

Trust-Enhanced Visibility for Personalized Document Recommendations

Claudia Hess

Klaus Stein

Christoph Schlieder

Laboratory for Semantic Information Technology
Bamberg University, Germany

{claudia.hess,klaus.stein,christoph.schlieder}@wiai.uni-bamberg.de

ABSTRACT

Documents are recommended by computer-based systems normally according to their prominence in the document reference network. Based on the requirements identified in a concrete use case for recommending scientific publications, the paper claims that merely measuring prominence is insufficient for high quality recommendations. We propose to use information from a trust network in addition to the document network in order to improve and to personalize recommendations. A trust-enhanced visibility measure integrates trust information and the classical reference based measures. A simulation study applies the new visibility measure to the presented use case.

Categories and Subject Descriptors

H.3.3 [Information Storage and Retrieval]: Information Search and Retrieval

Keywords

trust-enhanced document recommendations, recommender system, two-layer architecture, personalization

1. INTRODUCTION

The number of publications increases every day: new websites are appearing, new messages are posted in newsgroups, blogs are written, new scientific papers are published. Users face the problem to select the documents that correspond to their information need and to evaluate the documents regarding their usefulness and reliability. Different technologies are available to support users. A common way to select documents is to exploit the structure of the reference document network and to rank documents according to the prominence. Instead of analyzing the document network, recommender systems compute recommendations on the basis of user profiles and relationships between users. Trust-based recommenders have attracted much attention in the last years and have

proven to be successful. Well-known examples are ebay's reputation system and the Epinions¹ platform for consumer reviews.

To compute document recommendations, we present a two-layer network which comprises the document network as well as the trust network between users. Integrating citations and trust data in a single measure, the trust-enhanced visibility, permits us to provide personalized information for all documents in the document reference network.

The paper is structured as follows. Section 2 describes the use case and current technologies to calculate document recommendations. Section 3 presents the layered architecture on the basis of which section 4 introduces the trust-enhanced visibility measure. In section 5, the simulation study is explained and the findings are analyzed. The last section outlines future extensions and concludes the work.

2. DOCUMENT RECOMMENDATIONS

Scientists interested in an article have to decide whether they want to invest time in reading the article and in the case of a pay-per-view article whether to buy it. Normally the abstract supports the user deciding on the relevance of the article for her or his research. However, the information provided by the abstract often does not suffice. In research fields such as computing, abstracts contain information about the problem solved but often do not provide algorithmical or methodological details. In addition, reading abstracts is time consuming. Recommender systems dealing with requests such as 'can you recommend me this article' as well as with the ranking of search results would be very helpful. Current recommender systems use two different approaches, each of them analyzing one source of information, either the document network or user information.

Analyzing document reference networks, i. e. the directed graph of documents citing others, the prominence of a document can be determined. Measures exist in bibliometry, scientometrics, social network analysis, network physics and information retrieval.² Following Malsch and Schlieder [9], we refer to them as measures of social visibility. Visibility considers several criteria. An important one is the visibility of the publications citing the paper, i. e., a paper is the more visible the more important papers are citing it.

Visibility as single criterion is often insufficient. Recommending a paper proven to be forged would for example be incorrect, despite of being frequently cited. An example is the case published by the German Science Foundation in 2004³. Two publications

¹<http://www.epinions.com>

²see e. g. [8, 13, 12, 15, 2, 11, 10].

³http://www.dfg.de/aktuelles_presse/reden_stellungnahmen/2004/

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

SAC'06 April 23-27, 2006, Dijon, France

Copyright 2006 ACM 1-59593-108-2/06/0004 ...\$5.00.

were based on fabricated and manipulated information. Hence, they should not be cited any more. One of the articles has a rather high visibility because it was cited at least 71 times according to the NASA Astrophysics Database System⁴ from which the article can be downloaded. No information about the fraud is provided before download. Visibility-based systems would still recommend them and therefore be misleading. Less extreme but more frequent is that two scientific communities defend completely different approaches to solve a specific task. In this case, the trust network represents two groups with high trust values inside the group and low trust values towards members of the other group. Surely, there can be researchers not adhering to one specific group. Visibility-based systems would offer identical information to every user regardless of his or her scientific position.

Recommender systems have a different approach. In contrast to visibility measures, they personalize recommendations. They compute recommendations for users about different types of objects ranging from products⁵ to movies⁶. Collaborative filtering is one of the most popular technologies [4]. It considers user profiles and hence provides a personalization. Information about the users and previous actions such as buying or rating products must be available. These traditional recommending techniques have several shortcomings as for instance described in [1]. In a context like ski tours, recommendations must be trustworthy because they contain security related information such as snow conditions. Therefore Avesani et al. propose a trust-based recommending system, Moleskiing which permits users to make statements about other users' trustworthiness. Ski tour reviews will only be displayed to a user if written by someone who is considered trustworthy directly by the requesting user or indirectly by some user trusted by the requesting user.

Recommending documents, a trust-based recommender could provide useful advice. Users could ask experts in the research domain of the article whether they can recommend it. Recommendations are reliable and personalized because users obtain them by someone they consider as trustworthy. However, only those documents can be recommended that have been evaluated by some trusted user. Recommendations therefore risk to be closed up in the scientific community of the requesting user. Measures on the document network do not have this drawback. They permit recommendations about all available documents.

3. TWO-LAYER ARCHITECTURE

Trust-based recommending systems have many advantages in comparison with measures on the document network. However, they have disadvantages in exactly those parts in which visibility measures have their strengths. We present an approach for combining the advantages of both trust-based recommender systems and visibility functions by a two-layer network integrating a document reference network on the first and a trust network on the second layer. Both layers are irreducible, i. e., information from one layer cannot be computed from the other one.

3.1 The Document Reference Network

In 1976, Pinski and Narin [13] computed the importance (rank) vis_d of a scientific journal p_d by using the weighted sum of the ranks vis_k of the journals p_k with papers citing p_d . A slightly modified version of this algorithm (the pagerank algorithm) is used by

⁴[download/ha_jhschoen_1004_en.pdf](http://adsabs.harvard.edu)

⁵<http://adsabs.harvard.edu>, last access 2005-5-27.

⁶see for instance amazon.com for books

⁷see for instance MovieLens, <http://movielens.umn.edu>

the search engine google [12] to calculate the visibility vis_d of web-pages p_d (using the visibility function vf):

$$vis_d = vf(p_d) := (1 - \alpha) + \alpha \sum_{p_k \in R_d} \frac{vis_k}{|C_k|}$$

where R_d is the set of pages citing p_d and C_k is the set of pages cited by p_k .⁷ One important feature of this function is that the visibility of any document p_x depends on the visibility of other documents $p_i \in R_x$ citing it. Therefore changing the visibility of one document will influence the visibility of other documents.

The first layer of our approach consists of a document reference network on which visibilities vis_d of all documents p_d can be computed from the network structure. Any visibility function vf using the visibilities of the citing documents R_d to calculate the visibility vis_d of p_d can be applied.

3.2 The Trust Network

The second layer encompasses scientists making trust statements about colleagues in the sense that the trusted user applies similar criteria to the evaluation of a document. Trust statements thus refer to trust in the user's ability to make recommendations. This type of trust is called recommendation trust [3] in contrast to direct trust which denotes someone's capability to write 'good' papers. In contrast to the document network with visibility assigned to nodes, trust statements are attached to edges. Trust scores between indirectly connected users, also called actors, are inferred by trust metrics⁸. Trust statements are related to a specific research topic because the reliability of reviews by one scientist can differ across various research domains. Research topics can be organized in some way that permits to infer that someone who is trusted in a research topic A will also be trusted in topic B if A and B are related. Kinatader and Rothermel [7] propose to model topics in a semantic network with weighted edges expressing the semantic closeness between two topics. In the examples and the simulation study, trust statements refer to a single research topic.

Trust and document networks are connected via 'reviews'. Scientists evaluate the articles they have read regarding their usefulness and correctness. The document network includes evaluated documents as well as further articles, for instance in the computer sciences literature all articles indexed by the ACM digital library⁹. However, our two-layer approach is not restricted to scientific publications as discussed in the use case but works with every type of document in document reference networks such as websites.

4. TRUST-ENHANCED VISIBILITY

4.1 Interpolation

Starting from a base case, two interpolation problems with increasing complexity can be solved in two-layer networks. In the base case shown in Fig. 1, recommendations for *person 1* are derived merely from the document network by calculating the visibility for every document in the network. The visibility of the requested *document 4* is thus determined by the visibilities of the documents citing it. The property of being able to compute recommendations for all available documents will be maintained in the

⁷For n pages this gives a linear system of n equations. Solving this equation system is possible but (for large n) very expensive, so an iterative approach is used. First all vis_i are set to some default value and then the new values r'_i are calculated repeatedly until all vis_i converge (for a discussion of convergence problems in leaves see [12]).

⁸see e. g. [5, 16, 6, 14].

⁹<http://portal.acm.org/dl.cfm>

extensions of the visibility function while personalizing the recommendations.

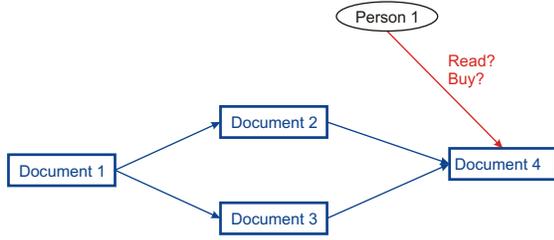


Figure 1: Recommendations on the Document Network

A first extension of the base case includes the requesting users’ reviews of documents in the recommendations and thus personalizes them (Fig. 2). Reviews as statements about the quality of documents influence their visibility. Reviews are in $[0, 1]$ with 0 meaning that a document is not worth to be read or bought and 1 denoting an absolutely recommendable one. A new visibility including the reviews is computed and propagated in the document network. Reviews of documents that directly cite the document in question have the highest impact on its new visibility.

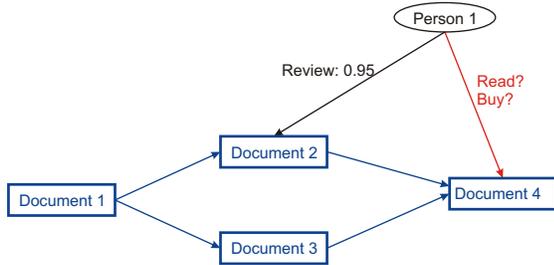


Figure 2: Recommendations including Reviews

Quality and personalization of recommendations can be further improved by including reviews not only by the requesting user but by all users deemed by her or him as trustworthy (Fig. 3). Trustworthy reviewers are determined on the basis of the trust values expressed between users. Trust values indicate that the user assigning the trust value believes that the trusted user applies similar criteria to the evaluation of a document than her- or himself. Trust values are in $[0, 1]$ with 0 for no trust and 1 for full trust. Due to trust propagation, trustworthy reviewers could also be such persons whom the user in question trusts only indirectly. This is to say that users trust to some degree (which is regulated by the applied trust metric) the “friends of their friends”, i. e. the users trusted by the actors they trust. As trust in a reviewer means to trust him or her to apply similar criteria to the reviews than oneself, we can take the trust value as trust in the reviews made by this person. This results in having for each requesting user a set of trust values for the other users t_{r_d} indicating her or his trust in the respective review r_d of a document p_d .

The degree of trust in a person influences the impact of this person’s reviews on the trust-enhanced visibility. Reviews of untrustworthy persons should not considerably influence the visibility whereas reviews of highly trustworthy persons are very significant.

4.2 Personalized Document Recommendations

Now we are able to enhance the visibility function vf (see section 3.1) on the document reference network by using the trust-

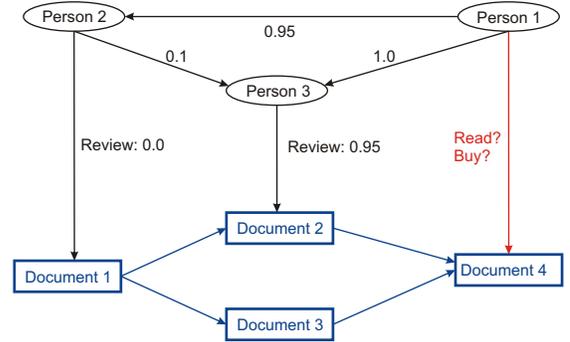


Figure 3: Recommendations including Trust-Weighted Reviews

weighted reviews (TWR) from the trust network to compute document recommendations. For each user, we have the trust expressed by him or her about other users. These trust values determine the user’s trust in the trusted persons’ reviews. This gives a TWR-enhanced visibility function

$$\text{vf}^{\text{TWR}}(p_d) = t_{r_d} r_d + (1 - t_{r_d}) \text{vf}(p_d)$$

with r_d being the review of p_d and t_{r_d} being the trust in this review (for documents without review: $t_{r_d} = 0$). vf^{TWR} has three interesting properties:

- As described in section 3.1 the visibility of each document depends on the visibility of the documents it is cited from, so one review indirectly changes the visibility of the whole surroundings.
- The recommendations are highly personalized: the visibility of a document depends on the reviews of persons deemed to be trustworthy by the requesting user.
- A review influences the overall visibility of a document depending on the degree of trust that the requesting user has in the reviewer. If the trust value to the reviewer is maximal, the recommendation will completely depend on the review. In contrast, if the trust value is minimal, the review will be neglected and the recommendation will be based on the computed visibility.

5. SIMULATION

The impact of the trust-weighted reviews is shown by a simulation study. We focus on the aspect of personalization. As expected we could show in a basic simulation that trust-weighted reviews modify the visibility of reviewed documents as well as of documents cited by the reviewed documents, surely to a lesser degree than the reviewed documents (not shown here). In the following we present the results of a simulation corresponding to the use case described in section 2 with users from two rivaling scientific communities.

5.1 Simulation Setting

To show the influence of trust-weighted reviews on document recommendations we set up a two layer network (T, D_i) with T the trust network shown in Fig. 4 and D_i a document reference network.

The trust network has two communities $C_1 = \{0, \dots, 4\}$ and $C_2 = \{5, \dots, 9\}$ of actors, each with three outgoing trust statements, two of which point to actor’s within the same community expressing

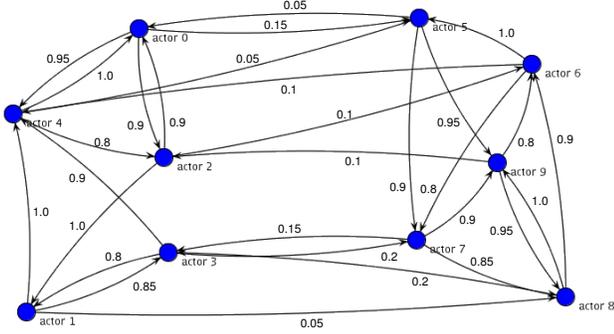


Figure 4: Trust network with two communities $C_1 = \{0, \dots, 4\}$ and $C_2 = \{5, \dots, 9\}$ of actors with high trust edges within and low trust edges across the communities.

high trust, and one of which assigns a low trust value to someone from the rivaling scientific community. Indicated trust values were completed with inferred trust values for each actor. For trust propagation we used the trust metric proposed by Golbeck et al. [5].

The simulation was performed with five different document networks D_1 to D_5 , each with slightly more than 1000 documents and with each document referencing 2 to 4 others. Both layers were connected with 100 randomly set reviews (each actor reviewing 10 documents) being positive (“read p_d ”: $r_d = 1$), neutral ($r_d = 0.5$) and negative (“do not read p_d ”: $r_d = 0$) with equal probability (33.3%), so about 10% of the messages in the document reference network are reviewed. We calculated the recommendations for all documents with vf^{TWR} as defined in section 4.2 and vf being PageRank (section 3.1).

5.2 Results

In each network (T, D_i) for each actor $a \in T$ the personalized TWR-enhanced recommendation $vis_k^{\text{TWR}(a)}$ of each document p_k was computed. Now for any $p_k \in D_i$ and for any pair of actors $a_1, a_2 \in T$

$$\Delta vis^{\text{TWR}(a_1, a_2)}(p_k) = \left| vis_k^{\text{TWR}(a_1)} - vis_k^{\text{TWR}(a_2)} \right|$$

tells how the views of a_1 and a_2 on p_k differ.

$$\overline{\Delta vis^{\text{TWR}(a_1, a_2)}(D_i)} = \frac{\sum_{p_k \in D_i} \left| vis_k^{\text{TWR}(a_1)} - vis_k^{\text{TWR}(a_2)} \right|}{|D_i|}$$

gives the average difference of the personalized recommendations for the whole document network D_i .

Figure 5 shows the summarized results: the average difference $\overline{\Delta vis^{\text{TWR}(a_1, a_2)}(D_i)}$ with a_1, a_2 being members of the same community (both in C_1 or both in C_2) differ much less than for a_1, a_2 being members of different communities (one in C_1 and the other in C_2). The left three columns show the average differences for the subset of documents D_i^R being reviewed, the next three columns the average differences on the unreviewed documents D_i^U where the reviews’ influence is only indirect, and the last three columns the average differences on all documents (D_i). Obviously the difference in the views of two actors highly depends on their position in the trust network, which even shows up in the recommendations of the unreviewed documents. As shown in table 1 the results are similar on all document networks D_1 to D_5 .

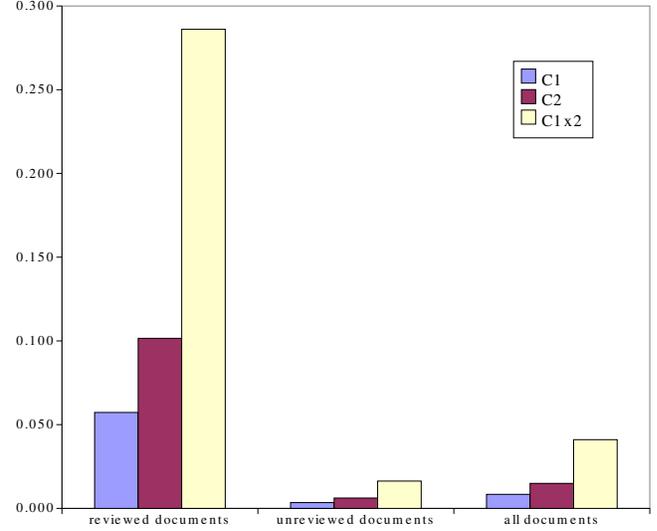


Figure 5: Simulation with Users from Two Rivaling Communities

$a_1, a_2 \in C_1 = \{0, 1, 2, 3, 4\}$:

	reviewed	unreviewed	all
D_1	0.0637	0.0041	0.0095
D_2	0.0589	0.0035	0.0083
D_3	0.0530	0.0029	0.0075
D_4	0.0465	0.0028	0.0069
D_5	0.0644	0.0040	0.0097
avg	0.0573	0.0035	0.0084

$a_1, a_2 \in C_2 = \{5, 6, 7, 8, 9\}$:

	reviewed	unreviewed	all
D_1	0.1015	0.0067	0.0153
D_2	0.0986	0.0055	0.0137
D_3	0.1122	0.0066	0.0163
D_4	0.0971	0.0053	0.0139
D_5	0.0984	0.0069	0.0155
avg	0.1016	0.0062	0.0149

$a_1 \in C_1 = \{0, 1, 2, 3, 4\}, a_2 \in C_2 = \{5, 6, 7, 8, 9\}$:

	reviewed	unreviewed	all
D_1	0.3012	0.0189	0.0445
D_2	0.2879	0.0150	0.0390
D_3	0.2820	0.0154	0.0399
D_4	0.2626	0.0152	0.0384
D_5	0.2966	0.0171	0.0434
avg	0.2861	0.0163	0.0410

Table 1: Average differences in personalized visibilities of two actors a_1 and a_2 on document networks D_1 to D_5 .

Summing up the results, we can say that the document recommendations differ as expected only slightly for two members of the same community. For members of different communities however, the views on the document network heavily differ. This reflects the fact that papers very important within the own scientific community can be very unpopular in another community.

6. CONCLUSION

In the paper we introduced a two-layer architecture connecting a trust with a document reference network via reviews. Trust-weighted reviews enhance classical visibility measures. The proposed trust-enhanced visibility function permits to calculate personalized recommendations for all documents in the document reference network and not only for those documents reviewed by oneself or by a user deemed as trustworthy. This is achieved by propagating the trust-weighted reviews in the document reference network as part of the visibility function. Trust-weighted reviews therefore influence the visibility of directly reviewed papers as well as the visibility of papers cited by reviewed papers. The integration of trust data allows for a strong personalization of the recommendations: the personal trust values assigned to other users determine the personal view on the document reference network.

Our future work will be concerned with further simulation studies analyzing the influence of systematic modifications in the structure of the trust network. In addition, we want to evaluate the two-layer approach with further document networks, also with a varying number of references between the documents and of reviews.

7. REFERENCES

- [1] P. Avesani, P. Massa, and R. Tiella. A trust-enhanced recommender system application: Moleskiing. In *SAC '05: Proceedings of the 2005 ACM symposium on Applied computing*, pages 1589–1593, 2005.
- [2] A.-L. Barabási. *Linked: How Everything Is Connected to Everything Else and What It Means for Business, Science, and Everyday Life*. PLUME, 2003.
- [3] T. Beth, M. Borcharding, and B. Klein. Valuation of trust in open networks. In *Proceedings of the European Symposium on Research in Computer Security (ESORICS)*, LNCS 875, pages 3–18, Brighton, UK, 1994. Springer Verlag.
- [4] R. Burke. Hybrid recommender systems: Survey and experiments. *User Modeling and User-Adapted Interaction*, 12(4):331 – 370, November 2002.
- [5] J. Golbeck, B. Parsia, and J. Hendler. Trust networks on the semantic web. In *Proceedings of Cooperative Intelligent Agents*, Helsinki, Finland, August 2003.
- [6] R. Guha, R. Kumar, P. Raghavan, and A. Tomkins. Propagation of trust and distrust. In *WWW '04: Proceedings of the 13th international conference on World Wide Web*, pages 403–412, New York, NY, USA, 2004. ACM Press.
- [7] M. Kinader and K. Rothermel. Architecture and algorithms for a distributed reputation system. In *Proceedings of the First International Conference on Trust Management*, volume 2692, pages 1–16. Springer Verlag, April 2003.
- [8] J. M. Kleinberg. Authoritative sources in a hyperlinked environment. *Journal of the ACM*, 46(5):604–632, 1999.
- [9] T. Malsch and C. Schlieder. Communication without agents? From agent-oriented to communication-oriented modeling. In L. et al., editor, *Regulated Agent-Based Social Systems: First International Workshop, RASTA 2002*, pages 113–133, Bologna, Italy, 2002. Springer-Verlag.
- [10] F. Menczer. Growing and navigating the small world web by local content. In *Proceedings of the National Academy of Sciences of the United States of America*, volume 99, pages 14014–14019, 2002.
- [11] M. E. J. Newman. The structure and function of complex networks. *SIAM Review*, 45:167–256, 2003.
- [12] L. Page, S. Brin, R. Motwani, and T. Winograd. The pagerank citation ranking: Bringing order to the web. Technical report, Stanford Digital Library Technologies Project, 1998.
- [13] G. Pinski and F. Narin. Citation influence for journal aggregates of scientific publications: Theory, with application to the literature of physics. *Information Processing & Management*, 12:297–312, 1976.
- [14] M. Richardson, R. Agrawal, and P. Domingos. Trust management for the semantic web. In D. Fensel, K. P. Sycara, and J. Mylopoulos, editors, *Second International Semantic Web Conference (ISWC 2003)*, volume 2870 of *Lecture Notes in Computer Science*, pages 351–368, Sanibel Island, FL, USA, October 2003. Springer Verlag.
- [15] D. J. Watts. *Six Degrees: The Science of a Connected Age*. W. W. Norton & Company, Inc., 2003.
- [16] C.-N. Ziegler and G. Lausen. Spreading activation models for trust propagation. In *Proceedings of the IEEE International Conference on e-Technology, e-Commerce, and e-Service*, Taipei, Taiwan, March 2004. IEEE Computer Society Press.