living in the USA was five (Milgram 1967; Travers and Milgram 1969). Hence the 'small world' expression referring to these findings. Note that, when analyzing the DBLP (Ley 2002) bibliographic records, Elmacioglu and Lee (2005) reported that an average of six coauthors separate any two randomly picked authors in Computer Science. This is in line with Milgram's (1967) findings, meaning that the Computer Science coauthor network is similar to a natural social network.

We build on this concept to compute the *proximity* of two researchers through the graph of coauthors (Figure 2a). The associated $p : R^2 \rightarrow [0,1]$ function computes the inverse of the length of the shortest path between two given coauthors. This equals to 1 when two researchers are coauthors, decreases according to their coauthoring distance, and equals 0 if there is no path connecting them. In the example of Figure 2a, $p(r_1, r_2) = 1/\min(2,2,3,5) = 1/2$ and $p(r_1, r_4) = 0$. This function captures a social aspect of researchers' work by assessing the proximity among authors.

3.2.2 Connectivity in the graph of coauthors

The *connectivity* (i.e., connection strength) between two researchers is also defined according to the graph of coauthors (Figure 2a). It is computed by the $c : R^2 \rightarrow \mathbb{N}_+$ function, which returns the number of distinct shortest paths between two researchers (i.e., with length equal to their degree of separation). In the example of Figure 2a, $c(r_1, r_2) = 2$ and $c(r_1, r_4) = 0$. This function captures a social aspect of researchers' work by assessing the opportunity to reach one researcher through a large number of different intermediaries.

3.2.3 Meeting opportunities in the graph of venues

The *meeting opportunities* (or number of shared venues) between two researchers are defined according to the graph of venues (Figure 2b). A venue is the edition of a specific conference (e.g., SIGIR 2010). We do not consider journal papers since only conference papers give researchers the opportunity to gather together for presenting their work and meeting up with other authors. It is computed by the $m : R^2 \rightarrow \mathbb{N}_+$ function, which returns the number of shared venues between two researchers. In the example of Figure 2b, $m(r_1, r_2) = 1$ and $m(r_1, r_4) = 0$. This function captures a social aspect of researchers' work by assessing the meeting opportunities for two researchers; it computes the odds that they may already have met during a conference where they both published. This probability is higher with an increasing number of shared venues, as an evidence of shared membership to (informal) communities. The $m(r_1, r_2)$ value is obviously a upper bound since all the authors of a multi-authored conference publication may not have attended the conference.

3.3 Inter-researcher topical similarity measure

The *topical similarity* between two researchers is defined according to their publication titles since we do not consider abstracts or full-text but metadata. This is the collaborative filtering part of our SLRS. We model each researcher as a mega-document (Klas and Fuhr 2000): all the terms used in the titles of a researcher's papers are collected in a single document. Then, mega-documents are processed to compute topical similarity as it is commonly achieved in IR (see Micarelli et al 2007; Manning et al 2008, chap. 2) with the Vector Space Model (Salton et al 1975). This process is comprised of the following stages.

1. Publication titles are tokenized into words (i.e., strings are split on whitespace characters). Next, words from a stop list (e.g., Fox 1989; Dolamic and Savoy 2010) are filtered out as they convey no meaning (these are essentially determinants, pronouns, and so on). Then, words are stemmed into terms with Porter's (1980) stemming algorithm, for instance. Stemming intends to reduce vocabulary variability that would impair further matching otherwise (e.g., 'algorithm,' 'algorithms,' and 'algorithmic' are stemmed as 'algorithm'). Finally, the n distinct stemmed terms from all publications are collected. These constitute the *indexing language*.
2. A *vector space* is created with n-dimensions; each one represents a term of the indexing language.
3. Each mega-document d_i, which represents one researcher, is implemented as a vector $\mathbf{d_i} = \left(w_i^1, \ldots, w_i^n\right)$ in the n-dimensional vector space. A weight w_i^j is calculated for each dimension j. Classically, $w_i^j = tf(d_i, t_j) \cdot idf(t_j)$, where the weight of a term t_j in d_i is higher when t_j is frequent in the document (*tf* meaning *term frequency*) and at the same time rare in other documents (*idf* meaning *inverse document frequency*). The higher the weight w_i^j, the more characteristic of document d_i the term t_j is. Readers interested in this TF·IDF weighting scheme may refer to (Spärck Jones 1973; Salton and Buckley 1988).
4. In the end, the topical similarity between two researchers r_1 and r_2 is computed by the $t : R^2 \rightarrow [0, 1]$ function, such that $t(r_i, r_j) = \cos\left(\mathbf{d_i}, \mathbf{d_j}\right)$. The smaller the angle between the two vectors, the more similar the two researchers are.

We explain in the next section how social and topical similarity measures are jointly used by our SLRS for issuing socio-topical recommendations of researcher names and publications to the user, who is also a researcher.

3.4 Issuing socio-topical recommendations: combining social and topical clues

The retrieval of interesting researchers according to a user's published work (i.e., research interests) is achieved in two steps. First, as presented in Section 3.4.1, we compute the list of closest researchers for the four similarity measures. Second, as described in Section 3.4.2, we combine these results in order to satisfy requirements R1 and R2 stated in Section 3.2.

3.4.1 Generation of a result list for each similarity measure

For any given researcher r, we compute the list T of the $n-1$ most similar researchers based on their topics, that is $\mathsf{T} = \left\langle \left(r_i^1, s_i^1\right), \ldots, \left(r_i^{n-1}, s_i^{n-1}\right) \right\rangle$ where $s_i^j = t\left(r_i, r_j\right)$ is the topical similarity between r_i et r_j. We hypothesize that recommending a researcher's own work or direct coauthors' work is irrelevant since he/she may be aware of it, obviously. Hence, we filter out the user's name and his/her direct coauthors from list T.

This process detailed for function t results in a T list. It is repeated for aforementioned functions p, c, and m that result in lists P (for Proximity), C (for Connectivity), and M (for Meeting opportunities).

3.4.2 Result lists combination with normalized CombMNZ

We hypothesize that the reinforcement of topical recommendations with social clues leads to effectiveness improvement. We realize this socio-topical combination with one of the Comb* functions proposed by Fox and Shaw (1993). These were originally designed as part of a meta-search engine for combining the outcomes of several source search engines into a single result list. Comb* functions combine k result lists (document-similarity pairs) into a single result list. This latter is comprised of every retrieved document associated with a combined similarity value computed by aggregating the k source similarities. Many variations of this combination scheme were proposed. They vary according to the aggregation function $f : \mathbb{R}^k \to \mathbb{R}$ used: CombSUM is based on the sum function, CombMAX is based on the max function, and so on. Fox and Shaw (1993) showed that CombMNZ yields best performance on TREC[4] test collections. It computes a combined similarity S by aggregating the k source similarities $\{s_1, \ldots, s_k\}$ according to $S = N \cdot \sum_{i=1}^{k} s_i$, where N represents the number of non null similarity values s_i. Furthermore, prior to realizing the combination, Lee (1997) pointed that similarity values in result lists should be normalized (i.e., converted into the $[0,1]$ range) with Equation 1 for a fair combination, when similarities are defined on different ranges (e.g., in our case $p(r_i, r_j) \in [0,1]$ while $c(r_i, r_j) \in \mathbb{N}_+$).

$$\text{normalized\_similarity} = \frac{\text{unnormalized\_similarity} - \text{minimum\_similarity}}{\text{maximum\_similarity} - \text{minimum\_similarity}} \tag{1}$$

Lee (1997) showed on TREC test collections that *normalized* CombMNZ yields best performance compared to other Comb* functions with or without normalization. That is the reason why we opted for normalized CombMNZ to combine the results of social similarity

[4] TREC stands for the Text REtrieval Conference (see Voorhees and Harman 2005).

measures. The resulting S list is illustrated in Figure 3d, which represents the combination of three social result lists (Figure 3a–c). We may notice that r_4 is present in the three source lists with high scores, which explains that it is top ranked in the combined S list. On the contrary, r_7 was retrieved in two lists only, and is thus ranked lower in the final S list.

(a) Proximity (P)

Researcher	Similarity
r_4	0.50
r_3	0.33
r_5	0.25
r_1	0.16

(b) Connectivity (C)

Researcher	Similarity
r_6	12
r_1	8
r_4	8
r_7	6
r_2	2

(c) Meeting Opportunities (M)

Researcher	Similarity
r_6	5
r_4	5
r_7	3
r_1	2
r_2	1

↘ ↓ ↙

Normalized CombMNZ

↓

(d) Combined social result list (S)

Researcher	Combined similarity (with detailed computation)
r_4	$7.8000 = 3 \cdot \left(\frac{0.50-0.16}{0.50-0.16} + \frac{8-2}{12-2} + \frac{5-1}{5-1}\right)$
r_6	$4.0000 = 2 \cdot \left(0 + \frac{12-2}{12-2} + \frac{5-1}{5-1}\right)$
r_1	$2.5500 = 3 \cdot \left(\frac{0.16-0.16}{0.50-0.16} + \frac{8-2}{12-2} + \frac{2-1}{5-1}\right)$
r_7	$1.8000 = 2 \cdot \left(0 + \frac{6-2}{12-2} + \frac{3-1}{5-1}\right)$
r_3	$0.5000 = 1 \cdot \left(\frac{0.33-0.16}{0.50-0.16} + 0 + 0\right)$
r_5	$0.2647 = 1 \cdot \left(\frac{0.25-0.16}{0.50-0.16} + 0 + 0\right)$
r_2	$0.0000 = 2 \cdot \left(0 + \frac{2-2}{12-2} + \frac{1-1}{5-1}\right)$

Fig. 3 Example of result lists combination with normalized CombMNZ (Fox and Shaw 1993)

For achieving our main purpose, that is computing a socio-topical list of recommended researchers, we designed a two-step process (Figure 4). First, we combine the results of the three social similarity measures into the S list. From this list, we filter out the researchers that have no topics in common with the user (requirement R1) by computing $\mathsf{S}' \leftarrow \mathsf{S} \cap \mathsf{T}$. Second, we combine the topical results T with the latter social results (requirement R2) giving the final recommendation result ST (an example is shown in Figure 5a). Notice that the combinations involved in this two-step process are done with no hypothesis about the relative importance of the result lists given as input. For instance, in the second step, social and

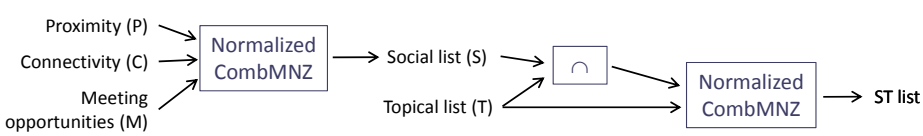

Fig. 4 Two-step process for combining social and topical recommendations

topical dimensions are equally combined. This represents an unsupervised (i.e., no learning is required for source weighting) approach that we evaluate in the next section.

We implemented this socio-topical similarity measure in an online SLRS. It takes a researcher name as input and issues recommendations (other researcher names) ranked by decreasing score. Recommendations are computed on the DBLP dataset that we used for the experiment presented in the next section.

4 Checking similarity measures against researchers' perception of relatedness

In this section, we compare topical and socio-topical inter-researcher similarity measures with researchers' perception of relatedness to the recommended researchers. Our hypothesis is: the combination of topical and social similarities better simulates researchers' perception than topical similarities only. We present the experimental protocol that we designed, its application with 71 subjects, our analyses, and related findings.

4.1 Experimental protocol and data analysis

Our aim is to validate the proposed similarity measures by having researchers assess the accuracy of issued recommendations. Hence, researchers taking part in the evaluation will be called *subjects*. Prior to their participation, subjects' ST and T lists shown in Figure 5 are computed. Then, each subject assesses the recommendations presented in list T (i.e., top topically related researchers). Note that we randomized T list in order to ensure that the rank of a recommendation does not influence subjects' perception. The subject rates each recommended researcher on a 7-point Likert (1932) scale ranging from zero to six (e.g., 2/6 with a star metaphor ★★☆☆☆☆). This mark is given according to the question: *'Would meeting this researcher help you to improve your research?'* Notice that the zero mark means that the subject is not interested at all in the recommended researcher: this recommendation is not relevant. On the contrary, a recommendation is all the more relevant as the subject's mark is high. These marks constitute the HP list (i.e., human perception) illustrated in Figure 5b.

(a) Socio-topical ST list

Researcher	Sim
Alan Turing	5.42
Donald E. Knuth	4.80
⋮	⋮
Bill Gates	0.10

$\longleftrightarrow^{?}$

(b) Human perception HP list

Researcher	Sim
Donald E. Knuth	6
Alan Turing	4
⋮	⋮
Bill Gates	0

$\longleftrightarrow^{?}$

(c) Topical T list

Researcher	Sim
Bill Gates	0.75
Alan Turing	0.71
⋮	⋮
Donald E. Knuth	0.24

Fig. 5 Illustrations of ST, HP, and T lists as considered in the experimental protocol

Here we draw a parallel between evaluating SLRSs combination outcomes and evaluating search engines in IR. A search engine retrieves a result list of documents in response to the user's query conveying his/her information need. The evaluation of search engine effectiveness is commonly realized with Cleverdon's (1962) 'Cranfield paradigm' (see also

Voorhees 2002; Sanderson 2010). It has been notably implemented at TREC for evaluating search engines (Buckley and Voorhees 2005) contributed by multiple participants (i.e, IR groups from worldwide universities, companies like Google and Yahoo). We illustrate it with an example in Figure 6. At TREC, evaluating the effectiveness of a search engine s for a given query q is done by comparing its outcome with relevance judgments j from human assessors. This comparison is achieved with an effectiveness measure m such that $m(s,q,j) \in [0,1]$. The zero value means that the search engine result list is totally irrelevant, whereas a value of one is obtained when the system retrieved every relevant document for q present in the corpus, and ranked them at the top of the result list. This measure supports inter-system comparisons: $m(s_1,q,j) > m(s_2,q,j)$ shows that s_1 yields more relevant results than s_2 does. Several m measures were defined, such as *ap* (i.e., *Average Precision*) for binary relevance judgments (i.e., documents are either relevant or non-relevant: $j \in \{0,1\}$) or Järvelin and Kekäläinen's (2002) *ndcg* (i.e., *Normalized Discounted Cumulative Gain*) for gradual judgments (i.e., documents are non-relevant or more or less relevant to the query: $j \in \mathbb{R}_+$). These two measures (along with several others) are implemented in trec_eval,[5] TREC's official software for evaluating participants' search engines.

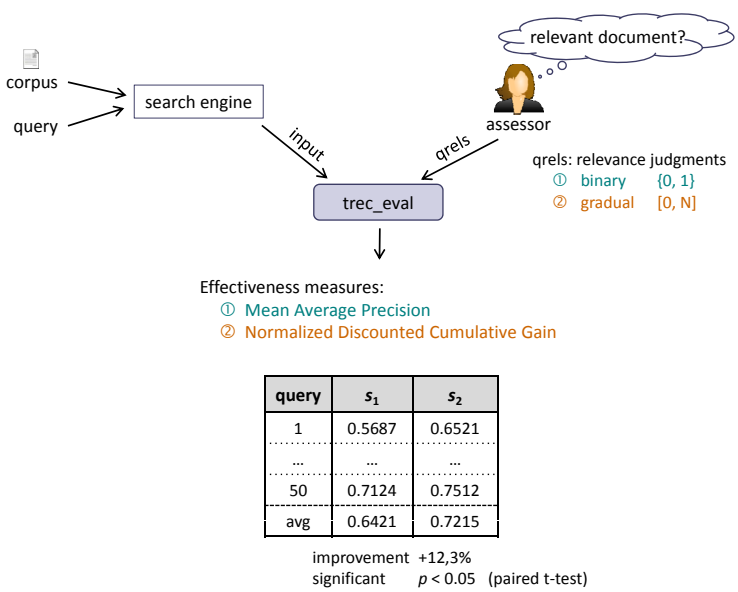

Fig. 6 Overview of search engine evaluation as realized at TREC (Buckley and Voorhees 2005)

Controlling for query effect (i.e., ensuring that no performance variation occurs when dealing with different queries) is usually done by evaluating the search engine with several queries and then averaging the individual effectiveness scores in a global score. The robustness of this practice was shown by Buckley and Voorhees (2000) with at least $n = 25$ queries, although $n = 50$ is more robust for statistical analyses (this is the usual number of queries at TREC). Finally, as shown at the bottom of Figure 6, the improvement in effectiveness

[5] Available for download at `http://trec.nist.gov/trec_eval`.

between s_1 and s_2 is computed by Equation 2.

$$\frac{\sum_{i=1}^{n} m(s_2, q_i, j)}{\sum_{i=1}^{n} m(s_1, q_i, j)} - 1 \tag{2}$$

The statistical significance of an improvement is represented as a p-value with Student's (1908) paired bilateral t-test (differences are computed between paired values $m(s_1, q_i, j)$ and $m(s_2, q_i, j)$). Although requiring a normal data distribution in theory, Hull (1993) points that it is robust to violations of this requirement in practice. Moreover, Sanderson and Zobel (2005) show this test to be more accurate than other ones, such as Wilcoxon's (1945) signed rank test. When $p < \alpha$, with $\alpha = 0.05$ the difference between the two systems is deemed to be statistically significant (Hull 1993). The smaller the p-value, the more significant this difference is. We refer the interested reader to (Sanderson 2010, chap. 5) for a detailed description of significance tests applied to IR.

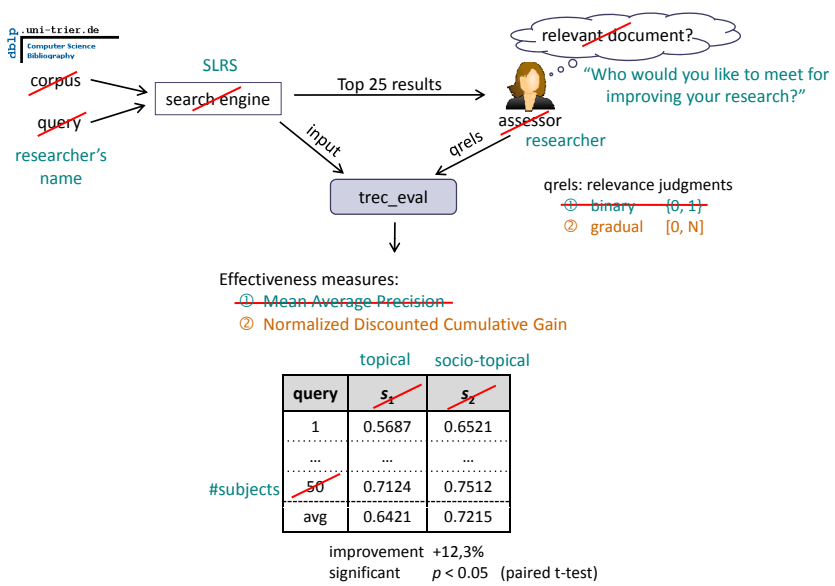

Fig. 7 Proposed framework for SLRS evaluation, as based on IR evaluation (see Figure 6)

Just as a search engine s retrieves documents for a query q, our SLRS extracts recommendations from a scientific literature digital library (e.g., DBLP) for a given researcher, who generally is the user. We thus build on the Cranfield paradigm for evaluating inter-researcher similarities. Consequently and as illustrated in Figure 7, we consider list T as the result of search engine s_1, and list ST as the result of search engine s_2. Moreover, a subject/researcher r is considered as a query q, and relevance judgments j are produced by the same researcher r for capturing his human perception (HP) of relatedness. Each subject rates the top 25 recommended researchers on the 7-point Likert scale. Finally, the computation of $m(\mathsf{T}, r, \mathsf{HP})$ and $m(\mathsf{ST}, r, \mathsf{HP})$ is realized with measure $m = ndcg$ since human perception HP is gradual (marks range from zero to six). Instead of evaluating with n queries, we evaluate with n subjects. This implies a constraint on the number of subjects: at least $n = 25$ researchers will be necessary to achieve statistically valid conclusions.

4.2 Implementing the experimental protocol with the DBLP digital library

The evaluation of similarity measures with the proposed protocol required the use of a scientific literature digital library. We opted for DBLP, which collects publications of the Computer Science field (Ley 2002). The rationale for this choice is threefold. First, DBLP is large as it indexed 1,392,143 publications authored by 811,787 researchers as of May 11th, 2010. It comprises papers published by major publishing groups, such as the ACM, Elsevier, IEEE, Springer, and Wiley. Second, metadata about indexed publications are publicly released as a 713 MB `dblp.xml` file.[6] Computing recommendations from this dataset will enable us to evaluate similarity measures while showing the feasibility of our metadata-based SLRS proposal. Third, DBLP has already been used for research purposes, such as Lotka's Law empirical validation (Elmacioglu and Lee 2005), personal name language identification (Biryukov 2008), expert search (Deng et al 2008), community detection (Huang et al 2009), and research field coverage measurement (Reitz and Hoffmann 2010). We analyzed data provided in `dblp.xml` and designed the associated UML class diagram representing available metadata and its organization, as shown in Figure 8. Class names refer to the well-known BibTEX format (Mittelbach and Goossens 2005, chap. 13). Note that our SLRS only processes gray background classes for issuing recommendations, other data being dismissed.

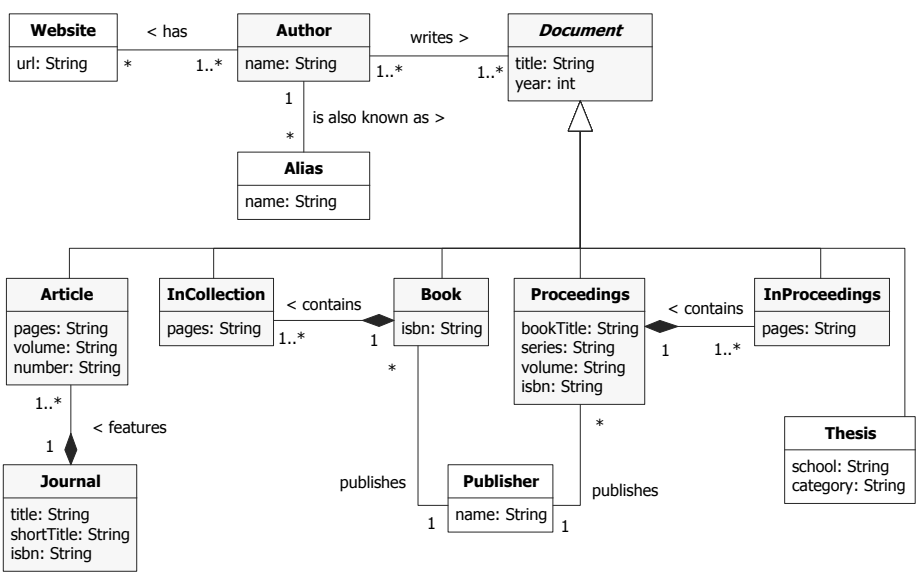

Fig. 8 Proposed UML class diagram for DBLP data, inferred from `dblp.xml`. Classes with gray background are processed by the SLRS for issuing recommendations

We involved researchers as subjects for our experiment in order to validate the similarity measures according to their human perception of relatedness. Through an online experimental setting (Reips 2002, 2007), we relied on crowdsourcing (Alonso et al 2008) in order to constitute a subject sample with varied research profiles and backgrounds. In July 2010,

[6] `http://dblp.uni-trier.de/xml`

we contacted 90 authors by sending individual emails, provided that they had at least two publications in DBLP. This sample of researchers was extracted from our contacts on the LinkedIn[7] business-oriented social networking website. Notice that this is not a random sample (Kelly 2009, chap. 7) of researchers since most of them are computer scientists, and they all are connected to the author's social network. The invitation email described the proposed threefold task, which duration was estimated to five minutes:

1. Provide age and seniority in research (number of years).
2. Rate the relevance of 25 recommended researchers presented in random order.
3. Report feedback about the assessment task and comment on the SLRS usefulness.

The email also contained a hyperlink to the website that we designed for experimentation purpose.[8] The aim of the experiment was clearly stated (comparing various inter-researcher similarity measures) while our hypothesis (social clues improve topical recommendations) was unrevealed in order not to influence the participants. Moreover, no reward was offered to researchers who had to work voluntarily.

4.3 Data analysis of the acquired 71 participations

4.3.1 Characteristics of the 71 subjects who took part in the experiment

Further to our email, 74 volunteer researchers began the experiment. This corresponds to a 82% response rate. In the end, 71 of these participants finished the experiment; they will be referred to as *subjects* from now on. This corresponds to a 4% dropout, which is far less than the 45% dropout measured by Reips (2007) for unrewarded web experiments. This may be due to three aspects of the experiment that we designed.

1. The estimated duration was short: five minutes on average.
2. We involved acquaintances who were certainly keener to help than strangers that we could have reached by mailing lists with larger audience, such as Reips and Lengler's (2005) Web Experimenter List.
3. We offered researchers the opportunity to get potentially interesting recommendations. This served as an incentive since several subjects reported that they found the recommended researcher names and associated publications helpful.

[7] `http://www.linkedin.com`

[8] A demonstration can be seen at `http://www.irit.fr/~Guillaume.Cabanac/expeSimT`.

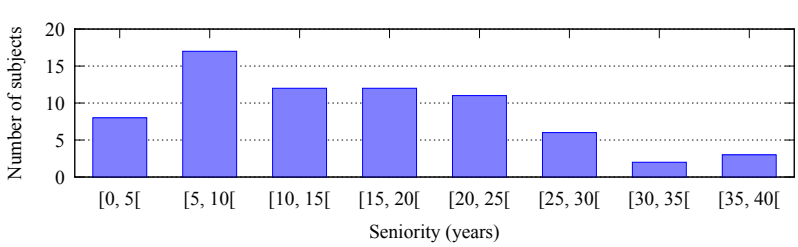

Fig. 9 Distribution of the 71 subjects according to seniority

The 71 subjects were between 24 and 61, with an average of 39 years old (standard deviation is 10). Seniority is reported in Figure 9. Subjects had between 1 and 39 years of research experience, with 14 years on average (standard deviation is 9 years).

The analysis of their scientific production shows that the average subject published 21 papers recorded in DBLP (Figure 10). The standard deviation of 21 publications underlines a natural variation of productivity between subjects. This is an interesting property of our sample, as it allowed us to run category-specific analyses according to their number of papers.

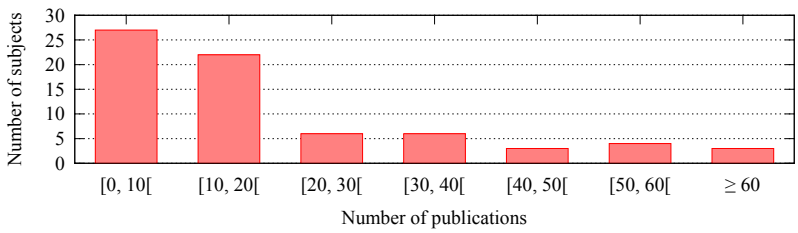

Fig. 10 Distribution of the 71 subjects according to their number of publications in DBLP

4.3.2 Effectiveness of the proposed similarity measures and improvement with social clues

We call *baseline* the SLRS producing topical results (i.e., T list). The contribution presented in this paper is the SLRS producing socio-topical results (i.e., ST list). Notice that both operate on paper metadata. Results of the experiment are reported in Table 1, while a more visual representation is offered in Figure 11.

Category	Case	Criterion	$ndcg(\mathsf{T}, r, \mathsf{HP})$	$ndcg(\mathsf{ST}, r, \mathsf{HP})$	Gain (%)
Global	①	all	0.7665	0.8316	8.49*
Number of publications	②	$<$ 15 publications	0.6960	0.7683	10.39*
	③	$\geqslant$ 15 publications	0.8313	0.8897	7.03*
Seniority	④	$<$ 13 years	0.7456	0.7941	6.50*
	⑤	$\geqslant$ 13 years	0.7857	0.8660	10.22*

Table 1 Differences in effectiveness between T and ST for the 71 subjects. Statistical significance (i.e., t-test with $p < 0.05$) is denoted by the '*' symbol in column Gain

We recall that $ndcg(\mathsf{ST}, r_i, \mathsf{HP}) \in [0,1]$ measures the effectiveness of the SLRS producing list ST (for researcher r_i) according to HP (cf. Section 4.1). Data analysis for the global level ① (i.e., considering the 71 participations) shows a strong baseline performance (0.7665). This means that issuing basic topical-based recommendations based on paper metadata (as presented in Section 3.3) is effective. Nevertheless, issuing socio-topical recommendations as we proposed in Section 3 overcomes topical recommendations, as it yields a 8.49% improvement that is statistically significant ($p < 0.05$). This significant improvement of the order of 5–10% is 'noticeable' according to Spärck Jones's (1974) interpre-

tations of levels of improvement.[9] These results validate our hypothesis: complementing topical clues with social clues for computing similarity measures leads to improved performance. In other words, the human perception of inter-researcher similarities is better simulated.

We go into our analysis in greater depth by constituting subject categories according to their number of publications (② and ③) and their seniority (④ to ⑤). Each category is divided in two equally sized groups according to the median (15 for publication count and 13 for seniority). We report the results accordingly. First, improvement is higher for researchers having published less ② than for researchers having published to a larger extent ③ (10.39% versus 7.03%). Second, improvement is higher for senior researchers ⑤ than for junior researchers ④ (10.22% versus 6.50%). Overall, we may explain these findings as follows. Junior researchers—and those who published few—had fewer collaboration opportunities with several other researchers, hence the limited effect of integrating social clues into similarity measures in this situation. On the contrary, senior researchers and those who published to a greater extent ($\geqslant$ 15 publications) had certainly more social opportunities (e.g., co-venues), hence the observed improvement.

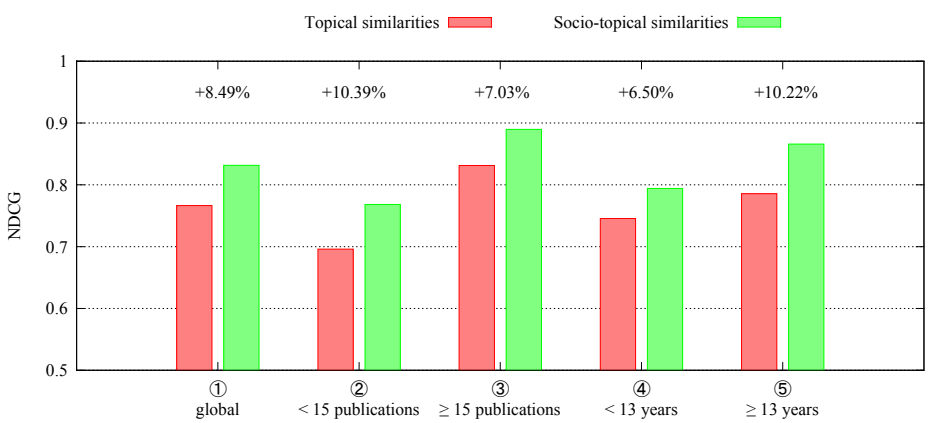

Fig. 11 Evaluation of inter-researcher similarity measures showing significant improvement of socio-topical recommendations over topical (only) recommendations

We showed that integrating social clues in inter-researcher similarity measures yields significant performance improvement for all considered categories. Improvement is all the higher as subjects are experienced researchers. Recall, however, that junior researchers (or even master students beginning with no publications yet) may also get relevant recommendations by querying the SLRS with their advisor's, supervisor's, or mentor's name.

In the next section, we analyze results in greater details in order to assess the relative influence of similarity measures on recommendation accuracy.

[9] "in the absence of significance tests, performance differences of less than 5% must be disregarded ... broadly characterize performance differences, assumed significant, as noticeable if the difference is of the order of 5–10%, and as material if it is more than 10%." (Spärck Jones 1974) as cited by Sanderson (2010, p. 313).

4.3.3 Effects of the five similarity measures on socio-topical recommendation accuracy

We studied the effects of the inter-researcher similarity measures involved in our SLRS on recommendation effectiveness. A parameter is introduced for each similarity measure used in the two-step process designed for issuing recommendations (cf. Section 3.4.2 and Figure 4). Let us consider $I = \{0.1, 0.2, \ldots, 0.9, 1.0\}$ as the domain of each parameter.

During phase one, list S is computed by combining lists P, C, and M. Originally, these are combined equitably (i.e., each list has the same importance as the two other ones). We now realize a weighted combination by introducing three parameters $(\pi, \chi, \mu) \in I^3$ such that $\pi + \chi + \mu = 1$. Each parameter controls the varying influence of the associated list by adjusting its normalized similarity values, as shown in Equation 3.

$$\mathsf{S} \leftarrow \mathit{CombMNZ}\left(\pi \cdot \mathit{normalized}\left(\mathsf{P}\right),\ \chi \cdot \mathit{normalized}\left(\mathsf{C}\right),\ \mu \cdot \mathit{normalized}\left(\mathsf{M}\right)\right) \tag{3}$$

As explained in Section 3.4.2, list S is then filtered out (i.e., S' $\leftarrow$ S $\cap$ T) such that only topically relevant researchers are kept. During phase two, list ST is computed by combining lists S' and T. Originally, these are combined equitably. Now, we introduce two parameters $(\sigma, \tau) \in I^2$ such that $\sigma + \tau = 1$. This intends to set up a varying influence for each list, as shown in Equation 4.

$$\mathsf{ST} \leftarrow \mathit{CombMNZ}\left(\sigma \cdot \mathit{normalized}\left(\mathsf{S'}\right),\ \tau \cdot \mathit{normalized}\left(\mathsf{T}\right)\right) \tag{4}$$

We generated the 726 parameter configurations complying with the aforementioned constraints on the five parameters: $\pi + \chi + \mu = 1$ and $\sigma + \tau = 1$. We computed the accuracy (*ndcg*) of each configuration and we report in Table 2 the five most effective configurations, as well as the five least effective ones. Among these 726 possible configurations, we analyzed worst and best performing configurations and report our findings below.

1. Worst performance (0.7665) is achieved when considering topical clues only (i.e., when list T is given full influence whereas list S' is dismissed). This means that integrating social clues never impaired but always improved the accuracy of recommendations, which is an interesting outcome.

Normalized CombMNZ weights					$ndcg(\mathsf{ST}, r, \mathsf{HP})$	Improvement over baseline $ndcg(\mathsf{T}, r, \mathsf{HP}) = 0.7665$ (%)
Social lists			Socio-topical lists			
P	C	M	S'	T		
π	χ	μ	σ	τ		
0.0	**0.3**	**0.7**	**0.9**	**0.1**	**0.8524**	**11.21**
0.0	0.3	0.7	1.0	0.0	0.8523	11.19
0.0	0.4	0.6	1.0	0.0	0.8521	11.17
0.0	0.4	0.6	0.9	0.1	0.8520	11.15
0.1	0.4	0.5	0.8	0.2	0.8518	11.13
0.0	0.0	1.0	0.0	1.0	0.7665	0.00
0.0	0.1	0.9	0.0	1.0	0.7665	0.00
0.0	0.2	0.8	0.0	1.0	0.7665	0.00
0.0	0.3	0.7	0.0	1.0	0.7665	0.00
0.0	0.4	0.6	0.0	1.0	0.7665	0.00

Table 2 Five most and least effective normalized CombMNZ weight configurations (5 parameters $\pi \rightarrow \tau$)

2. Best performance (0.8524) represents a material 11.21% improvement over the topical baseline (0.7665). Among the involved social and topical dimensions, the former is far more influential (0.9) than topical dimension (0.1). Note that the *meeting opportunities* function as defined in Section 3.2.3 is the most influential (0.7) parameter among the three social clues ($0.0 + 0.3$). This observation about the influence of the *meeting opportunities* social clue also holds for other best performing configurations. This is an important result as it shows that, besides being an original proposal, it is proven to be highly effective.

To sum up: the evaluation with 71 volunteer researchers allowed us to make the following three main findings. First, the usual topical similarity measure is quite accurate even when computed on paper metadata (i.e., baseline $ndcg = 0.7665$). Second, introducing social clues as we proposed in this paper yields a noticeable 8.49% performance improvement (0.8316). Third, best recommendations (0.8524) overcome the topical baseline by 11.21% when favoring social clues over topical clues. Among social clues, *meeting opportunities* are most influential.

5 Conclusion and future work

Literature review is supported by SLRSs relying on inter-researcher measures of similarity. State-of-the-art approaches implement collaborative filtering, content-based filtering, or an hybridization of these. We highlighted two key issues regarding these approaches. First, systems rely on rich theoretical models, but they are hardly feasible as they require fetching and analyzing full-text publication contents. Second, SLRSs only achieve loose simulation of researchers' cognitive process when evaluating inter-researcher similarities. For instance, social interactions between researchers (e.g., attending similar conferences) are not considered.

Our contribution is threefold. First, we proposed to process *publicly available metadata* related to publications in order to recommend researchers matching a researcher's scientific interests. Second, we designed inter-researcher similarity measures based on *topical* and *social clues*. Social clues comprise the proximity and strength of a relationship between researchers, as well as the meeting opportunities they experienced while attending conferences. Third, we implemented these inter-researcher similarity functions in our socio-topical SLRS. We then designed and conducted an evaluation with 71 volunteer researchers that allowed us to validate our proposal. We observed a strong topical baseline ($ndcg = 0.7665$) that we achieved to overcome by integrating the aforementioned social clues, yielding a 11.21% material improvement ($ndcg = 0.8524$). These results mean that socio-topical recommendations better simulate human perception of relatedness between researchers than topical recommendations alone.

Several directions for future work were identified. In the short term, we have to confirm these results with more participants stemming from a random sample (Kelly 2009, chap. 7) of researchers from various scientific domains. We should also consider more experienced researchers, as they represent a small extent of the 71 subjects we studied. Such further evaluations will also give us the opportunity to experiment with others models for computing inter-researcher similarities, such as Author-Topic models (Rosen-Zvi et al 2004, 2010) relying on Latent Dirichlet Allocation. Experimenting with other combination functions than normalized CombMNZ (Fox and Shaw 1993), such as machine learning techniques, will also be of interest.

In the medium term, we may compare recommendation accuracy when issued according to paper metadata versus full-text papers, as to assess the potential value added of full-text. Regarding the topical dimension, we may index publication titles with concepts instead of terms (e.g., for considering 'information retrieval' as a concept instead of two separate terms) and collect semantically identical expressions (e.g., 'IR,' 'information retrieval,' and 'document search'). This may be achieved with part-of-speech taggers (Janas 1977) for extracting non phrases, as well as conceptual analysis (Hurtado Martín et al 2010), or conceptual indexing (Hubert and Mothe 2009) if a domain ontology is available. Another idea is to promote innovation by the identification of pioneering researchers (i.e., those who introduced new terms massively adopted later, such as 'ontology') for promotion in recommendations. Regarding the social dimension, we may need to characterize and harness the sociability of researchers: does their publishing activity match a solitary pattern, a tribal pattern (often with the same group of coauthors), or a scattered pattern (with several changing coauthors)?

More generally, temporal clues about publications may also be worth considering for issuing relevant recommendations. For the topical dimension, researchers explore various topics or even domains and may lose interest in their initial subject. For the social dimension, the recommended researchers may not have published for a long time; such a recommendation may not be relevant for some people or tasks. Moreover, recent social interactions may have to be favored over older ones.

In the long term, we may work on a typology of researchers' needs associated with literature access. We believe that the recommendation process has to be adapted accordingly. For instance, the expected recommendations may be extremely different when we wish to be notified of events happening in our close scientific environment (i.e., awareness task) or when we wish to discover innovative authors and cutting edge subjects (i.e., task related to literature review).

Acknowledgements The constructive criticisms and suggestions of the referees are warmly acknowledged. I am also grateful to the 71 volunteer researchers who took part in the experiment reported in this paper. Their feedback, comments, and insightful advice have been a source of stimulating thinking. Finally, I am indebted to Anaïs Lefeuvre for her involvement in this work as a research assistant.

References

Adomavicius G, Tuzhilin A (2005) Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. IEEE Trans on Knowl and Data Eng 17(6):734–749, DOI 10.1109/TKDE.2005.99

Agarwal N, Haque E, Liu H, Parsons L (2005) Research paper recommender systems: A subspace clustering approach. In: Fan W, Wu Z, Yang J (eds) WAIM'05: Proceedings of the 6th International Conference on Web-Age Information Management, LNCS, vol 3739, Springer, pp 475–491, DOI 10.1007/11563952_-42

Alonso O, Rose DE, Stewart B (2008) Crowdsourcing for relevance evaluation. SIGIR Forum 42(2):9–15, DOI 10.1145/1480506.1480508

Balabanović M, Shoham Y (1997) Fab: Content-based, collaborative recommendation. Commun ACM 40(3):66–72, DOI 10.1145/245108.245124

Belkin NJ, Croft WB (1992) Information filtering and information retrieval: Two sides of the same coin? Commun ACM 35(12):29–38, DOI 10.1145/138859.138861

Ben Jabeur L, Tamine L, Boughanem M (2010) A social model for Literature Access: Towards a weighted social network of authors. In: RIAO'10: Proceedings of the 9th international conference on Information Retrieval and its Applications, CDROM

Biryukov M (2008) Co-author network analysis in DBLP: Classifying personal names. In: MCO'08: Proceedings of the 2nd international conference on Modelling, Computation and Optimization in Information

Systems and Management Sciences, Springer, Communications in Computer and Information Science, vol 14, pp 399–408, DOI 10.1007/978-3-540-87477-5_43

Bogers T, van den Bosch A (2008) Recommending Scientific Articles Using CiteULike. In: RecSys'08: Proceedings of the 4th ACM conference on Recommender systems, ACM, New York, NY, USA, pp 287–290, DOI 10.1145/1454008.1454053

Buckley C, Voorhees EM (2000) Evaluating Evaluation Measure Stability. In: SIGIR'00: Proceedings of the 23rd international ACM SIGIR conference, ACM, New York, NY, USA, pp 33–40, DOI 10.1145/345508.345543

Buckley C, Voorhees EM (2005) Retrieval System Evaluation. In: Voorhees and Harman (2005), chap 3, pp 53–75

Cazella SC, Campos Alvares LO (2005) Modeling user's opinion relevance to recommending research papers. In: UM'05: Proceedings of the 10th International Conference on User Modeling, Springer, LNCS, vol 3538, pp 327–331, DOI 10.1007/11527886_42

Cleverdon CW (1962) Report on the testing and analysis of an investigation into the comparative efficiency of indexing systems. ASLIB Cranfield Research Project, Cranfield, UK

Deng H, King I, Lyu MR (2008) Formal Models for Expert Finding on DBLP Bibliography Data. In: ICDM'08: Proceedings of the 8th IEEE International Conference on Data Mining, IEEE Computer Society, pp 163–172, DOI 10.1109/ICDM.2008.29

Dolamic L, Savoy J (2010) When stopword lists make the difference. J Am Soc Inf Sci Technol 61(1):200–203, DOI 10.1002/asi.21186

Easley D, Kleinberg J (2010) Networks, Crowds, and Markets: Reasoning About a Highly Connected World. Cambridge University Press, New York

Elmacioglu E, Lee D (2005) On Six Degrees of Separation in DBLP-DB and More. SIGMOD Rec 34(2):33–40, DOI 10.1145/1083784.1083791

Fox C (1989) A stop list for general text. SIGIR Forum 24(1-2):19–21, DOI 10.1145/378881.378888

Fox EA, Shaw JA (1993) Combination of Multiple Searches. In: Harman DK (ed) TREC-1: Proceedings of the First Text REtrieval Conference, NIST, Gaithersburg, MD, USA, pp 243–252

Garfield E (1955) Citation indexes for science: A new dimension in documentation through association of ideas. Science 122(3159):108–111, DOI 10.1126/science.122.3159.108

Garfield E (1996) What is the primordial reference for the phrase 'Publish or perish'? The Scientist 10(12):11, URL http://www.the-scientist.com/article/display/17052

Garfield E (2006) The history and meaning of the Journal Impact Factor. J Am Med Assoc 295(1):90–93, DOI 10.1001/jama.295.1.90

Glenisson P, Glänzel W, Janssens F, Moor BD (2005a) Combining full text and bibliometric information in mapping scientific disciplines. Inf Process Manage 41(6):1548–1572, DOI 10.1016/j.ipm.2005.03.021

Glenisson P, Glänzel W, Persson O (2005b) Combining full-text analysis and bibliometric indicators. a pilot study. Scientometr 63(1):163–180, DOI 10.1007/s11192-005-0208-0

Goldberg D, Nichols D, Oki BM, Terry DB (1992) Using collaborative filtering to weave an Information Tapestry. Commun ACM 35(12):61–70, DOI 10.1145/138859.138867

Gori M, Pucci A (2006) Research paper recommender systems: A random-walk based approach. In: WI'06: Proceedings of the 5th IEEE/WIC/ACM International Conference on Web Intelligence, IEEE Computer Society, Los Alamitos, CA, USA, pp 778–781, DOI 10.1109/WI.2006.149

Herlocker JL, Konstan JA, Terveen LG, Riedl JT (2004) Evaluating collaborative filtering recommender systems. ACM Trans Inf Syst 22(1):5–53, DOI 10.1145/963770.963772

Hirsch JE (2005) An index to quantify an individual's scientific research output. Proc Natl Acad Sci USA 102(46):16,569–16,572, DOI 10.1073/pnas.0507655102

Hirsch JE (2010) An index to quantify an individual's scientific research output that takes into account the effect of multiple coauthorship. Scientometr 85(3):741–754, DOI 10.1007/s11192-010-0193-9

Huang Z, Yan Y, Qiu Y, Qiao S (2009) Exploring Emergent Semantic Communities from DBLP Bibliography Database. In: Memon N, Alhajj R (eds) ASONAM'09: Proceedings of the 1st International Conference on Advances in Social Network Analysis and Mining, IEEE Computer Society, pp 219–224, DOI ASONAM.2009.6

Hubert G, Mothe J (2009) An adaptable search engine for multimodal information retrieval. J Am Soc Inf Sci Technol 60(8):1625–1634, DOI 10.1002/asi.21091

Hull D (1993) Using Statistical Testing in the Evaluation of Retrieval Experiments. In: SIGIR'93: Proceedings of the 16th annual international ACM SIGIR conference, ACM Press, New York, NY, USA, pp 329–338, DOI 10.1145/160688.160758

Hurtado Martín G, Cornelis C, Naessens H (2009) Training a Personal Alert System for Research Information Recommendation. In: Carvalho JP, Dubois D, Kaymak U, Sousa JMC (eds) IFSA/EUSFLAT'09: Proceedings of the Joint 2009 International Fuzzy Systems Association World Congress and 2009 European

Society of Fuzzy Logic and Technology Conference, pp 408–413
Hurtado Martín G, Schockaert S, Cornelis C, Naessens H (2010) Metadata impact on research paper similarity. In: Lalmas M, Jose J, Rauber A, Sebastiani F, Frommholz I (eds) ECDL'10: Proceedings of the 14th European Conference on Research and Advanced Technology for Digital Libraries, LNCS, vol 6273, Springer, pp 457–460, DOI 10.1007/978-3-642-15464-5_56
Janas JM (1977) Automatic recognition of the part-of-speech for english texts. Inf Process Manage 13(4):205–213, DOI 10.1016/0306-4573(77)90001-2
Järvelin K, Kekäläinen J (2002) Cumulated gain-based evaluation of IR techniques. ACM Trans Inf Syst 20(4):422–446, DOI 10.1145/582415.582418
Karoui H, Kanawati R, Petrucci L (2006) COBRAS: Cooperative CBR System for Bibliographical Reference Recommendation. In: ECCBR'06: Proceedings of the 8th European Conference on Advances in Case-Based Reasoning, Springer, LNCS, vol 4106, pp 76–90, DOI 10.1007/11805816_8
Kelly D (2009) Methods for evaluating interactive information retrieval systems with users. Found Trends Inf Retr 3(1–2):1–224, DOI 10.1561/1500000012
Klas CP, Fuhr N (2000) A new Effective Approach for Categorizing Web Documents. In: Proceedings of the 22th BCS-IRSG Colloquium on IR Research
Lee JH (1997) Analyses of Multiple Evidence Combination. In: SIGIR'97: Proceedings of the 20th annual international ACM SIGIR conference, ACM Press, New York, NY, USA, pp 267–276, DOI 10.1145/258525.258587
Ley M (2002) The DBLP computer science bibliography: Evolution, research issues, perspectives. In: Laender AHF, Oliveira AL (eds) SPIRE'02 : Proceedings of the 9th international conference on String Processing and Information Retrieval, Springer, LNCS, vol 2476, pp 1–10, DOI 10.1007/3-540-45735-6_1
Likert R (1932) A technique for the measurement of attitudes. Arch Psychol 22(140):5–54
Manning CD, Raghavan P, Schütze H (2008) Introduction to Information Retrieval. Cambridge University Press
McNee SM, Albert I, Cosley D, Gopalkrishnan P, Lam SK, Rashid AM, Konstan JA, Riedl J (2002) On the recommending of citations for research papers. In: CSCW'02: Proceedings of the 2002 ACM conference on Computer supported cooperative work, ACM, New York, NY, USA, pp 116–125, DOI 10.1145/587078.587096
McNee SM, Kapoor N, Konstan JA (2006) Don't look stupid: avoiding pitfalls when recommending research papers. In: CSCW '06: Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work, ACM, New York, NY, USA, pp 171–180, DOI 10.1145/1180875.1180903
Micarelli A, Sciarrone F, Marinilli M (2007) Web Document Modeling, LNCS, vol 4321, Springer, pp 155–192. DOI 10.1007/978-3-540-72079-9_5
Milgram S (1967) The small-world problem. Psychology Today 1(1):61–67
Mimno D, McCallum A (2007) Mining a digital library for influential authors. In: JCDL'07: Proceedings of the 7th ACM/IEEE-CS joint conference on Digital libraries, ACM, New York, NY, USA, pp 105–106, DOI 10.1145/1255175.1255196
Mittelbach F, Goossens M (2005) LaTeX Companion, 2nd edn. Pearson Education
Montaner M, López B, de la Rosa JL (2003) A Taxonomy of Recommender Agents on the Internet. Artif Intell Rev 19(4):285–330, DOI 10.1023/A:1022850703159
Naak A, Hage H, Aïmeur E (2009) A Multi-criteria Collaborative Filtering Approach for Research Paper Recommendation in Papyres. In: MCETECH'09: Proceedings of the 4th International Conference on E-Technologies: Innovation in an Open World, Springer, LNBIP, vol 26, pp 25–39, DOI 10.1007/978-3-642-01187-0_3
Porcel C, López-Herrera AG, Herrera-Viedma E (2009) A recommender system for research resources based on fuzzy linguistic modeling. Expert Syst Appl 36(3):5173–5183, DOI 10.1016/j.eswa.2008.06.038
Porter MF (1980) An algorithm for suffix stripping. Program 14(3):130–137
Powley B, Dale R (2007) Evidence-Based Information Extraction for High Accuracy Citation and Author Name Identification. In: RIAO'07: Proceedings of the 8th conference on Information Retrieval and its Applications, CID, CDROM
Reips UD (2002) Standards for Internet-Based Experimenting. Exp Psychol 49(4):243–256, DOI 10.1026//1618-3169.49.4.243
Reips UD (2007) The methodology of Internet-based experiments. In: Joinson AN, McKenna KYA, Postmes T, Reips UD (eds) The Oxford Handbook of Internet Psychology, Oxford University Press, New York, NY, USA, chap 24, pp 373–390
Reips UD, Lengler R (2005) The *Web Experiment List*: A Web service for the recruitment of participants and archiving of Internet-based experiments. Behav Res Meth 37(2):287–292
Reitz F, Hoffmann O (2010) An Analysis of the Evolving Coverage of Computer Science Sub-fields in the DBLP Digital Library. In: Lalmas M, Jose J, Rauber A, Sebastiani F, Frommholz I (eds) ECDL'10: Pro-

ceedings of the 14th European Conference on Research and Advanced Technology for Digital Libraries, LNCS, vol 6273, Springer, pp 216–227, DOI 10.1007/978-3-642-15464-5_23
Resnick P, Varian HR (1997) Recommender systems. Commun ACM 40(3):56–58, DOI 10.1145/245108.245121
Rosen-Zvi M, Griffiths T, Steyvers M, Smyth P (2004) The Author-Topic Model for Authors and Documents. In: UAI'04: Proceedings of the 20th Annual Conference on Uncertainty in Artificial Intelligence, AUAI Press, Arlington, Virginia, pp 487–494
Rosen-Zvi M, Chemudugunta C, Griffiths T, Smyth P, Steyvers M (2010) Learning author-topic models from text corpora. ACM Trans Inf Syst 28(1):4:1–4:38, DOI 10.1145/1658377.1658381
Salton G, Buckley C (1988) Term-weighting approaches in automatic text retrieval. Inf Process Manage 24(5):513–523, DOI 10.1016/0306-4573(88)90021-0
Salton G, Wong A, Yang CS (1975) A vector space model for automatic indexing. Commun ACM 18(11):613–620, DOI 10.1145/361219.361220
Sanderson M (2010) Test collection based evaluation of information retrieval systems. Found Trends Inf Retr 4(4):247–375, DOI 10.1561/1500000009
Sanderson M, Zobel J (2005) Information Retrieval System Evaluation: Effort, Sensitivity, and Reliability. In: SIGIR'05: Proceedings of the 28th annual international ACM SIGIR conference, ACM, New York, NY, USA, pp 162–169, DOI 10.1145/1076034.1076064
Spärck Jones K (1973) Index term weighting. Inform Stor Retr 9(11):619–633, DOI 10.1016/0020-0271(73)90043-0
Spärck Jones K (1974) Automatic indexing. J Doc 30(4):393–432, DOI 10.1108/eb026588
Student (1908) The probable error of a mean. Biometrika 6(1):1–25, DOI 10.2307/2331554
Tang J, Zhang J, Yao L, Li J, Zhang L, Su Z (2008) ArnetMiner: Extraction and Mining of Academic Social Networks. In: KDD'08: Proceeding of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining, ACM, New York, NY, USA, pp 990–998, DOI 10.1145/1401890.1402008
Travers J, Milgram S (1969) An experimental study of the small world problem. Sociometry 32(4):425–443, DOI 10.2307/2786545
Tsatsaronis G, Varlamis I, Stamou S, Nørvåg K, Vazirgiannis M (2009) Semantic relatedness hits bibliographic data. In: WIDM'09: Proceeding of the 11th international workshop on Web information and data management, ACM, New York, NY, USA, pp 87–90, DOI 10.1145/1651587.1651607
Voorhees EM (2002) The philosophy of information retrieval evaluation. In: Peters C, Braschler M, Gonzalo J, Kluck M (eds) CLEF'01: Second Workshop of the Cross-Language Evaluation Forum, Springer, LNCS, vol 2406, pp 355–370, DOI 10.1007/3-540-45691-0_34
Voorhees EM, Harman DK (2005) TREC: Experiment and Evaluation in Information Retrieval. MIT Press, Cambridge, MA, USA
Wilcoxon F (1945) Individual comparisons by ranking methods. Biom Bull 1(6):80–83, DOI 10.2307/3001968
Yan E, Ding Y (2009) Applying centrality measures to impact analysis: A coauthorship network analysis. J Am Soc Inf Sci Technol 60(10):2107–2118, DOI 10.1002/asi.21128
Yang Z, Hong L, Davison BD (2010) Topic-driven Multi-type Citation Network Analysis. In: RIAO'10: Proceedings of the 9th international conference on Information Retrieval and its Applications, CDROM
Zamparelli R (1998) Internet publications: Pay-per-use or pay-per-subscription? In: Nikolaou C, Stephanidis C (eds) ECDL'98: Proceedings of the 2nd European Conference on Research and Advanced Technology for Digital Libraries, LNCS, vol 1513, Springer, pp 635–636, DOI 10.1007/3-540-49653-X_38
Zhou D, Orshanskiy SA, Zha H, Giles CL (2007) Co-ranking authors and documents in a heterogeneous network. In: ICDM'07: Proceedings of the 7th IEEE International Conference on Data Mining, pp 739–744, DOI 10.1109/ICDM.2007.57