

# *Newsgroup Crowds and AuthorLines:* Visualizing the Activity of Individuals in Conversational Cyberspaces

**Fernanda B. Viégas**  
MIT Media Laboratory  
Cambridge, MA 02139 USA  
+1 617 253 2450  
fviegas@media.mit.edu

**Marc Smith**  
Microsoft Research  
Redmond, WA 98052 USA  
+1 425 706 6896  
masmith@microsoft.com

## **Abstract**

*We discuss the design, implementation and evaluation of two related visualizations of authors' activities in Usenet newsgroups. Current Usenet news browsers focus on messages and thread structures while disregarding valuable information about the authors of messages and the participants of the various discussions. Newsgroup Crowds graphically represents the population of authors in a particular newsgroup. Authors are displayed according to the number of messages they contribute to each thread and the number of different days they appear in the space, illustrating and contrasting the interaction patterns of participants within the newsgroup. AuthorLines visualizes a particular author's posting activity across all newsgroups over a period of one year. This visualization reveals temporal patterns of thread initiation and reply that can broadly characterize the roles authors play in Usenet. We report the results of a user study that explored the value of these interfaces for developing high-level awareness of the activity and population in these conversational spaces. We suggest that interfaces that convey information about the social histories of populations and individuals may support better selection and evaluation of newsgroup content.*

## **1. Introduction**

On an average day in the year 2001 more than 80,000 unique authors contributed 700,000 messages to Usenet, the global, distributed database of conversation [11]. The implications of such an incredible amount of activity are two-fold: on the one hand, Usenet is a great resource for conversations on virtually every topic and a place where one can find answers to almost any question. On the other hand, such vociferous places easily become noisy and hard to navigate. As it turns out, one of the major problems in Usenet is that there are enough poor-quality messages to render the quest for valuable content too difficult and cumbersome to

pursue. Moreover, users face the risk of deception when interacting with strangers online [2].

Attempts to solve this signal-to-noise ratio problem include reputation systems where users rank other users. Such systems have been widely put to use on popular Web sites such as eBay [12] and Amazon [13]. One of the main problems faced by such systems is their extensive misuse. An interesting exception to this problem is the ranking/moderating model used at Slashdot [14] where registered users take turns ranking posts, assuring a better turnout of high-quality posts.

An alternative to the ranking approach is feeding collected user data and site activity information back to the community [6]. This approach tends to situate online activity within a much richer social context that encourages awareness and accountability on the part of contributors [4] without requiring extra work (such as explicit ranking) from the rest of the community. Moreover, it has been suggested that metrics such as the longevity and frequency of participation, the number of newsgroups to which authors contribute messages, and the average size of the threads to which authors tend to contribute, strongly correlate to users' subjective assessment of the qualities and value of different authors in newsgroups [5]. That is to say that newsgroup users already rely on their personal knowledge of others' behavior in online environments to guide their choices of who to interact with and who to ignore. It seems desirable, therefore, that we leverage the evaluation activities performed by users and design tools that make these metrics more explicit and accessible.

Most news browsing interfaces, however, display minimal, if any, information about the authors of messages. When reading a post, users get no sense of the author's history; how active they are in the particular newsgroup, how long they have contributed to the community, in what other conversations they have engaged in the past, etc. Instead, current systems for newsgroup browsing focus on the message unit and the message structure of conversational threads;

thereby forcing users to pay attention to the organization of conversations instead of their social significance.

Our approach is to provide tools to support social awareness in online spaces by combining metrics of authors' activity with data visualization that feed these patterns back to community. We contend that large social cyberspaces such as Usenet are already so packed with information that social metrics need to be presented in ways that lower, rather than raise the cognitive load of users navigating these spaces. By utilizing visualization techniques that highlight the patterns of authors' behavior in newsgroups, we hope to harness the enormous power and subtlety of the human visual system as a pattern seeker [9]. As we discuss later in the paper, the two visualizations presented here turn out to be rather effective and quick ways of giving users a socially richer understanding of their online peers at a glance.

## 2. Related Work

There has been previous work in the fields of HCI and information visualization that explores the role that authors play in shaping conversational spaces. In a sense, the closest piece to our approach is PeopleGarden [10], which visualizes message boards in terms of their authors' activities. Each flower in PeopleGarden represents a user in the conversational space and its petals represent his/her postings. PeopleGarden also shows the amount of replies to a user's post by displaying pistil-like circles on top of a petal to denote responses. Even though PeopleGarden's focus was not Usenet newsgroups, the fact that it strived to reify a conversational space in terms of its people makes it conceptually close to the work presented here.

The original Loom project [3][1] focused on visualizing social patterns within Usenet newsgroups. It highlighted saliencies such as rowdy, vociferous users as well as the number of participants in different threads over time. It also visualized the difference between initiated posts and replies. Although its focus was not on the authors per se, Loom managed to uncover interesting author dynamics found in newsgroups – for instance the marked difference between the average number of participants per thread in technical versus social newsgroups.

Conversation Map [7] also looks at Usenet newsgroups and touches on the issue of the people present in the conversational space. It computes a set of social networks detailing who is talking to whom and who is citing whom in the newsgroup. The other main feature in Conversation Map is its visualization of the centrality degree of users in the newsgroup where the

social network of each newsgroup allows us to understand which users are more central than others to that group's discussions. Here as in Loom, remarkable patterns emerge that are related to people's interactions in the conversational space.

In retrospect, it becomes clear that even the projects that have the most in common with Newsgroup Crowds and AuthorLines do not, in themselves, set out to explore conversational spaces in order to better understand the people; instead, they try to better understand the spaces through the people found there. In a sense, authors are used as a means to an end, not as an end in themselves. In PeopleGarden, for instance, we get an understanding of the behavior of message board users but the system doesn't allow us to learn more about each of the authors it visualizes; we don't know who each one of the authors on the screen is (there are no names/ email addresses displayed). Both Loom and Conversation Map focus more on the dynamics of the newsgroup in general rather than on the dynamics of any specific person within a newsgroup.

## 3. Data Source

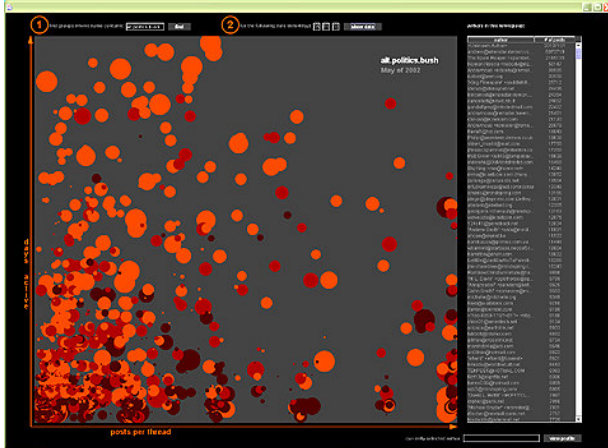
The Netscan [8] project data mines an on-going collection of Usenet newsgroups and creates aggregated meta-data for supporting sociological studies of these spaces. The system generates and displays detailed metrics about the activity of each Usenet newsgroup, author, and conversation thread. Profiles include the list of newsgroups to which each author has contributed, a list of the threads in which they have participated in each newsgroup, how much he/she has contributed to each thread, and whether he/she initiated any of the threads.

Newsgroup Crowds and AuthorLines build on the metrics created by the Netscan system.

One of the limitations in Netscan's user interface is that all of the data are displayed as endless tables of numbers. This format makes it hard for users to quickly grasp overall patterns and outlying values in the data. Our visualizations make these metrics more accessible to end users who, unlike researchers, might not be interested in painstakingly evaluating the tabular presentation of social metrics but who can profit extensively from expressive presentations of this information when browsing newsgroups.

## 4. Newsgroup Crowds

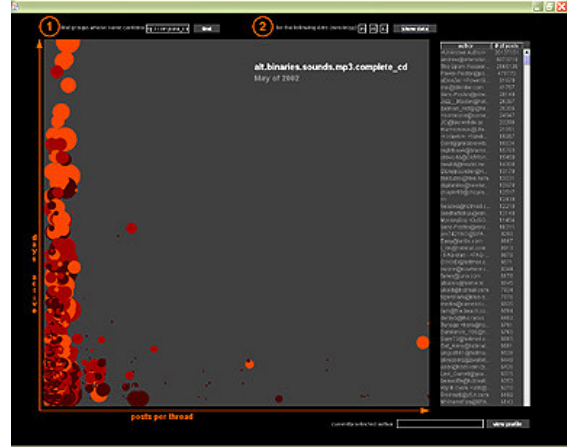
Newsgroup Crowds [Fig. 1] is a graphical interface that shows the activity of participants in a given newsgroup over a specific period of time. The visualization is a scatter plot of all authors who were active (i.e. posted messages) in the chosen newsgroup



**Figure 1.** Newsgroup Crowd visualization of *alt.politics.bush*; authors are distributed all over the scatter plot attesting to the diversity of contributors and the conversational character of the newsgroup.

during the month being visualized. Each author is represented by a circle whose placement is determined by two axes: number of days an author has been active during the chosen month – vertical axis – and the author’s average number of posts per thread in the newsgroup – horizontal axis. Our aim is to convey, at a glance, how densely inhabited, active and conversational a given newsgroup is. The visualization also makes other patterns explicit such as which authors are consistent contributors, and the color of circles represents how recently authors have been active in the newsgroup and their overall posting activity in Usenet as a whole. Newsgroup Crowds contains a table where the email addresses of all authors are displayed. The table displays, next to each address, the number of posts each author has contributed to Usenet newsgroups overall. Users can click on an address and the corresponding author circle is highlighted in the visualization panel. Users may also click directly on a circle in the visualization panel to see it highlighted. Whenever an author circle is highlighted, a small, semi-transparent information window is displayed next to the chosen circle containing more detailed information about the author [Fig. 5]: (a) author’s email address, (b) author’s number of posts during the chosen month in the newsgroup being visualized, (c) author’s total number of posts ever in all of Usenet, (d) first day this author was seen in this newsgroup, (e) last day this author was seen in this newsgroup, (f) author’s “top five” newsgroups (newsgroups to which the author posts the most) [Fig. 5].

This information window transforms the author’s generic circle into an individualized marker. The “top five” newsgroups, for instance, help us understand what kinds of topics/subjects are of interest to this



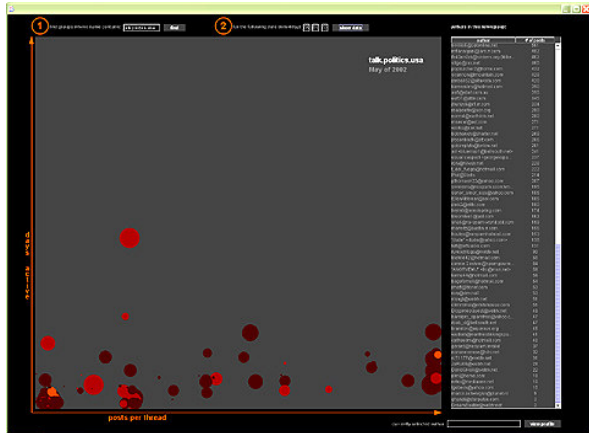
**Figure 2.** The newsgroup *alt.binaries.sounds.mp3.complete\_cd* displays almost all authors stacked are on the left of the scatter plot indicating that this is not a conversational newsgroup

person. They also indicate how focused on a specific topic the author might be; for example, if author A’s top five newsgroups all start with “microsoft.public.vb...” we might infer that author A is highly interested in Visual Basic. If, on the other hand, author B’s top five newsgroups range from political to philosophical and religious topics, we get the sense that this person has a more varied range of interests than author A. Moreover, clicking on various authors in the same newsgroups allows users to get a sense of whether there’s a lot of “top five newsgroup” overlap or not. It also allows users to determine whether this is a tightly knit and highly focused community or simply a place where people with wildly different interests come together.

#### 4.1 Author Metrics as Visualization Dimensions

Fiore and colleagues [5] report that metrics such as longevity, frequency of participation and the amount of messages an author contributes to each thread correlate highly with readers’ subjective evaluations of the author. In other words, users of newsgroups seem to employ these metrics informally and implicitly when interacting with others online in order to weigh and contextualize messages from different authors. We chose some of these metrics to function as dimensions in Newsgroup Crowds. Our intention was to build visualizations that make explicit dimensions already utilized by users when interacting with others in newsgroups.

Posts per thread – how densely packed posts are in a collection of threads – turns out to be a reliable metric to determine the degree of “conversational concentration” of an author in a given newsgroup. The higher the posts per thread ratio of an author, the more conversationally concentrated he/she is. For instance, if we were to look at an author who posted 30 messages



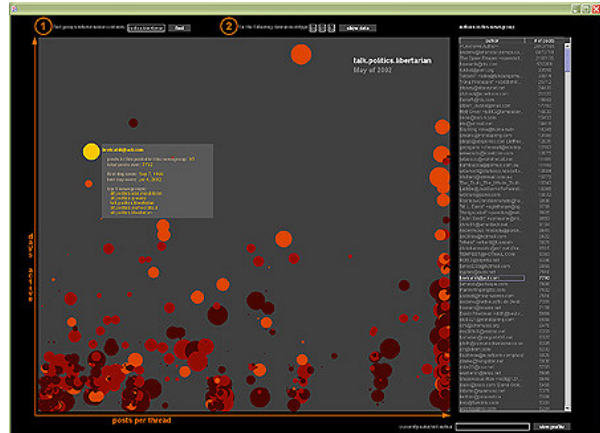
**Figure 3.** Visualization of *talk.politics.usa*, showing a lack of consistent contributors (all circles are at the bottom of the scatter plot) or recent contributions (there are no bright orange circles).

in the last week, all of which were posted to the same thread we would consider her a highly concentrated author. If, however, this same author had posted 30 messages to 30 different threads, we would not consider her a conversationally concentrated person. In Newsgroup Crowds posts per thread is the horizontal dimension of the scatter plot. This means that the more to the right of the scatter plot an author is displayed, the more conversationally concentrated they are. Conversely, the more towards the left side of the visualization an author is displayed, the less deeply they touch each thread and, therefore, the less concentrated they are.

Frequency of participation is the vertical dimension in Newsgroup Crowds. The higher an author is placed in the visualization panel, the more frequently she has posted to this newsgroup during the month being visualized. Authors displayed at the very top of the scatter plot have posted to the newsgroup every single day of the month. In opposition, authors who have been active for only one day show up at the very bottom of the visualization.

These two dimensions cause certain areas of interest to emerge from the distribution of authors within newsgroups. Generally speaking, there are four areas that turn out to be good descriptors of different patterns of behaviors for authors:

- . **Answer person or “Pollinator”** (upper left quadrant): high number of days active, low posts per thread ratio
- . **Debater** (right-most side of visualization): moderate to high number of days active, very high posts per thread ratio
- . **“Bursty” contributors** (lower right quadrant): low number of days active, moderate to high posts per thread ratio



**Figure 4:** Author with highlighted “information box” next to her circle; circles staked to the right of the interface reveal the highly conversationally concentrated behavior of several authors in this newsgroup: *talk.politics.libertarian*.

. **Newcomers and question askers** (bottom left corner): very low number of days active, low posts per thread ratio – in every newsgroup we analyzed this has always been the most densely populated area

The size of each author’s circle is determined by the amount of posts this person has contributed to Usenet as a whole – irrespective of newsgroup. This means that authors who have been consistently active for several years in the Usenet space, get shown as big circles whereas if they are new to Usenet, they get shown as small dots. This turns out to be an interesting piece of information in situations where “experts” drop in a newsgroup for the first time; it is their first time in this community, but they obviously have been around Usenet for a long time. It is also interesting to note when an author has a small dot that shows up fairly consistently (i.e. high in the vertical axis) in a given newsgroup; this could mean that this newsgroup is a core community for the author.

The color of each author’s circles reflects how recently this person has posted to this newsgroup: the brighter the circle, the more recent the posting activity. This metric makes highly active newsgroups [Fig. 1] look starkly different from more stale ones [Fig. 3].

Finally, there is a “View Profile” button and text field on the lower right corner of the Newsgroup Crowds visualization. This is where users choose to view a more highly detailed profile of a specific author. Should the user choose to press the view profile button, this would take them to the AuthorLines visualization, where much more detailed data about the chosen author is displayed.

## 5. Authorlines

After displaying authors within the context of a specific newsgroup and providing users with some



**Figure 5.** Detail of semi-transparent author information box with data points about this author.

information about each one of them, a natural next step seemed to be to focus on a single author and get a much deeper understanding of this person's posting activity.

AuthorLines functions very much like a histogram showing intensity of posting activity over time [Fig. 6]. It is a visualization of an author's posting behavior across newsgroups over an entire year. It shows a horizontal timeline with vertical monthly dividers. Month names are displayed at the top of the visualization panel. Vertical lines of circles represent weekly activity: each circle stands for a conversation thread to which the author has contributed during that week. In other words, a vertical lineup of 20 circles means that the author has contributed to 20 different conversations during that week. The size of the circle represents the number of messages contributed by the author to that thread; the more messages the author posted, the bigger the circle is. Whenever a circle gets too big to fit in the space allocated for a week it becomes semi-transparent so as not to obscure other circles around it.

AuthorLines differentiates between threads that were initiated by the author and those that were not. Orange circles placed above the timeline represent threads that were initiated by the author whereas yellow circles underneath the timeline are threads to which the author has contributed but which were not initiated by him/her. The disposition of circles is determined by their size: bigger circles are drawn closer to the timeline; this ensures that the conversations to which the author has devoted the most energy always show up around the center of the screen.

Because authors' activity is presented over time, it becomes very easy to spot periods of intense posting as well as weeks and months when there was no posting activity at all. In figure 6, for instance, the week of Sept. 11th 2001 is the only one that has hardly any posts, reflecting a clear exception to that author's

posting behavior. Moreover, the size of the circles makes patterns of posts per thread explicit, revealing whether the author tends to engage in deep debates on specific threads – big circles – or whether the author tends to touch threads lightly – small circles.

The visualization reveals posting patterns that illustrate different patterns of behaviors for authors:

- . **Answer person or "Pollinator"** [Fig. 7]: High number of days active, mostly responds to threads started by other authors with one or just a small number of messages sent to each thread.

- . **Debater** [Fig. 9] High number of days active, mostly responds to threads started by other authors with large numbers of messages sent to each thread.

- . **Spammer-like behavior** [Fig. 8] Moderate to high number of days active, almost entirely initiates threads which then receive no follow-up messages from this author.

- . **"Balanced" Conversationalist** [Fig. 6] Initiates about as many threads as he/she replies to and shows about the same posts per thread ratio on both initiated and non-initiated threads.

Users can click anywhere on the visualization panel and this causes the correspondent week to be highlighted by a vertical, semi-transparent rectangle. The date of the week is displayed at the top of the selection rectangle. The week selection causes the two tables to the right of the visualization to display the subject lines of the threads that were touched by the author during that week. The top table refers to the circles on top of the timeline – threads initiated by the author. The bottom table refers to the circles below the timeline – threads initiated by others. Say, for instance, that a user selects a week where the author has initiated five conversations and where he has replied to 20 conversations initiated by others; the top table would display five subject lines and the number of posts the author has contributed to each one of those conversations. Similarly, the bottom table would display 20 subject lines together with the number of posts the author has contributed to each one of them. Both tables can be sorted either by number of posts or by the alphabetical order of subject lines.

Users may also select an individual thread by clicking on one of the subject lines in the tables. This selection causes the corresponding circle in the selected week to become highlighted in red and the subject line of the thread to be displayed at the top of the visualization next to the week date. In case the author has contributed to this thread for more than a week, all other occurrences of this same thread are highlighted in red and all these red circles are connected by red lines [Fig. 9]. The lines become more important the smaller

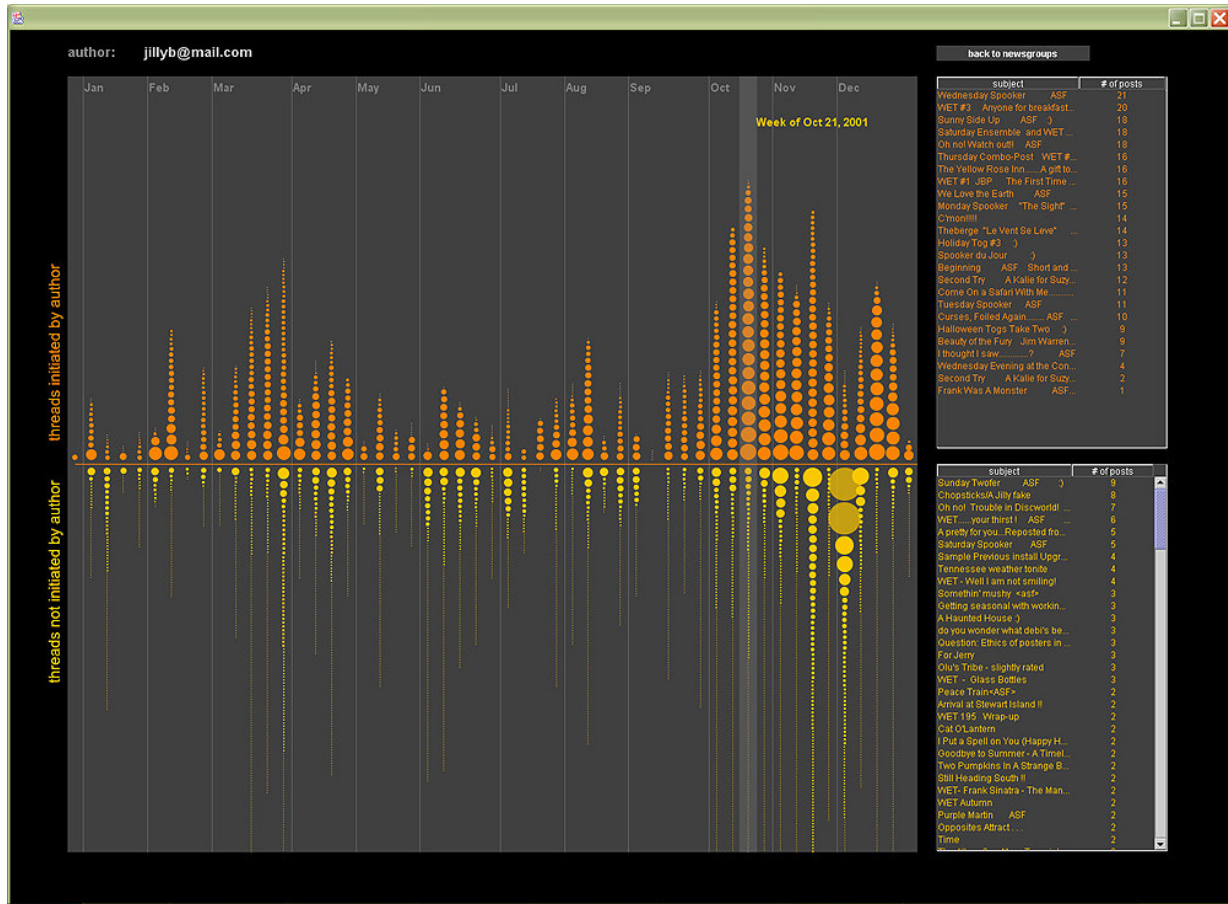


Figure 6. AuthorLines visualization of an author that starts about as many threads as the ones they respond to. This author has been active for practically every single week of the year with the most amount of activity happening towards the end of the year.

a circle is because a lot of times it is hard to spot the small circles which have been highlighted in red.

This visualization of an author's posting activity across newsgroups and over time reveals rather detailed patterns about a person's behavior. The sheer shape described by the circles around the timeline and their overlapping sizes make it very easy for users to grasp moments of intense activity and patterns of debating depth in any given thread. Also, by exploring the visualization one gets a good sense of the kinds of subject matters to which the author contributes and how much he/she contributes to each one of them over time.

## 6. User Study

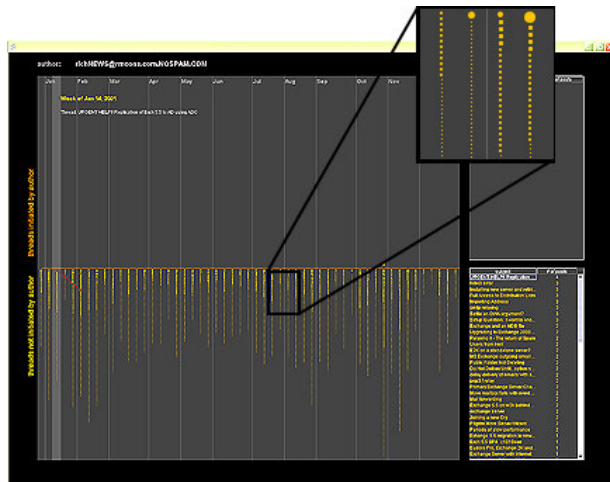
Because the user interface design and the functionality of our visualizations take such a different approach to news browsing, there was no point in constructing a study to compare and contrast our interface to those available in the market today. Therefore, we set up an exploratory evaluation rather than a comparative one. Basically, we wanted to investigate whether our visualizations could impress on

users some of the different behavior patterns of authors and, whether these impressions supported users' understanding of contributors in these conversational spaces. However, in order to get to such high level inquiry, more basic user interface elements had to be tested as well. For this reason, our user study included tasks that covered two main areas of interaction design:

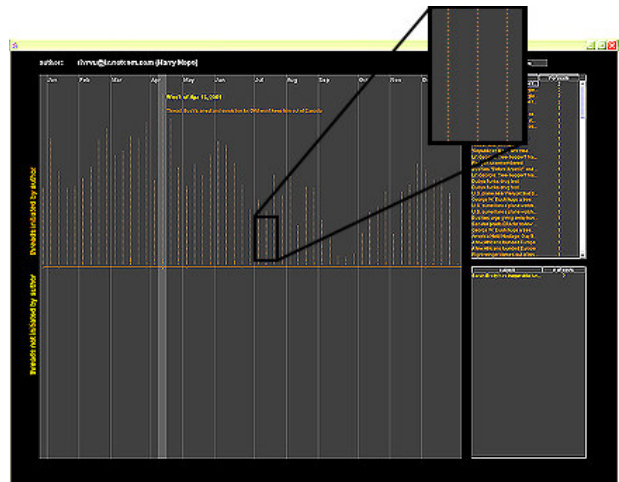
**1. usability:** Are these visualizations easy to use? Do the dimensions (axes, colors, sizes) and interaction areas (buttons, tables, clickable panels) make sense to users? Which interaction models in our visualizations frustrate users?

**2. usefulness:** How do users interpret the data they are shown here? Do the visualizations give users an at-a-glance understanding of the authors in these spaces? Do users understand the implications behind such user interfaces?

The results presented here stem both from our observation of how users interacted with the visualizations as well as from a survey that participants filled out at the end of the user-testing sessions. We chose quotes that were representative of users'



**Figure 7.** Overall view of “Question-answerer” author and detail box showing close up of small yellow circles; small circles indicate that the author contributes minimally to each thread, rarely posting more than three or four messages in each conversation.



**Figure 8.** Overall view of spammer-like behavior and detail box showing close up of small orange circles; the author contributes just one (initial) message to each of the threads suggesting that he/she is either a spammer or, at the very least, not a highly conversational person.

reactions; all quotes come from the written survey filled out by participants.

We administered a preliminary online survey within our company asking people about various aspects of their newsgroup experience. From the respondents we selected a set of 15 heavy newsgroup users – we defined “heavy” users as anyone who consistently read or posted to Usenet newsgroups at least once a week throughout the year. We invited participants to our lab for study sessions of 1.5 hours. During each session we briefly explained the functionality of both visualizations and led participants through a few practice tasks. After this introduction, participants were given a list of tasks to perform. They were asked to think aloud as they performed the tasks. At the end of the test, participants were asked to fill out an online survey about their experience with the interfaces.

## 6.1. Demographics

Of the 15 participants selected for the user study, eight were female and seven were male, with ages ranging from 24 to 53. Experience with Usenet newsgroups was as follows: 60% had over two years of experience; 25% had between one and two years of experience; and the other 15% had up to a year of experience. Reasons participants had for using newsgroups were: 86.7% for technical support, 66.7% for hobby, 40% for political discussions, 20% for emotional support and 33% used newsgroups for other reasons.

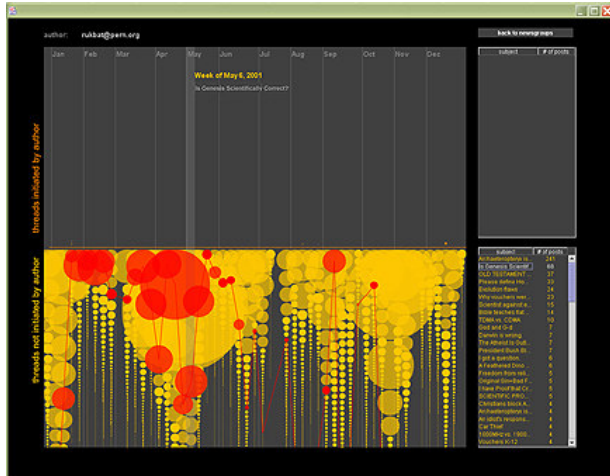
## 6.2 Procedure

(i) **Newsgroup Crowds:** Participants were asked to examine five different newsgroups:

1. alt.politics.bush
2. talk.origins
3. microsoft.public.vb.general.discussion
4. talk.politics.usa
5. rec.sport.tennis.

All of the newsgroups were visualized for the month of May 2002, which was the latest set of aggregate data we had available to us.

For each one of these newsgroups, participants were asked to identify a couple of authors by their email addresses (which could be done using the table with all authors’ email addresses). They were also asked to find authors based on how consistently they had returned to the specific newsgroup and how conversationally concentrated they were. For each one of these tasks, participants were asked to explain their choices – we wanted to determine whether users could understand the axes in the visualization or whether they had to make a big effort to comprehend the placement of authors in the scatter plot. Finally, participants were asked to report on the information displayed about each one of the authors they had highlighted during the study – the information contained in the author’s “information box” that shows up next to the author’s highlighted circle – they were asked to tell us what kind of contributors they thought these authors were based on the information they had about them. Finally, in a subset of newsgroups, they were asked to find out whether there was a lot of overlap among the “five top



**Figure 9.** “Debater” behavior: the size of the circles indicates that the author contributes significant amounts of messages to a lot of threads. The highlighted thread (red connected circles) lasts for almost the entire year.

newsgroups” lists of different authors within the same newsgroup. This task was performed for the following newsgroups: microsoft.public.vb. general.discussion, and alt.politics.bush.

The former newsgroup has, in its core contributors, a high degree of overlap – most of the core authors contribute to a lot of the same Visual Basic newsgroups outside of this one. In the second newsgroup, however, overlap is not as evident; rather, authors seem to have a wider range of interests spanning such topics as philosophy, religion, vegetarianism, cars, and guns, etc.

**(ii) AuthorLines:** In this visualization study, participants looked at the profiles of six different authors from the five newsgroups listed above. We chose these authors to represent a wide variety of posting behaviors. AuthorLines displayed authors’ data for 2001.

Tasks included: identifying periods of intense posting activity, identifying specific weeks in the year, identifying specific threads within given weeks; determining whether authors tended to participate in the same threads for over a week, recognizing threads initiated by the author and those that weren’t. Moreover, users were asked to elaborate on what the distinct activity patterns displayed by the different authors might mean: some authors displayed spurts of activity followed by months without posting anything, others posted every single week of the year, some posted a whole lot of messages to the same threads while others would lightly touch a lot of threads every week.

## 6.3 Results

Overall, users’ response to both visualizations was positive. Most users found the visualizations highly useful and did not have any major difficulties interacting with them. In general, users were also impressed at how fast they could learn about previously unknown authors in the newsgroups presented to them. Most users were also quick in forming opinions about the kinds of authors they were looking at; most of the time, after looking at data about five different authors, users had become comfortable with the metaphors used in the visualizations (colors, placement, sizes, etc) and were able to determine patterns that they thought reflected “regular” posting activity and patterns that they deemed unusual – for instance, most participants in our study were very surprised when presented with data from an author whose posting pattern looked like the one on figure 9, which showed very intense reply activity over the entire year; sometimes hundreds of messages to the same thread in a single week.

### *(i) Newsgroup Crowds*

**a) Usability:** For the most part, users found both axes in the scatter plot to be clear (81% found them clear, 10% moderately clear, 9% did not find them clear). Most users found it easy to find specific authors in the visualization (84% found it easy, 16% did not find it easy). In general, users also felt that they got a good sense of “interaction dynamics” by looking at the scatter plot (93% got a good sense, 7% did not get a good sense of the interaction dynamics of newsgroups).

**b) Usefulness:** For the most part, users found the visualization successful in portraying differences between newsgroups and most users said they would be interested in using this tool again for news browsing (50% were extremely interested, 21% were interested, 29% were moderately interested). Users’ responses about favorite features in the visualization include: (a) being able to tell, at a glance, how dynamic a newsgroup is and who the top contributors are, (b) ability to tell whether a particular newsgroup was inhabited by consistent contributors or whether it was a place where people “come and go” (e.g. “It gave me a sense of the community, responsiveness and participation types in the group immediately. I found myself starting to draw boxes around groups of people and saying: these guys over here behave like this whereas these other people behave like that”), (c) being able to, very quickly, get a sense of the community and what kinds of things interest people in the same newsgroup. Users’ responses about least favorite features include: (a) not being able to do side-by-side comparison of multiple newsgroups, (b) not being able



to get to the content of the messages exchanged by the people in the newsgroups (e.g. “You can't tell the quality of the posts from the visualization, only the quantity”), (c) not being able to zoom into the more dense areas of the scatter plot so that one could more easily identify overlapping authors.

Finally, when asked whether having access to a visualization like Newsgroup Crowds might affect their choice of newsgroups in which to participate, 54% of users answered Yes, 23% answered Maybe and another 23% answered No.

### *(ii) AuthorLines*

**b) Usability:** It became clear to us that users had no problems getting a sense of the times of the year when authors had the most posting activity (78% found it extremely easy, 22% found it easy). For the most part, users found it easy to find specific weeks during the year (43% found it extremely easy, 29% easy, 28% found it moderately easy). Users also didn't have major problems finding a specific thread within a selected week (14% found it extremely easy, 43% easy, 36% moderately easy, 7% not easy). About 40% of respondents were frustrated to find out that they could not click on the circles of a highlighted week as a way of selecting a single thread in that week. These repeated attempts made it clear that we should enable all selection actions to take place on the visualization panel in addition to any selection actions that might be possible on the tables.

**b) Usefulness:** For the most part, users found AuthorLines successful in expressing basic differences in author behavior (52% found it extremely successful, 32% successful, 16% moderately successful). Most users said they would be interested in having this tool available to them for news browsing (51% said they would be extremely interested, 21% interested, 21% moderately interested, 7% were not interested).

Users' responses about favorite features include: (a) being able to get a sense of author's behaviors without actually having to read all of their postings, (b) ability to, at a glance, grasp behavior over a long period of time, (c) being able to highlight threads and see how authors continue to contribute to the same threads over time, (d) seeing the different patterns in threads initiated by the author and threads initiated by others. Users' responses to least favorite features include: (a) inability to do side-by-side comparison of authors, (b) inability to play with the time dimension so that one could see posting patterns more clearly over a month or even over a week as opposed to simply looking at activity over a year, (c) not being able to get to the content of posts and threads from the visualization, (d)

inability to tell how big a thread is (e.g. “If I see that someone posted to a thread 20 times, I want to know if the thread has 40 posts total, or 2000 posts, total -- this affects my impression of the author's behavior”).

When asked whether having access to a visualization like this one might affect their choices of which messages and threads to read, 72% said Yes, 14% said Maybe and another 14% said No.

### *(iii) Privacy*

A deep discussion about privacy issues and their sociological implications in Newsgroup Crowds and AuthorLines is beyond the scope of this paper. Nevertheless, the nature of the work presented here is such that a few remarks on privacy seem not only relevant but also prudent if we are to carry out responsible research in this area. While some of our participants marveled at the possibility of, as one of them put it “seeing a particular author as the kind of person who is a vegetarian, drives a Volvo and sympathizes with environmentalist causes”, others felt this much insight into someone else's life style could be a dangerous thing. We feel that they are right.

The goal of this research is not only to impress on users a vivid image of authors' behaviors and activity patterns but also to give users some inkling at how much data is available out there about each one of us as we interact online.

Users' concerns about privacy demonstrated that the work presented here was effective in rapidly creating a “picture” of authors. We explained to our users that all data being visualized were public; nothing we showed them was, at any level, private. Having said that, we also understand that aggregating public data causes different privacy implications to emerge: people act a certain way in public spaces when they know their behavior is not being recorded. People act a different way in public spaces when they realize their actions are being recorded. Our take on privacy in this project is that, because computer-mediated conversational spaces such as Usenet are intrinsically recordable, user interfaces that bring up this archivable quality to the forefront of the user experience actually do users a favor. They remind users at all times that this is a mediated space, one where people can collect your data and profile your behavior. Because the interfaces here do not hide this mediated reality from users, they act as constant reminders and give users a better chance of adjusting their behavior and interactions accordingly.

## **7. Future Directions**

Even though users' response to both visualizations was decidedly positive, one of the problems users

pointed out in both interfaces was the fact that they currently don't allow you to get to the actual content of the messages; basically, we show users form (quantities of postings, intensity of activity over time) but not content. This is a serious limitation of the current system and we plan to add the ability of getting to thread structures and the body of messages via the visualizations. A useful scenario would be for users to bring up an entire thread – complete with message bodies – when they highlight a specific thread circle in AuthorLines.

Users also felt the need to compare multiple newsgroups and authors on the same screen. Our intention is to support side-by-side comparisons in the next version of the system. Moreover, we also feel it would be interesting to see where a particular author is placed on the multiple newsgroups to which he/she contributes in the Newsgroup Crowds visualization. This would allow users to see how the same author behaves in different newsgroups.

Because AuthorLines shows authors' behavior across newsgroups, it seems important that we give users a better sense of all the different newsgroups to which an author contributes. To that end, every time a user selects a thread in AuthorLines, a list of the newsgroups to which the thread was posted could be displayed next to the highlighted thread.

Finally, one major limitation of both systems right now is the lack of readership data. By displaying solely posting activity patterns, we automatically call attention to certain kinds of newsgroup users while completely omitting a whole category of participants: the readers. Data about "who is reading what" is crucial for transforming online conversational environments into richer, more legible social spaces.

## 8. Conclusions

The visualizations presented here feed collected author behavioral data back into communities of Usenet newsgroups in ways that make overall activity patterns easy to understand. Previous work in this area has shown that quantitative behavioral metrics, in particular those that capture aspects of an author's tenure in a newsgroup and level of interactivity with other authors, serve as reliable predictors of subjective evaluations of the author's social and informational value to people in the community. By extension, these metrics also serve as predictors of readers' satisfaction with messages from others. We contend that, once users are able to quickly grasp authors' patterns of posting activity and behavioral histories, they will be better equipped to decide which authors and messages might be of interest to them – something that is not

supported by current news browsing interfaces. Our study showed that, when viewing authors' data in our systems, users were able to quickly differentiate between the different behavior of authors. Moreover, the information users extracted from the visualizations was viewed by them as an important guide to further exploration of the conversational spaces as well as an effective means of navigating such voluminous spaces such as Usenet newsgroups.

Visual feedback mechanisms such as the ones discussed here might prove to be an important initial step towards designing social spaces that foster a higher level of social legibility and, possibly, accountability.

## 9. Acknowledgements

We are very grateful to the participants in our study. We would also like to thank the entire team from the Microsoft Research Community Technologies Group: Paul Johns, Duncan Davenport, Dany Rouhana, Bas Hugenholtz.

## 10. References

- [1] boy, d., Lee, H., Ramage, D., Donath J., Developing Legible Visualizations for Online Social Spaces. Proceedings of HICSS 2002.
- [2] Donath, J., Identity and Deception in the Virtual Community. In Communities in Cyberspace, London, Routledge Press, 1999.
- [3] Donath, J., Karahalios, K., Viégas, F., Visualizing Conversations. Proceedings of HICSS 1999.
- [4] Erickson, T., Smith, D. N., Kellogg, W. A., Socially Translucent Systems: Social Proxies, Persistent Conversation, and the Design of "Babble". Proceedings of CHI 1999.
- [5] Fiore, A. T., LeeTiernan, S., Smith, M., Observed Behavior and Perceived Value of Authors in Usenet Newsgroups: Bridging the Gap. Proceedings of CHI 2001.
- [6] Kelly, S. U., Sung, C., Farnham, S., Designing for Improved Social Responsibility, User Participation and Content in On-Line Communities. Proceedings of CHI 2002.
- [7] Sack, W., Conversation Map: A Content-Based Usenet Newsgroup Browser. Proceedings of IUI 2000.
- [8] Smith, M., Invisible Crowds in Cyberspace: Measuring and Mapping the Social Structure of USENET. In Communities in Cyberspace, Routledge Press, 1999.
- [9] Ware, Colin. Information Visualization: Perception for Design. Morgan Kaufman Publishers, 2000.
- [10] Xiong, R., Donath, J., PeopleGarden: Creating Data Portraits for Users. Proceedings of UIST 1999.
- [11] Metrics generated by the Netscan Usenet analysis system: <http://netscan.research.microsoft.com>
- [12] eBay Web site: <http://www.ebay.com/>
- [13] Amazon Web site: <http://www.amazon.com>
- [14] Slashdot: <http://slashdot.org/>