A Visual Inspection of Online Health Communities

Diana MacLean, Jeffrey Heer Stanford University {malcdi, jheer}@cs.stanford.edu

ABSTRACT

Online health communities (OHCs) are a growing source of public medical knowledge; they facilitate several healthrelated tasks including searching for and acquiring new medical information, and seeking emotional support. However, many underlying attributes of OHCs, such as forum discussion type and leadership hierarchy, are not readily apparent during simple forum browsing. Knowledge of these attributes might comprise useful decision making tools for a spectrum of community participants: from leaders, who maintain OHC forums, to potential new members, who might "window shop" forums before picking a best fit. Prior work demonstrates that visualization is an effective technique for discovering and exploring underlying online community attributes. We present a preliminary study using visualization techniques to interpret and compare community dynamics from selected MedHelp OHCs. We find that our approach successfully uncovers several OHC attributes that vary between disease forums; for example, some OHCs are characterized by Q&A interaction patterns, others by discursive interaction patterns.

Author Keywords

Health, Visualization, Online Community Management

ACM Classification Keywords

H.5.2 Information Interfaces and Presentation: Miscellaneous

INTRODUCTION

As people rely increasingly on the Internet as a source of medical knowledge, OHCs are becoming more and more prevalent. This shift is attributed mostly to changes in the health care system (lower access to healthcare professionals and higher costs of health care) and increased technological literacy in the general population [9]. Although the question of whether OHCs provide members with *actual* health benefits remains open [5], OHCs have several advantages over traditional health information systems. These include cost effectiveness [9]; always-available, unrestricted access[8]; patient empowerment [14]; providing comfortable venues

for discussing sensitive issues [14]; and enhanced social support stemming from interactions with people suffering from the same illness [9, 10, 14].

Despite this, OHCs also carry several disadvantages. Sites such as PatientsLikeMe, MedHelp, and others, cater to the growing population of OHC participants primarily by providing interactive discussion forums. However, manual evaluation of information quality in these forums [8, 14], or the trustworthiness and personal agenda of participants [8] is prohibitively tedious. Moreover, the vast quantity of peer-topeer contributions make OHCs difficult to navigate, as users must sift through archives of posts in a search for relevant information [8, 14].

Prior work indicates that visualizing online communities can be an effective technique for uncovering a range of subtextual community attributes, such as member demographics [12], conflict patterns [2], and types of social roles played by community members [12]. In this paper, we present exploratory, visual summaries of OHC data. Our immediate goal is to discover a set of salient attributes on which OHCs vary, with a long-term vision is to inform and enhance people's interactions with OHCs.

RELATED WORK

For a thorough overview of prior work on online communities, see Iriberri and Leroy's work on online community life cycles and evolution [7].

Online Health Communities

To our knowledge, no prior work exists on creating visual synopses of OHCs. However, of note is the work by Eysenbach et al. [5] and Rogers and Chen [10], who investigate the effect of OHC participation on illness prognoses; and a longitudinal study of a knee-injury support forum, in which Maloney-Krichmar and Preece determined several significant findings about OHCs. They found that external community governance is rarely required, and that communities have distinctive member roles and subgroups [8]. Finally, Brownstein et al. argue that OHCs comprise a viable data source for medical knowledge discovery [3].

Visualizing Online Communities

Our work follows closely that of Viégas and Smith, in which they present two visualization techniques designed to reveal attributes – such as temporal posting trends and community role demographics – of Usenet newsgroup members [12].



Figure 1. Node-link community diagrams & question/answer log scatterplots by forum. In the node-link diagrams, an edge is placed from one node (n1) to the other (n2) if n1 participates in a thread initiated by n2. Red nodes depict community leaders; orange nodes depict doctors; dark blue nodes indicate community leaders who have been awarded barn stars for commendable answers. Isolates occur when users initiate a thread that never receives a response. The question/answer log scatterplots depict the log question (x-axis) to answer (y-axis) ratio for each community member. Bubble size represents number of days between that member's first and last posts.

Their visualizations convey community types (discussionoriented or Q&A) as well as member demographics (new users, members participating in other newsgroups, consistent contributors, etc.). They were also able to convey several user role patterns within newsgroups, including those of answer person, spammer, and conversationalist.

Welser et al. present visual methods for categorizing "role signatures" in online communities [13]. Although they focus on identifying "answer people" in particular, their work suggests that visualization techniques combined with regression analyses can be used to build successful models for role prediction. Other related work includes Smith and Fiore's research on visualizing online discussions for improving user navigation through online community spaces [11]. The primary focus of this work is on content analysis and categorization, which we do not address in this paper.

Knowledge Sharing in Online Communities

In a study of knowledge sharing in Yahoo Answers forums, Adamic et al. investigate participation and interaction patterns of users across an array of online forums [1]. Based on user interaction attributes such as post length and crosscategory posting, they were able to classify forums into "discussion" and "Q&A" types. They also found that user role demographics were likely to be different based on community types. Along similar veins, Fiore et al. present a method of employing behavioral descriptors (posts, replies etc.) to estimate compatibility between forum users [6].

DATA

MedHelp¹ is a free, online health community website designed to aid users in the diagnosis, exploration, and management of personal illnesses. The site boasts a wide array of tools and services, including over 200 "Medical Support Communities", where users discuss medical conditions amongst themselves. Forums are structured in typical bulletinboard style: users reply to an initial post. Responses are not inlined, so detecting conversations between participants in longer threads is not possible without text analysis.

For this preliminary study, we gathered and analyzed data from a small, but highly diverse, set of forums. After browsing MedHelp's OHCs, we selected the following 5 because of their representative diversity in terms of disease type (mental, physical), prognosis (terminal, non-terminal), ease of diagnosis, and rarity: Asthma, Breast Cancer, Depression, Lupus, and Lyme Disease. We used the BeautifulSoup² library to crawl all forum posts from the start date of the forum (very first post) to June 2010. Table 1 summarizes the results.

VISUAL EXPLORATIONS

Figures 1, 2 and 3 present our visualizations. We discuss our observations in the context of particular OHC attributes highlighted in prior work below.

¹http://medhelp.com/

²http://www.crummy.com/software/BeautifulSoup

Forum	# Posts	# Threads	Thread Length Dist.	# Members	% Isolates	Post Frequency Trends
Asthma	3,903	1,369	lh	1,606	19.6	10/07/~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Breast Cancer	33,640	9,181		9,049	5.5	9/06
Depression	19,849	4,563	dh	5,125	6.0	11/06
Lupus	3,603	836	İ h	862	8.9	9/07 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Lyme Disease	9,762	1,630	, 111, 1 ,	879	4.1	9/07

Table 1. Forum Summary

Community Types & Entry Barriers

Viegas and Smith note that different online communities have different types, or structures, in terms of how information disseminates amongst their members [12]. In the node-link diagrams of Figure 1 we see that some communities have clear information hierarchies, in which leaders (either formal or informal) systematically respond to forum posts, while other communities entail involvement from many participants. In the Asthma forum, for example, users with a single edge have asked one-time questions, and bunch around people dispatching responses. The Lyme Disease forum, comparatively, is much flatter and denser, as one would expect of a forum with long debates. Indeed, the clustering coefficients of the Asthma and Lyme Disease communities are 0.21 and 0.55, indicating that Lyme Disease community members are much more likely to have communicated with each other than Asthma community members; their average thread lengths are 2.85 and 6.0, respectively.

The likelihood of response to a post by a new participant also varies across communities. For example, in Figure 1, we see that the isolate proportion in the Asthma community is large, while that of the Lyme Disease forum is smaller. Indeed, according to Table 1, the percentage of isolates in the Asthma and Lyme Disease communities is 19.6% and 4.1%, respectively. While isolate proportions might indicate one of several contexts (community exclusivity, high spam levels, disease rarity etc.), they give a coarse notion of the "entry barrier" for potential community members.

Roles & Sub-Communities

Prior work indicates that online community users have distinctive roles [12], and that user role demographics differ across online communities [1]. For example, certain communities may have a larger proportion of leaders than usual. Investigating user roles along two axes—membership tenure and forum contribution volume—we find that OHCs have distinctive community demographic role signatures, as shown in Figure 2. While in Figure 1 we noted that some communities have no officially-designated leaders, large bubbles near the top of Figure 2 suggest that *all* communities have someone filling a leadership role. Indeed, in the Lyme Disease forum (no officially allocated leaders), we see 3-4 unofficial leaders making significant forum contributions. The "official" Asthma community leaders (depicted by the 3 larger, middle dots) do not contribute nearly as much. In the previous section, we also noted that community knowledge exchange patterns may constitute particular forum types. Figure 2 suggests that member subsets may also comprise typed "sub-communities". Consider, for example, the Breast Cancer forum. The stack of tiny dots that runs almost all the way up the visualization indicates members who visit infrequently and contribute little: typical Q&A behavior. The clump of several larger bubbles in the middle depicts members who contribute a fair volume of content: typical discussion forum behavior.



Figure 2. A "Community Chromatograph". Each bubble represents a community member. Bubble size is proportion to the number of contributions that a user made to the forum, and y-axis placement indicates days since the user's first and last post. Thus, users are visually separated according to their forum contribution volume and community membership tenure.

Asking vs. Answering

A community chromatograph does not differentiate between question and answer contributions. Figure 1 depicts simple scatter plots, in which the x and y-axes represent number of questions asked and answered, respectively, and dot size represents number of days since first and last post. We discard fringe community members by retaining only members who have asked and answered at least 1 question. The x and y-axes are log-scaled.

In the scatterplots in Figure 1 we see strong diagonal trends in the Breast Cancer, Depression and Lyme Disease forums, indicating that long-term members tend to both ask and answer questions. This trend may indicate that these forums are discussion and support based, and is not seen as strongly in the Asthma or Lupus data. The forum scatterplots also depict clear horizontal and vertical lines close to the x and y axes. Consider the strong, vertical blue line on the y-axis in the Depression forum scatterplot: this line represents people who have asked only one question. However, the dot sizes indicate that many of these people went on to answer several questions.

Cross-Forum Posting & Medical Co-Occurrences

Prior work notes that OHC data may contain previously unknown connections between illnesses and treatments[3]. Moreover, certain medical conditions commonly co-occur: for example, hyperthyroidism and anxiety disorders. Figure 3 highlights cross-community posting activity of OHC members. While making medical inferences from co-occurrence data is premature, understanding community overlap is nevertheless useful information for OHC interface design.

Figure 3(a) shows a common co-occurrence (Lyme Disease and Thyroid Disorders), as well as a common misdiagnosis for Lyme Disease (Multiple Sclerosis). Figure 3(b), showing the top 14-29 cross-post communities, suggests a significant link between Asthma and Fertility. While no proven connection exists between the two, a recent analysis of OHC data from CureTogether suggests they are strongly correlated [4].



Figure 3. Cross-community post bars. Each bar represents a MedHelp OHC. Bar length represents the number of primary community members who are members of the bar community; hue encodes the number of posts made to the bar community by those members (also noted in the right of each bar). Figure 3(a) shows the top 15 communities with cross-membership for Lyme disease, while Figure 3(b) depicts the top 14-29 communities with cross-membership from Asthma.

DISCUSSION AND FUTURE WORK

We have presented several visualizations of online health community data with the goal of eliciting community attributes that are not obvious from typical forum-browsing behavior, but which may prove useful to community members. We discovered several subtextual community attributes, and explored their variation across different communities. In particular, we found that OHCs have definitive knowledgeexchange types and role demographics, that communities have leaders (whether official or not), and that members regularly participate in multiple OHCs.

Our study suggests several ways in which OHC visualizations might prove useful to users. Using community chromatographs and log question/answer scatterplots, community leaders could track posting trends over time, or detect neglected posts deserving a response. Potential members could use our visual summaries to evaluate whether or not a particular forum is a good fit, or to seek out community leaders.

Immediate next steps include a quantitative analysis of our discovered attributes, as well as a user study designed to gain more qualitative insight into observed patterns, using our visualizations as an elicitation tool. Additionally, comparative analyses with same-disease communities from different websites would provide insight into whether community attributes depend strongly on disease type. Another goal is to acquire forum data for a broader range of OHCs in order to determine whether they subscribe to some underlying categorization. Looking forward, we are interested in how incorporating visualizations into discussion interfaces may assist community members and facilitators to find and foster healthy forums.

REFERENCES

- L. Adamic, J. Zhang, E. Bakshy, and M. Ackerman. Knowledge sharing and yahoo answers: everyone knows something. In *World Wide Web*, pages 665–674, 2008.
- 2. U. Brandes and J. Lerner. Visualization of conflict networks. *Building* and Using Datasets on Armed Conflicts, 36.
- C. Brownstein, J. Brownstein, D. Williams, P. Wicks, and J. Heywood. The power of social networking in medicine. *Nature Biotechnology*, 27(10):888–890, 2009.
- A. Carmichael. Crowdsourced Health Confirms Infertility-Asthma Finding, 2009. http://curetogether.com/blog/2009/09/15/crowdsourced -health-confirms-infertility-asthma-finding.
- G. Eysenbach, J. Powell, M. Englesakis, C. Rizo, and A. Stern. Health related virtual communities and electronic support groups: systematic review of the effects of online peer to peer interactions. *British Medical Journal*, 328(7449):1166, 2004.
- A. Fiore, S. Tiernan, and M. Smith. Observed behavior and perceived value of authors in usenet newsgroups: bridging the gap. In ACM CHI, page 330, 2002.
- A. Iriberri and G. Leroy. A Life-Cycle Perspective on Online Community Success. ACM Computing Surveys, 41(2), 2009.
- D. Maloney-Krichmar and J. Preece. A multilevel analysis of sociability, usability, and community dynamics in an online health community. ACM TOCHI, 12(2):232, 2005.
- L. Neal, G. Lindgaard, K. Oakley, D. Hansen, S. Kogan, J. Leimeister, and T. Selker. Online health communities. In *Extended Abstracts,* ACM CHI, page 447. ACM, 2006.
- S. Rodgers and Q. Chen. Internet community group participation: psychosocial benefits for women with breast cancer. *Journal of Computer-Mediated Communication*, 10(4):5, 2005.
- M. Smith and A. Fiore. Visualization components for persistent conversations. In ACM CHI, pages 136–143. ACM, 2001.
- F. B. Viegas and M. Smith. Newsgroup crowds and authorlines: Visualizing the activity of individuals in conversational cyberspaces. In *Proc. Hawaii International Conference on System Sciences*, page 10, 2004.
- H. Welser, E. Gleave, D. Fisher, and M. Smith. Visualizing the signatures of social roles in online discussion groups. *Journal of Social Structure*, 8(2), 2007.
- M. White and S. Dorman. Receiving social support online: implications for health education. *Health Education Research*, 16(6):693, 2001.