

Trust-Based Recommendations for Publications - A Multi-Layer Network Approach

Claudia Hess

Laboratory for Semantic Information Technology, Bamberg University
claudia.hess@wiai.uni-bamberg.de

Abstract. Current recommender systems using network data focus on a single network, either a trust or a document reference network. However, in a number of applications, several types of networks need to be taken into account. The paper describes an architecture for a recommender system for publications that integrates information from different networks. Relationships in the trust network are modeled in a semantically rich way to reflect trust relationships from the real life in a concise way.

1 Introduction

Trust-based recommender systems have attracted much attention in the last years in research as well as in commercial applications. Well-known examples are the Epinions¹ platform for consumer reviews and ebay's reputation system². They use information about trust between users. Some systems infer recommendations from the network linking users by trust relationships (e.g. Guha, 2003, Kinatader and Rothermel, 2003, Montaner et al., 2002). Other recommendation systems evaluate citations between documents (e.g. Page et al., 1998). Current recommender systems analyze only data from one type of network, either a trust or document network. Trust relationships cannot automatically be extracted from the document network. Neither is it possible to infer citations from the trust network. The paper proposes a layered architecture which integrates different network data, collected independently from each other but analyzed jointly. To provide a realistic modeling of trust, semantically rich trust statements are introduced. Semantic Web technologies like ontological modeling provide the basis for formalizing complex, multi-criterial trust relationships.

The rest of the paper is structured in the following: section 2 describes one of the use case motivating the presented research proposal. Sections 3 and 4 present the central research areas. Both sections discuss related work and highlight areas for research addressed in the dissertation project. The former describes the multi-layer architecture whereas the later section introduces complex trust statements. Open research issues of the dissertation project are summarized in the last section.

¹ <http://www.epinions.com>

² <http://www.ebay.com>

2 Recommendations for Publications

An important use case in this context is that scientists often have to decide whether an article is worth to be read or whether a pay-per-view article is worth to be bought. The abstract is often insufficient to take a decision. In research fields such as computing, abstracts contain information about the problem solved but often do not provide algorithmical details. Users could also ask colleagues who are experts in the research domain of the article whether they have already read the article and can recommend it. To find the appropriate persons to ask, it would be necessary to know which articles they read and above all, that this person is available. This work claims that a trust management system integrating information from different networks can support the user in the decision by directly offering a trustworthy person's opinion or review of the article. Figure 1 shows a typical request and a way of answering it by exploiting the trust network.

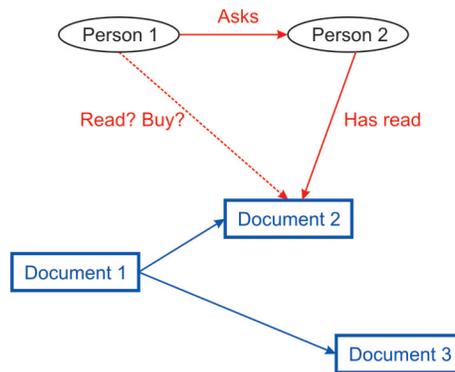


Fig. 1. Recommendations for Publications

Documents like scientific papers are linked via citations. Wouldn't it be sufficient to recommend the articles which are prominent in the scientific community? A prominent paper is for instance a paper which is read or cited very often whereas a paper with a low prominence is one that is very unlikely to be accessed. Several measures for the prominence of a paper exist in bibliometry, scientometrics, social network analysis, network physics and information retrieval. Following Malsch and Schlieder, 2002, we refer to such measures as measures of social visibility. A measure for 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. Another criterion is the author's visibility derived for example from his reputation in the scientific community.

Despite the various considered criteria, a high visibility is often insufficient for recommendations. It would for instance be false to recommend a highly visible paper if it was proven that it is manipulated. Such paper might be frequently cited until the manipulation was detected. An example is the case of scientific

misconduct published by the German science foundation (Deutsche Forschungsgemeinschaft) in 2004³. Two publications⁴ were based on fabricated and manipulated information. Hence, they should not be cited any more. However, a visibility-based system would still recommend them. One of the articles for instance has a rather high visibility because it was cited at least 71 times between 2000 and 2002 according to the NASA Astrophysics Database System⁵ from which the article can be downloaded. To rely only on the publisher's reputation is insufficient. The fraud was neither detected during peer-reviewing nor is there any information about the fraud before downloading the paper.

3 Multi-Layer Networks for High Quality Recommendations

Integrating different types of information in a multi-layer network would permit recommendations of a high quality. Visibility as single criterion is insufficient to provide high quality recommendations. However, merely trust-based recommender systems also have their disadvantages. Users will obtain recommendations only about a limited number of papers, namely those that have been read by their colleagues. Recommendations thus risk to be restricted to publications within the own scientific community. Coupling several networks, at least a document and a trust network, permits to combine the advantages of trust-based recommender and measures of social visibility: reliability and personalization is guaranteed by using the trust data and the coverage of recommendations is enlarged by considering the visibility of the documents.

3.1 A Multi-Layer Architecture

The paper proposes a layered architecture with three networks as depicted in Fig. 2. The first layer consists of a trust network between persons. Persons can be both readers who are using the trust management system to obtain recommendations and authors of articles. The trust statements are associated with the edges between the persons. Trust can be propagated in the trust network in order to express that people trust other people, even if they do not know them, to some extent when they are trusted by people they trust.

The second layer is the document network, which consists of the pay-per-view articles rated by the researchers and further articles, e.g., all articles indexed by CiteSeer⁶. Persons are connected to documents either by being its author or a

³ http://www.dfg.de/aktuelles_presse/reden_stellungnahmen/2004/download/ha_jhsc_hoen_1004_en.pdf

⁴ Schön, J. H., Berg, S., Kloc, Ch. and Batlogg, B. (2000): *Ambipolar Pentacene Field-Effect Transistors and Inverters*, Science 287, 1022 and Schön, J. H. , Kloc, Ch., Laudise, R. A. and Batlogg, B. (1998): *Electrical Properties of Single Crystals of Rigid Rodlike Conjugated Molecules*, Physical Review B 58, 12952

⁵ <http://adsabs.harvard.edu>, last access 27.05.05

⁶ <http://citeseer.ist.psu.edu>

reader. Having read an article, the reader has an opinion about it. The opinion could be written down in the form of a review.

The third layer is the organization network. Organizations and institutions are connected via cooperation relationships which often go along with trust relationships between the organizations. Persons are connected with organizations via membership, e.g. 'is professor at'. Layers are not arbitrarily connected but only with the preceding and the succeeding layer.

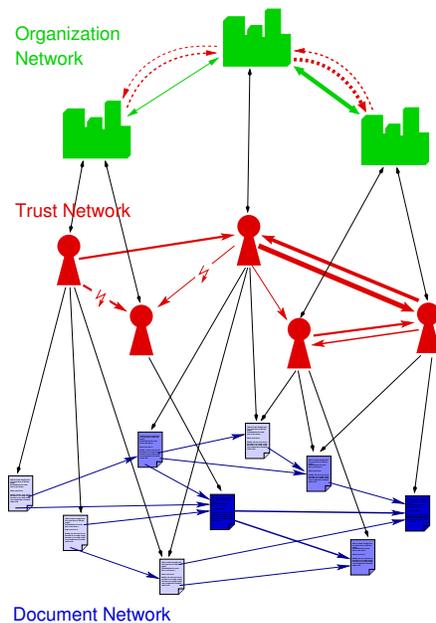


Fig. 2. Multi-Layer Network

Research in the last years focused on the understanding of the different kinds of networks, the development of applications and algorithmical support. Interdependencies between networks and their integration have only recently been addressed. An approach to the integration of document and author networks has been proposed by Börner et al., 2004. They focus is the simultaneous growth of coauthor and citation networks in time. Relationship types are co-authorship between authors, inception and reception of documents by authors and references between documents. The growth process is driven by authors who write papers together with other, randomly chosen authors working on the same topic.

3.2 Connecting Persons and Documents: Opinions or Explicit Reviews?

A trust management system based on multi-layers could answer requests in two different ways. Firstly, the user could obtain the name and contact address of colleagues determined by the system as well-suited to answer the user's question. The user could then directly ask the colleagues about their opinions on the pay-per-view article. Secondly, the requester could directly obtain the reviews made by the persons computed as trustworthy for the request. Reviews are made by those users who already bought the pay-per-view article. A paper would be evaluated by its readers according to criteria similar to those used in the reviews for conferences or journals. When defining the criteria used in the reviews, it must be considered that the review should easily be made in order to keep the additionally required time for the scientists as low as possible. The reviews could be presented separately or aggregated by the trust management system.

The first case has the advantage that the time and effort required by the scientists are minimal. Users only assign trust values to their colleagues and indicate which articles they have read. In contrast, the second alternative requires that users provide explicit reviews about the articles. The advantage of the second approach is that users could obtain recommendations by persons which could not, or at least not easily, be contacted. For instance, a student would rarely dare to contact some well-known experts and ask them about their opinions only in order to decide whether to read or to buy an article. Thus, it depends on the structure of the community whether the result should better be a list of persons recommended to ask about their opinions or a set of reviews by trustworthy persons. Written reviews will be the favored if users frequently change. In a stable network in which users remain over a long time period and in which users can directly contact each other, it would be sufficient to know whom to ask.

3.3 Computing Recommendations in Multi-Layer Networks

In order to automatically classify a faked article as 'strictly not recommended' as it would be appropriate in the above example, there must be at least one distrust statement. The source of the distrust statement could either be some trusted scientist or some trustworthy organization. The target of the distrust statement could be on the one hand the manipulated paper and on the other the author of the manipulated paper. To assign the distrust to the author is sensible under the assumption that someone who has cheated once will do it again or did it more often than in the known case(s). Thus, four different kinds of distrust statements are conceivable:

1. Distrust assigned by scientists in their reviews to the paper: Some scientist who is trustworthy ('expert') in the specific research topic of the manipulated publication would know about the manipulation and rate the paper as being of a bad quality. A problem might be that experts only provide reviews about articles they have read. An expert who has read and recommended

the article before having obtained the news that the article is manipulated could have rated the article as very recommendable and could forget to modify his review after having learned about the manipulation. Therefore, it would be better to use directly the information from the organizations that identify manipulations such as the organizations which are funding projects or departments.

2. Distrust assigned by scientists to the untrustworthy author: The distrust statements are directly made about the author of the paper which has been identified to be based on manipulated data sets. It is problematic that many scientists must insert new distrust statements or update their previous, probably high trust statements with a distrust relationship. It will take some time until the network is updated in this way and recommendations are based on the updated knowledge.
3. Distrust by an organization in its review to the article: Organizations like departments, universities and other societies (e.g. the DFG) could provide reviews about articles. In a multi-layer architecture this relation is not allowed. Connections between layers are only possible with the preceding and the succeeding layer. In the presented example, relations are only permitted between the organization layer and the person trust network and between the person trust network and the document network.
4. Distrust by an organization to the untrustworthy author: Organizations could directly assign the distrust to the author of the manipulated publication. Made by a trustworthy organization, such statement could have a considerable influence on the recommendations. As it is very unlikely that such organizations participate in a local trust management system (e.g. in the scope of a university), it would be desirable to extract the information provided by the organizations automatically from the organizations' publications about manipulated articles but also about awarded publications. Institutions like the DFG do not maintain a database with such information. However, it would for example be feasible that someone in the role of a moderator includes the information provided by institutions like the DFG or by universities into the multi-layer network.

According to the concrete situation, source and target of the distrust relationship must be chosen, also considering the interdependencies between the different layers. Distrust in authors can be propagated to their papers. Bad reviews can influence the trust value, too. Despite of the interdependencies, the semantics and the effects on the recommendations differ between trust statement about the author or reviews of the article. Trust or distrust relationships between authors cannot be extracted from the reviews and citations and vice versa.

Coauthors are also concerned by the distrust. This would exclude all future and past papers of these authors from being recommended. Therefore scientists should be able to assign trust not only to their colleagues reviewing papers, i.e. trust in their reviewing capabilities, but also to the authors of the papers, i.e. their trust in the author's capability of writing 'good' papers.

4 Trust Statements

The trust statements between persons are a central aspect of the trust management system. Most applications use rather simple trust statements like the research project FilmTrust⁷ which statements are numerical values or the platform Epinions which distinguishes only trust and distrust. For the presented trust management system, such trust statements are too restricted. An approach to modeling differentiated trust statements is presented in the following which on the one hand integrates different extensions to the current way of representing trust and on the other hand introduces complex trust concepts.

4.1 Trust and Distrust

Trust statement should not be restricted to trust relationships but include distrust. This would be necessary for example in order to express that a user distrusts an author who already manipulated publications. Until now, only few trust-based recommender systems deal with distrust values. Trust metrics for the propagation of trust between users become more difficult when integrating distrust. Most trust metrics like Golbeck et al., 2003, Ziegler and Lausen, 2004, are limited to positive trust values. Guha et al., 2004 discuss a number of research issues on distrust also concerning the development of a metric that is able to deal with trust and distrust. The question is on how to combine successive distrust values on a path from a source node to a target node: should two distrust statements cancel or intensify each other?

4.2 Recommendation Trust and Direct Trust

Readers and authors are integrated in a single network because each scientist can be reviewer for some articles and author of some other articles. This results in two different kinds of edges Beth et al., 1994. Direct trust in an author's capabilities, i.e., to write good and scientifically correct papers, can be distinguished from recommendation trust, i.e., trust in the quality of someone's trust statements and reviews. Explicit trust or distrust in an author can replace an individual review on each of the author's papers. Distrust statements concerning someone's recommendations - reviews or trust statements - could exclude this person's recommendations from the overall recommendation.

4.3 Trust Related To Topics

Trust statements should always be related to a specific research topic because users only rarely are experts in all available topics. Research topics should be organized in some way that permits to infer that someone who is trusted in some topic A is also trusted in some topic B in the case that topics A and B are related in some way. Topics can be organized for instance in a hierarchical

⁷ <http://trust.mindswap.org/FilmTrust>

structure (e.g. Chen and Singh, 2001). It can be inferred that someone who is expert in a subtopic has certain knowledge about the super-topic and to some limited extent about the siblings, too. Hierarchical classification schemes are used in the context of libraries to structure books and journals. In the computer sciences, hierarchical classifications exist like the computing classification system (CCS) from the ACM (Association for Computing Machinery)⁸. It provides a proper index of literature from the computer sciences according to their subject and hence facilitates retrieval. The CCS is even available as xml-file and could therefore easily be integrated in a trust management system. Furthermore, a considerable amount of scientific papers from the computer sciences is already classified according to the CCS because all authors publishing their papers with the ACM have to provide the categories of the paper and the general terms according to the classification.

Instead of organizing research topics in a hierarchy, they could be modeled in a semantic network with weighted edges, i.e. a dependency graph, as for example proposed by Kinateder and Rothermel, 2003. In a dependency graph, the semantic closeness between two topics is annotated. Trust in one topic hence influences the trust in another topic according to the weight of the relationships between the first and the second topic in the dependency graph. It would be possible to map the CCS hierarchical classification including the “terms” to a dependency graph. Weights between edges must be set manually.

4.4 Confidence in Trust Statements

Confidence values in trust statements permit to handle trust statements in a more elaborated way. The confidence values could be objective or subjective. Objective ones are computed on the basis of the interaction history between users. The number of interactions and their recency are indicators for the reliability of the trust statement. Subjective confidence values are indicated by the user. They represent the user’s individual opinion about the reliability of the assigned trust statements. It will be analyzed how the subjective confidence can be included in this approach and how it can be combined with the trust concepts.

4.5 Introducing Trust Concepts

The work proposes to use trust concepts instead of simple numerical values. Trust concepts offer a higher expressiveness in contrast to numerical values which only permit to indicate trust relationships on a continual scale but not to express two different, not directly comparable trust concepts. Concepts expressing a scientist’s trust in the expertise of a colleague could for example permit to distinguish the following criteria:

- It must be possible to express trust related to the level of detail of the colleague’s knowledge. For example, a scientist thinks that a colleague has

⁸ <http://www.acm.org/class>

only a general survey of the research topic or that the colleague has detailed knowledge in a specific research area. A trust statement could be for instance ‘high trust in the detail of knowledge’.

- The trust in a scientist’s expertise can refer to the up-to-dateness of the knowledge. Restrictions might be that the scientist thinks that a colleague has only knowledge about the current results or only about the “classic” publications in this specific area. A trust statement could be like ‘low trust in the up-to-dateness’.

Both example trust concepts are not directly comparable. A lattice-based structure organizes the trust concepts in a hierarchical fashion that permits to compare the concepts in some way.

5 Research Issues

Based on the described use case for supporting researchers whether to read an article or to buy a pay-per-view article, a number of research issues have been identified. The approach is not restricted either to a mere analysis of the document network like it is done by search engines or to an isolated consideration of the trust network like commercial recommender systems like Epinions but integrates knowledge from various sources into a multi-layer network.

The different aspects of a trust management system will be implemented in a prototype trust management system that realizes the use case. Users of the system will be the members research groups with an interdisciplinary background. Trust and review data that is inserted by the users can be used to directly evaluate the approach. The quality of recommendations and the added value by the trust management system can be evaluated with the help of user feedback.

As the presented dissertation project is rather at the beginning, there remain several open research questions that are addressed in the ongoing work. Research on the theoretical foundations of the presented trust management is twofold. On the one hand, interdependencies between the layers will be analyzed. The trust network cannot be viewed as independent from the document network. An author who has written many visible papers could be perceived by a reader as more trustworthy than another author of less visible papers even though the reader has been very satisfied with both authors. The trust network is coupled with an organization network, too. Inter-organizational trust is influenced by the trust between the members of the organization and vice versa. On the other hand, the modeling of the trust concepts is an important part. The modeling technique must be chosen. Ontologies seem to be a good approach. However, other modeling approaches will be evaluated. Having modeled trust statements as concepts, further research issues emerge, for instance on the trust metrics. Confidence cannot any longer be easily charged against complex trust concept. A trust metric for trust propagation has to be developed which can deal with concepts instead of the current metrics which calculate trust scores based on numerical values. However, research is not restricted to the theoretical aspects but novel ideas will be evaluated with the help of prototypical implementations.

Another research issue is to explore further application scenarios for trust management. In the context of digital libraries, personalization of services offered by digital libraries could be enhanced. Digital preservation could be another application domain: the selection which documents directly to convert to a new format or only on demand can be supported by a multi-layer network approach.

References

- Beth, T., Borcherding, M., Klein, B.: Valuation of trust in open networks. In: Proceedings of the European Symposium on Reserach in Computer Security (ESORICS). LNCS 875, Brighton, UK, Springer Verlag (1994) 3–18
- Börner, K., Maru, J.T., Goldstone, R.L.: The simultaneous evolution of author and paper networks. Proceedings of the National Academy of Sciences **101** (2004) 5266 – 5273
- Chen, M., Singh, J.P.: Computing and using reputations for internet ratings. In: Proceedings fo the 3rd ACM conference on Electronic Commerce, Tampa, USA (2001)
- Golbeck, J., Parsia, B., Hendler, J.: Trust networks on the semantic web. In: Proceedings of Cooperative Intelligent Agents, Helsinki, Finland (2003)
- Guha, R.: Open rating systems. Technical report, Stanford Knowledge Systems Laboratory, Standford, CA, USA (2003)
- Guha, R., Kumar, R., Raghavan, P., Tomkins, A.: Propagation of trust and distrust. In: WWW '04: Proceedings of the 13th international conference on World Wide Web, New York, NY, USA, ACM Press (2004) 403–412
- Kinateder, M., Rothermel, K.: Architecture and algorithms for a distributed reputation system. In: Proceedings of the First International Conference on Trust Management. Volume 2692., Springer Verlag (2003) 1–16
- Malsch, T. and Schlieder, C.: 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, Bologna, Italy, Springer-Verlag (2002) 113–133
- Malsch, T., Schlieder, C., Kiefer, P., Lübcke, M., Perschke, R., Schmitt, M., Stein, K.: Communication between process and structure: Modelling and simulating message-reference-networks with com/te. JASSS (2005) submitted.
- Montaner, M., López, B., Lluís de la Rosa, J.: Opinion-based filtering through trust. In Ossowski, S., Shehory, O., eds.: Proceedings of the Sixth International Workshop on Cooperative Information Agents. Volume 2446 of LNAI., Madrid, Spain, Springer Verlag (2002) 188–196
- Page, L., Brin, S., Motwani, R., Winograd, T.: The pagerank citation ranking: Bringing order to the web. Technical report, Standford Digital Library Technologies Project (1998)
- Ziegler, C.N., Lausen, G.: Spreading activation models for trust propagation. In: Proceedings of the IEEE International Conference on e-Technology, e-Commerce, and e-Service, Taipei, Taiwan, IEEE Computer Society Press (2004)

Appendix: Laboratory for Semantic Information Technology

Christoph Schlieder

Laboratory for Semantic Information Technology,
Chair for Computing in the Cultural Sciences, Bamberg University
`christoph.schlieder@wiai.uni-bamberg.de`

The University of Bamberg, founded in 1647 under the name of Academia Bambergensis, is a research-oriented academic institution with about 8000 students. In 2002, the Research Group on Computing in the Cultural Sciences and its home, the Laboratory for Semantic Information Technology were established. Currently, the group consists of 6 researchers from computer science headed by Professor Schlieder which adapt methods from semantic information processing for building information systems for the cultural sciences. The group focuses on two application domains: digital libraries with a special emphasis on computer-mediated communication and geoinformation systems including mobile applications. Researchers of the group are engaged in several projects on the national and European level.

The group has developed technology supporting mobile documentation, for instance a digital archive for medieval cathedrals. New methods and concepts for archiving and documentation constitute the basis for the monitoring of environmentally related damage on historic buildings. Furthermore, a video archive that permits to combine video sequences to complete movies according to the target group is developed. Currently, it is used to annotate and to manage video sequences about restoration techniques demonstrated in the dominican church in Bamberg. Video sequences will be retrieved by a semantic search and combined to movies for instance demonstrating students in the cultural heritage the application of the restoration technologies.

The project TeDUB (“Technical Drawings Understanding for the Blind”) is part of the Information Society Technologies (IST) program funded by the European Union. The objective of TeDUB is to make technical drawings in digital documents accessible to blind users. In the project, software was developed that interprets the diagrams and allows users to navigate through them.

Research in computer-mediated communication is undertaken in the context of the project Communication Oriented Modeling (COM) funded by the German science foundation (Deutsche Forschungsgemeinschaft) which comprises studying and modeling of large scale communication processes. Communication processes are simulated and monitored over long time spans (see Malsch and Schlieder, 2002). For example, it is examined why some communication threads are very active despite of changing actors. A central aspect is the analysis of social visibility, i.e., the question why some messages or documents are highly referenced or often accessed.

Claudia Hess works as a research assistant in the Research Group on Computing in the Cultural Sciences since 2004. Her dissertation project “Trust-management in Multi-Layer Networks” is in its first year and embedded in the group’s research activities on digital libraries. A close cooperation takes place with the COM project. Modeling trust relationships between persons is of interest for the COM project as communication processes are always influenced by the people publishing documents and the relationships between them. Trust is one of the social relationships that heavily influence the behavior of persons, for instance for co-authorships. Claudia Hess’ research therefore supports also the modeling of actors and of their behavior. The integration of document and trust networks is related to the COM project in the sense that search, selection and evaluation of documents or messages in large scale communication processes can be supported by the inclusion of trust information. Her experiences in semantic modeling (see Hess and Schlieder, 2004, Hess and de Vries, 2004), above all ontological modeling, obtained during her diploma thesis and in the ongoing dissertation project permit her to give input to the digital library projects such as the video archive.

Selected References

- Malsch, T. and Schlieder, C.: Communication without Agents? From Agent-Oriented to Communication-Oriented Modeling. In Lindemann et al., editor, *Regulated Agent-Based Social Systems: First International Workshop, RASTA 2002, Bologna, Italy*, Springer-Verlag (2002) 113–133
- Hess, C. and Schlieder, C.: Ontology-based Verification of Core Model Conformity in Conceptual Modeling. *Proceedings of the Workshop Standardization in the Cadastral Domain, Bamberg, Germany*, (selected for journal publication in *Computers, Environment and Urban Systems (CEUS)*) (2004)
- Hess, C. and de Vries, M.: From Models to Data: a Prototype Query Translator for the Cadastral Domain. *Proceedings of the Workshop Standardization in the Cadastral Domain, Bamberg, Germany*, (selected for journal publication in *Computers, Environment and Urban Systems (CEUS)*) (2004)