Using Social Metadata in Email Triage: Lessons from the Field

Danyel Fisher, A.J. Brush, Bernie Hogan, Marc Smith, and Andy Jacobs
Microsoft Research
One Microsoft Way
Redmond, Washington. USA
{danyelf, ajbrush, masmith, andyj}@microsoft.com, bernie.hogan@utoronto.ca

ABSTRACT
Email archives are full of social information, including how messages are addressed and frequency of contact between senders and receivers. To study the use of this rich metadata for email management, particularly email triage, we deployed SNARF, a prototype tool which uses social metadata to organize received email by correspondent, sort received email by past interactions, and filter email into multiple views. We discuss the lessons from a seven month deployment, including the value of organizing by personally addressed mail, the unexpected value of SNARF for email awareness, challenges with handling workflow, and ways to use social metadata in applications.

Categories and Subject Descriptors
H.1.2.a Human factors, H.4.3.c Electronic mail

General Terms
Performance, Design, Human Factors

Keywords
Email overflow, email triage, social metadata, social sorting

1. INTRODUCTION
The second time we meet a friend's dog it typically remembers us and doesn't growl. Remembering how we have interacted with others in the past is basic to social interaction. Much like in-person interaction, email communication is rich with social information. For example, how frequently, if at all, have you previously corresponded with a particular sender? Was the message addressed directly to you or to a mailing list? Are you the only recipient? Unfortunately, email clients typically ignore this *social metadata* —the information about both a persons' history of interaction with their correspondents, and the ways that a message is addressed— treating all messages equally without regard to whether they come from a frequent correspondent or new friend.

During the task of email *triage*, people seek to separate relevant messages from a swath of less relevant messages. Neustaedter et al. [10] suggested that many social aspects of an email message are important to users when deciding what email to read during email triage including the sender and how the email was addressed to the recipient. That research, as well as others [15], also found that many users choose to employ a multi-pass strategy to read their email, skimming through to select ones that are particularly easy to handle or particularly important. Additional passes—to choose the next most important message—follow as time permits. Unfortunately, most email clients have limited support for the multi-pass practice, presenting email in order of receipt, rather than relevance.

While several research systems [7, 13] use social metadata in various ways, contemporary email clients do not provide users a way to directly view or access social metadata, nor do they utilize interaction histories. (This information is sometimes indirectly used to help fight spam.) Indeed, current email tools provide few social cues: besides instant messaging icons, email tools rarely give any feedback about our correspondents at all.

Although email facilitates communication and collaboration, email management is an intensely personal practice. To study the value of social metadata for email triage in a real world environment over time we extended and deployed a prototype application, SNARF, the Social Network and Relationship Finder [11]. At the core, SNARF implements a simple idea: present email organized by sender in lists that are ordered by the strength of relationship between the sender and user. Email from a colleague one corresponds with frequently will appear higher in the list than email from an infrequent correspondent. SNARF, then, uses the frequency of communication as an (imperfect) proxy for relationship strength.

SNARF employs social metadata to present users with alternative views of their email in three ways. It *filters* email by whether it has been read and how it is addressed; it *organizes* email by sender; and it *sorts* senders based on their history of interaction with the user. By allowing the user to focus on contextually-appropriate messages, and by bringing messages from frequent correspondents to the fore, we believe that SNARF allows users to more easily locate relevant messages, especially when pressed for time. SNARF supports the advantages of a multi-pass strategy without requiring users to repeatedly scan their entire inbox.

Over the course of seven months and two iterations of SNARF we gathered feedback through surveys, usage logs, and user

suggestions from nearly a thousand people in our organization about their experience with SNARF. While the project was eventually deployed to the public, information was collected only from employees of our company for a range of privacy, sample quality, and data retention concerns.

SNARF, like any prototype, was not universally adopted. None the less, the amount of feedback (both positive and negative) that we collected from people using SNARF on their own email in everyday life taught us several lessons about the value of social metadata for email management. Most popular among SNARF users were the way to filter for messages explicitly addressed to themselves. Somewhat surprisingly, several SNARF users made unanticipated use of the tool to help with awareness of new mail. We also saw challenges for making use of social metadata around identity, particularly when people use multiple email addresses, and when people dramatically change their communication context, as when taking a new job.

In the rest of the paper we discuss related work (Section 2) and then describe the design of SNARF (Section 3) and our field deployment (Section 4). Section 5 presents our key findings from the field study and highlights how our experience with social metadata can help other designers and developers make informed decisions on how and when to incorporate social metadata in their applications (Section 6).

2. RELATED WORK

Assisting users in working with email has been addressed in many different ways in the HCI field from innovative interfaces [2, 7] to prioritizing systems [8]. In particular, many projects have recognized the challenges of email management and tried to help. Whittaker and Sidner [16] began a line of research in noting that email is used for a variety of personal information management tasks in addition to communication. They refer to the many different aspects of email—maintaining a calendar, keeping contacts, and driving a working memory—as "email overload."

Other projects have attempted to more generally help with email management and relieve stress caused by email. Priorities [8], for example, uses machine learning techniques to recommend which messages to read next. It uses a broad set of email features to evaluate messages. SNARF is more user-configurable, allowing users to create several views, each of which may help the user choose which messages to read.

IBM's ReMail [7] prototype includes a Correspondent Map, which organizes email by correspondent and organizational affiliation. This aspect was not a major focus of ReMail, which also displays conversation threads, displays groups of people by their organizational structure, and connects calendar information to email messages. Tyler and Tang [13] evaluates email histories to predict future availability and responsiveness. In comparison, SNARF explores in depth the use of social metadata in email for the task of triage.

Nardi et al. [9] suggest that contact management takes a substantial amount of effort, and propose a socially-based interface that shows information oriented around people. In SNARF, we adapt their notion of orienting an interface around correspondents (rather than messages) for handling email.

More generally, the notion of applying social metadata was used by Fiore et al. [5] to examine ordering of Usenet news messages. That study found that ordering messages by the people involved in them provided an effective mechanism for highlighting valuable content. Fisher and Dourish [6] explored another way of placing email within a social context by describing social roles that are visible in social networks of email messages. Both of these cases present interaction histories from a novel perspective. Both of them are oriented toward retroactive analysis of archives; by contrast, our approach uses archival information to help users make decisions in real time.

3. DESIGN OF THE SNARF PROTOTYPE

Takkinen and Shahmehri [12], followed by Venolia et al. [15] discuss multiple phases of handling email. SNARF is designed for email triage (or "busy mode"), a time when the user needs to locate particularly important messages and handle them quickly. The technical basis of SNARF's design was discussed in [11]. We modified and extended that version of SNARF.

Here, we focus on the ways that SNARF uses social metadata to present email and the features needed to support day to day use in the field deployment.

3.1 Use of Social Metadata within SNARF

SNARF uses information about past email behavior to display a user's email in three main ways.

3.1.1 Filtering Mail Into Views

The main SNARF window is divided into several panes, each presenting one *view* showing a subset of available email correspondents[1]. Views filter mail based on how it was addressed: they separate mail that is addressed directly to the user—mail that is more likely to invoke a new task—from mail that may not be as critical. In general, views are based on a collection of sorting and filtering settings. Several views are provided with SNARF and additional views can be created by the user.

Figure 1 shows the default configuration of SNARF, with three views. The top view, "Unread To/CC me," shows correspondents whose unread messages have explicitly included the user in the To or CC line. The middle view, "Unread Mail," shows unread messages from *all* correspondents (and thus may have been received through a mailing list). The bottom view, "This Week's Mail," shows all correspondents from whom the user has received messages in the last week, read or unread. We chose this view explicitly as a default for its confirmatory capabilities: even without unread mail, it ensured that users would see mail, and could see that SNARF was working. Note that the panes are not mutually exclusive; a particular message or person may appear in multiple panes if they match the criteria.

People may attach different social meaning to email sent to them or to a mailing list. Although SNARF can not automatically determine if a correspondent is a mailing list, it allows users to tag correspondents as Lists. The user can then display the built-in "Unread Lists" and "Lists" views by either adding another pane to SNARF, or replacing one of the default views.

The original design for SNARF [11] showed only one view at a time; users would re-configure the display to match their current task. However, as the design evolved it became apparent that it is

[1] An email address can refer to a person, a mailing list, or an automated generator. For simplicity, we refer to all senders and receivers as correspondents.

Figure 1. The SNARF Main Window with default views. (All names have been anonymized)

important to have multiple views so that *more* types of sorts could be shown. This reflects the difference between the relevance of a particular sender, and a sender's message. For example, Chris Adler may be very important to me, but messages from Chris sent only to me are often qualitatively different from messages sent by Chris to a mailing list we both subscribe to. The current display, which can accommodate three views of eight or so names without scrolling on a conventional screen, reflects this design.

3.1.2 Organizing by Correspondent

Messages are clustered together and linked to the sender of the message, in order to present a social perspective on the email. Each correspondent name is followed by the number of unread messages from them. For example, in Figure 1, the Unread To/CC me view shows Cynthia Allen has sent 2 unread messages. Highlights are provided as cues that convey the relative number of messages from the correspondent within a view: the highlight under Otto McCoy (4 messages) is longer than the highlight under Samantha Turner (1 unread message).

3.1.3 Sorting by Correspondence History

Each view presents names socially sorted: ordered by the degree of relation to the user that can be extracted from their interaction history. In the default view (and most uses of SNARF), names are ordered by the number of times that the user has sent mail to that person in the past year, although the time period is configurable.

Thus, frequent and recent correspondents bubble to the top, while infrequent correspondents sink to the bottom. In Figure 1, in the To/CC me view, the user has sent more email messages to *Darcy Adler* than to *Tim Andrews; Darcy* is therefore sorted higher on the display than *Tim.* The configuration options available on this display are discussed in more detail in [11]. The original design for SNARF displayed two bars for each person, with one of the bars was showing the sorting criteria. The current version now displays a single bar, with the value used for sorting available if the user hovers over a correspondent's name

3.2 Design for Field Deployment

In order to deploy SNARF, the prototype needed to be a tool that could support real day-to-day use. In particular, this required SNARF to respond to real-time updates and the ability to open mail and read it. SNARF therefore is a MAPI client, and connects itself automatically to the users' Outlook installation. The several views, like all aspects of SNARF, respond in real time to changes in email state. For example, if a message from a sender to the user is marked as read, the "Unread Messages" view is updated—it will no longer show that sender if no more unread messages remain.

When the user double-clicks on an entry showing just one message, SNARF opens the message in Outlook. If the correspondent is associated with more than one message—that is, there is more than one unread message from that person—then SNARF presents a message list in a second window. Each of the messages from the correspondent is shown with its author, subject, and date; they can each be individually opened in Outlook. SNARF also contains an optional "thread view", following the design by Venolia et al. [14] to show several messages in a thread.

4. FIELD DEPLOYMENT

We gathered data about SNARF's use in the field in two phases. The first phase was a structured field study that took place over a six week period in July and August of 2005 where we conducted pre-surveys and post-surveys of SNARF users. Once the structured field study concluded, we shifted to a second phase where we released an updated version of SNARF and collected usage logs.

4.1 Phase 1: Field Study

The field study began in July 2005 when we sent out an invitation to use SNARF to 1713 people who had registered their interest at an internal demonstration of the system. We also sent out an invitation to two high-volume mailing lists of employees at our company that specifically target people with interest in advanced and speculative projects. In total, 574 people ran SNARF at least once within the first two weeks after it was made available. We stopped accepting new people into the study after two weeks, although they could still use SNARF. During the study we explicitly chose to make only one change to SNARF to address a performance issue when opening the message list window.

4.1.1 Data Collection

All users that ran SNARF in the first two weeks of the study were asked to complete an optional pre-survey related to their

Table 1. Behavioral data collected in the SNARF Logs

SNARF USAGE LOG DATA (Daily):
• When SNARF was started and shut down
• What views were shown in SNARF
• How many times users clicked in the SNARF interface: opening messages, message lists, or thread views
• When incoming messages came to SNARF, and when messages changed status (e.g. from read to unread, or were deleted)
• Current folder structure, including number of read and unread messages, earliest message, and latest message per folder

experiences with email. The survey followed the outlines of [10]; however, it added more detailed information about email habits within the user's current client. The survey included questions asking users how long they thought they spent triaging email, how much of their email they read, and their experience of stress associated with handling email.

In August 2005, at the end of the six-week study period, we sent out a post-survey. Users were asked to respond regardless of whether they had filled out the pre-survey or had continued to use SNARF after its initial installation. All users who reported that they used SNARF 'for a few days' or more were asked about their use in more detail on the post-survey. Users were asked to respond regardless of whether they had filled out the pre-survey or had continued to use SNARF after its initial installation. Participants that filled out the post-survey were entered into a lottery for gift certificates.

SNARF is instrumented to log a substantial degree of detail about a how participants use the program as well as anonymized information about a user's email store. Unfortunately, the logging during the field study was not robust due to issues with logs on laptops and appropriately determining where users had clicked. Thus our log-based analysis relies only on newer logs from the ongoing study of the use of SNARF (4.2).

4.1.2 Study Experience

During the study we received extensive feedback from SNARF users. A total of 292 people completed the pre-survey, a response rate of 51% and 161 people filled out the post-survey (response rate of 28%), some of whom had not responded to the pre-survey. Post-survey respondents identified their job role as largely program managers (23%), developers (17%) and consultants (13%); sales (10%) and software testers (9%) made up the bulk of the remainder. The population was overwhelmingly male (92%).

Post-survey responses help us understand SNARF adoption. Of the respondents 4% of users reported that they could not get SNARF to work, 44% 'tried it once or twice', 29% tried it 'for a few days', 11% tried it 'for a few weeks' and 13% were 'still using SNARF'. Therefore 53% (N=84) of respondents told us about their experience in more detail. Due to logging problems, we can only approximate usage log statistics, but we received at least one log files from 510 people. Of these, we have logs from more than 5 days from 20% of them, while from 36% we received only one log.

Obviously not everyone who tried SNARF continued to use it. As with any prototype there are a variety of reasons that people might not adopt SNARF in the long term, ranging from performance (the primary reason given on the post-survey for discontinuing use) to expecting more functionality (another popular reason). Because user response ranged from those that had technical problems to those that enthusiastically adopted SNARF, the field study helped us understand positive and negative aspects of using social metadata for the task of email triage.

4.2 Phase 2: Ongoing SNARF Use

After the field study concluded, we released a series of updates to SNARF which improved performance and screen painting, fixed logging problems and addressed issues in the user interface that the first version revealed. In particular, later versions of SNARF gave users more flexibility to mark and delete messages from the interface, and the user interface was redrawn for a more elegant look.

In late November, we promoted and released this improved SNARF. 410 users tried the program at least once and provided log data during this second phase, which ran from late November to February of 2006. Usage data was logged and uploaded daily containing records of both any changes to the mail store and click data within the program itself. Table 1 outlines the general usage data which we collected. Similar to during field study, SNARF did not work for everyone. Of the 410 users, 25% used it on five or more days (the maximum was 103 days, almost seven days week). The usage data provides a complementary picture to the qualitative survey data.

5. LESSONS FROM THE FIELD

"[It] helps me quickly identify emails that need my attention – especially when I return from vacation."

In this section we describe the lessons we have learned from our seven month field deployment of SNARF, bringing together qualitative data gathered in the field study phrase and quantitative data from user logs as well as feedback provided directly to us by users. We classify our observations into a number of different aspects of using social history, including for *social filtering*, *social awareness*, and *email workflow*. We examine places where SNARF has succeeded—and ways in which it has not. Throughout, we use quotations from emailed and survey feedback to help illustrate SNARF use.

5.1 Social Filtering

SNARF uses social filtering to give users multiple views of their email. We learned that people are particularly focused on whether a message is addressed to them and they took advantage of having multiple views. We also saw that the interface must be careful to support multiple overlapping views in a way that is understandable to the user.

5.1.1 Mail That's Personal Matters

"[My favorite aspect of SNARF was:] Showing the email that was just to me or cc to me."

SNARF provides many different configurable views. Conceptually, we separate these views into two types. *Personal views* filter for messages that explicitly mention the user, such as "Unread To/CC me." In contrast, *aggregate views* show collections of messages are not addressed to the user, such as unread messages sent to lists ("Unread lists"). Personal views generally contain shorter lists of people; the latter type generally contains more people (e.g. average number of correspondents

listed in the Unread to/CC me view is 8.9, while average for Unread Mail view is 148).

If users were interested in keeping track of all their incoming email—as many users do by tracking their inboxes—we might expect to see that the aggregate views were most popular; were they trying to reduce their email load by skimming off high-value messages, we might instead see more use of personal views. On both the post survey and in the usage logs, we saw a preference for personal views

On the post survey, we asked users to evaluate which of five views they found useful. These included the three default views, the optional "Unread to Only Me" view (which shows mail that is addressed only to the user, and no other people) and the optional "List" view (which show messages sent to correspondents tagged as mailing lists). As shown in Table 1, survey respondents preferred personal views featuring unread messages to others: the personal views are rated higher than others (median values highlighted). Note that the "Unread to only me" view was not a default setting: users needed to explicitly choose it from a settings dialog. Yet for the users who found it, it was more likely to be rated as "indispensable" than any other view.

Data from the usage logs about where SNARF users click also favors personal views. The single most popular view was the "Unread to/cc me", garnering 3592 clicks from 237 distinct users, although unread mail overall also gathered a large number of clicks.

The survey responses and usage log data suggest that users were particularly interested in information about themselves: a message explicitly addressed to a person by name has some claim to social connection and personal content. The popularity of "Unread only to me" among those who found the view might be explained by the fact that it catches mail that can *only* be handled by the addressee: a message that he or she does not read will go unhandled.

5.1.2 Multiple Views Were Used

SNARF provides multiple views of a person's email. On the survey, users "agree" (mean = 1.17, N=69[2]) that "it is useful to filter mail into a number of panes." Of the 358 people that clicked once in any view, 225 of them clicked at least once in two or more views. This supports the idea that maintaining multiple perspectives on email is valuable. Although as the previous section showed some views were more popular than others.

Value	Personal Views		Aggregate Views		
	Unread to only me	Unread to/cc me (*)	Unread mail (*)	This week's mail (*)	List view
Used view (N=84)	59	76	74	70	52
It was distracting	3%	1%	3%	6%	4%
Not useful	5%	4%	11%	36%	14%
Somewhat useful	31%	33%	42%	46%	63%
Very useful	37%	45%	38%	10%	17%
Indispensable	24%	17%	5%	3%	2%
Logs (N=410)	22	237	221	208	24
Clicked at least once					
Click count	227	3592	2108	786	79

Table 1. Use of Views from the SNARF Survey and Logs. Survey medians are highlighted. (*) Default View

We believe that users found different views useful for different tasks. During the triage task, for example, users might concentrate on mail sent directly to them as they cleared out messages that directly needed their attention. Several users reported that list views were particularly useful for catching up on a high volume of mailing list messages. Skimming the top off the "Unread Mail" view allows a user to monitor interactions on mailing lists and discussion that involve their closet correspondents.

We saw interesting second order effects from multiple views among our research team. Some people who knew their correspondents were using SNARF began to alter their behavior when sending to mailing lists used by the team. Most mailing list messages are thought of as going to an undifferentiated set of recipients. Yet often, they are *intended* for one or two people, and *also sent* to lots of others. By putting the *intended* name in the "to" field of a message, SNARF will explicitly float this message into the "To/CC Me" view in the recipient's SNARF display making it more visible.

5.1.3 Challenges Showing Messages Multiple Places

On the other hand, there are complications to showing messages in multiple views. Some users complained that they would see the same message in several different places. For example, a message that was sent *both* to me and a mailing list I participate in, from someone I interact with often, might appear at the top of multiple lists: "Unread To/CC Me" might show it as personal mail; a List view might show it as aggregate mail; and "Unread Mail" might also bring it to the top.

The experience of reading one message and having several parts of the interface change can be distracting or confusing. While SNARF does not fully handle this situation, any system that repeats information in multiple places must make this clear.

5.2 Organizing by Correspondent

"[My favorite aspect of SNARF was:] Categorization by person."

SNARF supports the idea of organizing information entirely socially. When users are pressed for time, research [10] has suggested that they feel a need to "cherry-pick" their mail, often seeking out messages from people they know are relevant. SNARF cherry-picks for them, first separating messages that are directly to them, then ordering sets of messages by how often the user and sender correspond with one another.

Clustering by person seemed largely successful. While we feared that users might complain that it rolled together irrelevant ("Lunch!") messages with relevant ("Need report soon"), on the survey, users "agree" (mean = 1.08, N=73) that listing "correspondents rather than individual messages is useful." This suggests that while SNARF, most people were either consistent

[2] On a -2 ("Strongly disagree") to +2 ("strongly agree") scale. This scale is used for all survey responses unless noted. The words "agree," "strongly agree", and similar in quotation marks refer to the median choice.

with the topic of their messages, or were content to read occasional irrelevant messages. We are exploring adding additional context to messages in SNARF —displaying the subject line as a second level of peripheral information for each message, for example—might have reassured users who felt that the system was too sparse.

5.3 Social Sorting

"It is good to see mail from people I know regardless of the folder."

One important lesson from SNARF (and my friend's dog) is that the world is divided into two classes of people: those who I have previously interacted with and those who I have not ("strangers"). People who I have not sent mail to before are treated differently than people who I have. This is the concept behind white-lists, which are meant to hold off spam and bulk email; it can also be successful in handling email messages during triage. During those pressed times, emphasizing the top few messages from the most-contacted correspondents probably makes sense.

The goal of social ranking in SNARF is that the high-priority people a user interacts with often will float to the top of a list, while strangers would stay at the bottom. Survey respondents generally "agree" (mean = 0.95, N=58) that sorting correspondents socially is useful. One user commented, *"[My favorite aspect of SNARF was:] sorting mail by sender, first sender first."*

5.3.1 The Difference between 0 and 1

SNARF functionally presents two different sorts. One of them is the fine distinction presented by the gradations of the social sorting. A second, however, is the broad distinction between 0 and 1: people who I have sent mail to before and those who are new to me.

This is verified by the log files. It may be unsurprising that users clicked most often on people they had sent mail to before. Indeed, 94% of users clicked on more people who they had interacted with than they did on strangers. What is interesting is that when the name of a person who they had not previously interacted with was in a shorter list—that is, a more filtered set of people—users were more likely to click on it. We examined the 180 users who clicked at least once in a shorter list (up to 10 names) or at least once in a longer list (over 30 entries). In shorter lists, users averaged 38% of their clicks on strangers; on longer lists, only 15% of clicks were on strangers (one-tailed t-test, $t(318) = 7.08$, $p<0.001$). In other words, having email come directly addressed to them or otherwise filtered made users more willing to read messages (Note that this email had already passed through the corporate spam filter).

We incorporated this idea into later (post-study) versions of SNARF. The colored background behind a correspondent's name reflects the number of messages sent to them. Increasing numbers of messages are highlighted with more intense shades of blue; strangers, however, are coded with grey. In this view, both very close correspondents and strangers are immediately apparent (as illustrated in Figure 1).

The notion that one might want to ensure that some people would show up higher in lists than others again led to some second-order effects among the research team. One researcher started sending introductory messages in order to 'prime the pump' and have SNARF already set with names of relevant people. In a sense, this is a mechanism for catching SNARF up with events that have happened in the real-world, and turns a handshake into an email message.

It might be interesting to push this one step further: to identify "true strangers" as differentiated from people I have seen before. Often in the corporate setting, people come into our view piecemeal: through shared membership on a mailing list, introduction via third party, and only later do we interact in person. SNARF's archives could be interesting to help find that common ground. Of course, some jobs encourage contacts from strangers, journalists, sales, and technical support engineers, for example. Strangers can be given prominence in SNARF by reversing the sort order.

Inevitably, though, depending on an interaction history means that there will be a lag to context changes. One user complained, "*I had just moved teams but SNARF would bunch of the mails from my prev team and bring them up over my current teams."* One way around this would be to set their sorting threshold to be very short, so that messages from only the previous week or so were counted. This would give disproportionate social strength to recent messages; and SNARF would reflect the differences.

5.3.2 Organizational Information and Sorting

The default sorting in a SNARF view is based on the number of messages a user has sent to a correspondent. This method of sorting neglects to take into account important social information about organizational relationships. For example, a message to you sent from your manager's manager might be very important to quickly respond to, but if you do not typically email that person then the message might not bubble to the top of your view.

We have explored some ways of incorporating organizational information into SNARF, in part responding to feedback from users who wanted organizational information included. However, in thinking about how to "count" organizational information it quickly became clear that any method we chose would move away from the notion of simple and transparent ranking in SNARF. Taking into account the diversity of organizational roles—teammates, hierarchical ladders, and peers in sub-organizations—added more complexity than we were prepared to handle.

We have chosen not to incorporate organizational information into SNARF at this point. In practice, we have found that multiple views may handle issues of important email from an infrequent correspondent. As personal views are typically shorter (section 5.1.1), a personally addressed message from an important but infrequent correspondent is easily spotted.

5.4 Social Awareness: an unexpected gain

Perhaps the most surprising result of the deployment of the SNARF tool was the success of using SNARF as an awareness tool. The same minimalist display that addressed the triage task also meant that the application could be left active at a minimal cost of screen space.

Overall, post survey respondents were "neutral" when asked if SNARF was useful for awareness, (mean = -0.01, N = 74). However, we found that users who continued to use the tool were more excited about its use for awareness. The post-survey respondents that told us they were "still using" SNARF at the time of the survey "agreed" that the tool was useful for keeping aware

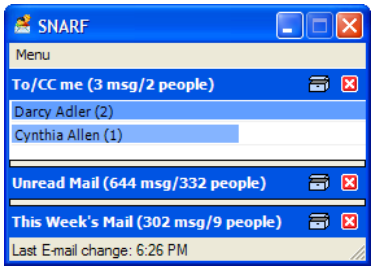

Figure 2. SNARF collapsed in "Awareness Mode". Double clicking any view title will open that view

(mean = 0.78, N=18). In contrast, users who used SNARF just for "a few days" before stopping "disagreed" that SNARF was useful for keeping aware (mean = -0.43, N=40), a lower rating than they gave to "finding important email" or "triaging."

Unfortunately, people that find SNARF valuable for awareness are a challenge for log file analysis: by their nature, they were less likely to click on messages in SNARF; rather they used SNARF to decide when to bring up Outlook. However, looking at logs for the 124 users who ran SNARF on 5 or more days, 87 of them (70%) did not click at all on half of the days they were running SNARF. Of course for these people who kept SNARF open, but didn't click on it, we cannot know whether they were using SNARF for awareness, or merely leaving it open. However, this pattern of behavior does suggest SNARF could have been helping handle the "flow" phase of email (to use [12] terminology): the time in which users keep up with incoming messages, is also a poorly-addressed feature in many current tools.

For example, in many common mail tools, a small notification window appears in the corner of the screen for a short time when new mail arrives, and disappears shortly after. This is a "push" mechanism: if the user is to read the message, they must quickly evaluate the window, reading the sender and subject, decide if they wish to read it, and click, all within a few seconds. In contrast, SNARF shifts to a "pull" model. A user can glance at SNARF to see if new email they might be interested in has arrived and then decide whether to read the message now or later, without the sound or a distracting [3] visible change.

Based on feedback during the field deployment we have made modifications to SNARF to better support awareness. In particular, clicking on the title bar of the view collapses it to one line that shows the number of messages and people in the view (Figure 2). This allows users to greatly collapse the size of the SNARF window, but still retain social awareness of new messages. Using SNARF as an awareness tool had a second advantage during the early deployment phase: when performance was particularly poor, users could still use the awareness aspects to tell themselves when to look to Outlook for new mail.

5.5 Email Management and Workflow

"Cool concept, just missing some of the UI that is core for my management style."

Email is, famously, a "habitat" [4]: users, living within it, have a well-established way of getting their tasks done and of solving their problems. Any new mail client must address the problems of handling all the users' challenges. A total redesign of a client is often impractical: some minor missing keystroke, macro, or formatting tool can block adoption [7; Venolia, Personal Communication]. A new mail client must respect a user's current structures, which can be deeply idiosyncratic.

5.5.1 Costs and Benefits of a Separate Application

"Simple UI, but it really needs to be integrated into Outlook."

SNARF, by acting as a supplement that addressed one major task—email triage—avoided some of those challenges, although it still did not work for all users. For some users, the mere fact of having a second application—and one that had to do its own indexing and mail reading—was a step too far. A common complaint on the post survey was that SNARF was disruptive to workflow. It required a second focus of attention, a second window to be maintained, and a second set of standards and displays to be learned. One solution is to *integrate* with normal use. ClearContext[3], for example, presents itself as a *part* of the Outlook environment: it provides new buttons, new controls, and a new ordering of messages, but it does not keep the user from their expected behavior.

However, certainly for the people that found awareness related benefits from SNARF, having a separate application could be desirable. If one builds a separate application that interacts with an email client, as SNARF does, we have found it is critical to think carefully about the integration between the two applications. For example, the version of SNARF used in the field study did not distinguish between folders, and so did not respect painstakingly-created user rules. One user told us: *"Can not set specific folder to monitor. I want to be able to create different view for different folder which sorted by Outlook rules."* People who were accustomed to getting important contextual information based on the folder in which an email is located, were denied this data in their views. We have since added the ability to specify the email folders used when selecting message for a view.

5.5.2 Challenges Filing and Recalling Mail

The email triage task, as explored by [10, 12, 15], focuses on deciding how to handle individual messages. The notion of a mailbox needing to be cleaned up or re-organized is left for a periodic "cleanup" sweep. In reality, users are accustomed to often being able to file or delete messages after they have been read, even if they often do not bother to do so. SNARF allowed users to do this, but only while they were reading the message itself through the Outlook message interface.

Many "one-touch" strategies for dealing with email (such as the popular "Getting Things Done" [1]) combine *reading* mail with immediate *filing* or *processing* of messages. From discussions with our users, many seemed accustomed to reading a message, closing it, and then filing or deleting it. Once the message was read in SNARF, it was no longer able to be deleted, as it had disappeared from the interface. Users who wanted to delete messages would have to go to Outlook to do so—or would have to do it from within the message. This meant users found that SNARF did not smoothly support the filing and processing phase of triage.

In addition, the inbox being an "overloaded" [16] "habitat" [4] means that *already read* messages can be important for browsing or reviewing for tasks. Once a message had been read, it

[3] Available commercially, http://www.clearcontext.com

disappeared from the SNARF interface. While this made SNARF excellent for triage, it was poor for reviewing "current tasks".

5.6 Social Metadata Challenges

SNARF, an experiment into using social metadata, presents cautionary tales as well as success stories. We found that resolving identity was surprisingly difficult; that changes in context are highly disruptive; and, most importantly, that social history must entail more then just email.

5.6.1 Knowing Me, Knowing You

One particular challenge turns out to be the issue of identity. While the concept of identity is a complex and rich one, SNARF needs only to resolve names and email addresses. In SNARF, identity becomes a locus for aggregation: a social history is assembled around unique names. As SNARF's most critical views are centered on messages addressed to the user, resolving which messages are "to me" becomes particularly important and surprisingly difficult. Several users, particularly once SNARF was publicly available, told us they had several different email addresses and combined messages addressed to them into a single account's contents; they wanted all of the messages to any form of their identity to appear in the SNARF interface.

In the version studied in the field deployment, users had to select a single name for "me". It is clear that the flexibility to specify "me" is critical to continuing to use SNARF, and other socially-sorted systems. Nor is this just a problem for knowing yourself. Rather, others might also have sent email from multiple addresses, such as work and personal accounts. Automatic identification and integration of content from multiple identifiers for a single individual is an open area for future research.

5.6.2 Changes in Context

As we note above (section 5.3.1), social histories by necessity track a user's past. Yet sometimes, a user's context changes—by changing jobs, for example, or by switching projects—and their history can become less valuable as predictors of the relevance of content from new correspondents. Indeed, it can be disruptive. Socially-sorted systems should consider ways to resolve this, either by reducing the importance of past interactions, by pre-populating information, or by incorporating organizational information.

5.6.3 Social History is More Than Email

SNARF treats the volume of past emails as a proxy for relationship strength. Obviously, this is a fairly limited metric—relationships and interactions with other people are far deeper, as anyone with a parent who dislikes email knows. SNARF's logs do not capture instant message conversations, telephone calls, or hallway interactions. Using a tool with a more unified back-end—one that can collect a variety of communication modalities—might help capture the first two of those; however, it should always be clear that measuring electronic communications may not, in the end, account for true social importance. Volume does seem to be a comprehensible way of operationalizing interaction.

6. Using Metadata in Different Contexts

SNARF shows one set of extreme experiments into social sorting, as applied to email triage. We believe, however, that these concepts of applying social metadata to end-user tools to be broadly applicable. We believe that using fairly simple metrics—such as the "number of email messages sent recently" that SNARF uses—can be a successful way of imposing a sort order.

Once a user's interaction history is known, a number of applications are possible. One of them is ranking results. When a user begins an email message, many contemporary mail programs auto-complete the name as they begin to type, based on contact book entries or other caches. Using social metadata suggests that it may be productive to order those choices by social history—the people who have been contacted the most are the most likely to be contacted next. Social metadata may also be applied to the various search tools across personal communication and desktops. As many forms of information carry the traces of collaboration—documents have been shared, email messages have been sent—it may be possible to use the people involved as an index for ranking results. When I search on an ambiguous name, I am more likely to mean people who are frequently a part of my communication than people who are not.

While we criticized animated pop-ups as a tool for awareness (Section 5.4), the more disruptive alert that it provides might be valuable for very important correspondents: a targeted notification, limited to both messages that are personally-directed and from highly-connected people, might be more valuable (as in [8]).

Finally, a number of users seemed particularly interested in using SNARF as a way to get a general overview of their email: who they have interacted with and how often, and to collect general statistics on their informational history. For example, one user commented *"[My favorite aspect of SNARF was:] being able to tell who the people are who send me the most mail—and having the evidence to persuade them to send me less!"* While much of this social metadata is collected in SNARF database, it is not yet reflected back to users in a general way.

7. CONCLUDING REMARKS

While SNARF may not be for everyone, deploying an application that people could use on their own email on a day-to-day basis allowed us to collect a wealth of data. The deployment illustrated that personally-addressed mail matters, that social metadata can be helpful for staying aware of incoming mail, and that strangers are importantly different from people you know. We also saw challenges in accurately representing people's identities and in handling the lag in social history when people switch communication contexts. These lessons can be applied to future systems that incorporate social metadata.

One of the most popular features of SNARF was the personal views. We are pleased that aspects of this can be directly approximated in other mail clients today. Many email clients support "search folders," dynamic folders that list all email that match given criteria. A search folder of "unread mail, sent to me" allows people to adopt one use of social metadata into their current email experience. Of course there are some drawbacks: the view provided in the search folder would not combine multiple messages by one person, nor would it sort those people by their names.

Finally, we feel one of the strengths of SNARF was the *simplicity* of the underlying metrics. The notion of "how many emails have I sent to this person" is a comprehensible one and the transparency of the social sorting algorithms allowed users to understand the system. As we begin to take use social metadata in other places,

as when auto-completing email addresses or ordering search results, emphasizing simple and straightforward metrics will prevent user confusion

8. ACKNOWLEDGMENTS

We would like to thank Gina Venolia for valuable survey and user interface design thoughts; in addition, we would very much like to thank our users.

9. REFERENCES

1. Allen, D. "Getting Things Done: The Art of Stress-Free Productivity." Penguin Putnam: New York, 2001.
2. Bellotti, V., Ducheneaut, N., Howard, M., and Smith, I. Taking email to task: the design and evaluation of a task management centered email tool. *Proceedings of CHI 2003, 345-352*.
3. Cutrell, E., Czerwinski, M., and Horvitz, E. Notification, Disruption, and Memory: Effects of Messaging Interruptions on Memory and Performance. *Proceedings of Interact 2001*. Tokyo, Japan. 2001.
4. Ducheneaut, N., & Bellotti, V. E-mail as Habitat: An Exploration of Embedded Personal Information Management. Communications of the ACM, 2001, 30-38.
5. Fiore, A., LeeTiernan, S., Smith, M. Observed Behavior and Perceived Value of Authors in Usenet Newsgroups: Bridging the Gap. *Proceedings of CHI 2002*, 323-330.
6. Fisher, D. and Dourish, P. Social and temporal structures in everyday collaboration. *Proceedings of CHI 2004.* 551-558.
7. Gruen, D., Rohall, S., Minassian, S., Kerr, B., Moody, P., Stachel, B., Wattenberg, E., and Wilcox, E. Lessons from the ReMail prototypes. *Proceedings of CSCW 2004*. 152-161.
8. Horvitz E., Jacobs A., Hovel D. Attention-Sensitive Alerting, Proceedings of UAI '99, 305-313.
9. Nardi, B., Whittaker, S., Schwartz, H. NetWORKers and their Activity in Intentional Networks. *The Journal of Computer-Supported Cooperative Work*. 11, 2002, 205-242.
10. Neustaedter, C., Brush, A., and Smith, M. Beyond "From" and "Received": Exploring the Dynamics of Email Triage. *Proceedings of CHI 2005*, 1977-1980.
11. Neustaedter, C., Brush, A., Smith, M., and Fisher, D.The Social Network and Relationship Finder: Social Sorting for Email Triage. *Proceedings of the 2005 Conference on Email and Anti-Spam (CEAS)*.
12. Takkinen, J. and Shahmehri, N. "Café: A Conceptual Model for Managing Information in Electronic Mail." *Proceedings of the Thirty-First Hawaii International Conference on Software Systems*, 1999.
13. Tyler, J., and Tang, J.C. When Can I Expect an Email Response? A Study of Rhythms in Email Usage. Proceedings of the European Conference on Computer Supported Cooperative Work (ECSCW 2003), ACM Press.
14. Venolia, G.D., and Neustaedter, C. Understanding Sequence and Reply Relationships within Email Conversations: A Mixed-Model Visualization. *Proceedings of Proceedings of CHI 2003*, 361-368.
15. Venolia, G.D., Dabbish, L., Cadiz, J.J., and Gupta, A. Supporting Email Workflow. *Microsoft Technical Report TR-2001-88.* 2001.
16. Whittaker, S. and Sidner, C. Email Overload: Exploring Personal Information Management of Email. *Proceedings of CHI 96*, 276-283.