

Assessing Differential Usage of Usenet Social Accounting Meta-Data

A.J. Bernheim Brush^{*}, Xiaoqing Wang⁺, Tammara Combs Turner^{*}, and Marc A. Smith^{*}

^{*}Microsoft Research
One Microsoft Way Redmond, WA 98052
[ajbrush, tcombs, masmith]@microsoft.com

⁺University of Pittsburgh
xiawang@katz.pitt.edu

ABSTRACT

We describe a usage study of Netscan\Tech, a system that generates and publishes daily a range of social metrics across three dimensions: newsgroup, author, and thread, for a set of approximately 15,000 technical newsgroups in Usenet. We bring together three interlinked datasets: survey data, usage log data and social accounting data from Usenet participation, to triangulate the relationship between various user roles and differential usage of social metrics in Netscan\Tech. We found our most frequent users focused on information related to individual authors far more than any other information provided. In contrast, users that visited less frequently focused more on information related to newsgroups and viewing newsgroup metrics. Our results suggest features that designers and developers of online communities may wish to include in their interfaces to support the cultivation of different community roles.

Author Keywords

Community, threaded discussions, newsgroups, Usenet, authors, posts, posters, assessment, reporting tools, social accounting metrics.

ACM Classification Keywords

H.5.3 Group and Organization Interfaces – Computer-supported cooperative work

INTRODUCTION

Usenet, a social cyberspace where millions of people interact within and across newsgroups, is a site that offers a unique opportunity for studying online social dynamics. The prevalence and growth of Usenet has facilitated understanding of social structure and social dynamics in online communities [1, 8, 15], and stimulated innovative ideas to develop systems in support of online interaction and collaboration [4, 14, 16, 17]. However, Usenet faces

both traditional and new challenges as the size of its population grows. Traditional problems of social groups such as the public goods dilemma and social loafing are evident in Usenet newsgroups [10]. Moreover, it is extremely difficult to keep track of social activity when newsgroup information is overwhelming in volume and lacks organization [8, 20].

The Netscan system is a resource for understanding the social context of Usenet newsgroups. Netscan is intended as an augmentation to existing Usenet interfaces with a primary focus on providing “social accounting meta-data” for activity in Usenet to help Usenet participants, newsgroups managers and community researchers understand the structure of Usenet newsgroups, conversations threads and authors. Netscan does not strive to serve as a complete interface to consume and contribute newsgroup content. For example, Netscan does not support posting to newsgroups.

The social accounting meta-data Netscan provides are metrics about the social dimensions of an online space, such as the number of messages and participants in a newsgroup, and information about the activities of each participant. Netscan generates social accounting metrics for three dimensions of online spaces: newsgroups, authors and threads. For large newsgroups with many messages, the information provided by Netscan can help users identify newsgroups and messages of interest [16]. The author metrics Netscan provides can also help increase mutual awareness of others’ presence and history of activity, which may facilitate the development of trust and identification of reliable authors and messages in newsgroups.

During 2003, over 125,000 unique users visited the public Netscan website and more than 168,000 visited in 2004. While the usage of Netscan suggests that it is providing value to visitors, we wanted to study how visitors make use of the social accounting data and visualizations, in order to evaluate and improve the Netscan system in particular, and community support systems in general. Therefore the first research question guiding this study is:

Q1) How is Netscan used? Are there differential information seeking behaviors among Netscan users?

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

CHI 2005, April 2–7, 2005, Portland, Oregon, USA.
Copyright 2005 ACM 1-58113-998-5/05/0004...\$5.00.

Social science theories suggest that there are different social roles within communities, and roles are associated with different goals, expectations, and behaviors [7, 9]. Netscan was developed to support a variety of different user groups, from new users to leaders, hosts and researchers interested in newsgroups communities. Therefore it is essential to examine whether and how users engaged in different social roles and with different needs make use of the system. Hence, our two additional research questions:

Q2) Can Netscan data provide empirical evidence of the different social roles in Usenet?

Q3) Do social roles in Usenet associate with different information seeking behavior on Netscan? Do different Usenet roles value different information?

Answers to these questions will lead to a better understanding of the needs of Usenet participants who we wish to support, enable us to evaluate the current state of Netscan, and also make design recommendations for similar tools supporting online community.

To explore these questions, we studied Netscan\Tech, a sub-site of Netscan that provides daily-updated metrics for approximately 15,000 technical newsgroups (13% of the newsgroups analyzed by the main “All Usenet” Netscan system). We collected three sources of data: usage logs of the Netscan\Tech website over two years, survey responses from the website visitors, and Usenet participation metrics for some of the survey respondents.

We chose to focus on the Netscan\Tech sub-site because of its daily update frequency, in contrast to the less frequent periodic updates on the main Netscan site, which made Netscan\Tech more likely to receive repeated use. The deployment of Netscan\Tech within our organization also allowed us to capture detailed usage information that we could easily link to data collected in user surveys. We felt the advantages of daily updates and improved user tracking made up for the drawback of only studying users internal to our organization. To test for generalizability, we have also done an initial exploration of the usage of the public Netscan site that suggests our findings for Netscan\Tech also apply to public users.

Our results showed that there are different patterns of searching and browsing behaviors on the Netscan\Tech website. Frequent users focused on author information much more often than information about newsgroups or threads. In comparison, users that came less frequently focused more on information about newsgroups and viewing individual messages. Our findings also contribute to the online community literature by empirically validating different roles in newsgroups using behavior metrics. We found that those who contributed heavily and frequently in Usenet newsgroups, regardless of their usage of Netscan\Tech, also focused on social metrics related to authors.

In the following, we first discuss related work and then describe the Netscan\Tech system. Next, we describe the data and procedure we used to analyze usage of Netscan\Tech. We then present our results organized by our research questions. We close with a discussion of our findings and directions for future research.

RELATED WORK

Existing research has built creative representations of the online social context in support of various types of interactions including chats [4, 5, 18], message boards [3], Usenet newsgroups [3, 16], and other large-scale online conversations [14, 19]. Donath and her colleagues developed tools using semantic visualization to bring the meaning and relevance of data to the audience [3]. By highlighting selective social and semantic structures of online interaction, visualization tools such as Chat Circle [4, 18] and Loom [4] can help users grasp the social context in the community, and facilitate navigation and participation.

Other researchers have focused on the challenge of large-scale conversations, and employed layered representations, context focusing visualizations, and automatic search engines to facilitate understanding and navigation of the large volumes of messages in such conversation spaces [14, 19]. On the other hand, the Babble system, intended for small to medium-sized corporate groups, emphasized the presence of subtle social cues in the visualization to support mutual awareness and accountability online [5]. This design principal of making visible others' presence and activity to shape collective activity is shared by the development of Netscan [16, 17].

Netscan is distinct from other community support tools in that it provides inter-related information on three social dimensions of Usenet: newsgroups, authors and message threads. Also, previous development and user studies of similar visualization tools, including those for Netscan [6, 17], tended to focus on the usefulness of a particular feature or interface design element for either threads or authors, which neglects the interrelation between them [3, 14]. Our study aims to fill this gap by examining the usage and evaluation of a variety of features on Netscan and exploring the relative importance of different social information as perceived by the users.

Online Social Roles

The prevalence of unequal participation in online communities has stimulated studies of different types of online participants [7, 12, 20]. Several studies have shown that there is commonly a small core group of people who account for most of the activity in online communities, while a large portion of the community members are peripheral [1, 8, 15]. Most research focuses on the distinction between lurkers, who remain inactive for a long period of time, and regulars who participate in discussions. Although the definitions of lurkers vary, previous studies

have consistently shown the percentage of silent participants within online community as about 90% [12, 16].

Social and computer scientists have made attempts at defining a broader range of roles in Usenet. Whittaker et al. [20], for example, made distinctions between repeat posters (persons who posted to a newsgroup more than one time) and single posters (those who posted to a newsgroup only once). Several researchers have conducted qualitative studies of social roles in Usenet including Golder [7] and Kim [9]. Golder explored the variation of participation behaviors in Usenet, and identified and compared roles such as: Newbie, Celebrity, Elder, Lurker, Flamer, Troll, and Ranter. Golder suggested that these different roles are associated with different goals and needs, and thus different participation behaviors. Particularly, celebrities have strong influence over newsgroup maintenance and change.

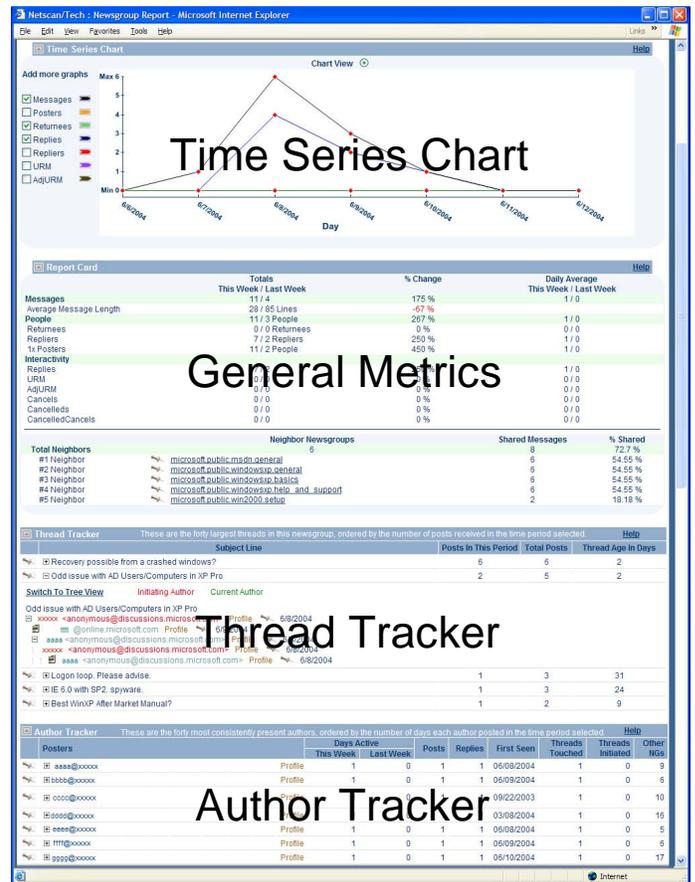
From a different perspective, targeting community managers interested in “building” communities, Kim defined roles in a membership lifecycle involving five successive stages: Visitors, Novices, Regulars, Leaders and Elders. She suggested a formula for creating the right mix of people (i.e. social roles) in online and offline communities. These research studies extended the simple distinction between lurkers and regulars and offered a framework for developing better community tools. However, no empirical work has systematically validated these roles, examined the difference among them in terms of participation behaviors within Usenet, or developed tools supporting their differential needs.

NETSCAN/TECH

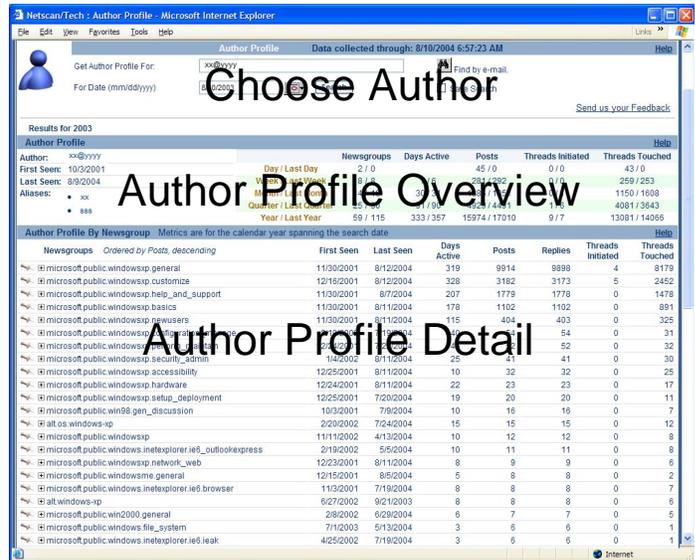
Before presenting our study, we first describe the main features of Netscan/Tech. Netscan/Tech is a sub-site of the Netscan system that provides metrics on a daily basis for a subset of Usenet focused on newsgroups on technical topics. In particular, the newsgroups that Netscan/Tech examines are all those named with the patterns microsoft.*, and *comp*, along with a selection of other newsgroups that crosspost strongly into these core sets. Users can view detailed reports on the activity in these newsgroups, the authors that participate in them, and the conversation threads that emerge from their activity. We briefly describe the main feature pages of the Netscan/Tech website: Newsgroup Search, Newsgroup Report Card, Author Profile, and My Usenet Portal.

Newsgroup Search

Newsgroup Search is the home page of Netscan/Tech. From this page, users can search for technical newsgroups of interest to them using key word search against the names of newsgroups. The search results provide summary statistics about different newsgroups; selecting a search result connects to a Newsgroup Report Card. By clicking the “push-pin” icon next to a newsgroup’s name in the search



Newsgroup Report Card



Author Profile

Figure 1. The Newsgroup Report Card and Author Profile features. The Newsgroup Report Card shows metrics for the windows.public.windowsxp newsgroup for the week of 6/12/2004. The Author Profile shows metrics about the contribution of a single author across all newsgroups in Netscan/Tech as well as detailed metrics for each newsgroup.

results, users can add the newsgroup to their personal My Usenet Portal page.

Newsgroup Report Card

The Newsgroup Report Card, shown in Figure 1, contains detailed information about a specific newsgroup for a user selected time period. It is meant to provide an overview of activity within the newsgroup by showing changes over time, and by highlighting the top participants and conversations within the newsgroup. The report card includes a *Time Series Chart*, *General Metrics*, *Thread Tracker* and *Author Tracker* sections.

The options on the Time Series Chart allow the user to plot the number of messages, authors, returnees, replies, repliers and unreplyed-to messages over a selected time period. The General Metrics section shows aggregated statistics about messages and people, and also lists up to five neighbor newsgroups that share the most messages (i.e. crossposts) with the selected newsgroup.

The Thread Tracker component of the Report Card lists up to forty of the largest threads in the selected newsgroup in terms of total messages contributed in the selected time period. The threads represented in the Thread Tracker can be rendered via a graphical *Tree View* or text outline *Thread View*.

The Author Tracker component on the Newsgroup Report Card shows metrics for up to forty authors active on the most number of different days in the selected time period in the newsgroup. Users can see the number of days an author was active, the number of times they contributed a message, the number of the messages that were replies, the first time the author ever contributed to the selected newsgroup and other metrics to help assess the pattern of activity the author has implicitly created. Clicking on the name of an author takes the user to the author's Author Profile page.

Author Profile

The Author Profile feature, also shown in Figure 1, provides both *Overview* and *Detail* information about an individual author. The Author Profile Overview shows information about an author's activity across all technical newsgroups in our dataset including: the number of newsgroups the author participated in, the number of days on which the author contributed at least one message, the number of posts the author contributed, the threads that were initiated by the author and the number of threads to which the author contributed. The Author Profile Detail section provides similar metrics for each newsgroup the author participated in, along with a list of up to twenty threads the author contributed to in that newsgroup. The Author Profile can also aggregate information from up to four different display names¹ the author may have used while contributing to Usenet.

My Usenet Portal

Netscan\Tech allows users to maintain a page called My Usenet Portal that displays a selected collection of authors, messages, threads and newsgroups. The portal displays metrics for each newsgroup, author, and thread specified by the user.

ANALYZING USAGE OF NETSCAN\TECH

We studied usage of Netscan\Tech from late January 2002 to June 2004 when it was deployed only within our organization. Netscan\Tech then became available to the public in June 2004. In this section we describe the methods we used to gather data from three sources: the Netscan\Tech usage logs, user surveys, and quantitative measures of users' Usenet posting behavior.

Usage Log

We instrumented the Netscan\Tech site to log the actions that our users performed. The usage data described in this paper covers almost 2 ½ years from January 29, 2002, when the site was deployed, to June 9, 2004, when a snapshot of the usage log was collected for this study. Metrics on the Netscan\Tech site were updated weekly until December 2003 and daily from then on. The log tracks the feature pages on the website visited by a user, such as the Newsgroup Report Card or Author Profile page, along with specific actions performed on a page, such as viewing the text of a Usenet message or listing the threads involving a particular author. Because all users were authenticated to our corporate network we also collected user IDs which allowed us to contact our internal users to request that they complete a survey.

User Surveys

To complement the usage log information, we surveyed 715 active users and 6460 past users of Netscan\Tech in mid-June 2004. We surveyed active users about why they visited Netscan\Tech, their satisfaction with the site, preferred features (current and potential) and their participation in Usenet newsgroups. We received 64 responses to the active user survey for a response rate of 10.6%, ignoring the 110 people who were unreachable due to vacation or change in email address.

We also contacted past users of Netscan\Tech with a separate survey that focused on why they stopped using Netscan\Tech, preferences among potential features and their participation in Usenet newsgroups. Of the past users contacted, we received a total of 215 responses for a response rate of 4.4%, ignoring the 1,532 people that were unreachable. Of these 215 responses, 26 respondents said they were still using Netscan\Tech so we asked them to take the user survey instead. Another 28 respondents told us they were unfamiliar with Netscan\Tech. This left 161 valid past user responses for further analysis.

¹ A display name looks like: "Name" <email@y.com>.

Behavior in Usenet

To understand how survey respondent's participation in Usenet newsgroups correlated with their use of Netscan\Tech, we asked respondents on both surveys to share the email address(es) they used when participating in Usenet. For the 127 respondents who provided email address(es), we used Netscan\Tech to gather social accounting meta-data about their participation in Usenet including the number of newsgroups they participated in, their days active in Usenet, the number of total messages sent, the number of replies and the number of threads they started. We combined this data with the survey data to validate different roles in Usenet, and with the Netscan\Tech usage log to examine whether different patterns of participation in Usenet correlated with different patterns of usage of the Netscan\Tech site.

Both the Netscan\Tech usage log and Usenet behavioral log data are highly right-skewed. This indicates a small portion of highly active individuals and a high proportion of people with low activity. Therefore, we conducted both parametric and non-parametric tests for the analyses, and found the results similar. As a result, we only reported parametric testing results in this paper.

RESULTS

We now discuss the results of our analysis organized by our three research questions.

Q1: Usage of Netscan\Tech

Our first research question explores how Netscan\Tech is used and valued by its users. In particular, we investigated whether there are different types of users of Netscan\Tech and differential information seeking behavior.

Are there different types of users of Netscan\Tech?

The Netscan\Tech users we studied formed distinct groups defined by high and low usage associated with distinct patterns of behavior. During our study, there were 7,311 unique visitors to the internal Netscan\Tech website. As Table 1 shows, we classified visitors to the Netscan\Tech site into three main groups:

- **Active users** (715, 10%): Users who visited the Netscan\Tech website at least once in the six months prior to the data collection date, and three or more days overall.
- **Past users** (6,460, 88%): Users who visited fewer than 3 times or had not visited the website in the six months prior to our study. Of the past users, most visited 1 or 2 days (86.5%, 5,589), while 13.5% (871) had visited more than three times, but not within the 6 months prior to June 2004.
- **New users** (136, 2%): Users that first visited the website in the month prior to the data collection date. New users were excluded from our analysis since their usage patterns may not have stabilized.

Website User Type (N = 7311)	Avg. Days Visited (SD, Median)	Avg. Visits per Month (SD, Median)	Avg. Actions per Visit Day (SD, Median)
Heavy - Active (231)	37.3 (36.9, 22)	2.6 (0.09, 1.6)	11 (8.2, 9)
Light - Active (484)	5.1 (2.1, 4)	1.2 (0.08, 0.4)	8.7 (7.8, 7)
Active - Total (715)	15.5 (25.9, 7)	1.8 (0.09, 0.6)	9.4 (8.0, 7.7)
Past (6460)	1.9 (2.7, 1)		8 (13.8, 4)
New (136)	1.3 (0.8, 1)		6.3 (10.5, 3)

Table 1. Visitors to Netscan\Tech Website. Of the active users, heavy users visited more than 10 times, while light users visited between 3 and 10 times. Past users visited less than 3 times or have not visited in the six months prior to June 2004. New users came for the first time within the last month. Average actions per visit day shows the average amount of user interaction with the website on each visit.

Comparing active and past users we found, as shown in Table 1, that active users visited the website significantly more often ($t(716) = 14.06$, $p < .001$) and performed significantly more actions per visit than past users ($t(1242) = 4.14$, $p < .001$).

Due to our interest in how active users were using Netscan\Tech, we further classified the active users into:

- **Heavy Active users** (231, 32%): Active users that visited more than 10 days.
- **Light Active users** (484, 68%): Active users that visited between 3-10 days.

We found that heavy users visited the website significantly more often than light users ($t(231) = 13.24$, $p < .001$) and had significantly more average actions per visit date than light users ($t(713)=3.64$, $p < .001$).

Does usage information vary by user type?

We explored the visiting pattern of the different types of users by ranking the actions they performed on the Netscan\Tech website based on the frequency of those actions. Table 2 lists the top 4 actions performed by each user type. For all users of Netscan\Tech, not surprisingly, visiting the home page for the site (the Newsgroup Search page) was the most common action. However, the rankings after the home page show that the Netscan\Tech user types differ in terms of their rate of usage of different features. Heavy users' second most visited feature was the Author Profile, followed by the Newsgroup Report Card, and viewing the text of messages. Light users, on the other hand, visited features related to newsgroups and messages more often than the Author Profile. Actions by past users,

Heavy User		Light User		Past User	
Top Actions	Mean (Med.)	Top Actions	Mean (Med.)	Top Actions	Mean (Med.)
NG Search	128.0 (55)	NG Search	13.6 (9)	NG Search	5.3 (2)
Author Profile	84.5 (21)	NG Report Card	7.3 (4)	View Messages	4.8 (0)
NG Report Card	67.0 (31)	View Messages	5.5 (0)	NG Report Card	2.7 (1)
View Messages	59.1 (9)	Author Profile	4.6 (1)	Expand Thread	1.6 (0)

Table 2. Actions (e.g. visiting the Newsgroup Search page) preformed most frequently by each Netscan\Tech user type.

when they did visit, showed an even stronger focus on information about messages and newsgroups.

Do feature preferences vary by user type?

On the active user survey we asked respondents about their experience with Netscan\Tech and to rate the usefulness of existing features. Overall, the sixty-four respondents to the user survey felt positively about Netscan\Tech with a median response of “Agree” to questions about liking the site, their satisfaction with the site, and whether the site was easy to use. Ninety-five percent of the respondents (61) agreed or strongly agreed that they planned to visit the site in the future. Table 3 shows the average rating of features based on their usefulness broken out by heavy and light active users, and ranked by the preference of the heavy users. Features were rated on a 4 point scale and only respondents familiar with a feature were asked to rate it.

As Table 3 shows, survey respondents that were heavy active users gave their top ratings to the features related to authors and the Newsgroup Report Card. The Author Tracker component on the Newsgroup Report Card was the highest rated feature overall, followed by the Author Profile Detail feature and the Author Profile Overview. Survey respondents that were light active users had a different order preference for the features. Most strikingly, light active users gave Newsgroup Search their highest rating, compared to heavy users who ranked it fifth. While one author feature, Author Profile Detail, was rated second highest by light active users, their two other highly rated features were both related to viewing threads and messages in threads.

While heavy active users generally rated features as more valuable than light active users, their ratings of the three author related features were all significantly higher than the ratings by light users at the $p < .01$ level based on t-tests². Heavy users also rated Newsgroup Report Card and Time

² Author Tracker: $t(49)=4.6$; Author Profile Detail: $t(48)=2.96$; Author Profile Overview: $t(49) = 3.26$;

Feature	Active User Survey Respondents (64)			
	Heavy Users (30)		Light Users (34)	
	N	Mean (SD, Med.)	N	Mean (SD, Med.)
Author Tracker**	27	3.6 (0.6, 4)	24	2.8 (0.6, 3)
Author Profile Detail**	28	3.5 (0.6, 4)	22	2.9 (0.7, 3)
Author Profile Overview **	28	3.4 (0.7, 4)	23	2.8 (0.6, 3)
Newsgroup Report Card*	23	3.4 (0.7, 4)	24	2.9 (0.8, 3)
Newsgroup Search	26	3.3 (0.7, 3)	24	3 (0.8, 3)
Time Series Chart*	25	3 (0.8, 3)	25	2.6 (0.8, 2)
Thread Tracker	20	2.8 (0.8, 3)	21	2.8 (0.8, 3)
Viewing Messages	23	2.7 (0.9, 3)	24	2.8 (0.8, 3)
Thread Tracker: Thread View	19	2.6 (0.9, 3)	24	2.9 (0.8, 3)
My Usenet Portal	21	2.6 (0.7, 3)	11	2.2 (0.6, 2)
Thread Tracker: Tree View	19	2.4 (1, 2)	22	2.3 (0.7, 2)

Table 3. Ratings of Netscan\Tech features by heavy and light users. Only respondents familiar with a feature rated it (hence different N’s). Ratings were on the scale 1=not useful, 2=somewhat useful, 3=Very useful, and 4=can’t live without. **Ratings of feature by heavy and light users were significantly different with $p < .01$. *Ratings of feature were significantly different with $p < .05$.

Series Chart significantly higher at the $p < .05$ level ($t(45) = 2.17$ and $t(48) = 2.03$ respectively). For other features, there is no significant different between the ratings given by heavy active users and light users.

The survey results revealed that heavy active users valued author-related features significantly more than light active users, while light active users mostly valued newsgroups and message related features. These findings reinforced the results from previous usage log analysis.

Why do past users no longer visit Netscan\Tech?

As part of understanding how users value Netscan\Tech and the social accounting meta-data it provides, we wanted to understand why past users no longer found Netscan\Tech useful. On the past user survey we asked respondents to rate several possible reasons for why they stopped using Netscan\Tech on a 5 point Likert scale from “Strongly Disagree” to “Strongly Agree”. While all options received a median response of “Neutral,” the statement “I no longer needed Netscan\Tech for my work” received the highest percentage of “Agree” and “Strongly Agree” responses (48.4%). Several respondents also commented that changes in job responsibilities led them to stop using Netscan\Tech.

Examples of such comments included: “Changed teams, stopped being involved in newsgroups,” and “I changed roles and have not gotten back in to community participation in my new role.”

After work related reasons, the next most popular reason to stop using the site was “The Netscan\Tech site was no longer useful for me” (35.5% “Agree” and “Strongly Agree” responses), followed by “The Netscan\Tech website was difficult to use” and “The Netscan\Tech website did not provide the features that I needed” (both 25%). We were somewhat surprised that only 19.6% of the respondents “Agreed” or “Strongly Agreed” that they stopped using Netscan\Tech because it did not allow them to write messages to newsgroups. We thought this might be a much more popular reason to stop using Netscan\Tech.

Q2: Social Roles In Usenet

Our second research question asked whether there is empirical evidence of different social roles in Usenet based on behavioral metrics provided by Netscan\Tech. To understand how survey respondent’s participation in Usenet newsgroups correlated with their use of Netscan\Tech, we included questions about Usenet participation on both active user and past user surveys. The questions asked the users to describe their roles in and relationships to the newsgroups and many provided their Usenet author names which allowed us to contrast self-report and observed behavior.

Of the 225 respondents to the active user and past user survey, 127 reported they participated in Usenet (Table 4). Of the active user respondents, 69% participated in Usenet, compared to 52% of the past users. When asked how many newsgroups they sent messages to in the last month the median response for both active and past users was 1-5 newsgroups. When asked about the number of newsgroups they read in the last month, the median response for active users was 6-10 newsgroups, while for past users it was 1-5 newsgroups.

Based on the survey responses, we classified the respondents that participated in Usenet into five groups: Key Contributors, Low Volume Repliers, Readers, Questioners and Disengaged Observers.

- **Key contributors:** Respondents who “Agreed” or “Strongly Agreed” that they considered themselves key contributors to at least one newsgroup.
- **Low volume repliers:** Respondents who “Agreed” or “Strongly Agreed” that they provided answers to others, but did not consider themselves to be key contributors.
- **Questioners:** Respondents who “Agreed” or “Strongly Agreed” that they asked questions in newsgroups, but did not consider themselves repliers or key contributors.
- **Readers:** Respondents who “Agreed” or “Strongly

Survey Participants in Usenet by Type	Active Users (N=44)	Past Users (N=83)	Total (N=127)
Key Contributor	16 (36%)	17 (21%)	33 (26%)
Low Volume Replier	16 (36%)	38 (46%)	54 (43%)
Questioner	1 (2%)	1 (1%)	2 (2%)
Reader	10 (23%)	25 (30%)	35 (28%)
Disengaged Observer	1 (2%)	2 (2%)	3 (2%)

Table 4. Active and Past user survey respondents that participate in Usenet grouped by participant types.

Agreed” that they read messages in newsgroups, but did not ask questions or provide answers in newsgroups.

- **Disengaged Observers:** Respondents who reported they participate in Usenet, but answered “Neutral”, “Disagree” or “Strongly Disagree” to all the questions that asked about whether they read or post messages.

Table 4 shows the breakdown of survey respondents into Usenet Participant types. Low volume repliers (54) were the most common type of self described Usenet participants among our respondents and most of them were past users of Netscan\Tech. The thirty-five readers came from both past and active users, while the thirty-three Key Contributors were more likely to be active users of Netscan\Tech.

In order to extend beyond self-reported categories, we tried to validate these different self-reported roles with quantitative data. Of the 127 survey respondents that participated in Usenet, 71 (56%) provided a Usenet addresses they used, enabling us to link the survey data with the Usenet behavior data collected by Netscan\Tech for that identity.

Data collected about respondent’s behavior in Usenet was consistent with their self-reported survey data, giving us confidence in our categorization of users. Compared to Low Volume Repliers, the Key Contributors were present in newsgroups more often ($t(25) = 3.25, p < .005$), and sent more messages ($t(23) = 2.14, p < .05$) of which more were replies ($t(23) = 2.28, p < .05$).

The survey questions we used in categorizing participants into groups also correlated well with the metrics generated by Netscan\Tech. Responses to “I consider myself one of the key contributors in a newsgroup(s)”, had a positive correlation to the number of days active in newsgroups ($r(67) = .497, p < .0001$), the number of posts ($r(67) = .415, p < .0001$), the number of replies ($r(67) = .420, p < .0001$) and the number of threads touched ($r(67) = .398, p = .001$). Responses to “I often provide answers to other people’s questions” were also positively correlated to days active ($r(68) = .370, p < .002$). Lastly responses to “I often ask

questions” were positively correlated with the number of threads started ($r(68) = .418, p < .0001$).

Q3: Usenet Roles and Use of Netscan\Tech

Our final research question focused on whether different social roles in Usenet had different information seeking behavior on Netscan\Tech. To explore if different Usenet roles seek and value different information, we examined the relationship between a person’s self-reported role in Usenet and their use of Netscan\Tech. For the three Usenet participant types where we had a large enough number of survey responses, Key Contributor, Low Volume Replier and Reader, we examined how these groups used the Netscan\Tech site focusing on the features visited most on average.

Key Contributors in Usenet most commonly visited the Newsgroup Search Home Page and Author Profile. This suggests that people that were active posters in Usenet visit the Author Profile page frequently. For Key Contributors that were heavy users of Netscan, viewing threads related to an author was also among the top three most visited pages. In contrast, Key Contributors that were light or past users of Netscan\Tech instead included the Newsgroup Report Card among their top three.

Among the survey respondents self-identified as Low Volume Repliers in Usenet, only those who were heavy users of Netscan\Tech included Author Profile as one of their three most visited pages. Low Volume Repliers in Usenet that were light or past users of Netscan\Tech visited the Newsgroup Search page, viewed messages and visited the Newsgroup Report Card most on average. This suggests that people who posted less in Usenet and visited Netscan\Tech less frequently preferred viewing messages and information related to newsgroups.

For survey respondents we identified as Readers in Usenet, similar patterns emerged. Only heavy users of Netscan\Tech included the Author Profile page in their top three most commonly visited pages. Readers that were light or past users focused more on newsgroup and message related features.

In summary, we found that self-identified Key Contributors focused on social metrics related to authors, regardless of their extent of usage of Netscan\Tech. This suggests that more frequent participants in Usenet are also very interested in the other authors in their communities. Low volume repliers and readers that were not heavy users of Netscan\Tech, on the other hand, preferred information on newsgroups and messages.

DISCUSSION

We now discuss the differential usage of Netscan\Tech and the steps we have taken toward developing a Usenet participant typology.

Differential Usage of Netscan\Tech

Netscan was developed with a focus on providing social metrics about newsgroups and the conversations within them. The newsgroup is the fundamental organizational element in Usenet and as a result, our initial belief was that social metrics related to newsgroups would be the most popular. For example, metrics that showed the number of messages and number of participants in a newsgroup. Hence, several features were deployed to convey this information and we gave newsgroup related metrics the most prominence in our system.

However, our usage log data and survey responses showed that the heavy users of Netscan\Tech focused on features providing metrics about authors, while light and past users focused more on newsgroups and messages. When exploring the relationship between Usenet participant types and use of Netscan\Tech we also found that survey respondents identified as Key Contributors in Usenet, regardless of their usage of Netscan\Tech (heavy, light or past), also appeared to focus on information about authors. In contrast, survey respondents identified as Low Volume Repliers and Readers in Usenet, that were not heavy users of Netscan\Tech, visited features related to newsgroups and messages more often than those related to authors.

Although perhaps surprising at first glance, the differential preferences for author-related information compared with newsgroup or message related information by regular and casual users of Netscan\Tech as well as different Usenet roles may be explained by how much users care about the social context within Usenet. Authors in newsgroups are primary resources of knowledge, material, social and emotional support [1], and they enable the social structure and dynamics within newsgroups [9]. Thus, people who care about the long term, ongoing development of newsgroups may look at the people who participate in those newsgroups, either to identify useful or interesting messages to follow [6], or to gain a better understanding of the authors as reflected by their behaviors. On the other hand, casual users of Netscan\Tech may be satisfied by the macro-level social metrics about newsgroups.

A further research question to ask would be: whose information are these users looking at: themselves or other members in the community? The availability of social accounting information creates a reputation mechanism for individuals, conversation and newsgroups, which may facilitate selection of content [13]. If Netscan\Tech mainly serves as an implicit reputation system for authors, visitors of the website would focus on identifying authors who they would or would not like to interact with, and thus would mainly look for information on others. On the other hand, if the social accounting metrics reflect social standing or identity of authors in newsgroups [2, 9], users of Netscan\Tech may be more interested in looking at information about themselves for social identification purposes. In future research we plan to collect more detailed information about the authors being looked at to

explore which theory better explains the extensive use of author-related features, and to guide the future design of visualization tools for author information.

Validation in Netscan

During our study, the user population of Netscan\Tech was internal to our organization. One possible explanation for the interest in author metrics might be tied to an organizational interest in following particular newsgroup contributors. Might behavior be different for outside users without this particular interest? To address this concern, we performed an analysis of usage logs we collected for the public Netscan website to see if there was a similar finding in an external, potentially non-work setting.

The Netscan usage log we examined contained data from June 2001 to June 2004, and had visits from 219,356 unique IP addresses³. We identified 8,242 active users (3.8%) and 200,646 past users (91.5%) in the Netscan usage logs. Among the active Netscan users, we found a similar focus on metrics about authors as our internal Netscan\Tech users. Active Netscan users visited the Author Profile page the most on average, even more often than the newsgroup search home page. Therefore, the Netscan usage log data reinforced our findings in Netscan\Tech: people who often used Netscan focused on author related metrics.

We also examined the usage of public Netscan by past users and saw they favored the newsgroup report card and thread expanding features more than the heavy active users. These data seem to confirm that participants that were past users of Netscan were more interested in newsgroups and messages than author information, similar to our findings for Netscan\Tech.

Implications

Our analysis of usage of Netscan\Tech and Netscan suggests that designers and developers of online communities may find value in providing social accounting metrics to their participants. We encourage community sites to incorporate metrics based on behavior into their sites to help participants understand the activity at the site.

More specifically, our findings, in particular that frequent visitors to Netscan\Tech and Key Contributors in Usenet favor author related information, suggest that interfaces to online communities, which typically focus heavily on newsgroups and threads, may want to shift focus to providing additional information about authors. For example, allowing users to see metrics about particular authors and making it easy to find all contributions by an author. This may be particularly important for online

communities that have a core set of frequent contributors that would be interested in other contributors.

To take a more concrete example, the Slashdot news site (Slashdot.org) has a population of frequent users that it relies heavily on for moderation. Slashdot also tracks user behavior to award moderation status [11]. However, the site reflects relatively little about a users' behavior to others visiting the site. It would potentially be valuable to augment the current information provided about a user with additional social meta-data analogous to what Netscan provides, such as days active, average score for comments, and average replies to a users' comments. This could help both frequent and infrequent visitors to the site learn more about a user's history of participation at the site and perhaps serve as an implicit reputation system.

However, in general as we did see differential usage of the metrics, it is important for designers and developers to identify which types of users visit their sites. For online community sites primarily devoted to populations that come infrequently, first developing metrics related to messages or newsgroups may be a higher priority than author metrics.

Steps Toward Building a Usenet Participant Typology

By utilizing both survey data and other quantitative data about Usenet participation, we believe we have begun the process of identifying Usenet participant types that will be valuable for our research community to discuss, debate, and perhaps build on to develop a Usenet participant typology that could be used across online community studies. In this study, we identified three main participant types: Key Contributors, Low Volume Repliers, and Readers.

Clearly there are many interesting directions to proceed toward a more complete understanding of the variety of types of participants in Usenet and related conversational media. For one thing, additional validation across a wider range of users would be valuable as our survey respondents are from a single organization and grouped mainly into three types. There are several other types of Usenet participants that researchers have identified that we did not have data to explore including Flamers and Ranters [7]. We are in the early stages of a research project to explore clustering Usenet authors to identify different participant types, beyond those identified in this study.

We are also interested in understanding Usenet participants that do not often send messages. Among our survey respondents we had a considerable number of Readers, participants that primarily read, but do not contribute. One potential way to better understand the behavior of this group of Usenet participants is to examine message access event logs in future studies.

CONCLUDING REMARKS

We have presented a systematic analysis of the usage of Netscan\Tech, a community system that generates social

³ Individuals can not be identified in the usage logs for public Netscan, instead we log IP addresses, which may aggregate the use by several people or fragment the behavior of an individual into several "IP" identities.

accounting metrics for technical newsgroups in the public Usenet feed.

Our main contribution is the finding that frequent visitors to Netscan\Tech and users that contribute often in Usenet placed a strong emphasis on viewing metrics related to authors in favor of newsgroup or thread centric reports. We also observed that among our less frequent users, including those that participate in Usenet less often, there was a greater focus on information about newsgroups and messages than authors. Of course we recognize that our data comes primarily from the study of users within one organization, but we believe that our initial analysis of the public usage of Netscan lends support to the broader applicability of these findings.

We believe that our findings suggest that designers and developers of online communities may wish to make information about each user's history of participation within their communities available. However, since we observed different types of users have different preferences, it is important for designers and developers to identify which types of users visit their sites and to provide information appropriate to these user types.

In this research we also utilized survey data and other quantitative information about Usenet participation to identify several types of Usenet participants. We look forward to building on this in the future with a broader study that examines all authors in Usenet with the goal of developing a typology of Usenet participants that could be used across online community studies.

ACKNOWLEDGMENTS

We thank our reviewers for their time and helpful suggestions for improving our paper. A special thanks to our survey respondents and the members of the Microsoft Research Community Technologies Group, in particular, Paul Johns, Steven Mozel, Dany Rouhana and Danyel Fisher.

REFERENCES

- Butler, B. S. Membership Size, Communication Activity, and Sustainability: A Resource-Based Model of Online Social Structures. *Information Systems Research*, 12, 4 (2001), 346-362.
- Donath, J. Identity and Deception in the Virtual Community. *Communities in Cyberspace*. P. Kollock and M. Smith. London, Routledge. (1996)
- Donath, J. A Semantic Approach to Visualizing Online Conversations. *Communications of the ACM*, 45, 4 (2002), 45-49.
- Donath, J. Visualizing Conversation. *Journal of Computer-Mediated Communication*, 4, 4 (1999).
- Erickson, T., Halverson C., Kellogg, W. A., Laff, M., and Wolf., T. Social Translucence: Designing Social Infrastructures that make Collective Activity Visible. *Communications of the ACM*, 45, 4 (2002), 40-44.
- Fiore, A. T., Tiernan, S. L. and Smith, M. A. Observed Behavior and Perceived Value of Authors in Usenet Newsgroups: Bridging the Gap. In *Proc. CHI 2002*, ACM Press (2002), 323-330.
- Golder, S. A Typology of Social Roles in Usenet. A thesis submitted to the Department of Linguistics, Harvard University. (2003).
- Jones, Q., and Rafaeli, S., and Ravid G.. Information Overload and the Message Dynamics of Online Interaction Spaces: A Theoretical Model and Empirical Exploration. *Information Systems Research*, 15, 2 (2004), 194-210.
- Kim, A. J. Community Building on the Web: Secret Strategies for Successful Online Communities. Berkeley, CA, Peachpit Press. (2000).
- Kollock, P. and Smith M. Managing the virtual commons: Cooperation and conflict in computer communities. *Computer-Mediated Communication: Linguistic, Social, and Cross-Cultural Perspectives*. S. C. Herring. Amsterdam, John Benjamin: (1996), 109-128.
- Lampe, C. and Resnick P. Slash(dot) and Burn: Distributed Moderation in a Large Online Conversation Space. In *Proc. CHI 2004*, ACM Press (2004), 543-550.
- Nonnecke, B. and Preece J. Silent Participants: Getting to Know Lurkers Better. *From Usenet to Cowebs: Interacting with Social Information Spaces*. C. Leug and D. Fisher, Springer Verlag: Amsterdam, Holland. (2003), 110 - 132.
- Resnick, P., Zeckhause, R., Friedman, E., and Kuwabara, K. Reputation Systems. *Communications of the ACM* 43, 12 (2000), 45-48.
- Sack, W. Conversation Map: An interface for Very Large-scale Conversations. *Journal of Management Information Systems*, 17 3 (2000), 73-92.
- Schoberth, T., Preece, J., and Heinzl A., Online Communities: A Longitudinal Analysis of Communication Activities. In *Proc. HICSS 2003*, IEEE Press (2003).
- Smith, M. Tools for Navigating Large Social Cyberspaces. *Communications of the ACM*, 45 4 (2002), 51 - 55.
- Smith, M. A. and Fiore A. T. Visualization Components for Persistent Conversations. In *Proc. CHI 2001*, ACM Press (2001), 136-143.
- Viegas, F. B. and Donath J. S. Chat circles. In *Proc. CHI 1999*, ACM Press (1999), 9-16.
- Wattenberg, M. and Millen D. Conversation Thumbnails for Large-scale Discussions. *Extended Abstracts CHI 2003*, ACM Press (2003), 742-743.
- Whittaker, S., Terveen, L., Hill, W., and Cherny, L. The Dynamics of Mass Interaction. In *Proc. CSCW 1998*, ACM Press (1998), 257-264.