

The Social Dimension of Collaborative Information Spaces

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 increasingly large collaborative information spaces is one of the big challenges of our time. Although it is clear that the social and the information layer of such systems are inherently coupled and thus inseparable, the question how ranking schemes in information spaces are influenced by the structure and dynamics of the social systems that create them has been addressed at most partially. In this talk, we introduce results of a study in which we quantified the social influence on the semantic layer of two evolving collaborative information spaces. We first show that the position of contributors in the evolving collaboration structures of Open Source Software communities allows to predict whether the community categorizes the contributed information as useful. As a second example, we study a large-scale, time-stamped scholarly database of collaborations and citations for four scientific disciplines and show that the structure of collaborations is correlated with the patterns of future citations. Our preliminary results highlight the necessity to analyze, model and design information systems from a systems perspective that integrates both the social and the semantic layers of collaborative information spaces.

The extraction of relevant knowledge from increasingly large collaborative information spaces like the World Wide Web (WWW), scholarly publication databases, Online Social Networks (OSNs) or collaboration platforms 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 [1]. 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*. Although it is clear that the social and the information layer of such systems are inherently coupled and thus inseparable, the question how ranking schemes in information spaces are influenced by the structure and dynamics of the social systems that create them has been addressed at most partially. At the same time, 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. 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 [2, 3].

In this talk, we introduce results of a study in which we quantified the social influence on the semantic layer of two evolving collaborative information spaces: We first show that the position of contributors in the evolving collaboration structures of Open Source Software communities allows to predict whether the community categorizes the contributed information as useful [4, 5]. As a second example, we study a large-scale, time-stamped scholarly database of collaborations and citations for four scientific disciplines and show that the structure of collaborations is correlated with the patterns of future citations. We thus conclude that ranking measures that are based on the semantic layer of such collaborative information spaces are necessarily influenced by the structure and dynamics of the underlying social system. On the other hand, semantic structures of information systems (like e.g. citations and resulting notions of “scientific success”) influence future social structures and thus generate a feedback process. Our preliminary results highlight the necessity to analyze, model and design information systems from a systems perspective that integrates both the social and the semantic layers of collaborative information spaces. We will introduce associated research challenges as well as approaches that promise to be suitable to tackle them.

References

- [1] Jon M. Kleinberg. The convergence of social and technological networks. *Communications of the ACM*, 51(11):66, November 2008.
- [2] Lucas Introna and Helen Nissenbaum. The Politics of Search Engines. *IEEE Computer*, pages 54–62, 2000.
- [3] Felicitas Kraemer, Kees Overveld, and Martin Peterson. Is there an ethics of algorithms? *Ethics and Information Technology*, 13(3):251–260, July 2010.
- [4] 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.
- [5] 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) held at ICSE 2013*, May 2013.