08391 Abstracts Collection
Social Web Communities
— Dagstuhl Seminar —

Harith Alani[1], Steffen Staab[2] and Gerd Stumme[3]

[1] Univ. of Southampton, GB
h.alani@ecs.soton.ac.uk
[2] Universität Koblenz-Landau, DE
staab@uni-koblenz.de
[3] Universität Kassel, DE
stumme@cs.uni-kassel.de

Abstract. From September 21st to September 26th 2008, the Dagstuhl Seminar 08391 "Social Web Communities" was held in the International Conference and Research Center (IBFI), Schloss Dagstuhl. During the seminar, several participants presented their current research, and ongoing work and open problems were discussed. Abstracts of the presentations given during the seminar as well as abstracts of seminar results and ideas are put together in this paper. The first section describes the seminar topics and goals in general. Links to extended abstracts or full papers are provided, if available.

Keywords. Social Web Communities, Social Network Analysis, Collaborative Tagging

08191 Executive Summary – Social Web Communities

Blogs, Wikis, and Social Bookmark Tools have rapidly emerged on the Web. The reasons for their immediate success are that people are happy to share information, and that these tools provide an infrastructure for doing so without requiring any specific skills. At the moment, there exists no foundational research for these systems, and they provide only very simple structures for organising knowledge. Individual users create their own structures, but these can currently not be exploited for knowledge sharing. The objective of the seminar was to provide theoretical foundations for upcoming Web 2.0 applications and to investigate further applications that go beyond bookmark- and file-sharing.

The main research question can be summarized as follows: How will current and emerging resource sharing systems support users to leverage more knowledge and power from the information they share on Web 2.0 applications? Research areas like Semantic Web, Machine Learning, Information Retrieval, Information Extraction, Social Network Analysis, Natural Language Processing, Library and Information Sciences, and Hypermedia Systems have been working for a while

Dagstuhl Seminar Proceedings 08391
Social Web Communities
http://drops.dagstuhl.de/opus/volltexte/2008/1792

on these questions. In the workshop, researchers from these areas came together to assess the state of the art and to set up a road map describing the next steps towards the next generation of social software.

Joint work of: Harith Alani, Steffen Staab and Gerd Stumme

Extended Abstract: http://drops.dagstuhl.de/opus/volltexte/2008/1786

08191 Group Summary – The Berners-Lee Hypothesis: Power laws and Group Structure in Flickr

An intriguing hypothesis, first suggested by Tim Berners-Lee, is that the structure of online groups should conform to a power law distribution. We relate this hypothesis to earlier work around the Dunbar Number, which is a supposed limit to the number of social contacts a user can have in a group. As preliminary results, we show that the number of contacts of a typical Flickr user, the number of groups a user belongs to, and the size of Flickr groups all follow power law distributions. Furthermore, we find some unexpected differences in the internal structure of public and private Flickr groups. For further research, we further operationalize the Berners-Lee hypothesis to suppose that users with a group membership distribution that follows a power law will produce more content for social Web systems.

Keywords: Social group flickr powerlaw

Joint work of: Andrea Baldassarri, Alain Barrat, Andrea Cappocci, Harry Halpin, Ulrike Lehner, Jose Ramasco, Valentin Robu, Dario Taraborelli

Extended Abstract: http://drops.dagstuhl.de/opus/volltexte/2008/1789

08191 Group Summary – The Evolution and Dynamics of Research Networks

Existing collaboration and innovation in scientific communities can be enhanced by understanding the underlying patterns und hidden relations. Social network analysis is an appropriate method to reveal such patterns. Nevertheless, research in this area is mainly focused on social networks. One promising approach is to use homophily networks as well. Furthermore, extending the static to a dynamic network model enables to understand existing interdependencies in these networks. A mathematical description of possible analyses is given. Finally, resulting research questions are illustrated and the necessity of an interdisciplinary research approach is pointed out.

Keywords: Homophily networks, social networks, evolution, scientific community

Joint work of: Vladimir Batagelj, Bettina Hoser, Claudia Müller, Steffen Staab, Gerd Stumme

Extended Abstract: http://drops.dagstuhl.de/opus/volltexte/2008/1790

08191 Group Summary – Mining for Social Serendipity

A common social problem at an event in which people do not personally know all of the other participants is the natural tendency for cliques to form and for discussions to mainly happen between people who already know each other. This limits the possibility for people to make interesting new acquaintances and acts as a retarding force in the creation of new links in the social web. Encouraging users to socialize with people they don't know by revealing to them hidden surprising links could help to improve the diversity of interactions at an event. The goal of this paper is to propose a method for detecting *"surprising"* relationships between people attending an event. By *"surprising"* relationship we mean those relationships that are not known a-priori, and that imply shared information not directly related with the local context of the event (location, interests, contacts) at which the meeting takes place. To demonstrate and test our concept we used the Flickr community. We focused on a community of users associated with a social event (a computer science conference) and represented in Flickr by means of a photo pool devoted to the event. We use Flickr metadata (tags) to mine for user similarity not related to the context of the event, as represented in the corresponding Flickr group. For example, we look for two group members who have been in the same highly specific place (identified by means of geo-tagged photos), but are not friends of each other and share no other common interests or, social neighborhood.

Keywords: Serendipity, online activity, context, ubiquitous computing

Joint work of: Alexandre Passant, Ian Mulvany, Peter Mika, Nicolas Maisonneauve, Alexander Löser, Ciro Cattuto, Christian Bizer, Christian Bauckhage, Harith Alani

Extended Abstract: http://drops.dagstuhl.de/opus/volltexte/2008/1791

08191 Working Group Summary – Analyzing Tag Semantics Across Collaborative Tagging Systems

The objective of our group was to exploit state-of-the-art Information Retrieval methods for finding associations and dependencies between tags, capturing and representing differences in tagging behavior and vocabulary of various folksonomies, with the overall aim to better understand the semantics of tags and the tagging process. Therefore we analyze the semantic content of tags in the

Flickr and Delicious folksonomies. We find that: tag context similarity leads to meaningful results in Flickr, despite its narrow folksonomy character; the comparison of tags across Flickr and Delicious shows little semantic overlap, being tags in Flickr associated more to visual aspects rather than technological as it seems to be in Delicious; there are regions in the tag-tag space, provided with the cosine similarity metric, that are characterized by high density; the order of tags inside a post has a semantic relevance.

Keywords: Social Web Communities, Folksonomy, Tag, Semantics

Joint work of: Dominik Benz, Marko Grobelnik, Andreas Hotho, Robert Jäschke, Dunja Mladenic, Vito D. P. Servedio, Sergej Sizov, Martin Szomszor

Extended Abstract: http://drops.dagstuhl.de/opus/volltexte/2008/1785

Cross-linking Folksonomies

Harith Alani (Univ. of Southampton)

There is a great push toward opening up, and linking, social network sites. This talk is aimed towards highlighting this move, and the new research topics that can be built on it.

Keywords: Folksonomies

Measuring Success of Online Communities

Harith Alani (Univ. of Southampton)

Many businesses are rushing towards launching their own online communities. However, very few of them are equipped with any tools or knowledge to measure how successful their communities are in terms of meeting their business objectives. This talk argues for the need to research and study these issues to guide the designs of online communities and to increase their benefit.

Contribution to the workshop

Andrea Baldassarri (University of Rome "La Sapienza")

In my talk I describe how the ideal workflow for the study of complex systems can be applied to tagging systems and "folksonomies".

Leveraging the unprecedented amount of data about user activity in on-line social communities collected through on-line systems, the researcher can identify statistical observables aimed to quantify emergent features, i.e. features which can not be simply deduced by the microscopic action of each single user. Such

features should be observed in different systems and situations, calling for a deeper, more general understanding or explanation.

The very existence of such features is nowadays corroborated by observation of regularities in probability distributions and other finer statistical observables.

Formulation of simple, minimal models can suggest basic mechanisms reproducing the observed phenomenology, as well as evidence relation and dependance between different observations (null models are crucial for this).

But models are not just descriptive. They predict unobserved features, against which models are checked, and, whenever possible, they allow a mathematical, analytical description of the systems in study.

Finally, the higher level of understanding attained, feeds back in the design and control of on-line systems, closing a virtuous loop of increasing knowledge and exploitation.

In the case of folksonomy systems, measures and models has been focused on global statistics of user tagging activity, as well as on the emerging correlations between tags, revealed by tag co-occurrence network, whose study involves definition and scrutiny of similarity measures, of their effectiveness and meaning, and suggests the existence of a latent semantic structure of concepts.

Exploiting the full tripartite nature of folksonomy systems (tags, users, resources), similar analysis could be performed in order to reveal emergent classifications of resources, or underlying unexpected relations between users.

Keywords: Folksonomy tagging complexsystems

Collective dynamics of social annotation

Alain Barrat (Centre de Physique Théorique - Marseille)

The enormous increase of popularity and use of the WWW has led in the recent years to important changes in the ways people communicate. An interesting example of this fact is provided by the now very popular social annotation systems, through which users annotate resources (such as web pages or digital photographs) with text keywords dubbed tags. Understanding the rich emerging structures resulting from the uncoordinated actions of users calls for an interdisciplinary effort. In particular concepts borrowed from statistical physics, such as random walks, and the complex networks framework, can effectively contribute to the mathematical modeling of social annotation systems. Here we show that the process of social annotation can be seen as a collective but uncoordinated exploration of an underlying semantic space, pictured as a graph, through a series of random walks. This modeling framework reproduces several aspects, so far unexplained, of social annotation, among which the peculiar growth of the size of the vocabulary used by the community and its complex network structure that represents an externalization of semantic structures grounded in cognition and typically hard to access.

Keywords: Tagging, networks, random walks

Joint work of: Cattuto, Ciro; Barrat, Alain; Baldassarri, Andrea; Schehr, Gregory; Loreto, Vittorio

Analysis and visualization of large networks

Vladimir Batagelj (University of Ljubljana,)

Some new approaches to analysis of large networks are presented.

Keywords: Pajek, cut, island, ring, core, multiplication of networks, 2-mode networks, pattern search, clustering with relational constraint

Web Wide Context, Convergence, Professional Communities

Christian Bauckhage (Deutsche Telekom Laboratories)

The ever increasing popularity and diversification of web communities will continue to impact life and work in the information society. We present three emerging trends and point out questions regarding their technological, economical, and social consequences.

Tagging communities: dynamics and analysis

Andrea Capocci (University of Rome "La Sapienza")

Social tagging systems can be analysed both from the dynamical point of view, in order to uncover collective phenomena taking place on them, and from the semantical point of view.

Keywords: Social del.icio.us tagging interarrival

Information-Theoretic Models of Tagging

Harry Halpin (University of Edinburgh)

In earlier work, we showed using Kulback-Leibler (KL) divergence that tags form a power law distribution very quickly. Yet there is one major observed deviation from the ideal power law distribution for the top 25 tags, a large "bump" in increased frequency for the top 7-10 tags.

We originally hypothesized that the "bump" in the data could be caused by a preferential attachment mechanism. However, an experiment that tested both feedback and no-feedback conditions over tagging (200+ subjects) shows that the power law distribution arises regardless of any feedback effect. We hypothesize that an information-theoretic analysis of tags lead to a power law without feedback.

Keywords: Tagging information theory feedback

Extended Abstract: http://drops.dagstuhl.de/opus/volltexte/2008/1787

Full Paper:
http://www.ibiblio.org/hhalpin/homepage/presentations/dagstuhl

Exploiting Folksonomy Structure

Robert Jäschke (Universität Kassel)

Looking from a tag recommendation perspective, I realized that the folksonomy graph structure is often not sufficient to create good recommendations. Especially, when data is sparse, the user or resource to be annotated unknown. But there is much more data available: groups, friends, tag relations, time, place, order of tags, content information, click data, etc. How can this data be employed to generate better tag recommendations? E.g., when annotating a picture which contains geo coordinates, I would like to get some tags recommended from pictures taken at the same location. Extending this application-centric view, there are further questions which arise:

What can we learn from that data? How can we build a folksonomy (graph) model which takes this data into account?

BI-OVER-TEXT: Analyzing user-generated data using cloud computing techniques

Alexander Löser (TU Berlin)

More than 10 gigabyte of “consumer-generated content” is created in the World Wide Web daily. Examples for unstructured web-2.0-data are product reviews, online-forums, blogs, tags and wikis. Unfortunately, the producers of the content, e.g., end-consumers, or self employed vendors, are not able to drill down into that “knowledge” yet, since they do not have access to an affordable and simple business intelligence technology. Example queries cover areas, such as market and competitor analysis, buy decisions or the analysis of medical treatments. Our vision is to provide a business analysis engine that would allow “ordinary” people to run individual, spontaneous analysis queries over unstructured user generated data and within their community. We introduce to this vision and review recent developments in cloud computing techniques that will enable us to execute ad hoc queries over text data contributed by users.

Keywords: BI-Over-Text, cloud computing, user-generated content

Monitoring noise pollution in the City 2.0: the NoiseTube.net

Nicolas Maisonneuve (Sony CSL - Paris)

Noise pollution has become a serious social problem in many cities. Noisetube.net is an ongoing project aimed at monitoring the noise pollution exposure as experienced by urban citizens using a participatory approach. Thanks to GPS built-in mobile phones used as an environmental sensor to capture the decibel, localization but also user's impression inputs, any citizen concerned with urban noise can measure, share their daily noise exposures with the community and visualize the collective result on the city's map.

Of course there are technological challenges to develop tools for non-experts. In particular getting credible decibel measures from a cellphone whereas the use of professional sound level meter is normally required is a real one. Furthermore we also added a tagging feature on the mobile application to tag noise sources and add a layer of meaning on top of the physical measures. Mixing physical and semantic measures will also be a challenge. But beyond this technical aspect the use of mobile phone as an environmental sensor to gather environmental data empowers communities to build local campaign without waiting formal projects. Like the web 2.0 "user-generated content" was made possible thanks to new publishing tools (e.g. blog) for newbies, these "citizen-generated measures" open a new approach for the management of urban commons.

Keywords: Participatory sensing, noise pollution, community, mobility

A Short Note on Social-Semiotic Networks from the Point of View of Quantitative Semantics

Alexander Mehler (Universität Bielefeld)

In this extended abstract we discuss four related characteristics of semantic spaces as the standard model of meaning representation in quantitative semantics. We argue that these characteristics are challenged from the point of view of social web communities and the possibilities which they offer in terms of exploring semantic *and* pragmatic data. More specifically, we plead for a reconstruction of the weak contextual hypothesis in order to account for non-linguistic, pragmatic aspects of context. Finally, we mention two consequences of such a pragmatic turn, that is, in the area of named entity recognition and of language evolution.

Keywords: Semantic space, social web community, quantitative semantic weak contextual hypothesis

Extended Abstract: http://drops.dagstuhl.de/opus/volltexte/2008/1788

Social interaction analysis

Dunja Mladenic (Jozef Stefan Institute - Ljubljana)

Social Web communities produce trace of the interactions in electronic form and in many cases also some content data is available. Data analysis of such content and interactions can be performed with different methods coming from social network analysis, data mining, text mining, cross-modal data anlysis. The analysis enable addressing of interesting problems including analysis of opinion makers, understanding of social network properties, cross-modal data analysis for recommendation systems or user profiling, community dynamics etc.

Keywords: Data analysis, social network analysis, data mining, text mining

Object-centric social communities

Alexandre Passant (Nat. University of Ireland - Galway)

We believe that a next step regarding "Web 2.0" is to provide virtual communities related to a particular object (movie, event, ...) so that one can easily find all user-generated content and people related to it.

Still, we must solve the issues of interoperability between social media websites and link people and content not only to simple tags, but to those objects / concepts.

We think that Semantic Web technologies can help to achieve those two goals, and our current efforts in that direction include the SIOC project, a model and food-chain to represent activities of online communities, and MOAT, a process to bridge the gap between tagging and semantic indexing.

Keywords: Semantic Web, Online communities, Linked Data, SIOC, MOAT, Tagging

The complex dynamics of collaborative tagging and sponsored search markets

Valentin Robu (CWI - Amsterdam)

We briefly review some results from two directions of research: analyzing the complex dynamics of collaborative tagging systems, and the dynamics of sponsored search markets. The tagging presentation is based on [Halpin et al., 2007] - with some more recent results.

The sponsored search part covers some preliminary results from a project analyzing the dynamics of advertising (sponsored) links, performed in collaboration with Sander Bothe and Han La Poutre (CWI, Amsterdam). It involves analyzing the relationships between click-through rates of different keywords,

the effect of the position of the link on the click-through rate, as well as the distribution of the market share of different advertisers in such a market.

[Halpin et al., 2007] Harry Halpin, Valentin Robu, Hana Shepherd "The complex dynamics of collaborative tagging" Proc. of the 16th International Conference on World Wide Web, Banff, 2007, pp. 211 - 220.

Keywords: Collaborative tagging, sponsored search, complex systems, Pajek, folksonomy, community algorithms

Quantitative Linguistics seen by a physicist

Vito Domenico Pietro Servedio (University of Rome "La Sapienza")

After a brief introduction of myself and an overview of my expertises, I shall present an example of *quantitative linguistics* as might be intended by a physicist.

I'll show an overview of some interesting quantities, like dictionary growth and correlations, that can be calculated from a stream of words in case of a text, or a stream of tags in case of folksonomies. The hope is that those quantities may help to uncover the semantics embedded in words and tags.

Eventually, I'll sketch some possible ideas for future works.

Keywords: Linguistics, correlations, dictionary, growth, tags, words

Latent Semantics of Social Media

Sergej Sizov (Universität Koblenz-Landau)

The rapidly increasing popularity of Web 2.0 knowledge and content sharing systems and growing amount of shared data make discovering relevant content and finding contacts a difficult enterprize.

Typically, folksonomies provide a rich set of structures and social relationships that can be mined for a variety of recommendation purposes. We propose a novel formal model for characterizing users, items, and annotations in Web 2.0 environments. Our objective is to construct social recommender systems that predict the utility of items, users, or groups based on the multi-dimensional social environment of a given user.