Understanding and supporting search for scholarly knowledge

Marcos Baez, Daniil Mirylenka, and Cristhian Parra

Dipartimento di Ingegneria e Scienza dell'Informazione, University of Trento, Italy
Via Sommarive, 14 I-38123 POVO
{baez, dmirylenka, parra}@disi.unitn.it

Abstract. In the last decade, scholarly communication have been greatly transformed by the web, moving research dissemination away from printed papers in journals to digital content that can be easily posted on the Internet. This technical factor along with a larger scientific community makes it really hard to find relevant content for research in the ever growing sea of publications. With the goal of gaining insights on how researchers find relevant knowledge, we have interviewed a small group of researchers and then opened an online survey to a larger group, asking them to explain how they had found references for one of their papers. The results of this study suggest that finding scientific knowledge has a strong social component, with the different researchers' social networks (e.g., coauthors, people met at conferences) accounting for a important percentage of the source of the references. In this paper we report on this study and compare our results with the evidence found in a dataset of 5×10^6 authors with their publications and references. We take these results and analyze different approaches for incorporating the social component into search and recommendation of scientific publications.

Keywords: search, social search, scientific knowledge

1 Introduction

The notion of scientific paper as the main means of scientific knowledge dissemination and peer review as the main mechanism to guarantee quality have been, for a long time, the cornerstones of scientific knowledge advance. In the last decades the scientific world has met great changes, the Web being the changing factor pushing us to gradually move away from printed papers in journals to digital content that can be easily posted online. In this context it becomes really hard to find relevant content for research in the ever growing sea of publications. This phenomenon, referred to as "information overload", is a reality and a challenge for the scientific community. It requires an understanding of the problem in this domain and the development of models and tools to overcome its effect.

Motivated by this challenge, we started to study how researchers find scientific knowledge looking for ways to improve the support of this process, taking as a particular use case the problem of finding relevant references. In our study we have found that a third of all the references cited in a scientific paper comes from the authors' interaction with their social networks, including co-authors, project colleagues, and people met at conferences or other events. Researchers stumble upon relevant scientific resources and share them *within* and *among* these different social networks in different contexts. This comes to support the observation that "networking" is highly important in our community. This strong social component, however, has not been fully exploited to overcome the information overload problem.

Current approaches, such as social bookmarking sites and other social networking systems provide tools and services to share the knowledge within a single context of their systems. However, they do not consider that different incentives and tools are required to effectively capture this knowledge in different contexts. People share, discover and discuss papers, usually informally, at conferences, lectures, in the mailing lists of their research groups and projects. We argue that capturing this knowledge will allow us to understand what people consider important and relevant. The fact, for example, that somebody (and especially somebody we "trust") shares a paper tells us a lot on the value of this paper, more than a citation can do. This fact supports our intuition that, having this kind of information available, we will be able to use it to improve search [5].

In this paper we introduce Knowledge Spaces, an approach to capturing and supporting search for scholarly knowledge. Knowledge spaces (kspaces for short) are a metaphor, a set of models and processes, and a social web platform that help you capture, share and find scientific knowledge in all of its forms. The principle behind kspaces is to allow knowledge dissemination in the scientific community to occur in a way similar to the way we share knowledge with our colleagues in informal settings. The rationale behind this is that when we interact informally with a small team of colleagues dissemination is very effective. We are free to choose the best format for communicating our thoughts and results, we share both established results as well as latest ideas, we interact and carry on a conversation (synchronously or via email), we comment on other people's contributions and papers and observe relations among various contributions.

This is not in this paper (how the reviewers didn't noticed it)

As regards search, we present our preliminary work towards exploiting this social aspect. Firstly, we discuss how to reuse the valuable social metadata already available on the Web in a scientific metasearch engine. Then we give some insights on using researcher's social network for search and recommendation of scholarly publications.

The contributions of this work are the following:

- A study on the impact of researchers' social networks in finding references for publications, and on the importance of different kinds of networks,
- ~~A model and a social platform for capturing knowledge and enabling search,~~
- A scientific metasearch engine leveraging social metrics, and
- Initial ideas on using researchers' social networks, and in particular co-authorship network, for personalized search for scholarly publications.

In what follows we present our study on understanding scholarly knowledge search, describe Knowledge Spaces as its enabling factor, and describe our proposals of scientific metasearch and network-aware search.

2 Understanding search

If we are to improve the way we find scientific knowledge, the very first step to do so is to understand how this process naturally works in the mind of those who search. With the goal of reaching this understanding we have conducted a sociological study consisting of two phases:

1. **Qualitative Analysis** of researchers' comments on how they find scientific knowledge they later cite, with the goal of finding an small set of categories in which we could classify this process. For this purpose, we have interviewed 30 researchers from the University of Trento, asking them to explain how they had found references for one particular publication of their authorship. The answers to the question ranged from "my advisor suggested it" to "I searched for papers in topic X using google scholar" and "I found it while following citation links".
2. **Quantitative Analysis** of an online survey on the same subject, using the categories we have found on the first phase of the study. The online survey[1] would ask researchers to provide their names, which we would later use to search for their publications on a dataset extracted from Microsoft Academic Search[2] of 5×10^6 authors with their publications and references. After selecting one publication, the survey ask the same question as the interview, but this time providing as optional answers the categories we have found on the previous phase.

The results we discuss on the following sub-sections, led us to the conclusion that finding scientific knowledge has a strong social component, which is a clear motivation to incorporate this component in the way we search for scientific knowledge.

2.1 Qualitative Analysis Results

Based on the interviews, we have run a qualitative analysis by classifying textual transcripts for each analysed reference. From this analysis, we have classified the sources of references into three main categories:

– **social:** includes all the references that authors came to know thanks to the interaction with their social network including co-authors, project colleagues, people met at conferences or other events;

[1] The survey is still available at http://survey.mateine.org/
[2] http://academic.research.microsoft.com/

– **keyword search:** includes all the references found while searching for some topics or keywords using tools for that purpose (e.g. google scholar, dblp, specific digital libraries);
– **navigation:** includes all the references found by following citations or other type of references in papers and other resources.

Figure 1 shows the average percentage of references in a paper that follows each of the patterns explained before. In general, the same proportion of references comes from a social network of the authors and from specific searches they run on their own, reaching a 38%. References they got from navigating through knowledge represent the remaining 24%. When divided by seniority, results hold the same trend, with the social scoring higher than search both for professors and postdocs, and lower only for Ph.D. students. The later can be seen as a very intuitive result taking into account that the academic social network of a Ph.D. student is typically smaller than that of a senior researcher.

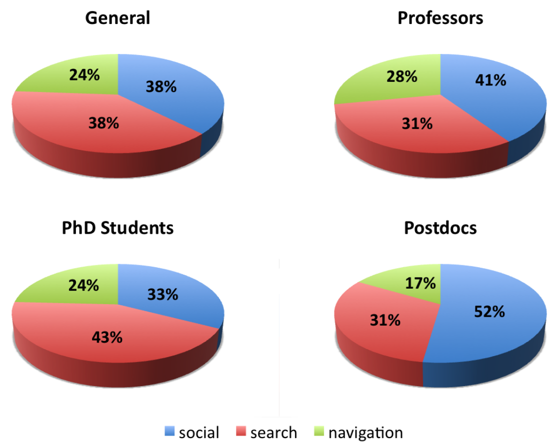

Fig. 1. Sources of references in a research paper, classified by seniority of the researcher

Furthermore, and given the high percentage of social related references, we have also classified the networks most commonly mentioned as sources of references. Figure 2 shows how many of the social references correspond to each of the following networks:

– **community/field** includes references that come from people or projects that can be considered as part of the same field or community around a certain topic, but whom the author has not necessarily met;
– **colleagues**, including peers whom the author has appointed as such. This is a very general term used by most of the interviewees and that might have a high intersection with other networks;

– **venues** includes references the author came to know through conferences or journals.
– **collaboration**, including people, groups or projects with whom the author has directly collaborated;
– **senior colleagues** includes mainly advisers or experts in a specific field;
– **coauthors**, including people who coauthored at least one article with the interviewee;
– **research group** includes people working in the author's department or research group;
– **acquaintances**, including people the author has personally met, which is also a general term used by interviewees that could also be included in other networks;
– **friends** includes people specifically appointed as such by the interviewee
– **educational** includes references the author got from courses, classmates or education related networks;

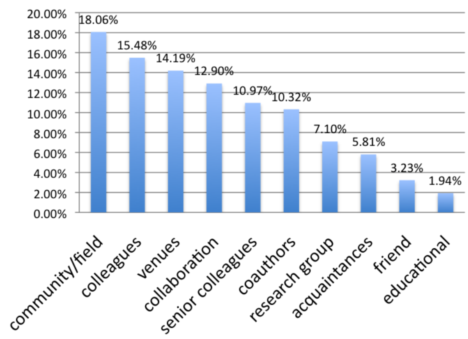

Fig. 2. Social networks of origin for references

In the same way, most citations found by navigation, were found while reading a paper or book, and then going deeper in the citation graph. Other subcategories of the navigation pattern include the follow up of one particular author, journal, conference or project. As for references found by keyword search, google and google scholar are the most used engines, while also dblp was mentioned (probably, due to the high number of computer scientist in the group of interviewees).

At the end of the study, we decided also to ask which of the analyzed references in the interview they liked the most. Even though this question was not in the original interview script, the trend we have found and that would later be confirmed by the online survey is that most liked references come mainly from authors' social networks, accounting up to a 41%.

Data: The analysis of this phase of the study is based on 351 references (without counting 43 self citations) spread across 30 interviews (18 Ph.D. students, 8 postdocs, 4 professors). Of these, 214 correspond to Ph.D. students, 64 are from professors and 71 from postdocs. We have removed self-citations from the analysis as the authors already knew them and had no need to find them. The interviews were conducted during May of 2011 resulting in 789 different notes that were later manually categorized to get the before mentioned results.

2.2 Quantitative Analysis Results

The second phase of our preliminary study consisted on conducting an online automated version of our interview, using the categories we have found on the first analysis.

Figure 3 shows the average percentage of references in a paper in each category. The numbers are similar to those of the first phase, with the exception of a significant drop in the percentage corresponding to navigation while the number of references for which authors selected the option **Do not remember** increased dramatically (especially for professors).

The reason behind the decrease in the navigation references percentage could be that the explanation in the online survey was not clear enough. Other reason might be that we are missing an important category, which we expect to discover as more people participate of the survey and provide feedback on possible missing categories.

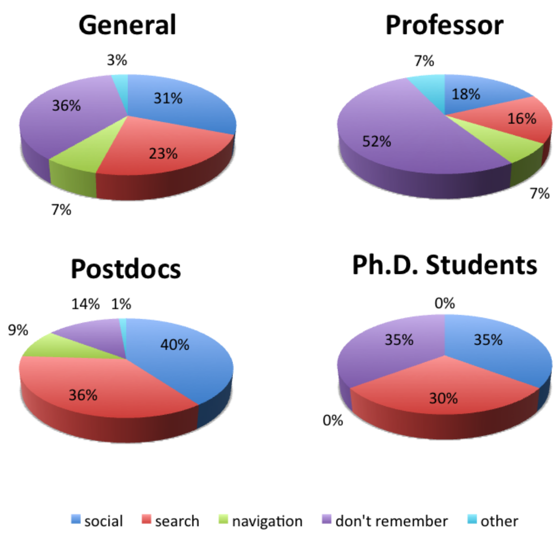

Fig. 3. Results of our online survey, classified by seniority of the researcher

For each of category, the online survey also included the option to further detail the answer by indicating (a) the social network from where the reference came from (for the social category), (b) the search engine or repository (for keyword search), and (c) while navigating what (for navigation)

Figure 4 shows the percentage of social references by social network. Although the community is again the mos important network, there is a significant increase in the percentage corresponding to coauthorship, which is not in line with the first phase analysis, implying that more research needs to be done in order to improve our understanding about the relevance of each of the many different social networks of a researcher. This however, we have gained interesting insights about which networks are the ones we have to investigate further.

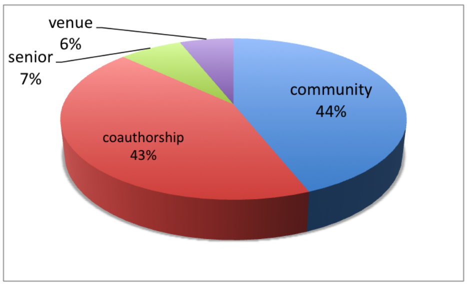

Fig. 4. Social references by social network

More details and results of this analysis are available online and will be constantly updated in the site of the survey[3].

Data: The online survey, currently ongoing, has gotten to this date, the reply of 28 different researchers, that responded about the source for a 226 references distributed over 23 different publications. Aggregating these responses confirm what we have already found in the interviews: the social component is stronger than any other.

3 Defining Network-aware recommendations

The goal of this work is to incorporate the social component of knowledge discovery into recommendation of scientific publications. More specifically, we aim to build a recommender system that suggests publications that researchers are likely to find through their social networks and that are related to their work.

Our model represents the graph of researchers and scientific publications connected with the relations of authorship and citation. We formalize the problem definition in the rest of the section.

[3] http://survey.mateine.org/results

3.1 Formal problem definition

Let $\mathcal{R}$ be the set of all researchers in the system, and $\mathcal{P}$ - the set of all publications. Relation $\mathsf{Authored}$ is defined by the set of pairs $(researcher, publication) \in \mathcal{R} \times \mathcal{P}$ such that *researcher* authored *publication*. Similarly, Cited includes all pairs of publications $(citing, cited) \in \mathcal{P}^2$ such that *citing* references *cited*. Publication p is cited by researcher r if it is cited by any paper of r:

$$\mathsf{Cited}(r, p) \iff \exists p' \in \mathcal{P}\,(\mathsf{Authored}(r, p') \wedge \mathsf{Cited}(p', p))\,.$$

For each researcher r we define a set of known and a set of unknown publications:

$$\mathsf{Known}(r) = \{p \in \mathcal{P} \mid \mathsf{Authored}(r, p) \vee \mathsf{Cited}(r, p)\}\,,$$

$$\mathsf{Unknown}(r) = \mathcal{P} \setminus \mathsf{Known}(r).$$

With each researcher r we also associate a network, which is set of researchers similar to r according to some similarity function:

$$\mathsf{Network}(r) = \{r' \in \mathcal{R} \mid \mathrm{sim}(r, r') > \delta\}\,.$$

Examples of such a network may include coauthors of r, or researchers publishing in the same conferences, or researchers citing the same papers.

The popularity of a publication p within a set of researchers s is defined as $\mathsf{popularity}(p, s)$ and its definition depends on the particular recommendation strategy. We introduce some strategies in the next subsection.

Given the definitions above, we formulate the problem of social recommendation of scientific publications: For a given researcher, find k publications unknown to him/her that have the highest popularity in his/her network:

1. $\mathsf{Rec}(r, k) \subseteq \mathsf{Unknown}(r)$,
2. $|\mathsf{Rec}(r, k)| = k$,
3. $\forall p \in \mathsf{Rec}(r, k)\ \forall p' \in (\mathsf{Unknown}(r) \setminus \mathsf{Rec}(r, k))$
   $(\mathsf{popularity}(p, \mathsf{Network}(r)) \geq \mathsf{popularity}(p', \mathsf{Network}(r)))$.

In what follows we explore different definitions of network and popularity to later evaluate and analyze their performance.

3.2 Defining the notion of network

Recommendations for scholarly knowledge depend on the context and the goal the user is trying to achieve. They are also strongly related to the type of network and the algorithms used to compute them. In this section we focus on defining different network configurations around researchers.

Coauthorship network We first introduce the co-authorship network based on our definition of network :

$$\mathsf{Coauthors}(r) = \{r' \in \mathcal{R} \mid \mathrm{sim}(r, r') > \delta\}$$

expressing in the similarity function the number of papers two researchers have written together normalized by the number of publications, and then applying $\delta > 0$ to create the network of all the coauthors a given researcher:

$$\mathsf{sim}(r, r') = \frac{\|\mathsf{Publications}(r) \cap \mathsf{Publications}(r')\|}{\|\mathsf{Publications}(r)\|}$$

where

$$\mathsf{Publications}(r) = \bigcup \{p \in \mathcal{P} \mid \mathsf{Auhtored}(r, p)\} .$$

Venue network Our definition of venue network tries to capture the likelihood of two researchers meeting at a venue, resulting in a future citation. We define the relation $\mathsf{VenueOf}$ as a set of pairs (*publication*, *venue*) $\in \mathcal{P} \times \mathcal{V}$:

$$\mathsf{Copublished}(r) = \{r' \in \mathcal{R} \mid \mathrm{sim}(r, r') > \delta\}$$

where the similarity function expresses the normalized number of venues two researchers have published together:

$$\mathsf{sim}(r, r') = \frac{\|\mathsf{Venues}(r) \cap \mathsf{Venues}(r')\|}{\|\mathsf{Venues}(r)\|}$$

given that

$$\mathsf{Venues}(r) = \bigcup \{\mathsf{VenueOf}(p \in \mathcal{P}) \mid \mathsf{Auhtored}(r, p)\} .$$

Topic network Topic-based network capture the notion of researchers working in the same field. We assume each publication p belongs to a set of topics $\mathsf{Topics}(p)$ thus:

$$\mathsf{Co-topic}(r) = \{r' \in \mathcal{R} \mid \mathrm{sim}(r, r') > \delta\}$$

where the similarity function expresses the normalized number of topics on which two researchers have both published:

$$\mathsf{sim}(r, r') = \frac{\|\mathsf{Afinity}(r) \cap \mathsf{Afinity}(r')\|}{\|\mathsf{Afinity}(r)\|}$$

given that

$$\mathsf{Afinity}(r) = \bigcup \{\mathsf{Topics}(p \in \mathcal{P}) \mid \mathsf{Auhtored}(r, p)\} .$$

3.3 Defining popularity functions

On the above we define different popularity functions that explore different views on the importance of a publication in the researcher's network:

- *network popularity*: The popularity of a publication p within the network of a researcher r is defined as the number of researchers in p who either authored or cited p:
$$\mathsf{pop_n}(p, r) = \frac{\|\{r' \in \mathsf{Network}(r) \mid p \in \mathsf{Known}(r')\}\|}{\|\mathsf{Network}(r)\|}.$$
- *work-weighted network popularity*: Expresses the popularity of a publication p in the network of a researcher r, weighted by her similarity with all researchers in the network:
$$\mathsf{pop_{wk}}(p, r) = \frac{\sum \{\mathsf{sim}(r, r' \in \mathsf{Network}(r)) \mid p \in \mathsf{Known}(r')\}}{\|\mathsf{Network}(r)\|}.$$
- *time-weighted network popularity*: Expresses the popularity of a publication p in the network of a researcher r, weighted by temporal similarity with other researchers in the network, considering the range $[y_{min}, y_{max}]$:
$$\mathsf{pop_{wt}}(p, r) = \frac{\sum \{\mathsf{sim_t}(r, r' \in \mathsf{Network}(r)) \mid p \in \mathsf{Known}(r')\}}{\|\mathsf{Network}(r)\|}$$
- *overall popularity*: The overall popularity of a publication p is defined as the number of all researchers who either authored or cited p:
$$\mathsf{popularity_o}(p) = \frac{\|\{r \in \mathcal{R} \mid p \in \mathsf{Known}(r)\}\|}{\|\mathcal{R}\|}.$$

Recommending for a topic The problem definition formulated above can be extended to the case where the recommendations are restricted to specific topics of interest. We assume each publication p belongs to a set of topics $\mathsf{Topics}(p)$. Let $\mathsf{FilteredBy}(ts)$ be a set of publications belonging to at least one topic from ts:
$$\mathsf{FilteredBy}(ts) = \{p \in \mathcal{P} \mid \mathsf{Topics}(p) \cap ts \neq\}\,.$$
For a given researcher r and a set of topics ts, we need to find a set of publications $\mathsf{Rec}(r, ts, k)$ such that

1. $\mathsf{Rec}(r, ts, k) \subseteq \mathsf{FilteredBy}(ts) \cap \mathsf{Unknown}(r)$,
2. $|\mathsf{Rec}(r, ts, k)| = k$,
3. $\forall p \in \mathsf{Rec}(r, ts, k)$
   $\forall p' \in (\mathsf{FilteredBy}(ts) \cap \mathsf{Unknown}(r)) \setminus \mathsf{Rec}(r, ts, k)$
   $(\mathsf{popularity}(p, \mathsf{Network}(r)) \geq \mathsf{popularity}(p', \mathsf{Network}(r)))$.

This extended problem definition will make possible the recommendations on a topic and the implementation of a network-aware search for scientific publications. This should be accomplished by mapping the search query specified by the user to the set of topics, and recommending the publications for this set of topics based on the user's network. In this work, however, we don't address the problem of topic-based recommendation.

4 Evaluation

In this section we present the evaluation of the social recommendations we introduced in the previous section.

4.1 Experiment definition

We obtained a crawled copy of the academic search database[4] containing data about publications, their authors and citation relations between them.

Our goal was to evaluate the ability of our recommender system to produce relevant recommendations for researchers depending on different popularity functions introduced in Section 3.3.

For the purpose of this experiment we assumed citation to be the indication of relevance. In other words, we considered publication p to be relevant for a researcher r at some moment in the year y if r cited p after the year y. We then evaluated the precision and recall of our recommendations. This evaluation was done by measuring how well our algorithms predicted researchers' citations after the year y based on citations of their network before that year. For a random sample of 1000 researchers we ran the experiment for different combinations of year y (ranging from 1999 to 2009), number of produced recommendations (1, 2, 4, 8 and so forth, following an exponential growth up to 512), and popularity function (Section 3.3), averaging the precision and recall metrics over the researchers in the sample.

4.2 Dataset Description

Our dataset contained 7 465 398 unique publications written by 5 726 226 different authors. As the design of our experiment required the year of publication to be known, we selected 3 937 907 papers for which this information was available. In order to improve the dataset, we approximated the publication year for 1 907 589 more publications by taking the maximal year of their references. Hence the total number of publications participating in experiment was 5 845 496.

4.3 Running and analyzing the experiments

Before running the experiment as described in the previous subsection, we analyzed the inherent quality of the coauthorship network as source of recommendations. According to our dataset, 20% of future citations for a researcher overlaps with the past citations from her coauthorship network. This percentage represents the maximum recall that any of our popularity functions could achieve.

In Figure 2 and 3 we present the precision and recall for each popularity function and year-cut.

In all methods the tendency points to a decrease in the precision by the year. This owes in part to the distribution of the dataset, but more importantly to

[4] http://academic.research.microsoft.com/

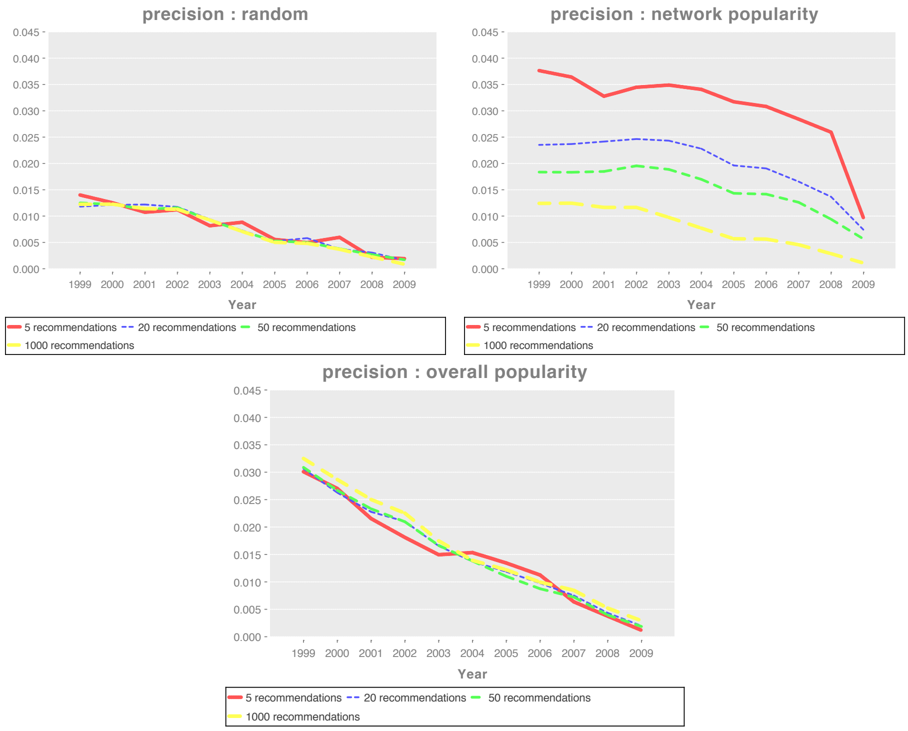

Fig. 5. Precision of the different popularity functions by year

the fact that the set of future citation declines with the year, decreasing the maximal number of relevant publications. In the case of the recall, the results are not necessarily related to the year, but to the number of recommendations.

Analyzing the performance in terms of number of recommendations, we can see that the precision of network-aware popularity gets better as the number of recommendation decreases. It means that the papers most cited by the researcher's network are much more likely to be relevant to the researcher. This effect requires further investigation in order to be fully explained. However, our preliminary hypothesis is that it may be due to the fact that the most cited papers in the researchers network belong to the topics relevant to the whole community of this network (which explains many citations) and thus likely to be relevant to the researcher, while the less-cited papers have topics relevant only to a part of the community (therefore, having smaller number of citations) to which the researcher is less likely to belong. This also explains why network-aware popularity outperforms random and overall popularity, especially for the small numbers of recommendations.

The random popularity shows very low precision but we can see that it is not sensitive to the number of recommendations. This can be explained by the

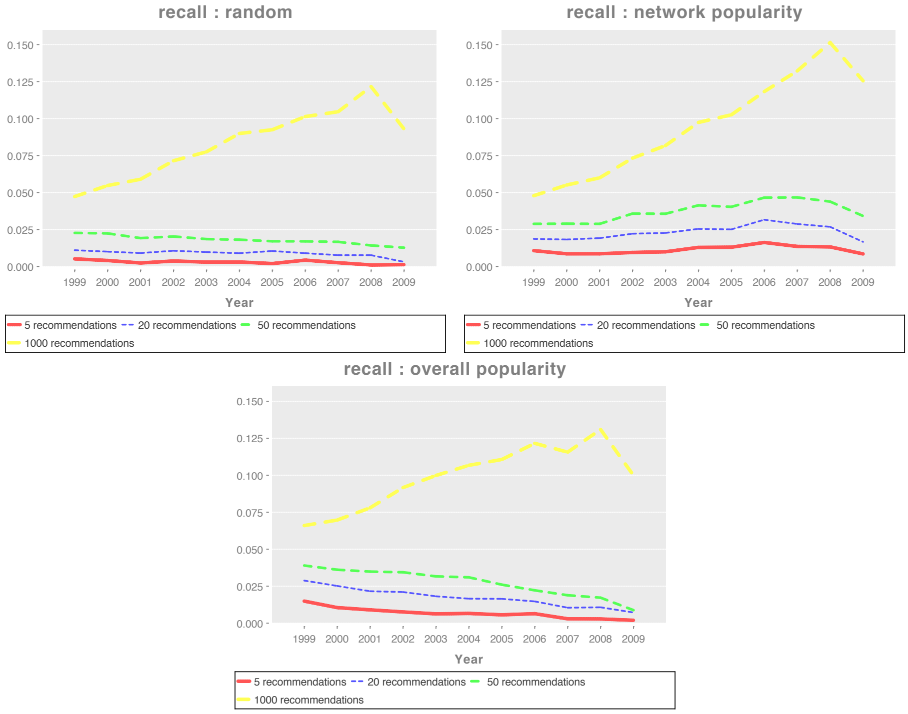

Fig. 6. Recall of the different popularity functions by year

fact that percentage of relevant papers in the random sample of papers does not depend on the sample size. The overall popularity function seems insensitive to the number of recommendations in terms of precision too, while explaining this phenomenon needs further analysis. We can also see that there is always a quality measure in the selection of references, and this becomes evident in how the overall popularity outperforms the random popularity.

The recall of the three described popularity metrics shows the strong tendency to grow as the number of recommendation increases, with the network-aware recommendation generally outperforming the other two methods. The difference between the performances of the three methods in terms of recall is the largest on the small number of recommendations and almost vanishes as the number of recommended papers reaches 1000. This is due to the fact that the different rankings of the publications (known to researcher's network) do not change set of recommendations as the number of recommended papers tends to the total number of papers in the network.

Finally, the fact that the popularity in the network performs better than the overall is an indicator of the importance of considering the network.

4.4 Discussion

Our results allow us to infer that the social awareness approach can provide some improvement in the recommendation of scholarly publications, but further exploration is still needed to understand how different networks would affect the results and, in particular, which of these networks (or combination of) have the best recommending power. It is interesting to see that the average recall of our co-authorship network-aware algorithm (being in the range of 1 and 15 percent) is similar to the percentage of papers coming from this network according to our study (almost 11% of the overall 40% of social-originated references).

Furthermore, a major improvement to be made is to include topic analysis within our recommendation logic. Our intuition is that such a logic can generate a better and more relevant final ranking of resources. More experiments also need to be performed to find relevant citation patterns in the networks we have available in our dataset.

Finally, the ranking scores we have used in this work were chosen for their relative straightforward implementation, while more complex ranking scores remain untested. Ranking recommendations following some notions of weighted co-authorship (based either on the number of coauthored resources or the recency of the last collaboration) might provide better results. Unfortunately due to both a lack of time and the characteristics of the available dataset, such rankings remain untested.

A beta prototype of our recommendation system is available for testing and playing at http://discover.mateine.org/.

5 Related Work

Classical research on user context and search tasks focused on understanding the patterns used in web search [9], resulting in well known taxonomies. Recent research focused on the final user goal on underlying the search, defining classification closer to the user needs [8]. In our study, however, we go domain-specific trying to understand how researchers find scientific knowledge, focusing on the impact of the social aspect.

As for capturing knowledge sharing, an increasing number of social bookmarking and annotation services have become available in the scientific communities. Following the success of other popular but generic social bookmarking sites (such as delicious.com), Zotero, CiteULike, Connotea, Mendeley are examples of scientific social bookmarking services that focus on sharing and organizing academic references. These tools and services deal with sharing and collecting materials targeting groups and individuals. However, they are of general purpose, and thus, their effective usage is limited to a reduced number of scenarios. In Knowledge spaces, we take a different approach going vertical to every scenario in order to lower down the barriers to share. An interesting work in this line is Mail2Tag [14], a system that explores the use of mail as a tool for sharing and organizing news in environments where the email is the main communication channel.

Other studies consider the user social network to provide personalized search results (e.g., [11] [12]). The most relevant to our work is [11], in which the authors propose a network-aware search for social bookmarking sites. This work introduces interesting techniques that can serve as baseline to this project, however, the modeling, analysis and optimization are specific to this domain, and require particular attention.

6 Concluding remarks

In this work we have proposed the personalized approach to recommending scientific publications based on researchers' social network. We formalized the problem, considering different definitions of networks and popularity metrics and formulated its topic-based version. Given the dataset of Microsoft Academic Search, we designed and conducted the validation experiment by evaluating precision and recall of three different recommendation strategies within our proposed approach with respect to researchers future citations. We analyzed the results, drew some preliminary conclusions regarding the applicability and the potential of network-based recommendation of scientific publications, and identified directions of future work. Finally, we implemented the recommender system relying on users' co-authorship network and deployed it within the prototype web application.

References

1. Towards the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions. IEEE TKDE 2005.
2. Bogers, T., Van Den Bosch, A. Recommending scientific articles using CiteILike. RecSys 2008. ACM Press.
3. Strohman, T., Croft, W. B., Jensen, D. Recommending Citations for Academic Papers. Evaluation. 2007. ACM Press.
4. McNee, S. M., Albert, I., Cosley, D., Gopalkrishnan, P., Lam, S. K., Rashid, A. M., Konstan, J. A., et al. On the recommending of citations for research papers. CSCW 2002. ACM Press.
5. Heymann, P. and Koutrika, G. and Garcia-Molina, H. Can social bookmarking improve web search?. WSDM 2008.
6. Tang, J., Zhang, J. A Discriminative Approach to Topic-Based Citation Recommendation. PAKDD 2009. Springer.
7. Agarwal, N., Haque, E., Liu, H., Parsons, L. Research Paper Recommender Systems: A Subspace Clustering Approach. WAIM 2005.
8. Rose, D.E. and Levinson, D. Understanding user goals in web search. WWW 2004. ACM.
9. Broder, A. A taxonomy of web search. ACM Sigir forum 2002. ACM.
10. Mitsche, N. Understanding the information search process within a tourism domain-specific search engine. Information and Communication Technologies in Tourism. 2005. Springer.
11. Yahia, S. A., Benedikt, M. Efficient Network-Aware Search in Collaborative Tagging. Networks. 2008. VLDB Endowment.

12. Carmel, D., Zwerdling, N., Guy, I., Ofek-Koifman, S., Har?el, N., Ronen, I., Uziel, E., et al. (2009). Personalized social search based on the user's social network. CIKM 2009, 1227. ACM Press.
13. Hwang, S.-Y., Wei, C.-P., Liao, Y.-F. (2010). Coauthorship networks and academic literature recommendation. Electronic Commerce Research and Applications, 9(4), 323-334. Elsevier B.V.
14. Nelson, L., Nairn, R., Chi, E. H. (2010). Mail2Tag : Lightweight Information Sharing Services Integrated with Email. Growth Lakeland, (June 2009), 541-542.