

Dealing With Community Data

Amy Bruckman¹ Thomas Erickson² Danyel Fisher³ Christopher Lueg⁴

ABSTRACT

As communities online grow, the set of techniques to store, analyze and understand their histories has not necessarily kept pace. This workshop is primarily designed to discuss what techniques are useful and interesting, and to share methodologies, for examining online communities. We will examine technical issues of storage mechanisms, examine both qualitative and quantitative analysis, and consider ways to feed back results of analysis into the community.

THEME

Online communities have expanded and multiplied in the last few years; large amounts of data have been accumulated, as every virtual organizer carefully logs their community's history. Privacy advocates worry about what happens to these logs; advertisers mine the logs looking for purchasing patterns. Researchers look at the activity patterns of their online groups [1], design new systems [3], and measure existing social systems [3]. Surveys of users allow studies that give background information on the communities [4]. Previous projects have provided interesting reviews of the data in a particular group [5], and have argued the coherence of social structures inside groups [6].

The processing of this data has been, so far, inconsistent; there is little notion of how to approach the logs; there are no standard measures for a group's vitality or activity. This leaves both community organizers and maintainers short on tools to store, process, and analyze these histories. The data is sometimes used for

This workshop will discuss a variety of perspectives on measuring and using logged community data. Emphasis will be given to a series of interlocking questions:

- What technical issues are involved in logging, storing, and measuring this data? It can be hard to find which data to store; a large community can generate potentially huge amounts of data.
- What quantitative measurements can be applied? What techniques can be used to measure the community? What measurements are interesting, important, or relate well to meaningful topics?

- What qualitative measurements are useful? What information can be understood from the logs without technical analysis? How can the logs be made easy to use for this sort of review? How can visualization techniques aid review?
- How can these results be used? What information does that data provide about the underlying community? How can the data be fed back into the community?
- What ethical issues emerge from these analyses? What should the data be used for? Are there forms of analysis that reveal too much about users?

A variety of computer science technologies, from databases to collaborative filtering[7], to social navigation, are intimately related to these questions. A number of social sciences, as social network analysis, social psychology, and discourse analysis, may have relevant insights.

There have been a number of past community workshops, many of them emphasizing notions of group vitality, asking how communities form, and discussing the questions of how to build a community infrastructure. We look forward to the opportunity to move onward to the dynamics of active communities.

ACTIIVITIES AND GOALS

The workshop will begin with a general orientation, as the organizers set a tone for the workshop and explain its organization. We will proceed with introductions by participants, who will be asked to each quickly address a question on communities that we will ask in advance.

Because we expect each participant to have read the position papers, there is little need to re-present them during the session. We will quickly begin a series of discussions intended to find the areas of common interest at the workshop, both as outlined above, and that the participants may bring.

We will then move to specific areas of concentration, and have three one-hour rounds of discussion. Each small group will come back and report their ideas; after three iterations, we expect some polished ideas will emerge.

By the end of the workshop, participants should be armed with a variety of techniques for online community

¹ College of Computing, Georgia Institute of Technology, Atlanta, GA 30332. asb@cc.gatech.edu

² IBM T.J. Watson Research Center, Yorktown Heights, New York, USA, snowfall@acm.org

³ Computer Science Department, University of California, Berkeley, USA. +1 510 642 8149 danyelf@cs.berkeley.edu

⁴ Department of Information Technology, University of Zurich. Zurich, Switzerland +41 1 63 54577. lueg@ifi.unizh.ch

assessment and maintenance. If all goes well, we look forward to first steps on the way to designing a set of basic standards for community measurement and understanding. Results will be promoted in a poster at CSCW; we will encourage participants to have their work published in an appropriate venue.

ORGANIZERS

Amy Bruckman is an Assistant Professor in the College of Computing at the Georgia Institute of Technology. She and her students in the Electronic Learning Communities (ELC) research group do research on online communities and education. Current projects include MOOSE Crossing (a text-based virtual world for kids) and The Turing Game (a game about identity and deception online). Amy received her PhD from the MIT Media Lab's Epistemology and Learning group. <http://www.cc.gatech.edu/~asb/>

Tom Erickson, IBM Research, is a designer and researcher of systems to make it possible for large numbers of people to interact coherently and productively over networks. His research group has designed and run a number of online systems, including the recent "Babble". http://www.pliant.org/personal/Tom_Erickson/index.html

Danyel Fisher has focused his graduate studies at UC Berkeley on internet interaction. His master's thesis was on WebPlaces, an IBM-based project for interaction on the WWW. Since then, he's worked on sociologically-related projects on visualizing newsgroup and community interaction. <http://www.cs.berkeley.edu/~danyelf>

Christopher Lueg is a senior research assistant at the Department of Information Technology, University of Zurich, Switzerland. This summer, he joined the University of Technology Sydney, Australia as Senior Lecturer. His doctoral thesis focused on supporting information seeking in electronic information environments. More recently, he began to explore social navigation as well as participation and identity shaping in the context of information spaces. Apart from academic qualifications, he is addicted to Usenet and has been running an academic news server for more than five years. <http://www.ifi.unizh.ch/staff/lueg>

PARTICIPANTS

Participants will be selected on the basis of position papers submitted prior to the workshop. Proposals should be no more than three pages in length and should consist of a description of some of the following:

- A currently-operative online communities,
- Theoretical ideas for the analysis,
- Practical discussion of logging and analysis issues,
- Methods for feeding back results into the community.

In general, the papers should explain how the author's work relates to the workshop theme, and should outline any measurements that the author already is taking on a community. We are particularly interested in seeing

references to measurements of group size, activity, and interconnection.

We expect a dozen participants, but are open to expanding to twenty and spending somewhat more time split off into small groups. We expect the mix to be heavily weighted toward online community maintainers—people who run or analyze communities at work or at home. However, we look forward to a wide variety of other participants, from community ethnographers and sociologists to system builders.

REFERENCES

1. Marc Smith, Shelly Farnham, Steven Drucker *The Social Life of Small Graphical Chat Spaces*. CHI 2000 Proceedings.
2. Thomas Erickson; David N. Smith; Wendy A. Kellogg; Mark Laff; John T. Richards; Erin Bradner. *Socially Translucent Systems: Social Proxies, Persistent Conversation, and the Design of 'Babble'*. CHI 1999 Proceedings.
3. Steve Whittaker, Loren G. Terveen, William C. Hill, Lynn Cherny: *The Dynamics of Mass Interaction*. CSCW 1998 Proceedings.
4. Blair Nonnecke, Jenny Preece *Lurker Demographics: Counting the Silent*. CHI 2000 Proceedings.
5. Erickson, Thomas. *Rhyme and Punishment: The Creation and Enforcement of Conventions in an On-Line Participatory Limerick Genre*. HICSS Proceedings, 1999.
6. Teresa L. Roberts. *Are newsgroups virtual communities?* CHI 1998 Proceedings.
7. Will Hill and Loren Terveen. Using Frequency-of-Mentioning in Public Conversations for Social Filtering. CSCW 1996 Proceedings.

RELATED PAST WORKSHOPS

Paolo Barbesino, Peter van den Besselaar, Fiorella de Cindio, Peter Mambrey, Douglas Schuler, Artur Serra, Erik Stolterman. *Designing Across Borders: The Community Design of Community Networks*. CSCW 1998

Lori Toomey, John Tang, Lia Adams, Gloria Mark *Designing Virtual Communities for Work*. CSCW 1998

Michael Muller and Jessica Friedman *Electronic Communities: Places and Spaces, Contents and Boundaries*. CHI 2000

Amy Bruckman, Thomas Erickson, Wendy Kellogg, Lee Sproull, Barry Wellman. *Workshop on Research Issues in the Design of Online Communities*. CHI 1999

Jenny Preece, Jonathan Lazar. *Online Communities: Supporting Sociability, Designing Usability*. UMD HCIL Open House, June 2000