Social Networks and the Semantic Web

Mika, P.

2007

document version
Publisher's PDF, also known as Version of record

Summary

Whether we changed the Web or the Web has changed us is difficult to distil even when equipped with the wisdom of hindsight.

While the process is a mystery, the changes are a fact. A recent large scale study on the Internet use of Americans has recorded the dramatic shift in the way that we approach the online world [Boase et al., 2006]. If we think of the Web as a giant billboard, we can say that the early days were spent with some affixing notes to this board, while most were merely passing by, carelessly surfing from one note to the next. These days however we do not just 'surf' anymore. We have learned to use the billboard to actively seek out others and made it a gathering place. Around the board we discuss matters with friends and unknowns, ask and give advice, share our tastes, experiences and ideas, and build relationships in the process.

The end result is clear from the survey: the ties we build are activated even in those situations that in the past we used to solve solely with the help of our closest allies, our intimate friends and family. The Net has changed the size and composition of our social networks: in particular, our networks have grown with an array of weak ties – a common name for those familiar faces on our blog rolls, buddy lists, chat groups, fora, mailing lists and the myriad other forums of our interactions.

Needless to say, the billboard had to change to adapt to its new function. What we now call Web 2.0 is a collective name for these evolutionary changes. The most important of these changes is conceptual: the Web has finally become the read/write Web that its inventor originally intended it to be. The popular 'places' of today are created by and for the people. We not only hang around the billboard and socialize, but use the space of the board to build a collective intelligence. The wisdom of the crowd is put to use in building and managing large repositories of knowledge such as Wikipedia, the online encyclopedia. And 'the crowd' is not only editing encyclopedias either, but we are also sharing photos and music, hunting for books, filtering news, writing stories, organizing digital collections through classification, and much more.

Technological change played the minor role in this process and it mostly had to with the adapting the Web to collaboration. New technologies such as AJAX-based development and a no-nonsense design of the websites have significantly improved the user experience while interacting with web applications, while RSS feeds and other technologies improved the connectivity between users and the content of the Web. The immense popularity of scripting languages hints at the way programming has been democratized and turned into a form of art, rather than engineering. The (sense of) collective own-

ership of both content and code continues to inspire creative experimentation with web content in the form of mash-ups, web applications that combine user generated content from multiple services.

In contrast to Web 2.0, the Semantic Web is an effort to carry out a more fundamental change in the architecture of the Web. Initiated by Tim Berners-Lee, the head of the World Wide Web Consortium (the standards organization behind the Web), the idea of the Semantic Web is to make the Web friendlier for machines. While at the moment most of the content in the online world is only accessible to human readers, the Semantic Web would provide additional layers of Web architecture for describing content using shared vocabularies called ontologies. This would allow computers to reason with the knowledge expressed in Web resources, in particular to aggregate relevant information from multiple sources and to come to conclusions in a manner that resembles human logic. While an infrastructure for machines, the knowledge that fills the Semantic Web and the rules of reasoning will in fact be provided by humans. In short, there is no semantics without humans and this makes the Semantic Web as much a social system as a technological one.

These developments are of interest to researchers in both the Social and Information Sciences, as well as to practitioners developing social-semantic software for the Web. On the one hand, the emergence of the Social Web opens up never foreseen opportunities for observing social behavior by tracing social interaction on the Web. Semantic Web technology comes to help by providing the means to aggregate the fragmented information about our online social networks. On the other hand, user generated content and metadata in social software requires a different treatment than other content and metadata. In particular, this knowledge comes with additional information about the social context in which it is conceived and this information (in particular, the social networks of users) is also accessible for our machines to reason with. This provides unprecedented possibilities in building socially-aware information systems.

In this book we provide two major case studies to demonstrate each of these opportunities. The first case study shows the possibilities of tracking a research community over the Web, combining the information obtained from the Web with other data sources (publications, emails) and analyzing the results (Chapter 9). Social network mining from the Web plays an important role in this case study for obtaining large scale, dynamic network data beyond the possibilities of survey methods. In turn semantic technology is the key to the representation and aggregation of information from multiple heterogeneous information sources. The methods of social network analysis are applied to the results in order to obtain network-based predictors of the performance of researchers.

Our methodology is more generally applicable than the context of our scientometric study, and thus we will spend significantly more time on describing our methods than discussing our results. We summarize the possibilities for (re)using electronic data for network analysis in Chapter 4 and evaluate two methods of social network mining from the Web in a separate study described in Chapter 8. We discuss semantic technology for social network data aggregation in Chapters 5 and 6. Lastly, we describe the implementation of our methods in the award-winning Flink system in Chapter 7. In fact these descriptions should not only allow the reader to reproduce our work, but to apply our methods in a wide range of settings. This includes adapting our methods to other

social settings and other kinds of information sources, while preserving the advantages of a fully automated process based on electronic data.

Our second study highlights the role of the social context in user-generated classifications of content, in particular in tagging systems known as folksonomies (Chapter 10). Tagging is widely applied in organizing the content in many Web 2.0 services, including the social bookmarking application del.icio.us and the photo sharing site Flickr. We consider folksonomies as lightweight semantic structures where the semantics of tags emerges over time from the way tags are applied. We study tagging systems using the concepts and methodology of network analysis. We establish that folksonomies are indeed much richer in semantics than it might seem at first and we show the dependence of semantics on the social context of application. These results are particularly relevant for the development of the Semantic Web using bottom-up, collaborative approaches. Putting the available knowledge in a social context also opens the way to more personalized applications such as social search.

As the above descriptions show, both studies are characterized by an interdisciplinary approach where we combine the concepts and methods of Artificial Intelligence with those of Social Network Analysis. However, we will not assume any particularly knowledge of these fields on the part of the reader and provide the necessary introductions to both (Chapters 2 and 3). These introductions should allow access to our work for both social scientists with an interest in electronic data and for information scientists with an interest in social-semantic applications.

Our primary goal is not to teach any of these disciplines in detail but to provide an insight for both Social and Information Scientists into the concepts and methods from outside their respective fields. We show some of the benefits that this understanding could bring in addressing complex outstanding issues that are inherently interdisciplinary in nature. Outside of the domain of Social Science we foresee further practical applications in areas where the automated, intelligent aggregation of personal electronic profiles and social networks plays an important role, mainly in digital lifestyle aggregation (connecting our fragmented online resources), but also in knowledge management, intelligence, emergency management and many others.

Our hope is also to inspire further creative experimentation toward a better understanding of both online social interaction and the nature of human knowledge. Such understanding will be indispensable in a world where the border between these once far-flung disciplines is expected to shrink rapidly through more and more socially immersive online environments such as the virtual worlds of Second Life. Only when equipped with the proper understanding will we succeed in designing systems that show true intelligence in both reasoning and social capabilities and are thus able to guide us through an ever more complex online universe.

The Author would like to acknowledge the support of the Vrije Universiteit Research School for Business Information Sciences (VUBIS) in conducting the research contained in this volume.