The Social Dimension of Information Ranking:
A Discussion of Research Challenges and Approaches

Ingo Scholtes, René Pfitzner, Frank Schweitzer

Chair of Systems Design
ETH Zürich
Zürich, Switzerland
{ischoltes,rpfitzner,fschweitzer}@ethz.ch

Abstract: The extraction of relevant knowledge from the increasingly large amount of information available in information repositories is one of the big challenges of our time. Although it is clear that the social and the information layer of collaborative knowledge spaces like the World Wide Web (WWW), scholarly publication databases or Online Social Networks (OSNs) are inherently coupled and thus inseparable, the question how the ranking and retrieval of information is influenced by the structure and dynamics of the social systems that create it has been addressed at most partially. In this talk, we will highlight associated research questions and challenges from an ethical, social and computer science perspective and introduce a multiplex network perspective that integrates both the social and the semantic layer of social information systems.

The extraction of relevant knowledge from the increasingly large amount of information available in information repositories like the World Wide Web (WWW), scholarly publication databases or Online Social Networks (OSNs) is one of the big challenges of our time. Apart from technical challenges in the storage, processing and retrieval of the associated huge amounts of data, the design of suitable *measures that rank information in terms of relevance* to the information need of a given user is of particular importance. In order to deal with the overabundance of information, a number of different approaches have been developed that intend to establish reasonable schemes to filter and rank information according to its relevance and the reputation of its source. Among the most successful ones are those measures that are based on a network perspective on linked information, like e.g. Web pages referring to each other or scholarly publications citing each other. Based on this perspective, *network-based ranking mechanisms* like, e.g., Google's PAGERANK algorithm or collaborative filtering schemes have proven to be extremely successful in the extraction of relevant information from large information repositories.

Nevertheless, the currently observable *convergence of social and technical* systems raises a number of important and novel issues [Kle08]. Knowledge spaces like, e.g., the WWW are created, organized and consumed in an increasingly collaborative fashion by groups of humans interacting on short time scales, a process commonly subsumed under the umbrella of *social computing* or *social information processing*. As such, the question how pieces of information are linked to each other, ranked and filtered not only affects the ability of individuals or organizations to access information in a timely, objective and transparent

manner. It is also of prime importance for society as a whole since notions of relevance in networks of linked information a) are increasingly influenced by social processes and b) can be an important driver of social dynamics themselves: In the context of citation networks, collaborations and other types of social interactions between scientists have a well-known effect on future citations [PGD04]. Similarly, how scientific works are linked in terms of citations is naturally correlated with the subsequent evolution of collaboration structures and thus the dynamics of the social organization of science. Similar observations about the link between collaboration structures and the success of individuals have been made in other collaborative systems as well [ZSTS13a, ZSTS13b]. The resulting feedback between the social and the semantic layer of collaborative knowledge spaces questions to what extent current network-based information ranking measures - even though they are defined algorithmically - can actually be seen as *objective* or *fair*. As a consequence, the design of information systems and the definition of information ranking methods has a social, political and ethical dimension that is often underestimated [IN00, KOP10].

Although it is clear that the social and the information layer of collaborative knowledge spaces are inherently coupled and thus inseparable, the question how the ranking and retrieval of information is influenced by the structure and dynamics of the social systems that create it has been addressed at most partially [KA13]. In this talk, we will highlight associated research questions and challenges from an ethical, social and computer science perspective and introduce a multiplex network perspective that integrates both the social and the semantic layer of social information systems. We further summarize recent findings in the modeling and analysis of dynamic social interaction networks which promise interesting opportunities for the development of refined measures of information relevance [PSG$^+$13]. We consider these research questions to be of particular interest and significance due to the recent trend towards "socially influenced" information retrieval systems, like e.g. "social search", trust-based recommender systems [WBS08] or collaborative filtering techniques.

References

[IN00]   Lucas Introna and Helen Nissenbaum. The Politics of Search Engines. *IEEE Computer*, pages 54–62, 2000.

[KA13]   Jon Kleinberg and Phil Trans R Soc A. Analysis of large-scale social and information networks. *Philosophical transactions of the Royal Society A*, 371(February):1987, 2013.

[Kle08]   Jon M. Kleinberg. The convergence of social and technological networks. *Communications of the ACM*, 51(11):66, November 2008.

[KOP10]   Felicitas Kraemer, Kees Overveld, and Martin Peterson. Is there an ethics of algorithms? *Ethics and Information Technology*, 13(3):251–260, July 2010.

[PGD04]   Olle Persson, Wolfgang Glaenzel, and Rickard Danell. Inflationary bibliometric values: The role of scientific collaboration and the need for relative indicators in evaluative studies. *Scientometrics*, 60(3):421–432, 2004.

[PSG+13]   Rene Pfitzner, Ingo Scholtes, Antonios Garas, Claudio Juan Tessone, and Frank Schweitzer. Betweenness preference: Quantifying correlations in the topological dynamics of temporal networks. *Physical Review Letters*, 110(19):198701, May 2013.

[WBS08]   Frank Edward Walter, Stefano Battiston, and Frank Schweitzer. A model of a trust-based recommendation system on a social network. *Autonomous Agents and Multi-Agent Systems*, 16(1):57– 74, October 2008.

[ZSTS13a]   Marcelo Serrano Zanetti, Ingo Scholtes, Claudio Juan Tessone, and Frank Schweitzer. Categorizing bugs with social networks: A case study on four open source software communities. In *In Proceedings of the 35th International Conference on Software Engineering, ICSE '13, San Francisco, CA, USA, May 18-26, 2013*, pages 1032–1041, 2013.

[ZSTS13b]   Marcelo Serrano Zanetti, Ingo Scholtes, Claudio Juan Tessone, and Frank Schweitzer. The rise and fall of a central contributor: Dynamics of social organization and performance in the Gentoo community. In *In Proceedings of the International Workshop on Cooperative and Human Aspects of Software Engineering (CHASE) helt at ICSE 2013*, May 2013.