

Contrasting Portraits of Email Practices: Visual approaches to reflection and analysis

Adam Perer

Human-Computer Interaction Lab &
Department of Computer Science
University of Maryland, College Park, MD 20742
adamp@cs.umd.edu

Marc A. Smith

Community Technologies Group
Microsoft Research
Redmond, WA 98052
masmith@microsoft.com

ABSTRACT

Over time, many people accumulate extensive email repositories that contain detailed information about their personal communication patterns and relationships. We present three visualizations that capture hierarchical, correlational, and temporal patterns present in user's email repositories. These patterns are difficult to discover using traditional interfaces and are valuable for navigation and reflection on social relationships and communication history. We interviewed users with diverse email habits and found that they were able to interpret these images and could find interesting features that were not evident to them through their standard email interfaces. The images also capture a wide range of variation in email practices. These results suggest that information visualizations of personal communications have value for end-users and analysts alike.

CATEGORIES AND SUBJECT CLASSIFICATION

H.1.2.a Human factors, H.4.3.c Electronic mail

KEYWORDS

Email, information visualization, personal communication

1. INTRODUCTION

Many people accumulate extensive email repositories that contain detailed patterns of their personal communications and relationships over time. The complex patterns include the intensity and duration of relationships with people and organizations, whose mail is never read and whose mail is always replied, and the times of the year the communication normally occurs. Locked up within standard email interfaces, patterns such as these are difficult to capture and

convey. The result is that rich information about one's self and relationships with others are not as apparent as they could be. We describe here a set of three related visualizations that attempt to visualize patterns automatically constructed from email repositories. Correspondent Treemaps provide users with an overview of their email store, making it possible to navigate to and manage emails from specific organizational groups and important contacts. Correspondent Crowds allow users to reflect on their communication dynamics with peers. Author Lines affords users with an annual window to track the evolution of conversations and navigate through their communication history.

These visualizations support two goals: 1) to enhance end-user ability to navigate and reflect on their email archives, and 2) to improve understanding of the range of variation among email user's practices. Our findings from initial explorations with eight highly disparate types of email users suggest that end users as well as analysts find value in these representations.

2. PATTERNS OF PERSONAL COMMUNICATION

An email repository is a complex multi-dimensional structure which can be observed at many scales. They may contain messages from months to years of time, each message potentially linked to others and to and from one or more people, email lists, and other automated entities. These messages are further sorted and deleted manually or systematically into folders or tagged with multiple keywords. No one view can capture every dimension, facet and scale of these archives. We sought to find visualizations that could capture the key hierarchical, correlational, and temporal patterns present in email stores. These are patterns that are difficult to capture using traditional interfaces and each has particular value for the social comprehension of relationships and history.

Many hierarchical structures are present in email datasets: folder hierarchies are obvious examples but the domain names of email senders also have an implicit hierarchical structure. Further, many corporate email systems save and publish organizational structures of who reports to whom which forms an additional source of data about social

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

AVI '06, May 23-26, 2006, Venezia, Italy.

Copyright 2006 ACM 1-59593-353-0/06/0005...\$5.00.

hierarchies and implicit social clustering. To capture some insight into these structures we created Correspondent Treemaps, an interactive treemap visualization based on the combination of email domain name hierarchies and organizational chart data. These maps were colored based on the state of read or unread email from that entity.

Hierarchies are not the only data structure present in email stores: dyadic ties based on sending and receiving messages are also present. Correlations between the numbers of messages sent-to or received-from an entity capture the differences in mutuality and balance between the user and their email correspondents. Not all relationships are marked by equal rates of initiation and reply (or reply at all) something that is difficult to assess using standard email displays. We linked the treemap displays to an interactive scatterplot visualization, Correspondent Crowds. These “crowds” capture the numbers of messages to whom a user sent emails against the number received from that person, with total number of messages either sent or received over a selected period of time represented by the size of the author circles.

The first visualization provides the user with a hierarchical map to browse contacts and the second visualization presents correlation between sent and received emails from each author. The third visualization, Author Lines, highlights temporal rhythms of initiation and reply in the form of a 52 column histogram displaying strips of stacked bubbles, each of which represent a thread started or replied to in a given week and sized in proportion to the number of messages in the thread. This captures the rhythm and quality of initiation and response activity and makes the size of the discussion an easily visible feature. The result is a set of three interlinked views that attempt to capture important patterns of the social life of email.

3. METHODOLOGY

We presented our three interlinked views to a set of subjects selected based on their diversity of email habits. We provided our subjects with a brief training and then asked them to interpret and engage the interactive visualizations.

We selected 8 subjects from a set of 480 employees from our company based on the structure of their personal email store. This information was gathered via SNARF [4] and was collected in a form that anonymized the names of authors, subjects, and folders. Data collected included attributes such as the number of correspondents present in each email store, the total volume of messages, and the number of folders into which the email was sorted. Using these dimensions, we used hierarchical clustering to identify collections of users with similar and distinct email practices, as shown in Figure 1.

We selected subjects who occupied extreme locations on these dimensions, ensuring that we had one subject from each cluster. These subjects were then contacted to request

		Low Message Volume	High Message Volume
Low # of Contacts	Low Folder Count	25 (5.2%)	7 (1.5%)
	High Folder Count	10 (2.1%)	4 (0.8%)
High # of Contacts	Low Folder Count	8 (1.7%)	6 (1.3%)
	High Folder Count	10 (2.1%)	3 (0.6%)

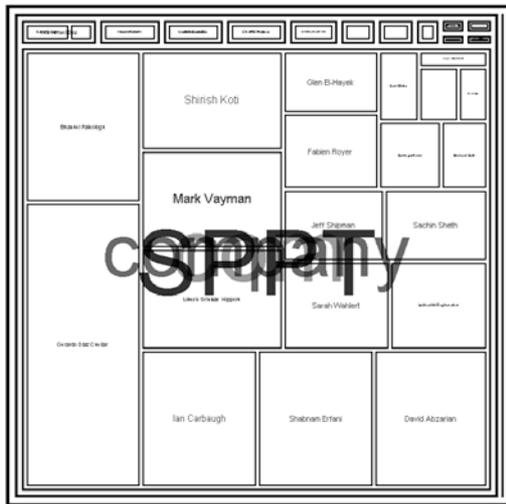
Figure 1. Our taxonomy of email practice diversity. Low means at least one standard deviation less than the average, and high means at least one standard deviation above the average. Inside each cell in the taxonomy are the number of people that fit this profile, with the percentage overall in parentheses.

an hour-long meeting to review visualizations of their email. Each subject who agreed to meet with us was sent a client application to run against their email store prior to our arrival. Subjects were also asked to complete a questionnaire before the interview. Because we were able to select subjects on the basis of detailed structural information about their email archives, these images also illustrate a range of variation present across a large set of email users.

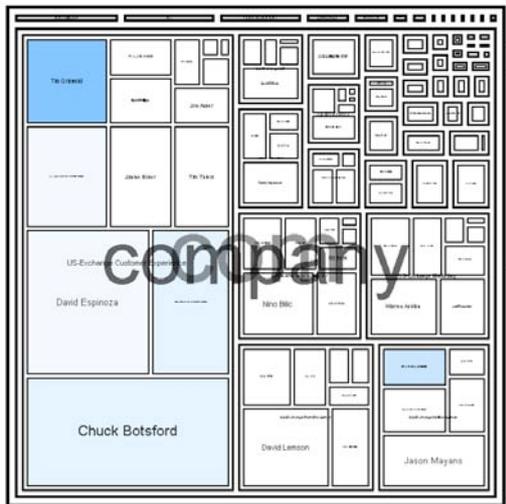
The email data analysis tool collected all email from mounted archives and processed them into metadata that could then be plotted in the form of three interlinked information visualizations. Subjects were then guided through the review of these visualizations while being questioned about their interpretations. Interviewees were shown data from the current month, previous month, and current year as separate views that they could then freely navigate and interpret. While the tasks we assigned to users were simple and straight forward, they served to ensure that users engaged with the visualizations and accurately interpreted their meaning. In fact, the heart of the interviews was to watch them explore task-free.

4. CORRESPONDENT TREEMAPS

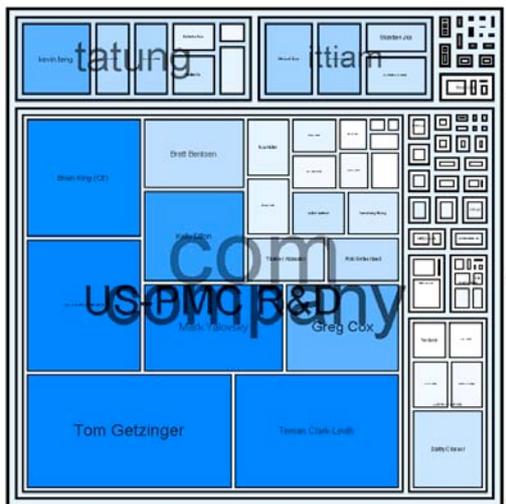
Our treemap visualization [7] displays each user’s correspondents organized into hierarchies based on the domain hierarchy implicit in email addresses. Treemaps are a hierarchically ordered set of boxes within boxes. For example, all .edu contacts would be constrained to an outer box which is comprised of smaller boxes representing each university, which are comprised of even smaller boxes representing contacts at that university. For internal contacts, we were able to make additional use of hierarchical data drawn from internally public organizational data so that additional sub-hierarchies could be constructed based on management structure and job titles. As a result the treemap could display workgroup structures centered on departments and job positions in addition to hierarchies expressed in the departmental domain names of many external email addresses (“@cs.umd.edu”, for example).



(a) Low # of contacts, Low # of messages, High # of folders



(b) High # of contacts, High # of messages, High # of folders



(c) High # of contacts, High # of messages, Low # of folders

Figure 2. Correspondent Treemaps for three different employees when visualizing their current month. The differences in organizational segmentation highlight the variation of contacts in in-groups and out-groups. Below each image, the user’s email type, rated according to our taxonomy in Figure 1, is listed.

The size of each correspondent’s node is proportional to the number of messages authored during the selected time period. The color is based on the number of unread messages currently in the email store, with a greater intensity of blue implying more unread messages.

Users could zoom-in to sub-hierarchies of the treemap to gain more resolution and isolate groups of interest. Right-clicking the node allows users to pivot to the selected person’s Author Lines (described below) or open the messages authored by the correspondent selected. During the interview, subjects were asked to locate their top correspondents that they listed on the pre-interview questionnaire. All subjects were able to complete the task.

Users expressed a great deal of satisfaction with this view, commenting particularly on its ability to automatically arrange email by organizational structures:

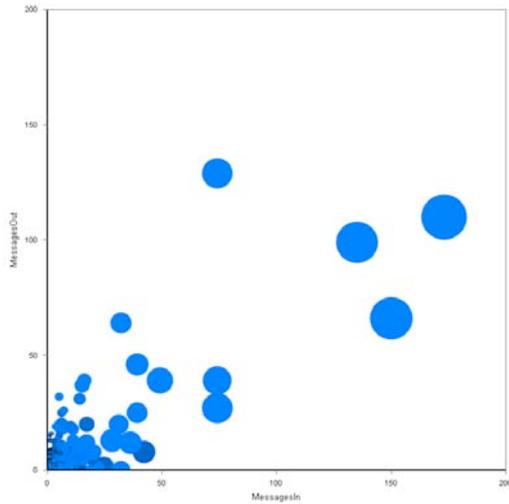
“The marketing team is separate. I love it. It represents my organizational scheme.” – Program Manager (Figure 2c)

“A good program manager is one who talks to her developers and testers. This view allows you to see that!” – Program Manager (Figure 2b)

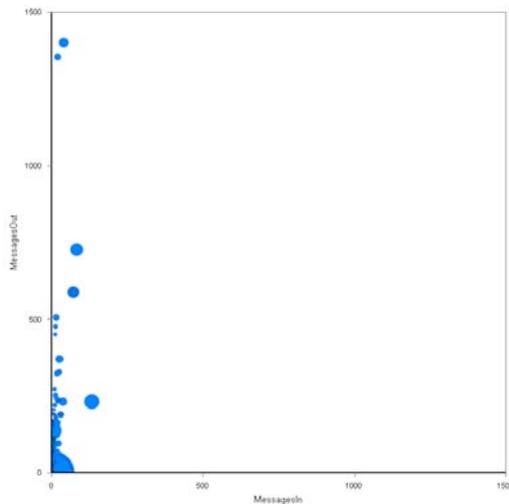
Treemaps organized by departmental hierarchy convey the proportion of the user’s email that is sent or received within or across official organizational boundaries. Figure 2a represents a manager who does not have many large volume correspondents outside of his organization chart defined work group. Figure 2b belongs to a manager whose duties sweep across departments, but still rarely contacts people outside of her company. Figure 2c reveals that the user has a lot of unread email, with contacts that span many internal departments, as well as external companies. This is consistent with this user’s job, as he needs to communicate with both internal marketing departments and external hardware providers.

Arranging contacts by hierarchy is compelling because it allows email users to put “blindners” on, and deal with a group of people whom they needed to attend to with higher priority. Users could zoom in, triage the unread emails from a specific group or class of employee, and then zoom back out. Several users claimed they manually author similar kinds of organizations with folders so that they can deal with certain types of emails all at once and separate from other classes of messages. Our workgroup clusters approximated those contexts and did not require active filing or authoring of rules.

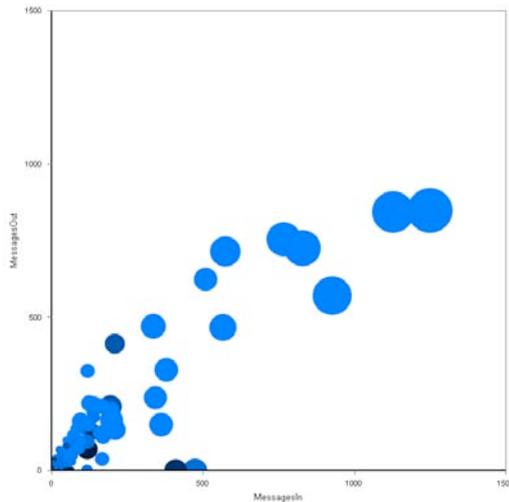
Internal contacts can also be alternatively grouped by job title, and some users felt this was more useful, particularly when their job spanned multiple departments that did not have clear boundaries. For instance, one program manager liked being able to see all of her developers and testers grouped together, because she knew responding to their needs was critical.



(a) High # of contacts, Low # of messages, Low # of folders



(b) High # of contacts, High # of messages, High # of folders



(c) Low # of contacts, High # of messages, High # of folders

Figure 3. Correspondent Crowds for three different employees. The horizontal axis measures emails received by the correspondent represented as a circle, whereas the vertical axis measures how many emails the user sent to that correspondent.

Since many email users within a work environment have an organization hierarchy, treemaps offer a compelling way to present a single screen overview of the whole hierarchy. Overviews are an important goal of our work, as are selective views: users want to “put the blinders on” because of high levels of organizational spam which overwhelm their email management time budget.

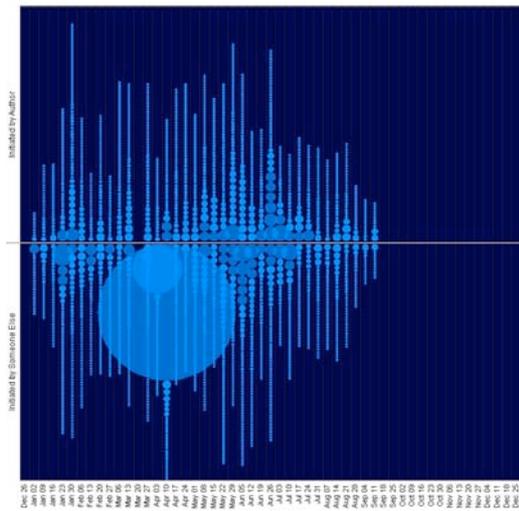
5. CORRESPONDENT CROWDS

Scatter plots (which we refer to as “Correspondent Crowds”) are generated based on the number of messages sent to the correspondent against the number of messages received from the correspondent. Each correspondent is represented as a circle whose diameter is proportional to the total number of messages received from that person in the selected time period. The color of the circle represents how much time elapsed since the last message was received from that correspondent, with saturation decreasing with age. Circles can be selected to gain detailed information on that person in a side panel (not pictured). Right-clicking the node allows users to pivot to the selected person’s Author Line (described later) or open the messages authored by the correspondent selected. During the interview, subjects were asked to locate the person who sent them the most mail, and the person to whom they sent the most mail. All subjects were able to complete the task.

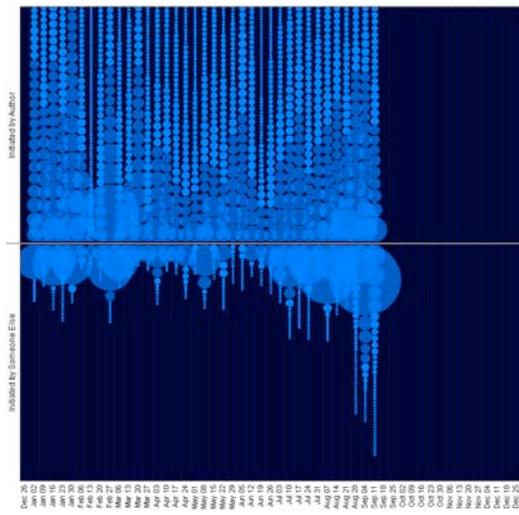
“It may help me think about how I communicate with my team. Email might not be the most effective way to get what you need out of some of them.”
–Development Manager (Figure 3c)

Email users who send their peers significant numbers of status tracking messages have patterns that skew to the upper left, as shown in Figure 3b. Participants that fit this pattern typically send lots of information, and requests for information, mostly to their direct reports. However, email users who typically respond to issues have patterns resembling Figure 3c. Then, there are patterns like Figure 3a, who have key correspondents in all corners of the image. In our interviews, these circles seemed to form meaningful clusters. For instance the correspondents who have more messages sent to them than received tend to be the people the user works for, whereas those on the opposite side of the image are the people that work for the user.

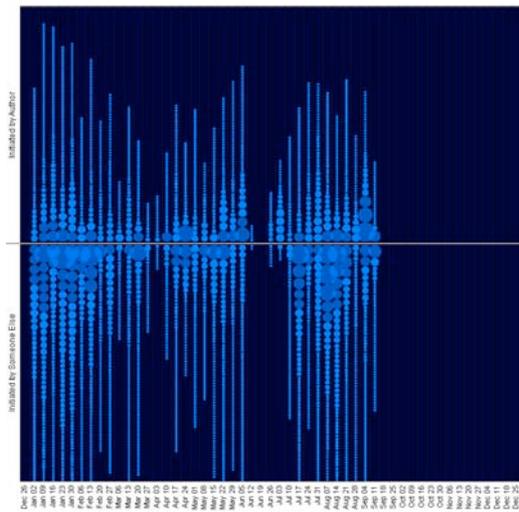
Several interview participants suggested that this would be a useful tool for monitoring who is taking up a lot of their time, and vice versa. Several managers saw this as a great way to see if email is an effective method to communicate with someone on their team. If one user never responds to a bulk of their email, perhaps a different method of interaction would be more suited to the tastes and habits of that individual.



(a) High # of contacts, High # of messages, Low # of folders



(b) High # of contacts, High # of messages, High # of folders



(c) Low # of contacts, High # of messages, High # of folders

Figure 4. Author Lines of the current year for three different employees. The visualizations were generated in September, so there is no data present for last the last quarter.

Correspondent Crowds provide users with a rare opportunity to assess their communication dynamics with their correspondents, information that is generally hidden in standard email interfaces.

6. AUTHOR LINES

Histogram timelines (which we refer to as “Author Lines”) capture weekly patterns of activity in terms of the initiation of new conversations (which were represented as bubbles above a center dividing line) and replies to message threads initiated by others (which were represented as bubbles below a center dividing line) [10]. Each image contains 52 strips of bubbles, one for each week in a fixed calendar year. The bubble’s size increases for every message authored in that week in that thread. The result is a double histogram that captures the patterns of initiation and reply for a single correspondent over a year. During the interviews, subjects were asked to locate the weeks where they initiated the most threads and responded to the most threads. All subjects were able to complete the task.

"I send a lot of emails for questions and that's how I manage things" – Program Manager (Figure 4b)

"Most of the time someone says something is wrong, and I have to figure out what went wrong" – Development Manager (Figure 4c)

Figure 4b portrays an email user whose job is to typically send out requests for information, so the denser activity in the upper half reflects this. Figure 4c belongs to an email user whose job is to respond to problems that emerge, which explains his dense activity in the bottom half of the visualization.

The interviewees typically had little trouble explaining why atypical drops in activity occurred, which were generally due to vacation and leaves of absence. Many participants pointed out that the biggest bubbles on the bottom half were usually big problems they had to deal with. The biggest bubbles on the top half were generally announcements for software releases (and the followup discussions) that get sent out to a wide range of people. The subject line of a thread was usually enough for them to tell a story about what problem concerned.

Users could select a thread bubble and see its activity highlighted in prior and subsequent weeks (not pictured). The ability to track a thread through time was considered to be a useful feature. For example, an employee involved in customer support could now track how long it took for a problem to be solved, since each problem was generally a unique thread.

Author Lines present users with an overview of their communication history, allowing them to follow the evolution of discussions and discern where their communication efforts were spent.

7. RELATED WORK

PostHistory and Social Network Fragments presented email owners with patterns illustrating the temporal evolution of email relationships and emerging social network ties [9]. In interviews, users were able to see meaningful patterns which they then used as artifacts for remembering and storytelling. Perer et al. showed that meaningful patterns could be constructed from temporal rhythms of contacts in an email archive based on the intensity of communication over time [5]. Fisher and Dourish recovered the temporal and social structures of email to build awareness tools to improve collaboration [2]

Visualizations of the thread structure of related messages were shown to provide users with a better understanding of how conversations evolve over time [6,8]. Fiore et al. showed that selecting content from online conversation repositories can be enhanced by considering dimensions of the pattern of each author's activity across a large message collection [1].

These projects highlight the general utility of visualizing patterns otherwise invisible within traditional interfaces to email and other online discussions. Our efforts are in this vein, seeking to find useful ways to capture salient patterns in conversation repositories.

8. DISCUSSION

We found that these visualizations were comprehensible to our interviewed users after a brief training. Participants were able to interpret patterns, explain the data by telling stories, and navigate through the visualizations. In addition to being comprehensible, the data from our interviews leads us to believe the visualizations are valuable as well.

Views have value for analysts.

The study of email use is enhanced by the use of these visualizations. They provide a mechanism for gaining an overview of multiple people's email stores. Our initial study of these views has already allowed us to clearly illustrate how widely different patterns of email engagement can be. These findings may have implications for the design of interfaces that can be tailored to the various styles of email use.

While the techniques in this paper have previously been applied to Usenet discussions we show here that they can also be usefully applied to personal conversation stores. When moved from a public forum to a private message store, the dimensions of these views have necessarily changed. In a newsgroup environment there is no "ego" who is at the center of all message traffic in the way that there is in personal email. This allows for an additional focus in email interfaces as content related to the user is likely to be of greater value.

Views have value for end users.

Correspondent Treemaps allow people to isolate varying groups of people based on their location in an organizational structure, in order to deal with one class of users at a time. Our approach requires no filing or special rules, but rather uses information gained automatically from email habits as inscribed in email archives. This visualization could be used during triage by zooming into a user's department, taking care of all of his peer's email, then zooming back out to take care of emails external to the company or across other organizational boundaries.

Correspondent Crowds allow people to interpret their interaction patterns with other people. With knowledge of how correspondents respond, users can reflect on their communication strategies to select the best mode for the target individual based on their prior history of interaction.

The Author Lines view allows users to reflect on their own email habits by looking at their pattern of activity for an entire year as well as their pattern of interaction with any selected author in their email store. It provides quick access to important threads and an overview of activity during a particular time period. It also allows users to visualize their correspondents' patterns and note that certain relationships ended or have distinct patterns of activity – which could be used to set expectations for future interactions.

Some argue that users already know which colleagues they send mail to and the strength of their relationships. However, this knowledge does not extend as clearly to the next tier of relationships, the more fleeting but still important weaker ties [3]. While people may be aware of their strong ties, their mid-tier relationships are much less understood or recallable (e.g. "who was that guy from the last project we worked on or met at that offsite location some months ago?").

Views may be useful for mobile devices

These images have some benefits over traditional textual lists of email: they can convey overviews in relatively small areas, making them potentially useful for managing email on mobile devices where display area is highly constrained.

Privacy issues

Given the personal and or commercial nature of email stores there are reasonable concerns about the privacy of such stores. Our approach in this work is to push analysis tools to the client system where the chance that personal information will leak is less likely and remains under the control of the end user. Moving such structural information into a public store could raise significant privacy concerns. Researchers and designers should be highly cautious in such circumstances. In our own study we collected data from our internal users but only after strongly redacting all

of the author names, subject lines, and folder names. None of the message content was copied. This approach allowed us to capture useful amounts of information about email habits and practices without risking the exposure of private information.

9. FUTURE WORK

While our implementation was not yet able to update images in real-time such a development shows promise as a way to convey email awareness information. Due to participant interest in using the views to triage and surface messages, we plan to integrate the views into a future version of SNARF, an email management and awareness tool [4].

Some metrics in the views can be customized to choose a specific metric for what the shapes, sizes, position and color represent in a view. We plan to further extend the current options to allow the user to have even more data to look at. Finally, we believe there are additional views that would be valuable, such as social networks created from patterns of reply. Social network data might also be used as an additional source of hierarchical clustering that can be visualized.

As the application is refined we plan a deployment to assess the longer term utility of these views of email. With a significant user population it should be possible to determine which aspects of these visualizations are useful and for what task.

10. CONCLUSION

In this paper we have introduced three portraits of email practices that capture the diversity of email user's message stores. Our views present hierarchical, correlational and temporal patterns that are comprehensible to users and often inspire stories and reflection. Correspondent Treemaps organize contacts for users automatically based on organizational data and their communication history, affording users a rare opportunity to see what is really in their email store. Correspondent Crowds allows users to reassess their communication strategies with peers. Author Lines provides an overview of a year's worth of discussions, highlighting lengthy conversations and their evolution over time. These views also have utility for analysts, who can use the views to highlight important variations in email practices. Our portraits of email practices suggest that information visualizations of personal communication are both appealing and useful.

11. ACKNOWLEDGEMENTS

We would like to thank the SNARF team: Danyel Fisher, AJ Brush, Andy Jacobs, and Bernie Hogan, Tony Capone,

Dave Levin, the participants of our interviews, and our reviewers for thoughtful comments on this work.

12. REFERENCES

1. Fiore, A., LeeTiernan, S., Smith, M. (2002). Observed Behavior and Perceived Value of Authors in Usenet Newsgroups: Bridging the Gap. Proceedings of the CHI 2002 Conference on Human Factors in Computing Systems. New York: ACM Press.
2. Fisher, D. and Dourish, P. (2004). Social and Temporal Structures in Everyday Collaboration. Proceedings of the CHI 2004 Conference on Human Factors in Computing Systems. New York: ACM Press.
3. Granovetter, M. (1973). The Strength of Weak Ties. *American Journal of Sociology*, 78, pp. 1360-80.
4. Neustaedter, C., Brush, A., Smith, M., and Fisher, D. (2005). The Social Network and Relationship Finder: Social Sorting for Email Triage. Proceedings of the 2005 Conference on Email and Anti-Spam (CEAS).
5. Perer, A., Shneiderman, B., and Oard, D. (2005). Using Rhythms of Relationships to Understand Email Archives. University of Maryland Tech Report HCIL-2005-08. To appear in the *Journal of the American Society for Information Science and Technology*.
6. Rohall, S. and Gruen, D. (2002). Remail: A Reinvented Email Prototype. ACM Conference on CSCW 2002.
7. Shneiderman, Ben. Tree visualization with tree-maps: 2-d space-filling approach (1992). *ACM Transactions on Graphics*, ACM Press: Volume 11, Issue 1.
8. Venolia, G.D., and Neustaedter, C. (2003). Understanding Sequence and Reply Relationships within Email Conversations: A Mixed-Model Visualization. Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI 2003), ACM Press.
9. Viegas, F., boyd, d., Nguyen, D., Potter, J., Donath, J. (2004): 'Digital Artifacts for Remembering and Storytelling: PostHistory and Social Network Fragments', Proceedings of the 37th Hawaii International Conference on System Sciences, January 2004, pp. 40109a.
10. Viegas, F., Smith, M.A. (2004): Newsgroup Crowds and AuthorLines: Visualizing the Activity of Individuals in Conversational Cyberspace. In: Proceedings of the 37th Hawaii International Conference on System Sciences, IEEE.